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.

COUNTING MONKEYPOX LESIONS IN PATIENT PHOTOGRAPHS: LIMITS OF AGREEMENT OF MANUAL COUNTS AND ARTIFICIAL INTELLIGENCE

Andrew J. McNeil, PhD, David W. House, BS, Placide Mbala-Kingebeni, MD, Olivier Tshiani Mbaya, MD, Lori E. Dodd, PhD, Edward W. Cowen, MD, M.H.Sc., Véronique Nussenblatt, MD, Tyler Bonnett, MS, Ziche Chen, BS, Inga Saknite, PhD, Benoit M. Dawant, PhD, Eric R. Tkaczyk, MD, PhD

PII: S0022-202X(22)01897-8

DOI: https://doi.org/10.1016/j.jid.2022.08.044

Reference: JID 3550

To appear in: *The Journal of Investigative Dermatology*

Received Date: 5 August 2022

Revised Date: 19 August 2022

Accepted Date: 23 August 2022

Please cite this article as: McNeil AJ, House DW, Mbala-Kingebeni P, Mbaya OT, Dodd LE, Cowen EW, Nussenblatt V, Bonnett T, Chen Z, Saknite I, Dawant BM, Tkaczyk ER, COUNTING MONKEYPOX LESIONS IN PATIENT PHOTOGRAPHS: LIMITS OF AGREEMENT OF MANUAL COUNTS AND ARTIFICIAL INTELLIGENCE, *The Journal of Investigative Dermatology* (2022), doi: https://doi.org/10.1016/j.jid.2022.08.044.

COUNTING MONKEYPOX LESIONS IN PATIENT PHOTOGRAPHS: LIMITS OF AGREEMENT OF MANUAL COUNTS AND ARTIFICIAL INTELLIGENCE

Andrew J. McNeil, PhD[1,2,3], David W. House, BS[1,2], Placide Mbala-Kingebeni, MD[4], Olivier Tshiani Mbaya, MD[4,5], Lori E. Dodd, PhD[6], Edward W. Cowen, MD, M.H.Sc.[7], Véronique Nussenblatt, MD[8], Tyler Bonnett, MS[5], Ziche Chen, BS[1,2], Inga Saknite, PhD[2,9], Benoit M. Dawant, PhD[3,10], Eric R. Tkaczyk, MD, PhD[1,2,10]

[1]Dermatology Service and Research Service, Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, TN, USA

[2]Department of Dermatology, Vanderbilt University Medical Center, Nashville, TN, USA

[3]Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA

[4]Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo

[5]Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[6]Clinical Trials Research Section, Division of Clinical Research, National Institute of Allergy and Infectious Disease, Bethesda, MD, USA

[7]Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), Bethesda, MD, USA

[8]Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA

[9]Biophotonics Laboratory, Institute of Atomic Physics and Spectroscopy, University of Latvia, Riga, Latvia

[10]Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

Corresponding author:

Eric R. Tkaczyk, MD, PhD

719 Thompson Lane, One Hundred Oaks

Suite 26300, Department of Dermatology

Nashville, TN 37204

Email: eric.tkaczyk@vumc.org

Twitter handle: @vdtrc

Tweet: Can AI recognise #Monkeypox? New study by McNeil et al in the #JIDJournal shows AI can count lesions on par with humans in patient photographs #dermtwitter #DermDataScience #vumc @vdtrc

ORCIDs

Andrew McNeil: 0000-0003-1496-9691

David W. House: 0000-0003-1650-0088

Placide Mbala-Kingebeni: 0000-0003-1556-3570

Olivier Tshiani Mbaya: 0000-0003-1908-5190

Lori E. Dodd: 0000-0002-3433-5429

Edward W. Cowen: 0000-0003-1918-4324

Véronique Nussenblatt: 0000-0002-9530-0131

Tyler Bonnett: 0000-0001-5545-6433

Ziche Chen: 0000-0003-4005-9261

Inga Saknite: 0000-0002-4000-5485

Benoit M. Dawant: 0000-0002-3804-8400

Eric R. Tkaczyk: 0000-0002-2850-4740

TO THE EDITOR

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

Monkeypox affects the skin in >99% of cases (Pittman et al. 2022) with substantial morbidity. Current WHO 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) (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 counts 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 with close agreement to manual counts.

We developed and tested the AI with a convenience series of photographs from an observational study, collected at the remote General Reference Hospital of Kole (Kole hospital) and the surrounding rainforest of the Congo River basin of the DRC. The observational study was a joint venture of the Institut National de Recherche Biomédicale and US Army Medical

Research Institute of Infectious Diseases (USAMRIID), approved by the Human Use Committee at the USAMRIID (FY05-13), the Headquarters, United States Army Medical Research and Development Command Institutional Review Board (IRB), and the Ethics Committee at the University of Kinshasa School of Public Health. 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.

Non-identifiable photographs were transferred to Vanderbilt University for use under local IRB approval (191042). From this set, all images amenable to unambiguous human counting were used for the AI training and testing. Photographs where counting in the field would not be performed (e.g., due to large confluent lesions or secondary infections), or where image quality prevented reasonable manual assessment (e.g., due to motion artifacts) were not used. The photograph set for analysis consisted of 66 photographs (median 3.5, interquartile range 2 to 4 photographs per patient) from eighteen patients (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 labelled as lesion or non-lesion. 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 in Supplement), without the raters knowing 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 since 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 maps (details in Supplement).

The primary clinical metric of interest was the lesion count performance, evaluated prospectively by comparing the predicted number of lesions for a given photograph to the ground truth number from rater 1. Simple linear regression and limits of agreement (LoA) (Bland-Altman) analysis were used to compare counts for each photograph. The width of 95% confidence LoA in this analysis is approximately four times the standard deviation 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 Figure S3. Performance in counting lesions by correlation and LoA analysis is shown in Figure 2. Relative to the ground truth counts (by rater 1), the AI had a mean bias of -5.86 (LoA 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 2 and Table S1). To demonstrate 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 to human

raters counting monkeypox lesions. As 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 trails.

Our cross-validation study of 18 monkeypox patients 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 disease.

DATA AVAILABILITY

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

CONFLICT OF INTEREST

The authors report no conflicts of interest.

ACKNOWLEDGEMENTS

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, National Institutes of Health, under Contract No. 75N91019D00024, and the NIH/Division of Intramural Research, NIAID. 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. We also thank Mr. Sahil Patel for creating the initial protocol for marking monkeypox lesions in photographs by using the GIMP 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: All authors

REFERENCES

Ghebreyesus TA. WHO Director-General's statement at the press conference following IHR Emergency Committee regarding the multi-country outbreak of monkeypox - 23 July 2022 [Internet]. 2022 [cited 2022 Aug 3]. Available from: 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

Mahy BWJ. CDC Public Health Image Library [Internet]. 1997 [cited 2022 Aug 3]. Available from: https://phil.cdc.gov/Details.aspx?pid=12763

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. Oxford University Press; 2017;216(7):824–8

Muyembe-Tamfum J-J. Clinical Aspects Of Monkeypox In DRC: Monkeypox epidemiology, surveillance, and laboratory capacity in DRC. WHO monkeypox Res. What are Knowl. gaps Prior. Res. Quest. 2022 [cited 2022 Jun 2]. Available from: https://cdn.who.int/media/docs/default-source/blue-print/day-1_jean-jacques-muyembe_drc_monkeypox-meeting_02june2022.pdf?sfvrsn=d133676a_3

NCDC. Nigeria Centre for Disease Control [Internet]. World Heal. Organ. Heal. Top. Monkeypox. 2022 [cited 2022 Aug 3]. Available from: https://www.who.int/health-topics/monkeypox

Nussenblatt V. Clinical trial to evaluate monkeypox treatments. WHO monkeypox Res. What are Knowl. gaps Prior. Res. Quest. 2022 [cited 2022 Jun 3]. Available from:

https://cdn.who.int/media/docs/default-source/blue-print/day-2_veronique-nussenblatt_vaccine-protocol_monkeypox-meeting_03june2022.pdf?sfvrsn=bf68fc01_3

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 Available from: http://www.ncbi.nlm.nih.gov/pubmed/35902115

Pittman PR, Martin JW, Kingebeni PM, Tamfum J-JM, Wan Q, Reynolds MG, et al. Clinical characterization of human monkeypox infections in the Democratic Republic of the Congo. medRxiv. Cold Spring Harbor Laboratory Press; 2022; Available from: https://www.medrxiv.org/content/early/2022/05/29/2022.05.26.22273379

Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 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 Heal. Frontiers Media SA; 2018;6 Available from: /pmc/articles/PMC6131633/

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):4278–84 Available from: http://arxiv.org/abs/1512.00567

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; Available from: http://www.ncbi.nlm.nih.gov/pubmed/35866746

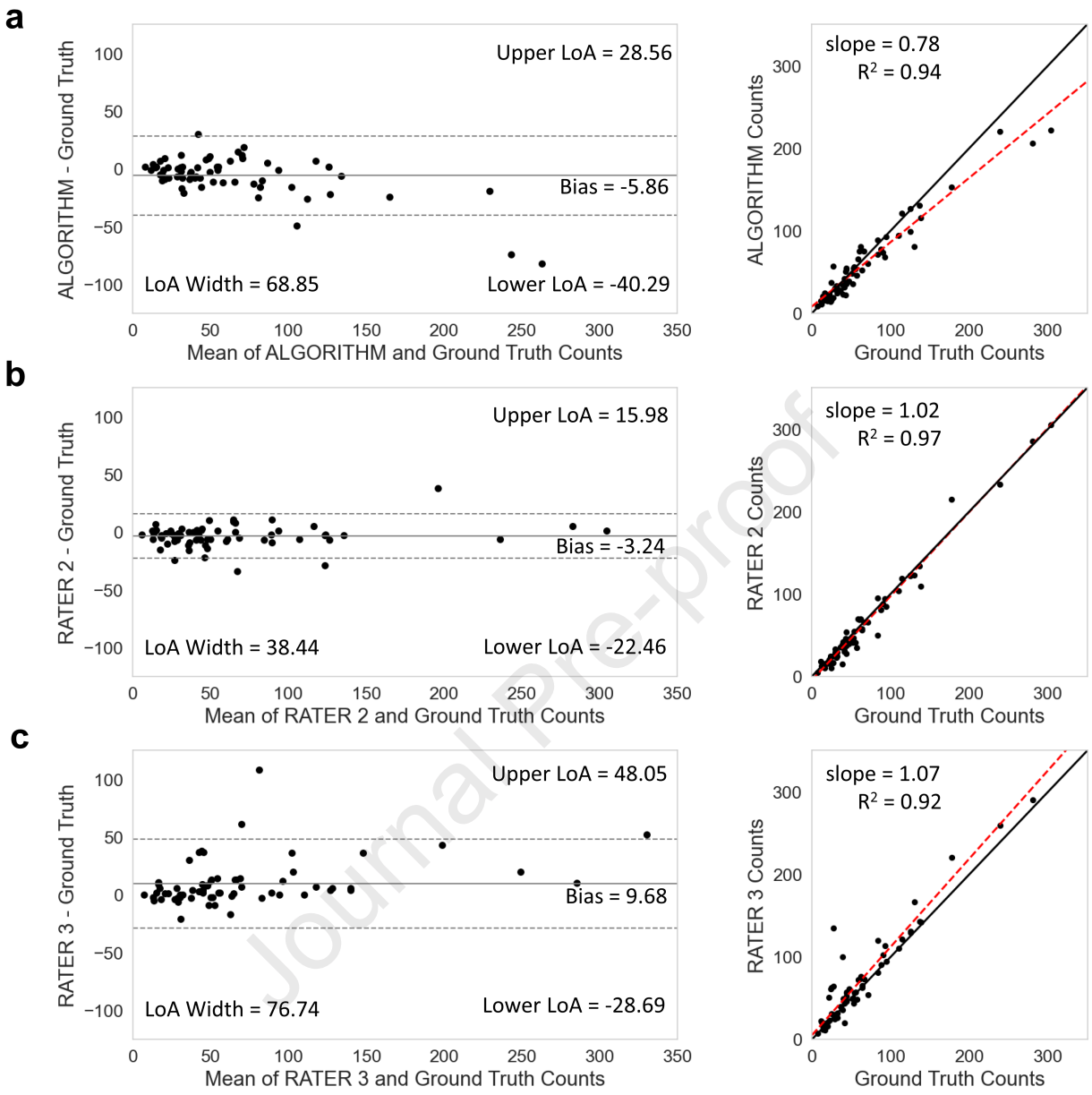
FIGURE LEGENDS

Figure 1. Representative monkeypox AI lesion predictions on photos of previously unseen patients. (a) Two patients from our photograph set. Green contours show true positive lesions, blue shows false positive lesions (outlined by the AI but not the human), and magenta shows false negative lesions (outlined by human but not the AI). Unmarked photos are on the left. Upper photo AI lesion counts: 220 lesions, manual counts from three human raters: 239 (rater 1), 233 (rater 2), and 259 (rater 3). Lower photo AI lesion count: 131. Manual counts: 137 (rater 1), 134 (rater 2), and 143 (rater 3). (b) Two patients from publicly available photographs. Predicted lesion contours by our AI model are shown in yellow. The AI model is the same used to test Patient ID 15 (N=17, n=61). Upper photo from the CDC Public Health Image Library (Mahy 1997). AI lesion counts: 58, manual counts by rater 1: 52. Lower photo from the Nigeria Centre for Disease Control, recently made available on WHO website (NCDC 2022), used with permission. AI lesion counts: 26, manual counts by rater 1: 29. Written informed consent was obtained for research and publication of photos from all patients.

Figure 2. Comparison of lesion count performance by AI and human raters. Limits of agreement (LoA, 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.

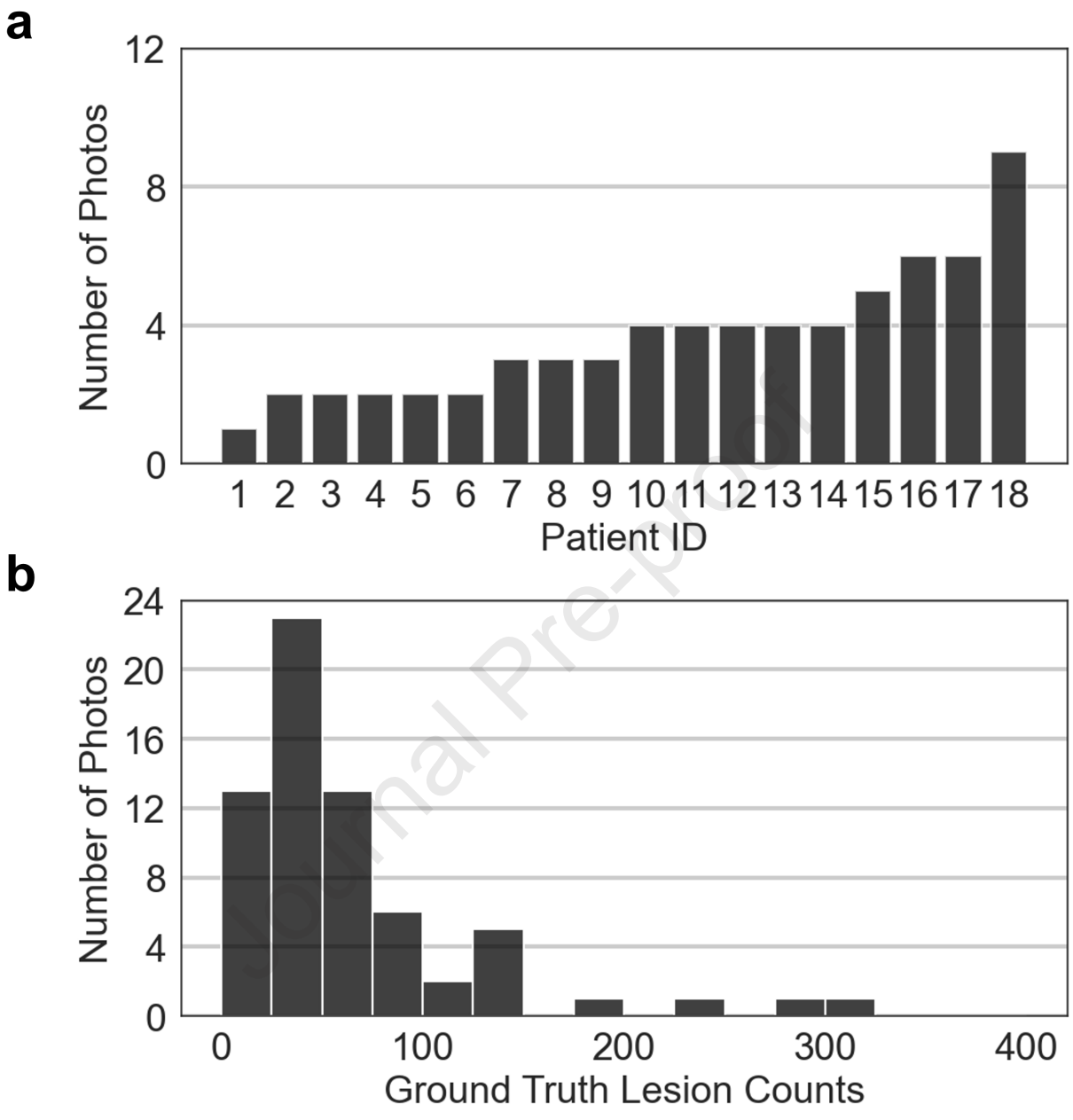
a

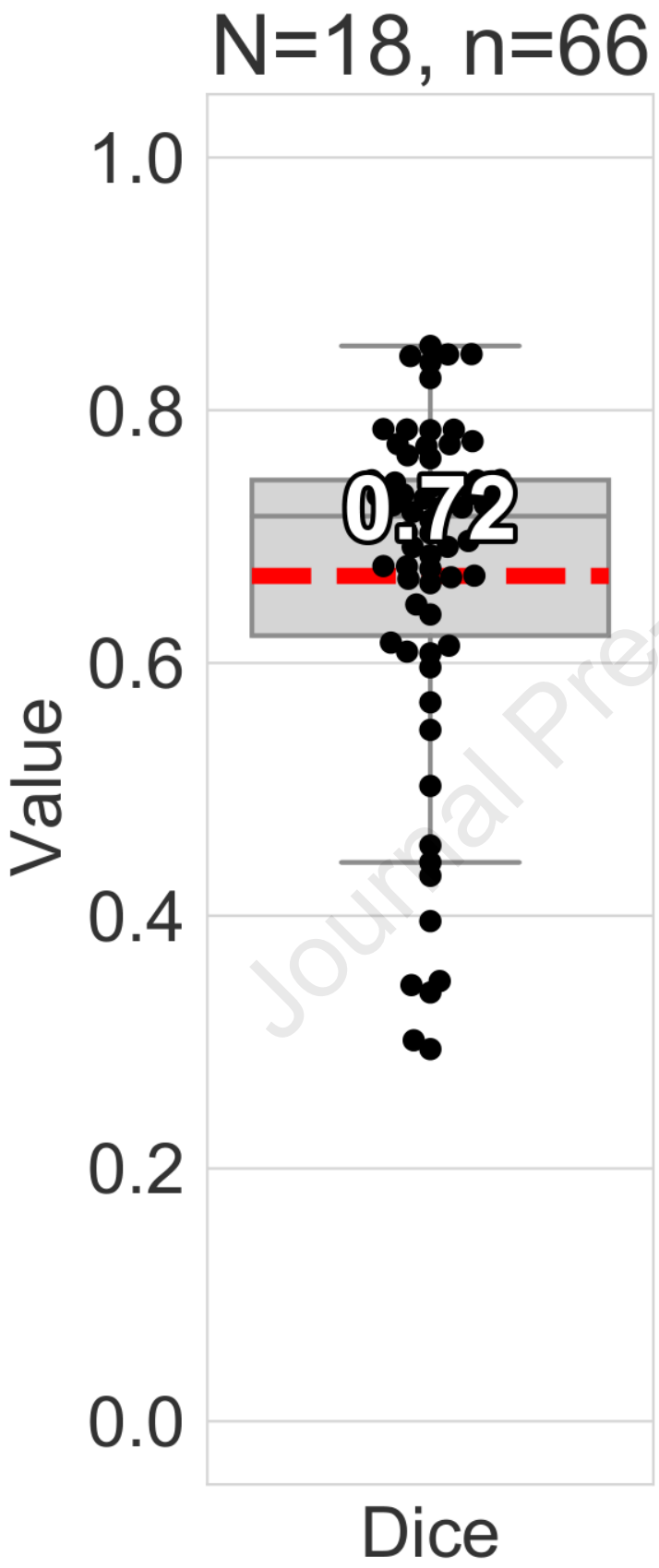
b



CLINICAL SEVERITY SCORE OF MPX BASED ON NUMBER/LESIONS (WHO)

- Mild illness(<25 skin lesions), no disability.
- (ii) Moderate illness (25-99 lésions), 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.





COUNTING MONKEYPOX LESIONS IN PATIENT PHOTOGRAPHS: LIMITS OF AGREEMENT OF MANUAL COUNTS AND ARTIFICIAL INTELLIGENCE

Andrew J. McNeil, PhD[1,2,3], David W. House, BS[1,2], Placide Mbala-Kingebeni, MD[4], Olivier Tshiani Mbaya, MD[4,5], Lori E. Dodd, PhD[6], Edward W. Cowen, MD, M.H.Sc.[7], Véronique Nussenblatt, MD[8], Tyler Bonnett, MS[5], Ziche Chen, BS[1,2], Inga Saknite, PhD[2,9], Benoit M. Dawant, PhD[3,10], Eric R. Tkaczyk, MD, PhD[1,2,10]

[1]Dermatology Service and Research Service, Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, TN, USA

[2]Vanderbilt Department of Dermatology, Vanderbilt University Medical Center, Nashville, TN, USA

[3]Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA

[4]Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo

[5]Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[6]Clinical Trials Research Section, Division of Clinical Research, National Institute of Allergy and Infectious Disease, Bethesda, MD, USA

[7]Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), Bethesda, MD, USA

[8]Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA

[9]Biophotonics Laboratory, Institute of Atomic Physics and Spectroscopy, University of Latvia, Riga, Latvia

[10]Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

Corresponding author:

Eric R. Tkaczyk, MD, PhD
719 Thompson Lane, One Hundred Oaks
Suite 26300, Department of Dermatology
Nashville, TN 37204
Email: eric.tkaczyk@vumc.org

SUPPLEMENTARY MATERIALS AND METHODS

Photograph acquisition

Photographs were collected with consumer-grade cameras prior to 2011 and, more recently, 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 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. 340 photographs from seventeen patients were excluded from analysis because the lesions would not have been counted in the field for WHO 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 of 66 photographs (median 3.5, interquartile range 2 to 4 photographs per patient, Figure S2) from eighteen patients (10 male; 3 female; 5 of unknown gender; 3 infants; 8 children; 6 adolescents; 1 adult). Fifteen of these patients had photographs between 2007 – 2011, and three in January 2022. Fourteen photographs (from 6 patients) contained pen markings adjacent to lesions.

Photograph Annotations

Manual segmentation masks were created for all photos using the free, open-source Gnu Image Manipulation Program (GIMP 2021). Rater 1 followed a predefined protocol (eProtocol 1). 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 AI algorithm.

Manual lesion counts were also collected manually for all photos, following a similar protocol using the GIMP software (eProtocol 2). 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 didn't 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 (2 and 3). This lesion counting process required less than 10 minutes per photograph.

No clinical outcomes or patient lesion counts were transferred to Vanderbilt for this study. All annotators were undergraduate or post-doctoral trainees of the Vanderbilt Dermatology Translational Research Clinic, with more than a year of research experience focused on dermatologic photograph analysis.

Algorithm Training

We report the AI development following the CLEAR Derm consensus guidelines (Daneshjou 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 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 256x256 resolution patches extracted from all photos of 17 patients. To ensure equal contribution to the training set for each patient, regardless of the number of photos, 1000 patches were extracted from each patient 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 non-lesional areas including background.

Each model was trained for 40 epochs with binary cross entropy loss using the Adam optimizer (Kingma and Ba 2014), 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 via 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 of 256x256 and 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 (non-touching) lesional area in the mask is counted as a single lesion.

Data Pre-Processing

All photos were collected with consumer-grade cameras (Samsung Digimax L70 and Canon Powershot A630) and resized to a resolution of 1024x768 pixels for archival, prior to transfer. 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 which 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 InceptionV4 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 prior to the application of the performance metrics (reference standards). The performance of each model was evaluated on all skin in each image (excluding background 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 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 to 0.74), where 0 represents no agreement and 1 represents perfect agreement.

No significant difference in performance was found between photographs with or 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 Dermatology. 2022;158(1):90–6

GIMP. GNU Image Manipulation Program [Internet]. 2021 [cited 2021 Dec 10]. Available from: https://www.gimp.org

Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 2014;1–15 Available from: http://arxiv.org/abs/1412.6980

Muyembe-Tamfum J-J. Clinical Aspects Of Monkeypox In DRC: Monkeypox epidemiology, surveillance, and laboratory capacity in DRC. WHO monkeypox Res. What are Knowl. gaps Prior. Res. Quest. 2022 [cited 2022 Jun 2]. Available from: https://cdn.who.int/media/docs/default-source/blue-print/day-1_jean-jacques-muyembe_drc_monkeypox-meeting_02june2022.pdf?sfvrsn=d133676a_3

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

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):4278–84 Available from: http://arxiv.org/abs/1512.00567

Yakubovskiy P. Segmentation Models Pytorch [Internet]. GitHub Repos. 2021 [cited 2021 Dec 2]. Available from: https://github.com/qubvel/segmentation_models.pytorch

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(4):716–24

TABLES

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 vs 1	-5.86	28.56	-40.29	68.85	0.78	0.94
2 vs 1	-3.24	15.98	-22.46	38.44	1.02	0.97
3 vs 1	9.68	48.05	-28.69	76.74	1.07	0.92
3 vs 2	12.92	53.88	-28.03	81.91	1.03	0.90

FIGURE LEGENDS

Figure S1. WHO Monkeypox infection severity guidelines, from "Clinical Aspects of Monkeypox In DRC" by Professor Jean-Jacques Muyembe-Tamfum (Muyembe-Tamfum 2022).

Figure S2. Dataset characteristics. Number of photographs per patient (n=66, N=18) and histogram of lesion counts per photo by rater 1 (ground truth).

Figure S3. AI algorithm segmentation performance by Dice index. Each point represents a single photo (n=66). Boxplot shows median (0.72) and interquartile range (0.62 to 0.74) and mean (0.67, dashed red line).