Toward the determination of sensitive and reliable whole-lung computed tomography features for robust standard radiomics and delta-radiomics analysis in a nonhuman primate model of coronavirus disease 2019

Marcelo A. Castro[a,*] **Syed Reza**[b] **Winston T. Chu**[b] **Dara Bradley**[b] **Ji Hyun Lee**[a] **Ian Crozier**[c] **Philip J. Sayre**[a] **Byeong Y. Lee**[a] **Venkatesh Mani**[a] **Thomas C. Friedrich**[d] **David H. O'Connor**[e] **Courtney L. Finch**[a] **Gabriella Worwa**[a] **Irwin M. Feuerstein**[a] **Jens H. Kuhn**[a] **and Jeffrey Solomon**[c]

[a]National Institutes of Health, National Institute of Allergy and Infectious Diseases, Integrated Research Facility at Fort Detrick, Frederick, Maryland, United States
[b]National Institutes of Health, Clinical Center, Radiology and Imaging Sciences, Center for Infectious Disease Imaging, Bethesda, Maryland, United States
[c]Frederick National Laboratory for Cancer Research, Clinical Monitoring Research Program Directorate, Frederick, Maryland, United States
[d]University of Wisconsin–Madison, School of Veterinary Medicine, Department of Pathobiological Sciences, Madison, Wisconsin, United States
[e]University of Wisconsin–Madison, Department of Pathology and Laboratory Medicine, Madison, Wisconsin, United States

Abstract

Purpose: We propose a method to identify sensitive and reliable whole-lung radiomic features from computed tomography (CT) images in a nonhuman primate model of coronavirus disease 2019 (COVID-19). Criteria used for feature selection in this method may improve the performance and robustness of predictive models.

Approach: Fourteen crab-eating macaques were assigned to two experimental groups and exposed to either severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or a mock inoculum. High-resolution CT scans were acquired before exposure and on several post-exposure days. Lung volumes were segmented using a deep-learning methodology, and radiomic features were extracted from the original image. The reliability of each feature was assessed by the intraclass correlation coefficient (ICC) using the mock-exposed group data. The sensitivity of each feature was assessed using the virus-exposed group data by defining a factor R that estimates the excess of variation above the maximum normal variation computed in the mock-exposed group. R and ICC were used to rank features and identify non-sensitive and unstable features.

Results: Out of 111 radiomic features, 43% had excellent reliability (ICC $>$ 0.90), and 55% had either good (ICC $>$ 0.75) or moderate (ICC $>$ 0.50) reliability. Nineteen features were not sensitive to the radiological manifestations of SARS-CoV-2 exposure. The sensitivity of features showed patterns that suggested a correlation with the radiological manifestations.

Conclusions: Features were quantified and ranked based on their sensitivity and reliability. Features to be excluded to create more robust models were identified. Applicability to similar viral pneumonia studies is also possible.

 [DOI: 10.1117/1.JMI.9.6.066003]

Keywords: computed tomography; COVID-19; animal models; radiomics; reliability; sensitivity.

Paper 22136GR received Jun. 2, 2022; accepted for publication Nov. 21, 2022; published online Dec. 8, 2022.

*Address all correspondence to Marcelo A. Castro, marcelo.castro@nih.gov

1 Introduction

As of May 8, 2022, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused 514 million confirmed coronavirus disease 2019 (COVID-19) cases and over 6 million COVID-19 confirmed deaths worldwide.[1,2] The efficacy of recently approved medical countermeasures for COVID-19 may be circumvented by emergent SARS-CoV-2 variants that are more transmissible and immune-evasive.[3] Data from patients during the first few months of the COVID-19 pandemic (in early 2020) showed that chest CT is sensitive to the detection of the radiographic lung abnormalities associated with COVID-19.[4] Independent of SARS-CoV-2 variants, pathogenesis is similar, and computed tomography (CT) continues to be an option for determining prognosis. However, the use of medical imaging as a means of evaluating medical countermeasure efficacy in randomized clinical trials is critically hindered by the lack of standardized quantitative image analysis methods and reliable animal models.[5] In animal models of severe COVID-19, quantitative image analysis methods enable accurate, quantifiable, unbiased, and reproducible measurements of COVID-19 pulmonary disease from medical images.[6] In particular, noninvasive quantitative imaging biomarkers that do not require serial euthanasia are essential to the characterization of disease severity, progression, and pathogenesis in animal models.[7–12] Quantitation of COVID-19-like lung abnormalities using multimodality imaging biomarkers, including volumetric assessment of radiodensity (CT), has been described in crab-eating (cynomolgus) macaques (*Macaca fascicularis*) exposed to SARS-CoV-2.[6]

In the past, toward fast and accurate clinical evaluation and prognostication, radiomics analysis of chest CT images was proposed to explore imaging correlates with non-imaging markers of the development, progression, severity, and outcomes of COVID-19. Textural features to assess the classification of lung abnormalities were analyzed using artificial neural networks[13,14] and machine learning techniques.[15] Radiomics, which is a method that extracts and analyzes a large number of features from medical images using data characterization algorithms, includes textural features that were initially used to characterize topography from satellite images in the early 1970s[16–20] and was first introduced about a decade ago.[21] The use of textural features enables the translation of medical images into quantitative data to phenotypically profile lung abnormalities.[22] During the last decade, the vast majority of lung studies using radiomics analyzed features extracted from segmented lesions and focused mainly on tumor characterization, phenotype differentiation,[23,24] and prognostication of recurrence and survival.[25,26] Radiomic analyses began to be used for COVID-19 in 2020, when chest CT was identified as a sensitive SARS-CoV-2 infection diagnostic tool.[4]

Cai et al.[27] proposed a model based on CT radiomic features that could predict a negative reverse transcription quantitative polymerase chain reaction (RT-qPCR) test for SARS-CoV-2 and could be used to recommend early patient discharge from hospitals. Other authors focused on the prediction of patient outcomes,[28–30] prediction of residual lung lesions after discharge,[31] diagnosis,[32,33] discrimination of stable and progressive disease,[34] and differentiating COVID-19 from other causes of pneumonia.[34–37] Areas of interest to compute radiomic features ranged from the manual delineation of bounding boxes around lesions[38] to semiautomatic segmentation of lesions[39] and whole-lung volumes.[34] Some clinical studies used data from just a single hospital,[4] whereas others included data from multiple locations with different acquisition protocols[40] and faced challenges in comparing differentially acquired image datasets. Although each study used different software for radiomics feature extraction, in general, they adhered to the Image Biomarker Standardisation Initiative (IBSI).[41] Given the critical nature of the pandemic, baseline (preinfection) reference scans have not typically been available in these studies. Furthermore, radiomic feature reliability has not always been addressed.[42–45] COVID-19 animal model research has been used to investigate both the natural history of the disease and the efficacy of medical countermeasures in preclinical studies; radiomic features may be used not only to predict outcomes and differentiate different pathologies but also as subject-specific imaging biomarkers of the disease when preexposure images are acquired and control groups are considered.

In the field of radiomics, several analytic terms (e.g., repeatability, reproducibility, reliability, robustness, stability, and sensitivity) are used across studies, but their meanings may depend on the scope of the research;[46] thus terminology should be clarified. Here, *repeatability* refers to

features that remain the same when the subject is imaged multiple times. *Reproducibility* refers to features that remain the same when images are acquired using different equipment, software, acquisition settings, and operators (e.g., in studies that include multiple hospitals).[47] *Reliability* is the extent to which measurements can be replicated under either similar or different conditions. Reliability, which is often regarded as a measure of *robustness*, reflects the correlation and agreement between measurements and represents the ratio of true variance over the true variance plus error variance;[48] it is useful for the analysis of intrasubject and intersubject variations.[48,49] In delta (Δ) radiomics, longitudinal data can be used to assess intra-individual reproducibility and relative differences in pre- and post-treatment radiomic features to predict outcomes and treatment response. Δ radiomics has been referred to as "patient-specific" radiomics[50] and was first proposed a few years ago to improve reproducibility and predictive power.[23] Δ radiomics has since been studied in clinical and experimental settings to assess recovery or response to treatment in cancer research.[25,51] However, to the best of our knowledge, Δ radiomics has not been applied to imaging studies related to COVID-19. In principle, Δ radiomics also has the potential to be used to characterize the evolution of an infectious disease when "normal" preinfection baseline information is available. Under this context, we will use *stability* to describe features that do not exceed the intrasubject normal range and refer to *sensitivity* as the range of a feature during the course of the disease relative to the intrasubject normal range.

The reproducibility of radiomics features is affected by different scanners and acquisition parameters,[52,53] and reproducible features can be grouped into a limited number of clusters due to redundancy of information. There may be a high demand for research in the areas of image acquisition, image postprocessing, volume-of-interest segmentation, image discretization, and feature calculation to select features with sufficient dynamic range among patients, intrapatient reproducibility, and low sensitivity to image acquisition and reconstruction protocols.[54,55] Intraobserver delineation variability, respiratory motion, and reconstruction kernels were also found to strongly affect feature reproducibility.[56–58] In our previous work, we found that B-kernel (smooth) reconstructions were more reliable than D-kernel (sharp) ones;[59] therefore B-kernels are used in this work. To the best of our knowledge, the reliability of features based on the intrasubject and inter-subject variability in animal models, the determination of ranges of normal variation, and the sensitivity to radiological manifestations have not been investigated in the past. This information may be used to increase the robustness of future analysis via standard radiomics and Δ radiomics.

In this work, crab-eating macaques were exposed to either SARS-CoV-2 or a mock inoculum. CT images were acquired prior to exposure and at multiple time points after exposure, whole-lung fields were segmented, and radiomic features were extracted. The animals included in this work were scanned with two identical scanners with the same acquisition protocols, and different reconstructions were not used interchangeably for radiomics analysis. The reliability of radiomic features was characterized by the intrasubject and intersubject variability, and stability and sensitivity to the disease were assessed by analysis of the change of features during the course of the disease with respect to the baseline scan. This information can contribute to building robust standard and Δ-radiomics signatures that correlate with nonimaging features to help identify disease stage and severity, evaluate the efficacy of candidate medical countermeasures, and predict clinical outcomes.

2 Methodology

2.1 *Animals and Virus*

Initially, the study was to use a total of 25 crab-eating macaques (*Macaca fascicularis*). However, 11 were excluded due to abnormalities at the baseline scan or not meeting the CT-score criterion for inclusion—a score of no more than two at every time point. (This criterion was set to better characterize the normal variation of radiomic features computed from the segmented lungs.) Thus, a total of 14 (four males and 10 females; age range: 4 to 7 years old, weight range: 2.56 to 6.83 kg) macaques were assigned to two experimental groups (Mock: NmTOT = 6

and Virus: NvTOT = 8). The macaques were anesthetized in accordance with standard procedures prior to all manipulations, including intrabronchial exposure, sample collection, and medical imaging. Animals in the Mock group were administered 2 mL of cell culture medium supplemented with 2% heat-inactivated fetal bovine serum into each bronchus followed by 1 mL of normal saline flush and 5 mL of air. Animals in the Virus group were exposed to 2 mL containing 9.13×10^5 PFU/mL of SARS-CoV-2 (isolate 2019-nCoV/USA/A12/2020, obtained from the US Centers for Disease Control and Prevention [CDC], Atlanta, GA) for a total dose of 3.6×10^6 PFU.[6] RT-qPCR analysis was performed to determine the presence of SARS-CoV-2 RNAs in the collected specimens.[6] All experiments were performed in a maximum (biosafety level 4 [BSL-4]) containment laboratory at the IRF-Frederick, a facility accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC). Experimental procedures were approved by the National Institute of Allergy and Infectious Diseases (NIAID) Division of Clinical Research (DCR) Animal Care and Use Committee (ACUC) and conducted in compliance with the Animal Welfare Act regulations, Public Health Service policy, and the Guide for the Care and Use of Laboratory Animals (Eighth Edition).

2.2 *Imaging of Crab-Eating Macaques*

The animals considered in this work had been assigned to one of three studies with identical exposure and imaging protocols; however, in the first study, the animals were scanned for a longer period of time after exposure. Animals in both groups were scanned before exposure and either eight or four times after exposure (at 2, 4, 6, 8, 10, 12, 19, and 30 days for those in the first study or at 2, 4, 6, and 8 days for those in the second and third studies) (Table 1). High-resolution chest CT scans were performed using the 16-slice CT component of either a Gemini TF 16 scanner (Philips Healthcare, Cleveland, OH) or a Precedence scanner (Philips Healthcare). Images were acquired in helical scan mode with the following parameter settings: ultra-high resolution, 140 kVp, 300 mAs per slice, 1-mm thickness, 0.5-mm increment, 0.688-mm pitch, collimation 16×, and 0.75-s rotation. CT image reconstruction used a 512×512 matrix size for a 250-mm transverse field-of-view (FOV), leading to a pixel size of 0.488 mm. CT images were produced with the standard B reconstruction kernel for smoother images because, in previous work, we showed that radiomic features extracted from CT images reconstructed with a bone-enhanced D reconstruction kernel for sharper images were less reproducible.[59] No contrast agent was administered. Each macaque underwent a 15 to 20 s breath-hold during acquisition. The pressure for the breath-hold was maintained at 150 mm H_2O. For imaging procedures, each macaque was anesthetized intramuscularly with 15 mg/kg ketamine following 0.06 mg/kg glycopyrrolate intramuscularly. Anesthesia was maintained using a constant rate intravenous infusion of propofol at 0.3 mg/kg/min. Macaques were placed on the scanner bed in a supine head-out/feet-in position and connected to a ventilator to facilitate breath holds. Vital signs were monitored throughout the procedure.[6] All images were visually inspected for possible signal loss and/or artifacts. Inclusion criteria were different for each group. Animals in the Mock group were required to have qualitatively normal scans on all scan days to accurately estimate the maximum normal variation of the radiomic features. Animals in the Virus group were required to have a normal baseline scan to avoid inaccurate estimation of changes in radiomic features during the course of the disease due to abnormalities present at baseline.

Table 1 Number of P.E. scans for all animals included in this study.

Animal ID	M#1	M#2	M#3	M#4	M#5	M#6	V#1	V#2	V#3	V#4	V#5	V#6	V#7	V#8
#P.E. scans	8	8	4	4	4	4	8	8	4	8	4	4	4	4

P.E.: postexposure with either mock inoculum (Mock) or SARS-CoV-2 (Virus)

2.3 *Whole-Lung Segmentation*

For training purposes, a total of 64 whole-lung CT scans (reconstructed using a B kernel of crab-eating macaques with the same imaging protocols) were used. The automated organ segmentation method, based on the convolutional neural network (CNN), used in this work has been described before.[60] The feature pyramids network (FPN), which produces a multiscale feature representation in which all levels, even the high-resolution levels, are semantically strong, was used in this work. The network was trained using input patches of size $64 \times 64 \times 64$ voxels, which were randomly extracted from both lung and nonlung areas with equal numbers. The output of the CNN was a probability map, which was resampled to the original image size and smoothed using a Gaussian filter. The quality of the segmentations was evaluated. Whole-lung masked CT images were generated.

2.4 *Radiomic Feature Extraction*

In this study, 90 whole-lung masked CT images from 14 crab-eating macaques were generated using the methodology described in the previous section. Radiomics feature extraction from the whole-lung masked CT images was performed using PyRadiomics 2.2.0.[61] For each image, 111 features were extracted: 17 3D shape features and 94 intensity features split into 19 first-order features and 75 second-order features. The latter were derived from five different matrices: (1) 24 features from the gray-level co-occurrence matrix (GLCM); (2) 14 features from the gray-level dependence matrix (GLDM); (3) 16 features from the gray-level run length matrix (GLRLM); (4) 16 features from the gray-level size zone matrix (GLSZM); (5) five features from the neighboring gray tone difference matrix (NGTDM). Images were discretized using a 25-HU bin width, resulting in ≈ 30 to 40 bins. A shift of 1024 HU was set for the first-order features to avoid negative attenuations. For each voxel, two neighbors were considered for each of the 13 directions corresponding to the first neighbors in the second-order features.

2.5 *Data Analysis*

In this work, we studied the reliability regarding intrasubject and inter-subject reproducibility in a normal population scanned under the same conditions, as well as the stability and sensitivity of features during the disease course.

First, all scans from the Mock group were used to compute the intra-class correlation coefficient (ICC) to assess the reliability of radiomic features when both intrasubject and inter-subject variations were present under the same scanning conditions. ICC estimates and their 95% confidence intervals were calculated using the R package IRR version 0.84.1, based on a single measurement ($k = 1$), absolute-agreement, two-way mixed-effects model. To manage the different number of scans among subjects, ICC was computed from two different subsets of five scans and averaged for each feature. Reliability of ICC values was considered as follows: 0.00 to 0.50 (poor), 0.50 to 0.75 (moderate), 0.75 to 0.90 (good), and 0.90 to 1.00 (excellent).[49] This information has the potential to be used to identify features not reliable for standard radiomic analysis. The reliability of each radiomic feature was assessed.

Second, the maximum normal intrasubject variation of each feature, along with a comparison of the intrasubject dynamic range of the feature within the course of the disease, was investigated. To estimate the maximum normal intrasubject variation Δ_f (%) of each radiomic feature f, only scans from the Mock group were considered, and for each animal m, the maximum percent change ${\Delta_f}^m$ with respect to that lowest measurement was computed. The maximum among all animals was used as an estimate of the maximum normal intrasubject variation: $\Delta_f = \max_m\{{\Delta_f}^m\}$. Afterward, for each animal v in the Virus group and each feature f, the lower and upper thresholds of the normal range f_L and f_U, respectively, were computed from Δ_f and the feature value at the baseline scan. The percent change at each postexposure scan was computed with respect to the baseline scan, and the dynamic range ${\Delta_f}^v$ was identified and

compared with Δ_f. For each animal, each feature was classified as follows: C1 = not sensitive (the feature value was between f_L and f_U at all postexposure scans); C2 = not stable (the feature value was predominantly above/below f_L/f_U but also beyond the opposite threshold at some time points); C3 = sensitive and increasing (the feature value was above f_U for at least one day and remained within normal values for the rest of the scans); and C4 = sensitive and decreasing (the feature value was below f_L for the at least one day and remained within normal values for the rest of the days). Only features in categories C3 and C4 were considered; those in category C1 were not sensitive for radiomic analysis, and those in C2 should be evaluated separately.

For each animal v in the Virus group and each radiomic feature in C3 and C4, the maximum variation ${\Delta_f}^v$ was identified, and the ratio ${R_f}^v = ({\Delta_f}^v - \Delta_f)100/{\Delta_f}^v$ was computed. Note that $|{R_f}^v|$ ranges between 0 and 100, both inclusive, where $|{R_f}^v| \approx 0$ means that Δ_f and ${\Delta_f}^v$ are comparable; e.g., ${R_f}^v = 50$ means that ${\Delta_f}^v$ is two times Δ_f. The average among all animals in the Virus group was computed, and a ranking was generated. This information has the potential to be useful in identifying features that are unstable and nonsensitive to the disease in Δ-radiomics analysis.

3 Results

3.1 *Whole-lung CT Images*

After exposure, SARS-CoV-2 infection was confirmed in all virus-exposed but not in mock-exposed animals (data not shown).[6] Initially, the Mock and Virus groups included a total of 13 and 12 animals, respectively. In the Mock group, seven animals were excluded because they did not pass the criterion to have a total CT score[6] of no more than two at every time point. This criterion was set to avoid overestimation of the maximum normal variation of radiomic features computed from the segmented lungs. In the Virus group, four animals were excluded because they had abnormalities at the baseline scan, although they did not reach the threshold to have a CT score above 0. This criterion was set to avoid underestimation of the changes in the radiomic features with respect to baseline in the virus group. Therefore, a total of 14 animals were included in the study (six in the Mock group and eight in the Virus group).

Selected axial slices exhibiting representative lesions from scans at the peak of the radiological manifestation (typically, at Day 2 and/or Day 4) of all animals in the Virus group are displayed in Fig. 1. Selected axial slices from arbitrary scans of all animals in the Mock group are shown in Fig. 2. Binary masks were created using a deep learning algorithm from CT images. All animals were scanned before exposure and either four or eight times after exposure (Table 1). Scans were reconstructed using a B kernel. All masks were visually compared with their corresponding CT images to assess their accuracies.

3.2 *Radiomic Features Reliability*

ICC estimates and their 95% confidence intervals were calculated from B-kernel reconstructions based on a single measurement ($k = 1$), absolute-agreement, two-way mixed-effects model. In previous work, we showed that the estimated ICC averaged over all features was greater for the B-kernel (0.819) than the D-kernel (0.722) and 93 features had a higher ICC when the B-kernel was used for reconstruction;[59] therefore, all results in this paper are based on B-kernel reconstructions. The number of features with ICC values with poor (0.00 to 0.50), moderate (0.50 to 0.75), good (0.75 to 0.90), and excellent (0.90 to 1.00) reliability is shown in Fig. 3(a). Poor reliability (ICC < 0.50) was observed in only two features (GLCM-Imc1 and NGTDM-Strength), and 48 features exhibited excellent reliability (ICC > 0.90) (Table 2). The reliability of features within each type is shown in Fig. 3(b). Figure 5 shows the ICC of all features along with the ratio R that compares their maximum variation due to the disease with the maximum normal variation (Secs. 3.2 and 3.3).

Fig. 1 Selected axial slices exhibiting representative lesions from scans at the peak of the radiological manifestation (typically 2 or 4 days after exposure) of all animals in the Virus group. The experiment was designed to mimic mild disease in humans. For each animal in the Virus group, four slices were chosen from locations where the most visually noticeable lesions appeared at the peak of the disease. The selected pictures show a range of abnormalities from mild to severe at a glance.

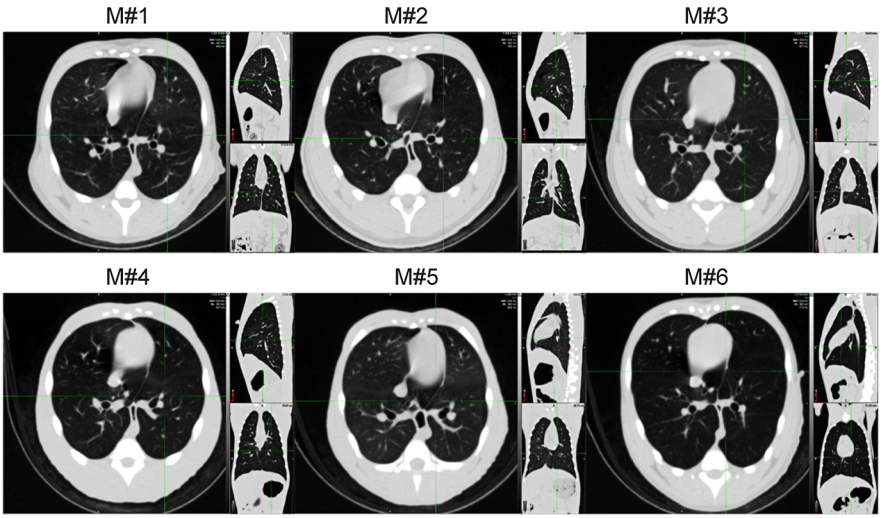

Fig. 2 Selected axial slices from the scans of all animals M#1 to M#6 in the Mock group.

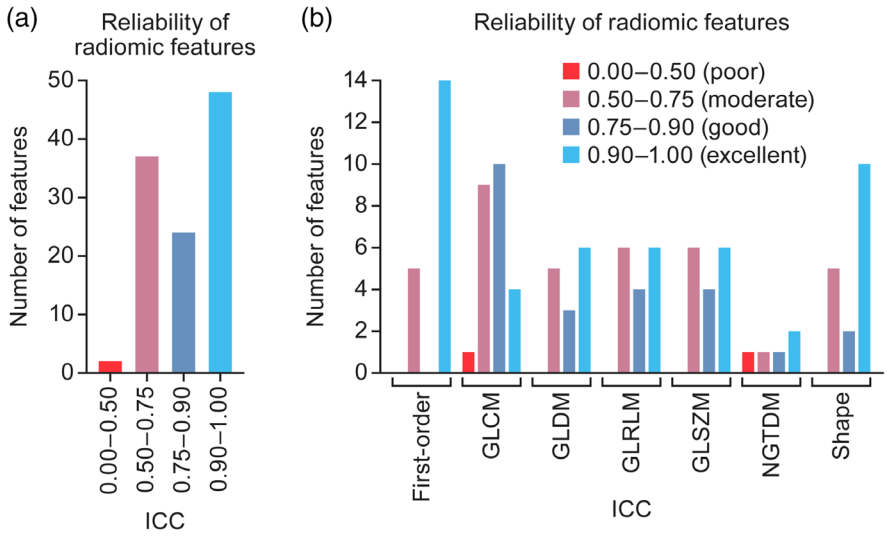

Fig. 3 (a) Number of features in the four ICC ranges for B-kernel reconstructions. Reliability of ICC values: 0.00 to 0.50 (poor), 0.50 to 0.75 (moderate), 0.75 to 0.90 (good), and 0.90 to 1.00 (excellent). (b) Reliability of each type of feature.

3.3 *Maximum Intrasubject Normal Variation Compared with Variations due to the Disease*

To characterize the maximum intrasubject normal variation (Δ_f) of each radiomic feature (f), only animals in the Mock group were considered: two animals were scanned nine times and the other four animals were scanned five times (Table 1). For each animal, the minimum value of each radiomic feature and the corresponding time point were identified; subsequently, all other scans were individually compared with the minimum values to arrive at the percent change for each feature. The maximum of those values s is called the maximum intrasubject normal variation Δ_f. The number of features with different ranges of Δ_f is shown in Fig. 4(a). In Fig. 4(b), the number of features is discriminated among feature types. Overall, $\Delta_{f\text{average}} = 66\%$ and $\Delta_{f\text{median}} = 29\%$. All 49 features with $\Delta_f < 25\%$ are shown in Table 3. Note that the maximum normal variation of a given feature as an indicator of stability must be combined with the variation of the feature during the course of the disease with respect to its baseline value to determine the usefulness of that feature. Figure 5 shows the ICC of all features along with the ratio R that compares their maximum variation due to the disease with the maximum normal variation.

Table 2 All 48 features extracted from B-kernel reconstruction with excellent reliability (ICC > 0.90).

TYPE	FEATURE	ICC
First-order	Minimum	1.000
GLRLM	RunLengthNonUniformity	0.998
GLDM	DependenceNonUniformity	0.997
Shape	SurfaceArea	0.997
Shape	LeastAxisLength	0.997
GLSZM	SizeZoneNonUniformity	0.995
Shape	MeshVolume	0.995
Shape	VoxelVolume	0.995
Shape	Maximum2DDiameterColumn	0.994
Shape	Maximum3DDiameter	0.993
Shape	MajorAxisLength	0.992
GLSZM	GrayLevelNonUniformity	0.991
GLSZM	SmallAreaHighGrayLevelEmphasis	0.991
GLRLM	GrayLevelNonUniformity	0.990
Shape	Maximum2DDiameterRow	0.990
Shape	MinorAxisLength	0.989
GLSZM	HighGrayLevelZoneEmphasis	0.989
GLDM	LargeDependenceHighGrayLevelEmphasis	0.988
First-order	10Percentile	0.988
First-order	Energy	0.983
First-order	TotalEnergy	0.983
NGTDM	Busyness	0.982
GLRLM	GrayLevelVariance	0.982
First-order	90Percentile	0.979
Shape	Maximum2DDiameterSlice	0.978
First-order	Mean	0.978
GLCM	ClusterProminence	0.978
GLDM	GrayLevelNonUniformity	0.978
First-order	RootMeanSquared	0.976
First-order	Median	0.973
First-order	MeanAbsoluteDeviation	0.970
First-order	Variance	0.968
First-order	StandardDeviation	0.968

Table 2 (*Continued*).

TYPE	FEATURE	ICC
NGTDM	Complexity	0.966
GLCM	ClusterShade	0.959
GLRLM	ShortRunHighGrayLevelEmphasis	0.950
GLDM	SmallDependenceHighGrayLevelEmphasis	0.949
GLRLM	HighGrayLevelRunEmphasis	0.946
GLDM	GrayLevelVariance	0.942
GLDM	HighGrayLevelEmphasis	0.939
GLCM	ClusterTendency	0.938
First-order	Range	0.935
GLSZM	ZoneEntropy	0.935
GLSZM	GrayLevelVariance	0.930
First-order	InterquartileRange	0.921
GLRLM	LongRunHighGrayLevelEmphasis	0.912
First-order	RobustMeanAbsoluteDeviation	0.907
GLCM	SumSquares	0.905

GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLZSM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix.

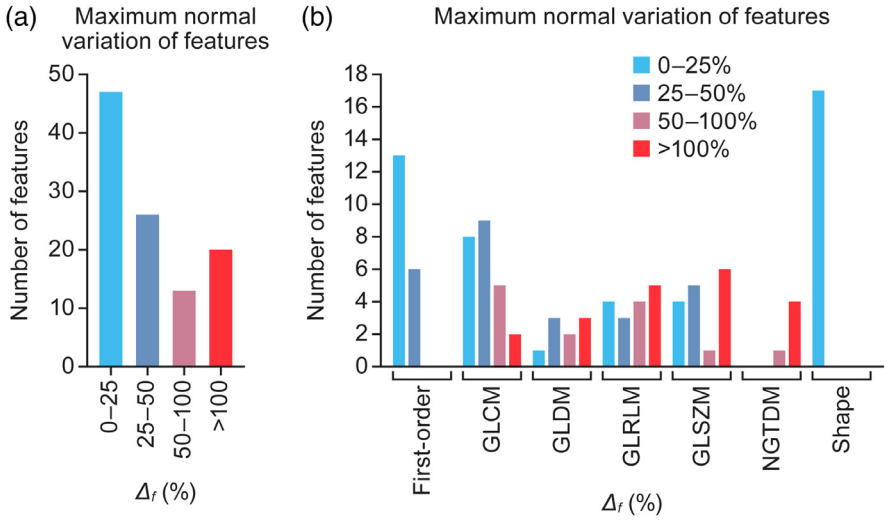

Fig. 4 (a) The number of features in the four ICC ranges for B-kernel reconstructions. Reliability of ICC values: 0.00 to 0.50 (poor), 0.50 to 0.75 (moderate), 0.75 to 0.90 (good), and 0.90 to 1.00 (excellent). (b) Reliability of each type of feature.

3.4 *Radiomic Features Insensitive to Radiological Manifestations*

For a feature f that is not sensitive to the radiological manifestation in animals in the Virus group, $R_f = 0$. The 19 nonsensitive features are listed in Table 4. An example of a nonsensitive feature is shown in Fig. 6(a).

Table 3 All 49 features extracted from B-kernel reconstruction with maximum normal variation $\Delta_f < 25\%$.

TYPE	FEATURE	Δf (%)
GLCM	IDMN	0.48
GLCM	IDN	1.87
GLSZM	ZoneEntropy	2.09
Shape	LeastAxisLength	2.80
GLSZM	SmallAreaEmphasis	3.17
First-order	Median	4.10
Shape	Maximum3DDiameter	4.48
First-order	Mean	4.50
First-order	10Percentile	5.18
Shape	Maximum2DDiameterColumn	5.21
Shape	Maximum2DDiameterRow	5.30
GLCM	InverseVariance	5.45
Shape	MinorAxisLength	5.69
Shape	MajorAxisLength	5.94
First-order	90Percentile	6.35
GLSZM	SizeZoneNonUniformityNormalized	6.42
GLCM	SumEntropy	6.61
Shape	SphericalDisproportion	6.75
Shape	Sphericity	6.75
Shape	Elongation	7.34
Shape	Flatness	7.45
GLDM	DependenceEntropy	8.14
Shape	SurfaceArea	8.33
First-order	StandardDeviation	8.58
Shape	SurfaceVolumeRatio	9.22
First-order	MeanAbsoluteDeviation	9.61
Shape	Maximum2DDiameterSlice	10.08
First-order	Entropy	10.23
Shape	Compactness1	10.30
GLRLM	RunEntropy	10.86
GLRLM	GrayLevelVariance	10.98
GLRLM	ShortRunEmphasis	12.96
Shape	MeshVolume	14.38

Table 3 (*Continued*).

TYPE	FEATURE	Δf (%)
Shape	VoxelVolume	14.39
GLSZM	GrayLevelNonUniformityNormalized	14.84
First-order	Skewness	15.27
GLCM	ClusterTendency	15.94
GLCM	JointEntropy	17.38
GLDM	GrayLevelVariance	17.79
First-order	Variance	17.89
GLDM	SmallDependenceHighGrayLevelEmphasis	18.13
First-order	RootMeanSquared	19.10
GLCM	SumSquares	20.65
GLCM	DifferenceEntropy	20.90
Shape	Compactness2	21.66
First-order	RobustMeanAbsoluteDeviation	21.66
First-order	Minimum	21.97
First-order	Kurtosis	22.52
GLRLM	GrayLevelNonUniformityNormalized	23.67

Δ_f, maximum normal variation of feature *f*; GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLZSM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix.

3.5 *Radiomic Features Sensitive to Radiological Manifestations but Unstable*

It was observed that a limited number of features sensitive to radiological manifestation can at the same time vary beyond the opposite threshold of the normal interval. The occurrence was only in 17 out of 1776 computations from three animals in the Virus group: V#1 (6/17), V#4 (2/17), and V#5 (9/17). An example is shown in Fig. 6(b), and the results are shown in Table 5. These features should be investigated separately.

3.6 *Radiomic Features Sensitive to Radiological Manifestations*

A total of 74 features had a variation beyond Δ_f with respect to the baseline scan for at least one animal in the Virus group. First, a ranking of $\langle R_f \rangle$ was generated to determine which features are more sensitive to the radiological manifestations in animals in the Virus group. An arbitrary threshold was set to differentiate those features f for which R_f varied <5% above or below $\Delta_f (\Delta_f \pm 5\%)$.

Figure 7(a) shows the results for features extracted along with the average for each feature over all animals in the group. As an example, a comparison of two animals' CT scans—one dominated by large consolidations and other lesions [Fig. 7(b)] and the other having smaller lesions with less attenuation [Fig. 7(c)]—is shown in Fig. 7(d). Also, as an example, the evolution along the course of the disease is shown for two sensitive features, one increasing above the normal range [GLRLM short-run high gray-level emphasis, Fig. 8(a)] and the other sensitive and decreasing below the normal range [first-order skew, Fig. 8(b)].

Figure 9 shows the lung involvement over time at a selected axial slice in the baseline scan and the eight postexposure scans for V#1. Although scans are visually different, the information

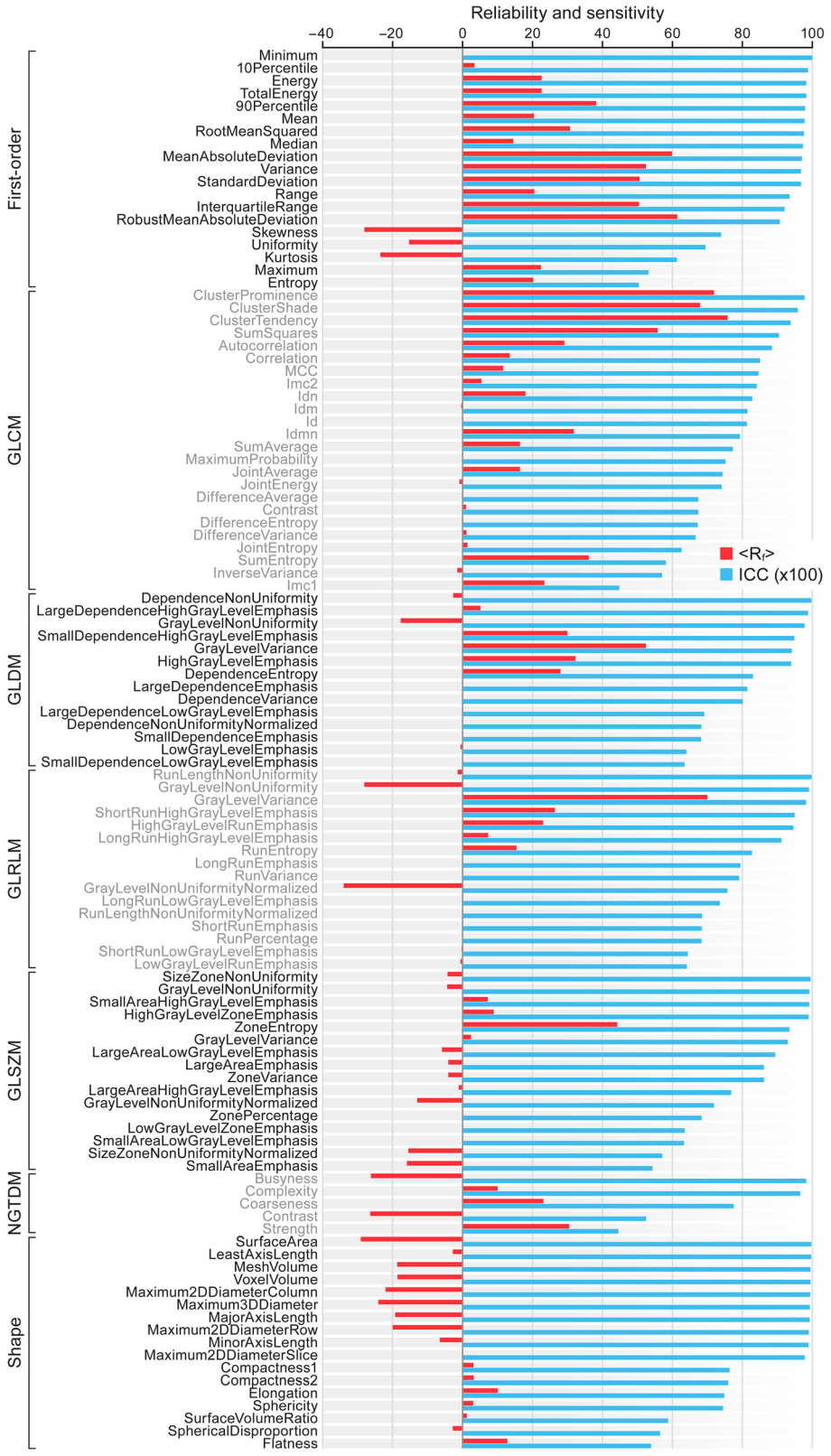

Fig. 5 ICC values for each feature (white and black) along with $\langle R_f \rangle$ as a measure of sensitivity (blue). Reliability of ICC values were considered as 0.00 to 0.50 (poor), 0.50 to 0.75 (moderate), 0.75 to 0.90 (good), and 0.90 to 1.00 (excellent). $\langle R_f \rangle$ with an absolute value near zero indicates that the feature f is not sensitive to the disease when compared with its baseline value. GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix.

Table 4 Features not sensitive to the radiological manifestations in the Virus group.

Type	Feature (f)
Shape	Maximum 2D diameter slice
GLCM	Maximum probability
GLCM	Difference entropy
GLDM	Small dependence emphasis
GLDM	Dependence nonuniformity normalized
GLDM	Large dependence emphasis
GLDM	Large dependence low gray level emphasis
GLDM	Dependence variance
GLDM	Small dependence low gray level emphasis
First-order	Minimum
GLRLM	Run variance
GLRLM	Long run emphasis
GLRLM	Short run emphasis
GLRLM	Run percentage
GLRLM	Long run low gray level emphasis
GLRLM	Run length non uniformity normalized
GLSZM	Zone percentage
GLSZM	Low gray level zone emphasis
GLSZM	Small area low gray level emphasis

GLCM, gray-level cooccurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix.

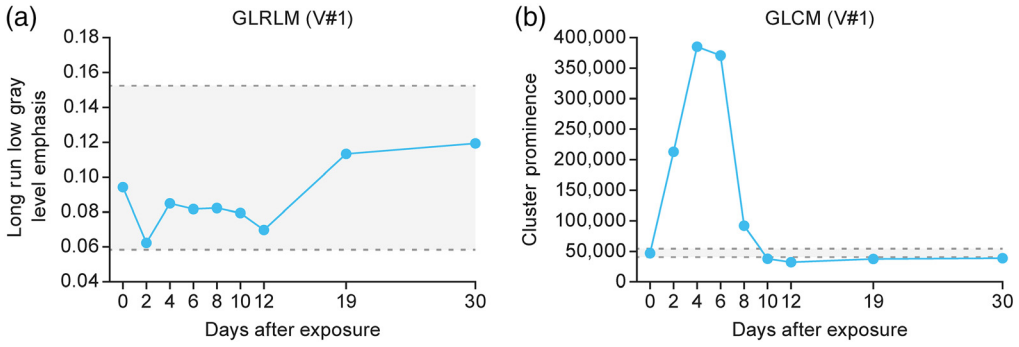

Fig. 6 Time evolution along the course of the disease in V#1 of radiomic features: (a) GLRLM long run low gray level emphasis (not sensitive and with $\Delta f = 61.5\%$) and (b) GLCM cluster prominence (sensitive but not stable and with $\Delta f = 15.4\%$). Dotted lines represent the lower and upper thresholds of the normal variation range. GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix.

Table 5 Features that are sensitive and unstable.

Type	Feature (f)	N_f	$\langle R_f \rangle$ (%)
GLCM	Cluster shade	3	67.9
GLCM	Cluster prominence	2	71.9
GLDM	Gray level variance	0	52.4
GLDM	Small dependence high gray level emphasis	1	30.0
First-order	Standard deviation	0	50.6
First-order	Range	1	20.5
First-order	Variance	0	52.4
First-order	Maximum	1	22.4
GLRLM	Gray level variance	1	70.0
GLSZM	Gray level nonuniformity normalized	0	−13.0
NGTDM	Strength	1	30.5

N_f: number of animals in the Virus group with an unstable feature f. $\langle R_f \rangle$: average among all animals v in the Virus group of ${R_f}^v = ({\Delta_f}^v - \Delta_f)/{\Delta_f}^v$ without considering those N_f animals. GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix.

obtained from a visual inspection helps to identify differences in some first-order features. However, changes in higher-order features are usually difficult to assess visually.

4 Discussion

Animal models of human disease are a critical part of biological research, including in the investigation of pathogenesis and the evaluation of candidate medical countermeasures, such as therapeutics and vaccines. Noninvasive quantitative imaging biomarkers that do not require serial euthanasia help with characterizing the progression of a disease severity and progression and understanding the underlying mechanisms. Measurable changes in imaging biomarkers throughout the course of the disease provide useful information when compared with baseline scans, which are typically not acquired in clinical settings. Furthermore, information from a control group is essential for determining the range of normal variation for imaging biomarkers. In recent years, radiomics has been explored as a tool, for instance, to investigate associations between both textural and nontextural features and survival rate, to predict outcomes, or for differential diagnosis. Although radiomics has been used to investigate COVID-19 in humans, to the best of our knowledge, application to animal models has not been reported yet. The characteristics and uniqueness of our data allow for the implementation of both standard radiomics and delta radiomics, particularly to quantify the progression of the disease, evaluate therapeutic options, and potentially predict outcomes.

Radiomics data are typically analyzed with statistical and machine learning methods that may depend on the disease context and image modality, among other factors. Machine learning techniques can capture complex interactions among features, feature combinations, and clinical biomarkers to build efficient prognostic and predictive models. However, the inclusion of radiomic features that are not reliable, not sensitive, and/or redundant may affect the robustness of those techniques. In particular, features with low intrasubject and intersubject repeatability may affect the statistical power, ability to interpret, and extrapolation to a more general application. Within the scope of Δ–radiomics, the identification and inclusion of specific features that are sensitive to radiological manifestations during the course of the disease may help to establish a connection between the number of features and their change above the normal variation during a given stage

Fig. 7 (a) Factor $R_f{}^V$ as defined in Sec. 2.5 of features sensitive to radiological manifestations computed from all scans of the eight animals in the Virus group for the B-kernel reconstruction along with the average over all animals in the group for each feature; (b) arbitrary axial slice of V#1 showing a large consolidation and other lesions; (c) arbitrary slice of V#7 dominated by smaller lesions with less attenuation; (d) comparison between $R_f{}^V$ in V#1 and V#7 only for features with R_f that vary <5% above or below Δf. GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix.

of disease. However, there is no "one-fits-all" solution; deciding which features to include and exclude depends on several factors—e.g., the disease, the type of lesion or abnormality under scrutiny, the imaging modality, and the area of interest (i.e., organ or lesion).

The animal-model experiments performed at the IRF-Frederick used two identical CT scanners with the same imaging protocol; therefore, no reproducibility study was required. Instead, reliability focused on intrasubject repeatability and intersubject normal variation when images from all scans of mock-exposed control animals without underlying abnormality were considered. The dynamic range of features extracted from CT images of virus-exposed animals during the course of the disease was analyzed along with the normal variation. If the dynamic range of

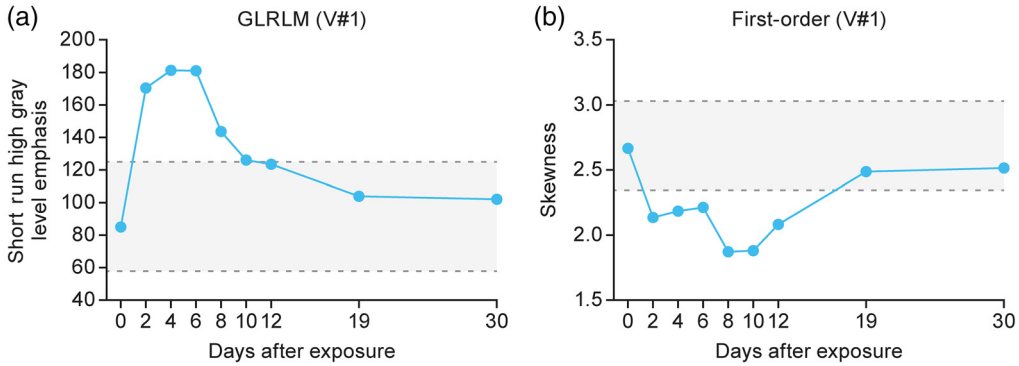

Fig. 8 Time evolution along the course of the disease in animal V#1 of radiomic features (a) GLRLM short run high gray level emphasis (sensitive and increasing above the normal range with $\Delta f = 47.1\%$) and (b) first-order skewness (sensitive and decreasing below the normal range with $\Delta f = 13.7\%$). Dotted lines represent the lower and upper thresholds of the normal variation range. GLRLM, gray-level run length matrix.

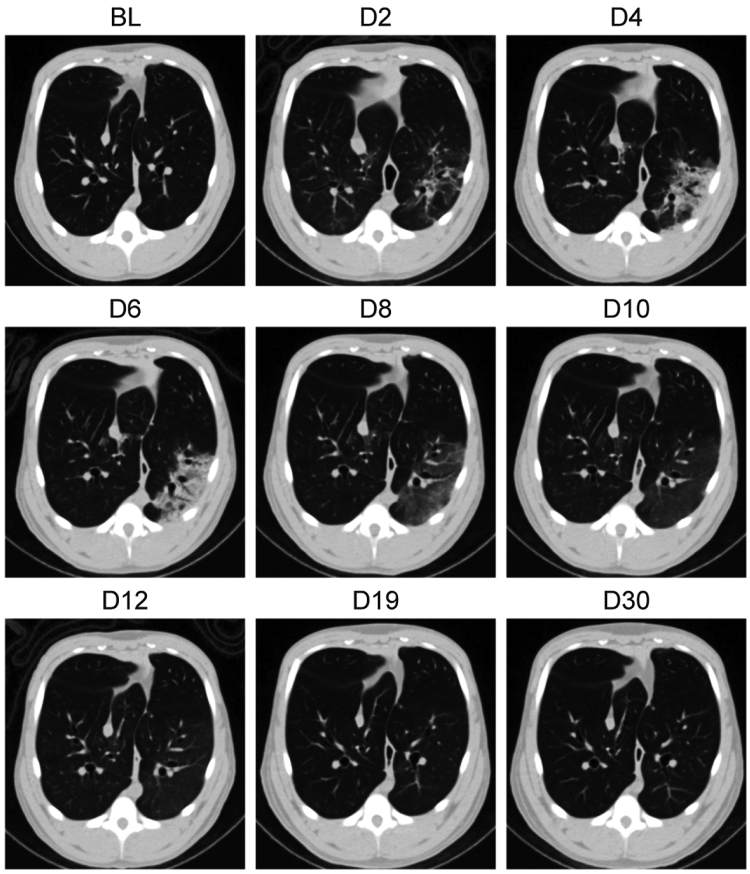

Fig. 9 Time evolution of the lung involvement at a selected axial slice along the course of the disease in animal V#1.

a given feature did not exceed the maximum intrasubject normal variation for all animals in the virus-exposed group, regardless of the radiological manifestation, that feature was considered not sensitive to the radiological manifestations, and therefore, that feature was not expected to provide any meaningful information. Sensitive feature values remained within the normal range unless near the peak of the disease.

The animals in the Virus group had a variety of radiological manifestations, and a given feature may be sensitive for some animals but not for others. To characterize the sensitivity

compared with the stability, a ratio R that takes into account the percent of the dynamic range that is above the normal variation was proposed to rank those features. A limited number of features varied beyond the normal range near the peak of the disease and later varied beyond the opposite threshold of the normal range when recovering. Those features were marked as sensitive to the disease but unstable. The meaningfulness of those features should be investigated in more detail and the opposite threshold should eventually be relaxed to avoid misclassification.

We focused on the B-kernel reconstructions because of their higher reliability. From the standard radiomics perspective, features with poor and moderate reliability (ICC < 0.75) should be excluded from further analysis. Otherwise, larger variations of features in both the mock-exposed control group and the virus-exposed group may occur. From the Δ-radiomics perspective, the aim would be to include only the features expected to vary beyond the normal range. A threshold to decide which features should be included has not been investigated in detail; however, we identified features with R_f that did not exceed ±5%. It is worth mentioning that 69% of the features with $\Delta_f > 50\%$ also had R_f below ±5%, and only 15% of the features with ICC > 90% had $\Delta_f > 50\%$.

The results presented in this work have the potential to be useful either to exclude irrelevant features for more accurate standard radiomics analysis[62] or to perform delta-radiomics analysis using changes of features with respect to their baseline values within a normal range. For example, five out of nine features based on low gray-level emphasis were not sensitive to radiological manifestations, whereas the other four had a ratio $R_f < 5\%$. On the other hand, eight out of nine features based on high gray level emphasis were sensitive to radiological manifestations ($R_f > 5\%$), whereas the remaining feature was sensitive but unstable because R_f fell below the lower threshold of the normal interval at some time points. Nevertheless, the study had some limitations. Both the average R and a set of features with a certain range of ratios R might eventually be characteristic of specific radiological manifestations. However, the total number of animals was not large enough to include a sufficient number of animals with the most common abnormalities; therefore, a characterization of abnormalities based on R was not pursued. As a preliminary result, it was found that R significantly varied between an animal with large areas of highly attenuated abnormalities and another with smaller areas with lower attenuation. Also, the use of the intrasubject dynamic range computed as the maximum percent change with respect to the baseline scan was useful to exclude not sensitive features when compared with the maximum intrasubject normal variation. However, the analysis of the time evolution of the overall abnormalities at every time point was lacking. Potentially, the stage of the disease might be assessed at each time point based on the set of sensitive features and their corresponding ratios.

In further analyses, more animals will be included to allow for a better association of changes in radiomic features and a radiological manifestation. Features from preprocessed images, such as, Laplacian of Gaussian and wavelets, will also be explored. The associations of delta-radiomic features with nonimaging biomarkers will be studied as well to pursue one of the main goals of radiomic analysis while concurrently addressing a significant need for animal models of COVID-19.

Disclosures

This work was partially based on the paper "Determination of reliable whole-lung CT features for robust standard radiomics and delta-radiomics analysis in a crab-eating macaque model of COVID-19: stability and sensitivity analysis" (https://doi.org/10.1117/12.2607154) published in the SPIE Medical Imaging Proceedings Volume 12036, Medical Imaging 2022: Biomedical Applications in Molecular, Structural, and Functional Imaging; 1203621, on April 4, 2022. Disclaimer: This work was supported in part through the Laulima Government Solutions, LLC prime contract with the U.S. National Institute of Allergy and Infectious Diseases (NIAID) under Contract No. HHSN272201800013C (M.A.C., P.J.S., and B.Y.L.) and Kelly Services' contract with NIAID under Contract No. 75N93019D00027 (V.M.). J.H.L., C.L.F., and J.H.K. performed this work as employees of Tunnell Government Services (TGS), a subcontractor of Laulima Government Solutions, LLC under Contract No. HHSN272201800013C.

This work was also supported in part with federal funds from the National Cancer Institute (NCI), National Institutes of Health (NIH), under Contract No. 75N910D00024, Task Order No. 75N91019F00130. (I.C. and J.S. were supported by the Clinical Monitoring Research Program Directorate, Frederick National Lab for Cancer Research, sponsored by NCI.) This project was also partially funded by the Center for Infectious Disease Imaging (CIDI), Clinical Center, National Institute of Health (NIH) (S.R. and W.T.C.). The research was partially completed as part of the NIH Intramural Research Training Award (IRTA) Program through the NIAID (D.B.). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Health and Human Services or of the institutions and companies affiliated with the authors. The study protocol was reviewed and approved by the NIH/NIAID/DCR/Integrated Research Facility at Fort Detrick Animal Care and Use Committee in compliance with all applicable federal regulations governing the protection of animals and research.

Acknowledgments

The authors declare no conflicts of interest. The authors want to thank Oscar Rojas and the Comparative Medicine team (NIAID IRF-Frederick) for handling the animals during the studies, Claudia Mani (NIAID IRF-Frederick) for reviewing the manuscript, Jiro Wada (NIAID IRF-Frederick) for figure preparation and layout, and Anya Crane (NAIAD IRF-Frederick) for critically editing the manuscript.

References

1. WHO, *COVID-19 Weekly Epidemiological Update, Edition 91*, pp. 1–10, World Health Organization (2022).
2. B. Hu et al., "Characteristics of SARS-CoV-2 and COVID-19," *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
3. R. K. Gupta, "Will SARS-CoV-2 variants of concern affect the promise of vaccines?" *Nat. Rev. Immunol.* **21**(6), 340–341 (2021).
4. T. Ai et al., "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology* **296**(2), E32–E40 (2020).
5. M. D. Johansen et al., "Animal and translational models of SARS-CoV-2 infection and COVID-19," *Mucosal Immunol.* **13**(6), 877–891 (2020).
6. C. L. Finch et al., "Characteristic and quantifiable COVID-19-like abnormalities in CT- and PET/CT-imaged lungs of SARS-CoV-2-infected crab-eating macaques (*Macaca fascicularis*)," bioRxiv, 1–48 (2020).
7. J. K. Willmann et al., "Molecular imaging in drug development," *Nat. Rev. Drug Discov.* **7**(7), 591–607 (2008).
8. L. Keith et al., *Preclinical Imaging in BSL-3 and BSL-4 Environments: Imaging Pathophysiology of Highly Pathogenic Infectious Diseases*, AAPS Advances in the Pharmaceutical Sciences, Vol. **10**, Springer, New York (2014).
9. M. A. Castro et al., "Quantification of axillary lymphadenopathy from CT images of filovirus infections in non-human primates: sensitivity and evaluation of radiomics-based methods," *Proc. SPIE* **11317**, 1131725 (2020).
10. M. A. Castro et al., "Evaluation of R2 *as a noninvasive endogenous imaging biomarker for Ebola virus liver disease progression in nonhuman primates," *Proc. SPIE* **11317**, 1131726 (2020).
11. J. H. Lee et al., "The use of large-particle aerosol exposure to Nipah Virus to mimic human neurological disease manifestations in the African green monkey," *J. Infect. Dis.* **221**(Supplement_4), S419–S430 (2020).
12. W. Schreiber-Stainthorp et al., "Longitudinal in vivo imaging of acute neuropathology in a monkey model of Ebola virus infection," *Nat. Commun.* **12**, 2855 (2021).

13. K. Nakamura et al., "Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks," *Radiology* **214**(3), 823–830 (2000).
14. X. W. Xu et al., "Development of an improved CAD scheme for automated detection of lung nodules in digital chest images," *Med. Phys.* **24**(9), 1395–1403 (1997).
15. Y. Zhu et al., "Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography," *J. Digit. Imaging* **23**(1), 51–65 (2010).
16. M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst. Man Cybern.* **19**(5), 1264–1274 (1989).
17. A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognit. Lett.* **11**(6), 415–419 (1990).
18. M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graph. Image Process.* **4**, 172–179 (1975).
19. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973).
20. C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Comput. Vis. Graph. Image Process.* **23**, 341–352 (1983).
21. P. Lambin et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer* **48**(4), 441–446 (2012).
22. P. Grossmann et al., "Defining the biological basis of radiomic phenotypes in lung cancer," *Elife* **6**, e23421 (2017).
23. H. J. W. L. Aerts et al., "Defining a radiomic response phenotype: a Pilot study using targeted therapy in NSCLC," *Sci. Rep.* **6**, 33860 (2016).
24. E. Linning et al., "Radiomics for classification of lung cancer histological subtypes based on nonenhanced computed tomography," *Acad. Radiol.* **26**(9), 1245–1252 (2019).
25. X. Fave et al., "Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer," *Sci. Rep.* **7**, 588 (2017).
26. T. Wang et al., "Radiomics signature predicts the recurrence-free survival in stage I non-small cell lung cancer," *Ann. Thorac. Surg.* **109**(6), 1741–1749 (2020).
27. Q. Cai et al., "A model based on CT radiomic features for predicting RT-PCR becoming negative in coronavirus disease 2019 (COVID-19) patients," *BMC Med. Imaging* **20**(118), 1–10 (2020).
28. H. Chao et al., "Integrative analysis for COVID-19 patient outcome prediction," *Med. Image Anal.* **67**, 101844 (2021).
29. H. Chen et al., "A CT-based radiomics nomogram for predicting prognosis of coronavirus disease 2019 (COVID-19) radiomics nomogram predicting COVID-19," *Br. J. Radiol.* **94**(1117), 20200634 (2021).
30. Q. Wu et al., "Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19," *Theranostics* **10**(16), 7231–7244 (2020).
31. J. Huang et al., "CT-based radiomics helps to predict residual lung lesions in COVID-19 patients at three months after discharge," *Diagnostics* **11**(10), 1814 (2021).
32. X. Fang et al., "Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2," *Eur. Radiol.* **30**(12), 6888–6901 (2020).
33. J. Guiot et al., "Development and validation of an automated radiomic CT signature for detecting COVID-19," *Diagnostics* **11**(1), 41 (2021).
34. L. Fu et al., "A novel machine learning-derived radiomic signature of the whole lung differentiates stable from progressive COVID-19 infection: a retrospective cohort study," *J. Thorac. Imaging* **35**(6), 361–368 (2020).
35. C. Xie et al., "Discrimination of pulmonary ground-glass opacity changes in COVID-19 and non-COVID-19 patients using CT radiomics analysis," *Eur. J. Radiol. Open* **7**, 100271 (2020).
36. H. Liu et al., "CT radiomics facilitates more accurate diagnosis of COVID-19 pneumonia: compared with CO-RADS," *J. Transl. Med.* **19**(29), 1–12 (2021).
37. N. Yang et al., "Diagnostic classification of coronavirus disease 2019 (COVID-19) and other pneumonias using radiomics features in CT chest images," *Sci. Rep.* **11**, 17885 (2021).

38. J. Qiu et al., "A radiomics signature to quantitatively analyze COVID-19-infected pulmonary lesions," *Interdiscip. Sci.* **13**(1), 61–72 (2021).
39. L. G. Sapienza et al., "Risk of in-hospital death associated with Covid-19 lung consolidations on chest computed tomography - a novel translational approach using a radiation oncology contour software," *Eur. J. Radiol. Open* **8**, 100322 (2021).
40. G. Wu et al., "Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study," *Eur. Respir. J.* **56**(2), 2001104 (2020).
41. ISBI, "Image biomarker standardisation initiative - reference manual," (2019). https://ibsi.readthedocs.io/en/latest.
42. Z. Ke et al., "Radiomics analysis enables fatal outcome prediction for hospitalized patients with coronavirus disease 2019 (COVID-19)," *Acta Radiol.* **63**(3), 319–327 (2022).
43. Y. Huang et al., "CT-based radiomics combined with signs: a valuable tool to help radiologist discriminate COVID-19 and influenza pneumonia," *BMC Med. Imaging* **21**(1), 31 (2021).
44. X. Zhang et al., "A deep learning integrated radiomics model for identification of coronavirus disease 2019 using computed tomography," *Sci. Rep.* **11**, 3938 (2021).
45. L. Wang et al., "Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable radiomics features," *Eur. J. Radiol.* **136**, 109552 (2021).
46. R. Cattell, S. Chen, and C. Huang, "Robustness of radiomic features in magnetic resonance imaging: review and a phantom study," *Vis. Comput. Ind. Biomed. Art* **2**(1), 19 (2019).
47. A. Traverso et al., "Repeatability and reproducibility of radiomic features: a systematic review," *Int. J. Radiat. Oncol. Biol. Phys.* **102**(4), 1143–1158 (2018).
48. C. Xue et al., "Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review," *Quant. Imaging Med. Surg.* **11**(10), 4431–4460 (2021).
49. T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropr. Med.* **15**(2), 155–163 (2016).
50. J. E. Park et al., "Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives," *Korean J. Radiol.* **20**(7), 1124–1137 (2019).
51. S. S. Alahmari et al., "Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening," *IEEE Access* **6**, 77796–77806 (2018).
52. R. Berenguer et al., "Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters," *Radiology* **288**(2), 407–415 (2018).
53. A. Midya et al., "Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility," *J. Med. Imaging* **5**(1), 011020 (2018).
54. M. Vallières et al., "Responsible radiomics research for faster clinical translation," *J. Nucl. Med.* **59**(2), 189–193 (2018).
55. S. Rizzo et al., "Radiomics: the facts and the challenges of image analysis," *Eur. Radiol. Exp.* **2**(36), 1–8 (2018).
56. D. Vuong et al., "Comparison of robust to standardized CT radiomics models to predict overall survival for non-small cell lung cancer patients," *Med. Phys.* **47**(9), 4045–4053 (2020).
57. D. Mackin et al., "Matching and homogenizing convolution kernels for quantitative studies in computed tomography," *Invest. Radiol.* **54**(5), 288–295 (2019).
58. S. Denzler et al., "Impact of CT convolution kernel on robustness of radiomic features for different lung diseases and tissue types," *Br. J. Radiol.* **94**(1120), 20200947 (2021).
59. M. A. Castro et al., "Determination of reliable whole-lung CT features for robust standard radiomics and delta-radiomics analysis in a crab-eating macaque model of COVID-19: stability and sensitivity analysis," *Proc. SPIE* **12036**, 1203621 (2022).
60. S. M. S. Reza et al., "Deep learning for automated liver segmentation to aid in the study of infectious diseases in nonhuman primates," *Acad. Radiol.* **28 Suppl 1**(Suppl 1), S37–S44 (2020).
61. J. J. M. van Griethuysen et al., "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.* **77**(21), e104–e107 (2017).

62. N. Papanikolaou, C. Matos, and D. M. Koh, "How to develop a meaningful radiomic signature for clinical use in oncologic patients," *Cancer Imaging* **20**(33), 1–10 (2020).

Marcelo A. Castro is a physicist and computational scientist serving as an imaging physicist (contractor) at the NIH NIAID DCR Integrated Research Facility at Fort Detrick, a BSL-4 facility in Frederick, Maryland. His specialization includes multi-modality quantitative image analysis, parametric mapping, radiomics, computational simulations, scientific programming, and data analysis and visualization for multiple models, organs, and diseases. Over 20 years he has published +55 scientific papers with +2,000 citations.

Syed Reza is a postdoctoral fellow at the NIH. His research focuses on machine-learning-driven computational modeling for medical image analysis, such as segmentation, classification, disease tracking, and growth prediction for affected organs in infectious disease analyses and brain lesions, tumors, and traumatic brain injury.

Winston T. Chu is a postdoctoral research fellow in the Department of Radiology and Imaging Sciences at the NIH Clinical Center and is associated with the NIAID Integrated Research Facility at Fort Detrick. His research focus is on the development of novel techniques driven by artificial intelligence to automatically segment and classify medical images of biosafety level 4 infectious diseases.

Dara Bradley is a master's student in medical physiology and biophysics at Case Western Reserve University. Prior to attending Case Western Reserve, she completed a postbaccalaureate fellowship at the NIH as an Intramural Research Trainee Award Fellow (2019 to 2021). Her research projects applied artificial intelligence and machine learning methods to investigate questions within medical imaging of infectious disease research.

Ji Hyun Lee serves as a medical physicist at the Department of Radiology and Imaging Sciences at the NIH Clinical Center. She has supported multi-disciplinary investigators with the highest quality consultation, study design, image acquisition, and analysis services through hands-on, in-depth knowledge of multimodal imaging techniques/applications. Her research interests are in quantifying translational imaging through novel acquisition strategies, developing post-processing techniques in physiological imaging, and facilitating the transfer of molecular imaging-based techniques to the bedside.

Ian Crozier is an infectious diseases clinician-scientist at the Frederick National, Lab providing chief medical officer support to the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, MD. His role bridges the human clinical bedside and animal models of emerging high-threat infectious diseases. He has extensive experience at the Ebola virus disease outbreak bedside, including in ongoing clinical research efforts in Western Africa and the Democratic Republic of the Congo.

Philip J. Sayre is a research imaging technologist (contractor) with Laulima Government Solutions in support of the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, Maryland. His research is focused on PET/CT imaging of infectious disease in a BSL-4 setting. This work includes COVID-19, Ebola, Lassa, Middle East respiratory syndrome (MERS), Marburg, Nipah, monkeypox, and cowpox diseases.

Byeong Y. Lee is a biomedical imaging analyst (contractor) at the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, Maryland. His research aims to develop surrogate imaging biomarkers using *in vivo* multimodal medical imaging techniques, such as MRI, PET, and CT, as well as advanced imaging analysis methods, to aid in the evaluation of viral infectious disease models and identification of pathophysiology underlying the diseases, evaluation of antiviral therapies, and diagnostics.

Venkatesh Mani serves as a senior imaging scientist (contractor) at the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, Maryland. He specializes in the use of multimodality imaging such as MRI, CT, PET, and SPECT to evaluate the molecular biology

and pathogenesis of, and medical countermeasure development against, WHO Risk Group 4 pathogens. He has published over 100 peer-reviewed papers and has an h-index of 50.

Thomas C. Friedrich is a professor at the University of Wisconsin–Madison Department of Pathobiological Sciences. He studies why and how immune responses sometimes fail to protect us from acute and chronic diseases.

David H. O'Connor is University of Wisconsin Medical Foundation Professor of Pathology and Laboratory Medicine at the University of Wisconsin–Madison and Professorial Fellow at the University of Melbourne. His research focuses on the interplay between viral pathogenesis, immunity, and host genetics. He has been involved in the movement to accelerate the dissemination of scientific information during the Zika virus and COVID-19 pandemics.

Courtney L. Finch is Director of Pre-Clinical, Research and Development at Sabin Vaccine Institute, where she oversees animal studies associated with the advancement of two filovirus vaccines. She has more than a decade of experience studying primarily vaccines and therapeutics, including extensive animal model experience across numerous animal species and viral pathogens. Her focus has been on high-consequence pathogens in high-containment laboratory environments. She has authored publications across multiple disciplines, including medical imaging.

Gabriella Worwa is a study director and associate supervisor (contractor) at the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, Maryland. She specializes in the development and use of animal models for the study of WHO Risk Group 4 viruses.

Irwin M. Feuerstein is a board-certified diagnostic radiologist with extensive experience in cardiovascular imaging, computed tomography, and infectious disease imaging. He has worked with the National Institutes of Health (NIH), U.S. Department of Defense (DoD), and U.S. Food and Drug Administration (FDA) in a number of diagnostic, research, and regulatory capacities.

Jens H. Kuhn serves as a principal scientist and director of virology (contractor) at the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, Maryland. He specializes in the molecular biology and pathogenesis of, and medical countermeasure development against, WHO Risk Group 4 pathogens, evolutionary virology and virus taxonomy, and bioweapons defense. He has published 277 journal articles, 79 book chapters, and three books.

Jeffrey Solomon is an imaging scientist (contractor) at the NIH NIAID DCR Integrated Research Facility at Fort Detrick in Frederick, Maryland. In this role, he leads an artificial intelligence team that implements novel techniques to create predictive models and automate segmentation of medical images based on machine-learning principles. Working directly with radiologist colleagues, he consults on best-of-class quantitative image analysis methods to employ in infectious disease imaging research.