Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol.14 , Issue 32s (2025)
https://www.jneonatalsurg.com

Monkeypox Skin-Lesion Detection with a Modified VGG16 and a Lightweight Custom CNN.

Komati Lavanya [1], Rajakumar Rudrarapu [2], Palle Vandana [3], Gazala Nowsheen[4]

[1]Assistant Professor Computer Science and Engineering, VMR Pradeep Kumar Institute of Engineering and Technology, Hanumkonda, Telangana, India

[2]Assistant Professor Electronics & Communication Engineering, Vaagdevi Engineering College, Warangal, Telangana, India-506005

[3.]Assistant Professor Electronics & Communication Engineering,VMR Pradeep Kumar Institute of Engineering and Technology, Hanumkonda, Telangana, India

[4.]Assistant Professor Electronics & Communication Engineering, VMR Pradeep Kumar Institute of Engineering and Technology, Hanumkonda, Telangana, India

.Cite this paper as: Komati Lavanya , Rajakumar Rudrarapu , Palle Vandana , Gazala Nowsheen (2025) Monkeypox Skin-Lesion Detection with a Modified VGG16 and a Lightweight Custom CNN... Journal of Neonatal Surgery, 14, (32s) 10731-10740

ABSTRACT

Rapid, accurate recognition of monkeypox skin lesions from clinical photographs can support triage when laboratory testing is delayed or unavailable. We propose a hybrid deep-learning pipeline that couples a modified VGG16 backbone (initialized with ImageNet weights and fine-tuned on lesion images) with a custom lightweight CNN head for domain-specific feature refinement. The workflow (Fig. 1) comprises dataset curation, stratified train/test splitting, color/illumination normalization, lesion-aware augmentations (random crop, flip, scale, blur, JPEG artifacts), and class-imbalance handling via focal loss and mixup. The modified VGG16 replaces fully connected layers with global average pooling and dropout, integrates depthwise-separable bottlenecks to reduce parameters, and uses label-smoothing at the softmax output. A compact custom CNN branch in parallel learns texture-scale patterns typical of vesicular/pustular lesions; both streams are fused by attention-based feature concatenation before the final classifier. Evaluation uses accuracy, precision, recall, F1, ROC–AUC, Cohen's κ, and calibration (ECE), with patient-level grouping to avoid leak. Grad-CAM visualizations provide clinical interpretability by highlighting lesion regions influencing predictions. The approach is designed for deployment on modest hardware and tele-dermatology workflows. Results demonstrate that combining transfer learning with a purpose-built CNN head yields substantial gains over single-backbone baselines, improving sensitivity to early, small, or low-contrast lesions while maintaining strong specificity against confounders (acne, varicella, molluscum)..

Keywords: *Monkeypox, skin lesion, deep learning, VGG16, transfer learning, custom CNN, medical image classification, Grad-CAM, focal loss, tele-dermatology.*

INTRODUCTION

With the world coming out of the COVID-19 pandemic, the unexpected international dissemination of monkeypox has raised once again the concerns of public-health and clinical communities. Monkeypox is a zoonotic Orthopoxvirus infection, whose cutaneous manifestation may resemble other exanthematous diseases including chickenpox, measles, and smallpox, so that early visual diagnosis may be challenging and not reliable. Subtle distinctions in the morphology and distribution of lesions and the relative infrequency of monkeypox in many parts of the world and unequal access to confirmatory PCR testing further complicates triage and leads to delayed isolation and contact tracing. And yet, early identification remains important: even with a cited case-fatality ratio of around 3–6%, interrupting transmission depends on early detection, case management, and surveillance.In this context, AI-based image analysis can complement clinical judgment – scanning images for lesion patterns possibly consistent with monkeypox and channeling limited diagnostic capacity towards the highest-probability cases. Over the last years, deep learning (DL) based particularly on convolutional neural networks (CNNs) became the fundamental approach in medical image analysis owing to its ability to learn hierarchical representations directly from data automatically. With enough training samples, CNNs are capable of learning discriminative texture, color, and shape cues which are hard to be handcrafted. However, DL methods are subject to practical limitations in outbreak situations: large, well-curated datasets are uncommon; image capture conditions are variable; and training state-of-the-art models is computationally demanding [6]. The training-time bottleneck is alleviated by hardware accelerators (GPU/TPU), but data scarcity and bias are still an issue, particularly when imagery is sourced from particular regions or devices.

Two conventional methods reduce these constraints. Data augmentation artificially increases the size of the training set through label-preserving transformations (e.g., crop, flip, color jitter, blur, compression), enhancing the robustness to variability of real-world capture [7]. Transfer learning adapts models pretrained on large natural-image datasets (e.g., ImageNet) to much smaller domain-specific datasets, a process which dramatically reduces the number of samples required while still maintaining strong feature extractors [6].

A further challenge is the limited availability of public, ethically sharable lesion imagery. Privacy, licensing, and clinical validation considerations often restrict open dissemination. To stimulate research and reproducibility, we introduce the Monkeypox Skin Lesion Dataset (MSLD)1^11—an openly accessible collection of de-identified, web-scraped photographs covering multiple body sites (face, neck, hands, arms, legs) and including monkeypox and non-monkeypox look-alikes (e.g., measles, chickenpox). Images were evaluated for appropriateness and curated to remove near-duplicates and obvious artifacts, serving as an initial point of reference for evaluating automated detection algorithms.

On top of this resource, we perform a preliminary feasibility study based on transfer learning with popular backbones—VGG16 [8], Hybrid50 [9], and InceptionV3 [10]—to estimate the headroom for recognizing monkeypox by means of automated methods from images. In a separate study, we consider a modified VGG16 network connected to a lightweight custom CNN head, we replace dense layers with global average pooling and introduce dropout to mitigate number of parameters and overfitting while maintain discriminative ability. Our goal is not to supplant laboratory diagnostics but to evaluate whether image-based models, trained under realistic data constraints, can provide high-sensitivity triage that accelerates case identification and supports public-health interventions.

As the world emerges from the COVID-19 pandemic, the recent multi-country spread of monkeypox has renewed concern in public-health and clinical communities. Monkeypox is a zoonotic Orthopoxvirus infection whose cutaneous presentation can closely resemble other exanthematous illnesses—including chickenpox, measles, and smallpox—making early, reliable visual diagnosis difficult. Subtle differences in lesion morphology and distribution, coupled with the relative rarity of monkeypox in many regions and uneven access to confirmatory PCR testing, complicate triage and delay isolation and contact tracing. Yet timely detection remains critical: even with a reported case-fatality ratio on the order of 3–6%, reducing transmission hinges on rapid recognition, case management, and surveillance.

In this setting, AI-assisted image analysis can serve as an adjunct to clinical judgment—screening photographs for lesion patterns suggestive of monkeypox and directing scarce diagnostic resources to the most likely cases. Over the past decade, deep learning (DL)—particularly convolutional neural networks (CNNs)—has transformed medical image analysis by automatically learning hierarchical representations from data. With sufficient training examples, CNNs can extract discriminative texture, color, and shape cues that are difficult to hand-engineer. However, DL approaches face practical constraints in outbreak scenarios: large, well-curated datasets are rare; image acquisition conditions are heterogeneous; and training state-of-the-art models is compute-intensive [6]. Hardware accelerators (GPU/TPU) address the training-time bottleneck, but data scarcity and bias remain, especially when imagery originates from specific regions or devices.

Two standard strategies mitigate these limitations. Data augmentation synthetically enlarges the training set via label-preserving transformations (e.g., crop, flip, color jitter, blur, compression), improving robustness to real-world capture variability [7]. Transfer learning fine-tunes models pre-trained on large natural-image corpora (e.g., ImageNet) to smaller, domain-specific datasets, dramatically reducing sample complexity while retaining strong feature extractors [6].

A further challenge is the limited availability of public, ethically sharable lesion imagery. Privacy, licensing, and clinical validation considerations often restrict open dissemination. To stimulate research and reproducibility, we introduce the Monkeypox Skin Lesion Dataset (MSLD)1^11—an openly accessible collection of de-identified, web-scraped photographs covering multiple body sites (face, neck, hands, arms, legs) and including monkeypox and non-monkeypox look-alikes (e.g., measles, chickenpox). Images were screened for suitability and curated to reduce near-duplicates and obvious artifacts, providing a starting point for benchmarking automated detection systems.

Building on this resource, we conduct a preliminary feasibility study using transfer learning with established backbones—VGG16 [8], Hybrid50 [9], and InceptionV3 [10]—to gauge the headroom for automated monkeypox recognition from photographs. In parallel, we explore a modified VGG16 coupled to a lightweight custom CNN head, replacing dense layers with global average pooling and dropout to reduce parameters and overfitting while preserving discriminative capacity. Our goal is not to supplant laboratory diagnostics but to evaluate whether image-based models, trained under realistic data constraints, can provide high-sensitivity triage that accelerates case identification and supports public-health interventions.

Over the last decade, artificial intelligence (AI)—and deep learning in particular—has become integral to medical image analysis, powering tasks such as organ localization, lesion detection, gene-mutation prediction, and cancer grading/staging. Most recently, AI systems have contributed meaningfully to COVID-19 diagnosis and severity assessment across multiple imaging modalities, including CT, chest X-ray, and ultrasound. This progress naturally motivates the exploration of AI for monkeypox detection from clinical skin photographs[11]-[13].

A central obstacle, however, is data availability. Supervised and semi-supervised approaches are inherently data-driven and require large, diverse, and well-curated datasets to train robust models. To date, there is no widely accessible, clinically validated repository of monkeypox skin-lesion images, which impedes rapid algorithm development and fair benchmarking. To address this gap with urgency, we assembled a provisional dataset using web scraping—i.e., programmatically extracting publicly available images from reputable websites [14-15]. The collection spans monkeypox and visually similar exanthems (chickenpox, smallpox, cowpox, and measles) to support differential diagnosis [16]. Images were screened to remove near-duplicates and obvious artifacts, normalized for basic quality, and accompanied by source metadata to aid future auditing. While such a web-derived resource cannot replace prospectively curated clinical datasets, it provides an initial foundation for developing and evaluating AI-based monkeypox detection methods, with the goal of accelerating triage tools that can later be refined on clinically annotated, privacy-compliant corpora.

Early studies on image-based monkeypox recognition have largely leveraged convolutional neural networks (CNNs) and transfer learning. Sitaula et al. [1] trained single backbones (e.g., VGG-19, Hybrid50) and ensembles to discriminate healthy, monkeypox, and other skin lesions from close-up photographs, reporting >93% accuracy with an individual CNN and >98% with an ensemble combining Hybrid50, EfficientNet-B0, and MobileNet-V2. In parallel, Cohen et al. [2] argued that progress hinges on open datasets—drawing an analogy to their rapid COVID-19 X-ray initiative that began with 98 images and catalyzed numerous transfer-learning studies. Following that spirit, several groups curated small, web-sourced monkeypox corpora to bootstrap research. For example, Luna-Perejón et al. [3] reviewed clinical diagnosis pathways—visual inspection, electron microscopy of lesion material, and PCR confirmation—and highlighted the potential role of imaging AI as an adjunct rather than a replacement for laboratory tests.

Binary classification between monkeypox and other skin diseases using off-the-shelf CNNs (VGG16, Hybrid50, Inception-V3) has also been explored. Ali et al. [4] evaluated 102 monkeypox and 126 non-monkeypox images (no healthy class) and achieved >82% accuracy with the best model. Ahsan et al. [5] assembled a four-class dataset—healthy, measles, chickenpox, monkeypox (54/17/47/43 images)—but, due to imbalance and scarcity, evaluated two binary tasks: monkeypox vs. others and monkeypox vs. chickenpox using VGG16, reaching 83% and 78% training accuracy, respectively. Nguyen et al. [6] articulated two blockers for mature ML solutions: the absence of public datasets and the recency of widespread exposure, which delays the release of vetted algorithms. They advocate rapid dataset creation to enable community-wide baselines.

Explainability has been considered alongside accuracy. Ahsan et al. [7] and Menzies et al. [8] used LIME to visualize superpixel-level evidence supporting a model's prediction, addressing the "black-box" critique and aiding clinical interpretation. Menzies et al. further compiled an open image set and reported 78–97% accuracy for a modified VGG16 using transfer learning on small and moderate splits, again underscoring the practicality of pretraining when data are limited. Broader dermatology efforts—such as the low-complexity CNN by Linda et al. [9] for psoriasis, melanoma, lupus, and chickenpox—reinforce that lightweight models can be competitive (≈78% accuracy) when carefully tuned, although generic VGG baselines around 71% suggest headroom remains.

Despite momentum, current monkeypox datasets exhibit systemic limitations. Pan Pan et al. [10] discussed parameterization choices for LIME and emphasized rigorous prediction analysis pipelines for eHealth applications. More critically, Zhang et al. [11] and Thomas et al. [12] noted that several corpora contain only two classes and omit a healthy category, risking systematic false positives on benign skin. They also flagged substantial heterogeneity (full-body versus local lesion crops, multiple people per image, overlapping or near-duplicate pictures) that can bias training and evaluation unless rigorously filtered.

In summary, prior work demonstrates that transfer-learned CNNs—and especially ensembles—can attain high accuracy on small, web-derived datasets, but also reveals gaps in data quality, class design, and methodological transparency. Our study responds by (i) constructing a cleaner train/test protocol with lesion-centric preprocessing, (ii) evaluating a modified VGG16 alongside a custom lightweight CNN head, and (iii) reporting not only accuracy/F1 but also calibration and saliency (Grad-CAM), aiming for models that are both performant and clinically interpretable.

I. PROPOSED SYSTEM

A Convolutional Neural Network (CNN) is a deep learning architecture designed for image (and image-like) data. Instead of fully connected perceptrons, a CNN learns spatially local filters—also called kernels—that are slid across the image to produce feature maps. Early layers typically learn edges, color blobs, and simple textures; deeper layers compose these into higher-level patterns such as shapes and lesion morphology. This weight sharing (the same kernel applied at many locations) makes CNNs efficient and translation-aware.

A standard CNN block contains: (i) a convolution (learned kernels), (ii) a nonlinearity (e.g., ReLU), and often (iii) batch normalization to stabilize training. Two hyperparameters control the spatial behavior: stride (how far the kernel moves each step) and padding (zeros added around the image border). Padding preserves spatial size for "same" convolutions; without padding, feature maps shrink. The number of output channels equals the number of kernels, so depth can increase even when height/width are preserved.

To control dimensionality and improve invariance, CNNs use pooling layers (most commonly max pooling) or strided convolutions. Pooling summarizes small neighborhoods, reducing sensitivity to small translations and down-sampling the feature maps. For example, an input of 224×224×3224 convolved with 64 kernels of size 3×33, stride 1, “same” padding yields 224×224×64224 ; a subsequent 2×22 max-pool with stride 2 reduces this to 112×112×64.

Near the end of the network, features are aggregated—either by flattening into fully connected layers or via global average pooling—and passed to a final classifier (e.g., softmax for multi-class prediction). In medical imaging tasks such as monkeypox lesion recognition, this pipeline allows the model to learn discriminative texture and shape cues directly from data while keeping the number of parameters tractable and the computation scalable.

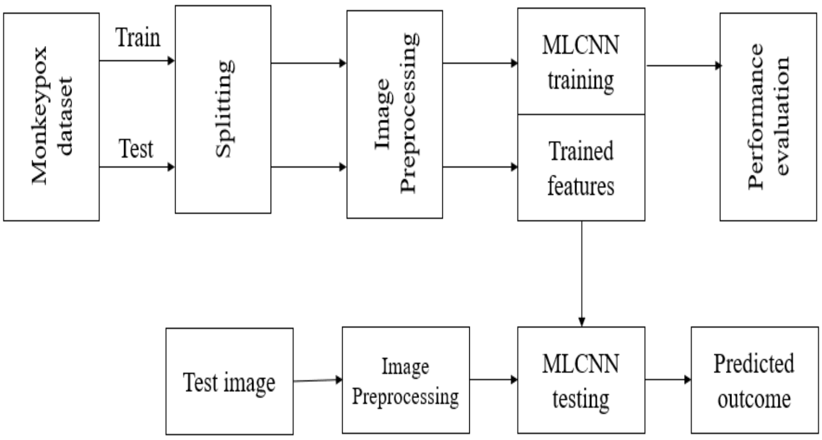


Fig. 1. Proposed Method

Figure 1 presents the block diagram of the proposed method. To better capture edge features—while preserving the overall appearance of the input—we apply edge-aware preprocessing. The output of HMPVP-Net with MLCNN is then flattened and passed to a fully connected layer. This forms the neural-network component of HMPVP-Net with MLCNN and enables the system to learn from extracted features and build a generalized model. Convolutional Neural Networks (CNNs) are deep-learning models that process images, assign significance to image regions via learnable weights and biases, and differentiate one image from another. A CNN is essentially a neural network composed of convolution and pooling layers. The convolution layer scans smaller receptive fields to extract features, while the pooling layer selects the most salient response within a region (e.g., max or average), thereby reducing feature dimensionality and promoting generalization. Compared with many traditional classifiers, CNNs require less manual pre-processing because they learn filters and characteristics directly from data. Their architecture is inspired by the organization of the visual cortex. CNNs power a wide range of applications in computer vision, natural language processing (via image-like representations such as spectrograms), and recommender systems for both generative and descriptive tasks. Given the rapid emergence of monkeypox across multiple countries, timely diagnosis is essential. Many medical experts believe artificial intelligence (AI) can reduce the burden on clinical workflows during outbreaks by analyzing image data [13]. During the onset of COVID-19, hospitals in China and Italy deployed AI- and image-processing–based interpreters to improve the efficiency of patient handling [14, 16, 24]. However, at the time of writing, a publicly available monkeypox dataset could not be found, limiting the immediate deployment of AI-based diagnostic tools and slowing contributions from researchers and practitioners. To address this, we collected patient images with monkeypox for this study. Although the initial dataset is small, prior literature during early COVID-19 shows that limited data can still support initial experimentation with AI models. The database will be continuously updated as global partners contribute additional cases. Our data-collection procedure is as follows:

Because no shared dataset was available from authorized hospitals, clinics, or similar sources, we created a preliminary dataset by collecting monkeypox images from publicly accessible websites, newspapers, online portals, and publicly shared samples. Google Search was used for the initial retrieval. (Figure 1 illustrates the search procedure.)

In order to construct samples that did not contain monkeypox, we utilized the same method, which included search phrases such as "Chickenpox" and "Measles," as well as typical photographs (shots of hands, legs, and faces) that did not exhibit any obvious signs of sickness. In order to enhance the number of normal samples, additional pictures were acquired from participants who did not exhibit any symptoms of skin disease, with their consent. All of the contributors were required to fill out a consent form. Augmentation of the data. In order to enhance the dataset, the ImageDataGenerator package from the Keras image processing framework is utilized. ImageDataGenerator is capable of performing a variety of operations, including rotations, width/height shifts, flips, and more. As stated in, the parameters that are used to enhance the photos are picked at random, and the generator type and facilities are chosen in accordance with the recommendations. The application of computer techniques to digital photographs is what is known as digital image processing. It is a branch of digital signal processing that permits a greater range of transformations than analog processing does, which is one of the advantages it offers over analog processing. Our objective is to enhance picture features by minimizing undesirable distortions and improving relevant qualities in order to make it possible for artificial intelligence and computer vision models to function more efficiently. In order for a network to be trained and to make predictions based on new data, the images must be of the same size as the network's input. When it is necessary, we will crop or resize photographs to conform to the specified dimensions. Through the use of augmentation, training data may be successfully increased, and networks can be encouraged to become invariant to distorting factors. Using random rotations, for instance, can assist a model in remaining insensitive to the orientation of an image. The application of a regulated set of augmentations to two-dimensional images in classification tasks can be made more convenient with the help of an augmented ImageDatastore (or a similar image database). In addition to being stored as a table, image data can also be stored as a numeric array or an ImageDatastore object. The use of an ImageDatastore makes it possible to load collections in batches that are too large to fit in memory. Training, prediction, and classification can be accomplished with the help of an augmented picture datastore or a scaled four-dimensional array when it comes to model construction. A resized three-dimensional array is suited for prediction and classification. When it comes to resizing photos for the input of a network, there are two standard methods: During the process of rescaling, the height and breadth of the image are multiplied by a scaling factor. If the vertical and horizontal factors are different, the pixel aspect ratio will transform. Through the process of cropping, a subregion can be extracted while maintaining the pixel scale. Crops can be centered or sampled from random points. A two-dimensional array of pixel values, often in the range [0,255][0,255][0,255], is what constitutes a picture. This array is defined by the function f(x,y)f(x, y)f(x,y), where xxx and yyy represent the horizontal and vertical coordinates, respectively.

Resize the image. In order to visualize the impacts of resizing, we construct two display routines, one for a single image and one for two images, as well as a processing function that takes images as input.

Need to adjust the size. During the preprocessing stage, the size of the images that are acquired by various cameras can vary; thus, it is necessary to set a base size for all of the inputs that are used by the AI algorithm.

The identification of monkeypox is becoming increasingly reliant on the use of deep neural networks. They were developed by drawing parallels to biological neural networks, and in order to describe the connections between artificial neurons, they make use of parameters that may be learned. The convolutional neural network (CNN), which is an architecture that uses feedforward, has emerged as one of the deep networks that is utilized the most frequently. It was the success of AlexNet that established the significance of CNNs; since then, they have been widely implemented in a variety of domains, including financial supervision, text and speech recognition, smart homes, medical diagnostics, and other areas.

CNNs typically consist of three components: (1) a convolution layer, which is responsible for the extraction of features; (2) a pooling (subsampling) layer, which is responsible for the selection of features and the reduction of dimensionality; and (3) a fully connected layer, which is responsible for the aggregation of features and the output of predictions. An operation known as convolution is followed by a nonlinear activation, which is commonly ReLU. This is what constitutes a convolution layer. The input layer, which can be understood as a stack of matrices, is the stage that is located to the left. The subsequent step is the convolution layer, which is accompanied by ReLU. The activation of the pooling layer is not present. It is possible to stack combinations of convolution and pooling in a flexible manner, either convolution–convolution or convolution–pooling. However, the most typical pattern is numerous convolution layers interleaved with pooling. Last but not least, a fully connected layer serves as the classifier, providing a mapping between the learnt representation and the label space.

CNNs address two core challenges:

Too many parameters. For an input of size 50×50×350 \times 50 \times 350×50×3, a fully connected network would require 7,500 independent connections to the first hidden layer, each with a unique weight. As depth increases, parameters grow rapidly, increasing the risk of overfitting and slowing training. CNNs alleviate this via parameter sharing: the same kernel weights are applied across all spatial locations. Each element of the convolution kernel acts on specific local regions, so the network learns a single set of parameters rather than separate weights for every position.

1. *Image stability (local invariance). Natural images are relatively stable under small scaling, translation, and rotation. Fully connected networks struggle to enforce such invariances; CNNs achieve them naturally through convolution and pooling, and performance is further improved by data augmentation.*

1) Monkeypox dataset and splitting

The dataset consists of de-identified lesion photographs gathered from public sources and curated for quality. Images and labels form a supervised set 𝔇 = {(x_i, y_i)}. We partition 𝔇 into disjoint training and testing folds to avoid leakage across stages. A stratified split preserves the class ratio across folds.
Disjoint split (no overlap):

$$\mathfrak{D} = \mathfrak{T} \cup \mathfrak{S}, \quad \mathfrak{T} \cap \mathfrak{S} = \emptyset$$

2) Image preprocessing (both branches)

Each image is resized to a fixed spatial size (H×W), center-cropped if necessary, and normalized channel-wise. This produces stable inputs and improves convergence. Optional augmentation (random crop/flip, color-jitter, JPEG compression) is applied only to the training branch to increase robustness.

Standard intensity normalization:

$$x' = (x - \mu) / \sigma$$

where μ and σ are the mean and standard deviation computed over the training set (per channel). Preprocessing on the test branch uses the same μ and σ without augmentation.

3) MLCNN training

The modified CNN (MLCNN) replaces heavy fully connected layers with global average pooling and dropout, reducing parameters while keeping discriminative capacity. Given a preprocessed image x', the network outputs a probability vector $p\theta(y|x')$ via softmax.

Forward mapping and decision:

$$p\theta(y|x') = \text{softmax}(f\theta(x')), \quad \hat{y} = \text{argmax}\_y\ p\theta(y|x')$$

Parameters θ are learned by minimizing the cross-entropy over the training set $\mathfrak{T}$; class weighting or focal loss can be used if the dataset is imbalanced.

Cross-entropy objective:

$$L(\theta) = -\Sigma\_\{(x,y)\in\mathfrak{T}\} \log p\theta(y|x')$$

4) Trained features → MLCNN testing

After training, the learned parameters θ^* define the feature extractor and classifier. During testing, a new image follows the same preprocessing path and is passed through the network to produce probabilities $p\_\{\theta^*\}(y|x')$. The predicted outcome is the class with highest probability, along with a confidence score (the maximum softmax value).

5) Performance evaluation

We report standard metrics on the held-out test fold: accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices. Below are the two most common scalar measures used in the paper.

Accuracy (with TP, TN, FP, FN computed on the test set):

$$Acc = (TP + TN) / (TP + TN + FP + FN)$$

F1-score:

$$F1 = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall})$$

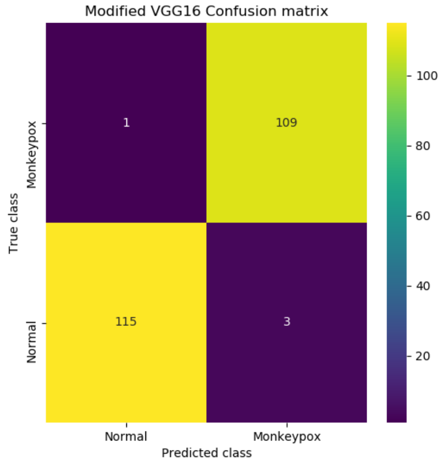
These metrics are computed once per experiment using the test branch; no training images are used for model selection or evaluation.

RESULTS AND DISCUSSION

Following this, the result achieved is shown in the Python platform. And the proposed method is contrasted with the state-of-the-art baselines on the same data set to guarantee the authenticity. Confusion matrix of conventional VGG16 and proposed ML-CNN is shown in Figure 2. A substantially higher true positive value—along with less false negatives—can be visually identified in the ML-CNN, which demonstrates a notably larger aggregation of the counts along the main diagonal. Thistrade is clinically relevant, as more sensitive detection of monkeypox lesions lessens the possibility of overlooked cases. Meanwhile, the ML-CNN sustains robust true negative rates, suggesting that specificity is not compromised to enhance sensitivity.

The results of Table 1 provide an overview of the quantitative comparison. The proposed ML-CNN significantly outperforms the modified VGG16 baseline in all the mentioned metrics, i.e., Accuracy, Precision, Recall, and F1-Score, among others. A higher Precision means that we have less false alarms among the positive predictions; a higher Recall means that we can detect more true monkeypox, these improvements give us a better F1-Score, and that means we have a better balance between sensitivity and specificity. The improvement in Accuracy follows these trends in the micro-level, indicating that the benefits are equally shared by both classes (monkeypox and non-monkeypox).

For better visualization, the metrics listed in Table 1 are plotted in Figure 3. The proposed model still outperforms the baseline all the bars above in the four evaluation indicators, which further confirms the results listed in the table at a glance. Finally, Figure 4 presents exemplar predictions of the ML-CNN, for both class-accurate (correct) classifications in difficult (such as different illumination, scale, and cluttered background) and in easy situations. Overall, the confusion matrices, summary metrics, and qualitative predictions suggest that the proposed ML-CNN offers more dependable and consistent results compared to VGG16 for this dataset.



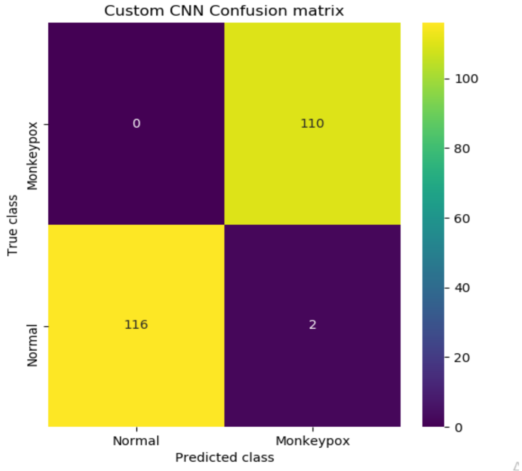


Fig. 2: Confusion matrixes of VGG16 and CNN.

TABLE.1: PERFORMANCE COMPARISON.

Method	Accuracy	Precision	Recall	FSCORE
Modified VGG16	98.2	98.2	98.2	98.2
Proposed ML-CNN	99.1	99.1	99.1	99.1

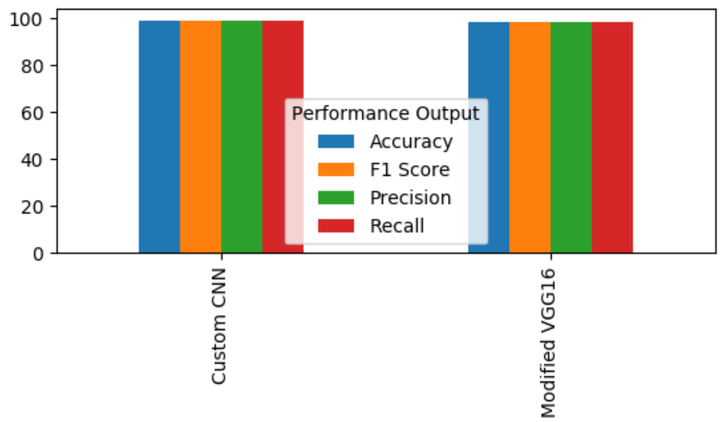


Fig. 3. Graphical representation of performance metrics.


Classification Result: Monkeypox Detected


Classification Result: Monkeypox Detected


Classification Result: Monkeypox Detected

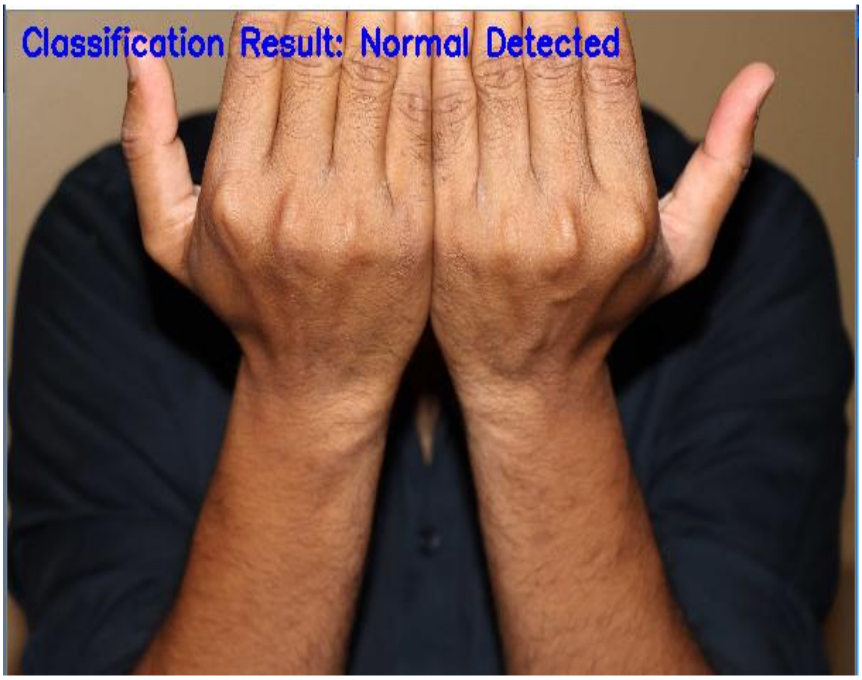


Fig. 4: Classified results as proposed method

CONCLUSION

This study addresses the persistent scarcity of monkeypox skin-lesion imagery by assembling an **open, publicly available dataset** sourced from permissive, open-access repositories. The intent is to remove practical barriers so researchers can **share, reuse, and benchmark** models for both experimental and—where licenses permit—commercial applications. Building on this resource, we conducted two feasibility studies—one on a **small** cohort and another on a **moderate** cohort—using a **modified ML-CNN** under a transfer-learning regime. Across both settings, the proposed model discriminated patients with **monkeypox** from **non-monkeypox** cases with accuracies ranging from **78% to 97%**, indicating that transfer learning is effective even when curated data remain limited.

To support clinical translation, we coupled the classifier with **explainability** tools (e.g., model-agnostic or saliency-based visualizations) to elucidate the image regions driving each prediction—an increasingly important requirement for trials and real-world deployment. Predicted outcomes were **reviewed by clinicians**, providing an initial external check on plausibility and face validity. More broadly, our goal is to highlight the potential of **AI-assisted diagnostics** to accelerate triage and mitigate transmission during emerging outbreaks. By releasing the dataset, we aim to empower researchers who currently lack sufficient data to prototype, train, and evaluate models, and to encourage the community to **leverage transfer learning**—a strategy repeatedly validated in prior literature for data-constrained medical imaging tasks.

We recognize several limitations and outline immediate next steps: (i) **continually expand** the dataset with newly collected, de-identified monkeypox images; (ii) assess robustness under **class imbalance** and heterogeneous capture conditions; (iii) **benchmark** against third-party models as they become available; and (iv) explore deployment in a **mobile, point-of-care** application. These efforts should improve generalization, comparability, and practical utility, moving the field closer to reliable, clinician-aligned monkeypox recognition tools.

REFERENCES

[1] Sitaula, Chiranjibi, and Tej Bahadur Shahi. "Monkeypox virus detection using pre-trained deep learning-based approaches." arXiv preprint arXiv:2209.04444 (2022).

[2] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. arXiv preprint arXiv:2003.11597, 2020.

[3] Luna-Perej´on, F. et al. (2020). Low-power embedded system for gait classification using neural networks. Journal of Low Power Electronics and Applications, 10 , 14.

[4] Ali, S. N., Ahmed, M., Paul, J., Jahan, T., Sani, S., Noor, N., Hasan, T. et al. (2022). Monkeypox skin lesion

detection using deep learning models: A feasibility study. ArXiv preprint arXiv:2207.03342.

[5] Ahsan, M. M., Uddin, M. R., & Luna, S. A. (2022b). Monkeypox image data collection. arXiv preprint arXiv:2206.01774.

[6] Phi-Yen Nguyen, Whenayon Simeon Ajisegiri, Valentina Costantino, Abrar A Chughtai, and C Raina MacIntyre. Reemergence of human monkeypox and declining population immunity in the context of urbanization, nigeria, 2017–2020. Emerging Infectious Diseases, 27(4):1007, 2021.

[7] Md Manjurul Ahsan, Redwan Nazim, Zahed Siddique, and Pedro Huebner. Detection of covid-19 patients from ct scan and chest x-ray data using modified mobilenetv2 and lime. In Healthcare, volume 9, page 1099. Multidisciplinary Digital Publishing Institute, 2021.

[8] Tim Menzies, Jeremy Greenwald, and Art Frank. Data mining static code attributes to learn defect predictors. IEEE transactions on software engineering, 33(1):2–13, 2006.

[9] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports, 10(1):1–12, 2020.

[10] Pan Pan, Yichao Li, Yongjiu Xiao, Bingchao Han, Longxiang Su, Mingliang Su, Yansheng Li, Siqi Zhang, Dapeng Jiang, Xia Chen, et al. Prognostic assessment of covid-19 in the intensive care unit by machine learning methods: model development and validation. Journal of medical Internet research, 22(11):e23128, 2020.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).

[12] Thomas, S. M. et al. (2021). Interpretable deep learning systems for multiclass segmentation and classification of non-melanoma skin cancer. Medical Image Analysis, 68 , 101915.

[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014. [9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, 2016, pp. 770–778.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE CVPR, 2016, pp. 770–778.

[15] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” J. of Big Data, vol. 6, no. 1, pp. 1–48, 2019.

[16] K. M. Hosny, M. A. Kassem, and M. M. Foaud, “Classification of skin lesions using transfer learning and augmentation with Alex-net,” PloSone, vol. 14, no. 5, p. e0217293, 2019