Elsevier has created a [Monkeypox Information Center](#) in response to the declared public health emergency of international concern, with free information in English on the monkeypox virus. The Monkeypox Information Center is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its monkeypox related research that is available on the Monkeypox Information Center - including this research content - immediately available in publicly funded repositories, with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the Monkeypox Information Center remains active.
To the Editor

The extent of cutaneous involvement is a key aspect of diagnosing and monitoring monkeypox disease, which is considered the most important orthopoxvirus in humans (Sklenovska et al., 2018). The spread of monkeypox cases in Europe and North America in May 2022 raised global public health concerns (Muyembe-Tamfum, 2022), leading to the World Health Organization declaring a public health emergency on July 23, 2022 (World Health Organization, 2022).

Monkeypox affects the skin in >99% of cases (Pittman et al., 2022) with substantial morbidity. Current World Health Organization guidelines assign severity according to the number of skin lesions: mild (<25 skin lesions), moderate (25–99 skin lesions), severe (100–250 skin lesions), or grave (>250 skin lesions) (Muyembe-Tamfum, 2022) (Supplementary Figure S1). Lesion counts are also a key parameter in monkeypox therapeutic trials. For example, the PALM007 randomized controlled trial of tecovirimat versus placebo requires counting lesions daily until resolution or day 28 (Nussenblatt, 2022). Counting skin lesions manually is labor intensive and presents logistical challenges, especially in remote regions prone to monkeypox outbreaks. We sought to develop an artificial intelligence (AI) algorithm to count monkeypox lesions in patient photographs. We hypothesized that the AI would count lesions in close agreement with manual counts.

We developed and tested the AI with a convenient series of photographs from an observational study, collected at the remote General Reference Hospital of Kole (Kole Hospital, Kole, Democratic Republic of the Congo) and the surrounding rainforest of the Congo River basin of the Democratic Republic of the Congo. The observational study was a joint venture of the Institut National de Recherche Biomédicale and the United States Army Medical Research Institute of Infectious Diseases, approved by the Human Use Committee at the United States Army Medical Research and Development Command Institutional Review Board; and the Ethics Committee at the University of Kinshasa School of Public Health (Kinshasa, Democratic Republic of the Congo). Initial clinical results and study population characteristics have been reported elsewhere (Mbala et al., 2017; Pittman et al., 2022). All patients provided written informed consent and were confirmed to have monkeypox virus infection by PCR.

Nonidentifiable photographs were transferred to Vanderbilt University Medical Center (Nashville, TN) for use under local institutional review board approval (191042). From this set, all images amenable to unambiguous human counting were used for AI training and testing. Photographs where counting in the field would not be performed (e.g., owing to large confluent lesions or secondary infections) or where image quality prevented reasonable manual assessment (e.g., owing to motion artifacts) were not used. The photograph set for analysis consisted of 66 photographs (median = 3.5, interquartile range = 2–4 photographs per patient) from 18 patients (Supplementary Figure S2). All patients were estimated as Fitzpatrick skin type VI by a board-certified dermatologist (ERT).

Two types of manual annotations were collected for each photograph. First, rater 1 provided segmentation masks for AI training, where every pixel in the photograph was manually labeled as lesion or nonlesion. Second, manual lesion counts were collected for each photograph by three human raters (raters 1–3) separately. Manual lesion counts were collected prospectively on unannotated photographs (details are provided in Supplementary Materials and Methods) without the raters knowing the AI outputs. We consider the lesion counts by rater 1 as the ground truth given the greater familiarity and annotating experience with this dataset. This reference standard was selected because clinical adjudication in prospective clinical trials will likely be based on manual counts from photographs of enrolled patients.

To identify and count lesions, we adopted a segmentation approach whereby every pixel in each photograph is classified as belonging to a monkeypox lesion or not. Our AI is based on the ubiquitous U-Net deep learning architecture (Ronneberger et al., 2015) with an Inception-v4 encoder (Szegedy et al., 2017). Prediction models were developed for each of the 18 patients in a leave-one-out experiment. For each model, lesion prediction maps were created for all photographs of the patient not seen during training. Lesion counts were estimated by the number of non-touching lesional areas in the prediction.
The primary clinical metric of interest was the lesion count performance, evaluated prospectively by comparing the predicted number of lesions for a given photograph with the ground truth number from rater 1. Simple linear regression and limits of agreement (Bland–Altman) analysis were used to compare counts for each photograph. The width of 95% confidence limits of agreement in this analysis is approximately four times the SD of the difference between predicted lesions and ground truth.

Figure 1a shows a representative image from Kole with manually identified lesions and AI output. Segmentation performance by the traditional computer vision metric of Dice index is shown in Supplementary Figure S3. Performance in counting lesions by correlation and limits of agreement analysis is shown in Supplementary Figure S2. Relative to the mean bias of the ground truth counts (by rater 1), the AI had a mean bias of −5.86 (limits of agreement width = 68.85) lesions. For the remaining two human raters, the bias from ground truth was −3.24 (38.44) for rater 2, 9.68 (76.74) for rater 3, and 12.92 (81.91) between raters 3 and 2 (Figure 2a-c and Supplementary Table S1). To show potential generalizability, we also applied the AI to publicly available images of monkeypox (Figure 1b).

Despite the small training dataset, our AI performed at a comparable level with human raters counting monkeypox lesions. Because monkeypox skin lesion counts are an important measure to stage and monitor disease severity, this approach could be used as a practical support tool in monkeypox trials that are imminently launching.

A limitation of our study is the presence of a single skin type (Fitzpatrick type VI), which may hamper direct application in other skin types. Our set also lacked images of anogenital or perineal skin, which is an important emerging disease site in the European and North American outbreaks (Patel et al., 2022; Thornhill et al., 2022). Practical protocols to capture standardized, high-quality photographs of large body regions in resource-limited regions will be a critical next step for AI image analysis to support monkeypox research. Classifying lesion types may also enable more advanced differential diagnosis and monitoring and objective confirmation of endpoints in monkeypox trials.

Our cross-validation study of 18 patients with monkeypox provides proof of principle for AI algorithms to provide reliable lesion identification and counting from photographs of patients with monkeypox. Ultimately, this could become a globally scalable solution to diagnose, stage, and monitor the disease.

maps (details are provided in Supplementary Materials and Methods).

The primary clinical metric of interest was the lesion count performance, evaluated prospectively by comparing the predicted number of lesions for a given photograph with the ground truth number from rater 1. Simple linear regression and limits of agreement (Bland–Altman) analysis were used to compare counts for each photograph. The width of 95% confidence limits of agreement in this analysis is approximately four times the SD of the difference between predicted lesions and ground truth.
Figure 2. Comparison of lesion count performance by AI and human raters. LoAs (shown with dashed lines) are the boundaries within which 95% of future measurement differences are expected to fall. LoA width = upper LoA – lower LoA. We also show the slope and coefficient of determination (R²) for the linear regression fit (red dashed line) between estimated counts for each pair. The solid black line is the line of agreement. (a) Bland–Altman and correlation plots for the AI against the ground truth (human rater 1). (b) Rater 2 against ground truth. (c) Rater 3 against ground truth. AI, artificial intelligence; LoA, limit of agreement.

Data availability statement

No public dataset is available owing to the limited size of the study. Data and analyses are available on reasonable request to the corresponding author.

ORCIDs

Andrew J. McNeil: http://orcid.org/0000-0003-1496-9691
David W. House: http://orcid.org/0000-0003-1650-0088
Placide Mbala-Kingebeni: http://orcid.org/0000-0003-1556-3570
Olivier Tshiani Mbaya: http://orcid.org/0000-0003-1908-5190
Lori E. Dodd: http://orcid.org/0000-0002-3433-5429
Edward W. Cowen: http://orcid.org/0000-0003-1918-4324
Veronique Nussenblatt: http://orcid.org/0000-0002-9510-0131
Tyler Bonnett: http://orcid.org/0000-0001-5545-6433
Ziche Chen: http://orcid.org/0000-0003-4005-9261

The authors state no conflict of interest.

Acknowledgments

This work was supported by a Career Development Award from the United States Department of Veterans Affairs Clinical Sciences R&D Service (IK2 CX001785); a Discovery Research Grant from Vanderbilt University, the Vanderbilt University Medical Center Departments of Medicine and Dermatology; the National Institutes of Health (K12 CA090625, R21 AR074589; and the European Regional Development Fund (1.1.1.2/ VIAA/4/20/665). This project has been funded in part with federal funds from the National Cancer Institute of National Institutes of Health (under contract number 75N91019D00024) and the National Institutes of Health/Division of Intramural Research and the National Institute of Allergy and Infectious Diseases. We also thank Sahil Patel for creating the initial protocol for marking monkeypox lesions in photographs using the GIMP (GNU Image Manipulation Program software).

Author Contributions

Conceptualization: AJM, ERT, BMD, LED, PMK, OTM, IS; Data Curation: AJM, DWH, ZC, ERT, EWC, PMK; Formal Analysis: AJM, ERT, BMD, IS; TB; Funding Acquisition: ERT, BMD, OTM, LED; VN; Investigation: AJM, PMK, IS, TB, ERT; Methodology: AJM, ERT, BMD; Project Administration: LED, PMK, OTM, ERT; Resources: PMK, OTM, LED, BMD; Software: AJM, IS; Supervision: ERT, BMD, LED, EWC; Validation: AJM, IS; Visualization: AJM, ERT, BMD, IS; Writing – Original Draft Preparation: AJM, ERT, BMD; IS; Writing – Review and Editing: AJM, DWH, PMK, OTM, LED, EWC, VN, TB, ZC, IS, BMD, ERT

Disclaimer

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Andrew J. McNeil1,2,3, David W. House1,2, Placide Mbala-Kingebeni1, Olivier Tshiani Mbaya4,5, Lori E. Dodd6, Edward W. Cowen7, Veronique Nussenblatt8, Tyler Bonnett5, Ziche Chen1,2, Inga Saknite9, Benoit M. Dawant10 and Eric R. Tkaczyk1,2,10,*

1Dermatology Service and Research Service, Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, Tennessee, USA; 2Department of Dermatology, Vanderbilt University Medical Center, Nashville, Tennessee, USA; 3Department of Electrical and Computer Engineering, School of Engineering, Vanderbilt University, Nashville, Tennessee, USA; 4Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo; 5Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA; 6Clinical Trials Research Section, Division of Clinical Research, National Institute of Allergy and Infectious Disease, Bethesda, Maryland, USA; 7Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland, USA; 8Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA; 9Biophotonics Laboratory, Institute of Atomic Physics and Spectroscopy, University of Latvia, Riga, Latvia; and 10Department of Biomedical Engineering, School of Engineering, Vanderbilt University, Nashville, Tennessee, USA

*Corresponding author e-mail: eric.tkaczyk@vumc.org

Supplementary Material

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2022.08.044
**A.J. McNeil et al.**

**Counting Monkeypox Lesions Using AI**

**REFERENCES**

Centers for Disease Control and Prevention. Public Health Image Library, https://phil.cdc.gov/Details.aspx?pid=12763; 1997 (accessed 3 August 2022).

Mbala PK, Huggins JW, Riu-Rovira T, Ahuka SM, Mulembakani P, Rimoin AW, et al. Maternal and fetal outcomes among pregnant women with human monkeypox infection in the Democratic Republic of Congo. J Infect Dis 2017;216:824–8.

Muyembe-Tamfum JJ. Clinical aspects of monkeypox in DRC. Monkeypox epidemiology, surveillance, and laboratory capacity in DRC, https://cdn.who.int/media/docs/default-source/blue-print/day-1_jean-jacques-muyembe_drc_monkeypox-meeting_02june2022.pdf?sfvrsn=d133676a_3; 2022 (accessed 2 June 2022).

Nussenblatt V. A randomized, placebo-controlled trial of the safety and efficacy of Tecovirimat for the treatment of patients with monkeypox virus PALM007, https://cdn.who.int/media/docs/default-source/blue-print/day-2_veronique-nussenblatt_vaccine-protocol_monkeypox-meeting_03june2022.pdf?sfvrsn=bf68fc01_3; 2022 (accessed June 3 2022).

Patel A, Bilinska J, Tam JCH, Da Silva Fontoura D, Mason CY, Daunt A, et al. Clinical features and novel presentations of human monkeypox in a central London centre during the 2022 outbreak: descriptive case series. BMJ 2022;378:e072410.

Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lect. Notes Comput. Sci 2015;9351(Cvd):12–20.

Sklenovska N, Van Ranst M, Sklenovská N, Van Ranst M. Emergence of Monkeypox as the most important orthopoxvirus infection in humans. Front Public Health 2018;6:241.

Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. 31st AAAI Conf. Artif. Intell. AAAI 2017. 2017;42(1):4278–4284.

Thornhill JP, Barkati S, Walmsley S, Rockstroh J, Antinori A, Harrison LB, et al. Monkeypox virus infection in humans across 16 countries - April–June 2022. N Engl J Med 2022;387:679–91.

World Health Organization. Monkeypox, https://www.who.int/health-topics/monkeypox; 2022 (accessed 3 August 2022).

World Health Organization. WHO Director-General’s statement at the press conference following IHR Emergency Committee regarding the multi-country outbreak of monkeypox - 23 July 2022, https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-the-press-conference-following-ihr-emergency-committee-regarding-the-multi-country-outbreak-of-monkeypox-23-july-2022; 2022 (accessed 3 August 2022).
SUPPLEMENTARY MATERIALS AND METHODS

Photograph acquisition

Photographs were collected with consumer-grade cameras before 2011 and were more recently collected with smartphone cameras. No standardized imaging protocol was followed, leading to a variety of lighting conditions, backgrounds, fields of view, imaging distances, body sites, and time points since the first symptoms. Occasional images had pen markings used to subdivide larger skin areas or track individual lesions during manual counts.

Photograph set characteristics

A total of 406 photographs from 35 patients were transmitted to Vanderbilt. A total of 340 photographs from 17 patients were excluded from the analysis because the lesions would not have been counted in the field for World Health Organization severity scoring. The specific issues precluding clinical scoring were image of scalp/eye/intraoral cavity (194 photographs), extensive confluent lesions not amenable to unambiguous human counting (24 photographs), poor image quality (e.g., motion artifact) (52 photographs), and duplicate images (70 photographs). The final set comprised 66 photographs (median = 3.5, interquartile range = 2–4 photographs per patient) (Supplementary Figure S2) from 18 patients (10 male, 3 female, 5 of unknown sex, 3 infants, 8 children, 6 adolescents, 1 adult). Fifteen of these patients had photographs between 2007 and 2011, and three had in January 2022. Fourteen photographs (from six patients) contained pen markings adjacent to lesions.

Photograph annotations

Manual segmentation masks were created for all photos using the free, open-source GIMP (GNU Image Manipulation Program) (GIMP, 2021). Rater 1 followed a predefined protocol (eProtocol 1, online only). All visible lesions were manually traced on a transparent annotation layer using the pencil tool. Lesions from all stages were demarcated in the same manner, marking the full extent of each lesion with the boundaries drawn to the edge of affected and normal-appearing skin. This required approximately 20–60 minutes per photograph, depending on the severity and number of lesions. Once completed, the annotation layer of each image was exported to create a binary segmentation mask of lesion pixels. These segmentation masks were used to train the artificial intelligence (AI) algorithm.

Manual lesion counts were also collected manually for all photos, following a similar protocol using the GIMP software (eProtocol 2, online only). The pencil tool used to mark a single spot at the center of each visible lesion on a transparent annotation layer. Touching and coalesced lesions were marked separately if defined structures could still be discerned. A pencil diameter of 2–5 pixels was used to ensure that each marking did not overlap even for small adjacent lesions. Each photo was assessed by the same annotator who had provided segmentation ground truth (rater 1), in addition to two other human raters (raters 2 and 3). This lesion counting process required <10 minutes per photograph.

Document segmentation masks were created using 256/C2 from 18 patients (median = 3.5, interquartile range = 2–4 photographs per patient) (Supplementary Figure S2) from 18 patients (10 male, 3 female, 5 of unknown sex, 3 infants, 8 children, 6 adolescents, 1 adult). Fifteen of these patients had photographs between 2007 and 2011, and three had in January 2022. Fourteen photographs (from six patients) contained pen markings adjacent to lesions.

Algorithm training

We report the AI development following the CLEAR Derm consensus guidelines (Danesjhou et al., 2022). We used a patient-level leave-one-out experiment to evaluate the AI performance, whereby a model was trained using all photos of 17 patients and then tested on the unseen held-out patient. Model training was performed on consumer-grade hardware (Nvidia RTX 2080 and RTX A4000 graphics processing units), with training times of approximately 8 hours per model. For each model, the training set was constructed using 256 × 256 resolution patches extracted from all photos of the 17 patients. To ensure equal contribution to the training set for each patient, regardless of the number of photos, 1,000 patches were extracted from each patient and distributed equally across all available photos. Photos were not taken in a clinical setting, leading to a wide variety of environments and backgrounds. The patch locations were therefore distributed using 80% centered on a lesion pixel (determined from the ground truth mask) and 20% randomly sampled from nonlesional areas, including the background.

Each model was trained for 40 epochs with binary cross entropy loss using the Adam optimizer (Kingma and Ba, 20142), with an initial learning rate of 0.0001, reduced to 0.00001 after 30 epochs. The best weights for each model were selected from the highest Dice value on the validation set. The validation set for each model was split from the training set as a random 10% subset of patches. During training, data augmentation was applied at each iteration through a random combination of elastic transformations, left/right flipping, gaussian blur, affine scaling and translation, perspective transforms, color temperature adjustments, and gamma adjustments.

Model testing was carried out using a sliding window of size 256 × 256 and a stride of 32. The segmentation map was reconstructed by summing the predictions of all patches for a given pixel, with the final segmentation mask calculated by thresholding the segmentation map at the optimal level determined from all photos in the training set. Lesion counts were estimated by a connected-component analysis of the segmentation mask, where each discrete (nontouching) lesional area in the mask is counted as a single lesion.

Data preprocessing

Historical photos were collected with consumer-grade cameras - Samsung Digimax L70 (Seoul, South Korea) and Canon Powershot A630 (Tokyo, Japan) - and resized to a resolution of 1,024 × 768 pixels for archival before transfer. Additionally, three patients were photographed with smartphones during the most recent outbreak. Any photos of faces were deidentified by placing black boxes across the eyes, central forehead, and periorbital region before any processing, training, or analysis was performed.

AI algorithm

We selected the ubiquitous U-Net deep learning architecture (Ronneberger et al., 2015). This network uses a symmetric encoder–decoder structure with a contracting path that captures context...
and expanding path, which enables precise localization. Long skip connections are used to concatenate the upsampled feature map in the expansive path with the corresponding feature map from the contracting path. Our algorithm uses an Inception-v4 network as the encoder (Szegedy et al., 2017), initialized with ImageNet weights, following the implementation in Segmentation Models Pytorch (Yakubovskiy, 2021).

**Algorithm performance assessment**

To ensure prospective evaluation, the collection of the AI algorithm output data was planned before the application of the performance metrics (reference standards). The performance of each model was evaluated on all skin in each image (excluding backgrounds such as clothing and buildings) for the held-out patient. This was done by manually masking off background pixels in black before analysis, retaining only the areas of the photo containing the patient’s skin.

Segmentation performance of the AI algorithm was done by the Dice coefficient, a widely used computer vision metric of spatial overlap (Zijdenbos et al., 1994). In photographs of patients not used in training, the AI algorithm achieved a median Dice index of 0.72 (interquartile range = 0.62–0.74), where 0 represents no agreement, and 1 represents perfect agreement.

No significant difference in performance was found between photographs with and without pen marks by Wilcoxon rank sum test ($P > 0.5$).

**SUPPLEMENTARY REFERENCES**

Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, et al. Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR Derm consensus guidelines from the international skin imaging collaboration artificial intelligence working group. JAMA Dermatol 2022;158:90–6.

GIMP. GNU Image Manipulation Program, https://www.gimp.org; 2021 (accessed 10 December 2021).

Muyembe-Tamfum JJ. Clinical aspects of monkeypox in DRC. Monkeypox epidemiology, surveillance, and laboratory capacity in DRC, https://cdn.who.int/media/docs/default-source/blue-print/day-1_jean-jacques-muyembe_drc_monkeypox-meeting_02_june2022.pdf?sfvrsn=d133676a_3; 2022 (accessed 2 June 2022).

Yakubovskiy P. qubvel/segmentation models. pytorch, https://github.com/qubvel/segmentation_models.pytorch; 2021 (accessed 2 December 2021).

Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging 1994;13:716–24.
CLINICAL SEVERITY SCORE OF MPX BASED ON NUMBER/LESIONS (WHO)

- Mild illness (<25 skin lesions), no disability.
- (ii) Moderate illness (25-99 lesions), unable to perform most physical activities but does not require nursing cares.
- (iii) Severe illness (100-250 skin lesions), unable to perform most physical activities and requires nursing cares.
- (iv) Grave illness (>250 skin lesions), unable to perform most physical activities and requires intensive nursing cares.

Supplementary Figure S1. WHO monkeypox infection severity guidelines, from “Clinical Aspects of Monkeypox in the DRC” by Muyembe-Tamfum (2022). DRC, Democratic Republic of the Congo; WHO, World Health Organization.

Supplementary Figure S2. Dataset characteristics. The number of photographs per patient (n = 66, N = 18) and histogram of lesion counts per photo by rater 1 (ground truth). ID, identification.
Supplementary Figure S3. AI algorithm segmentation performance by Dice index. Each point represents a single photo (n = 66). Boxplot shows the median (0.72), interquartile range (0.62–0.74), and mean (0.67, dashed red line). AI, artificial intelligence.

**Supplementary Table S1. Summary of Pairwise Comparisons between Different Human Raters and AI Algorithm**

| Rater Pair    | Bias | Upper LoA | Lower LoA | LoA Width | Slope | $R^2$ |
|---------------|------|-----------|-----------|-----------|-------|-------|
| AI versus 1   | -5.86| 28.56     | -40.29    | 68.85     | 0.78  | 0.94  |
| 2 versus 1    | -3.24| 15.98     | -22.46    | 38.44     | 1.02  | 0.97  |
| 3 versus 1    | 9.68 | 48.05     | -28.69    | 76.74     | 1.07  | 0.92  |
| 3 versus 2    | 12.92| 53.88     | -28.03    | 81.91     | 1.03  | 0.90  |

Abbreviations: AI, artificial intelligence; LoA, limit of agreement.