

Radiomic phenotyping of tuberculosis histopathology from computed tomography scans of non-human primates

Dmitry Cherezov

Emory University

Laura E. Via

National Institute of Allergy and Infectious Diseases

Zakariya Khaleel

Rutgers New Jersey Medical School

Edwin Klein

University of Pittsburgh

Robert Reiss

Rutgers New Jersey Medical School

Vincent Dong

University of Pennsylvania

H. Jacob Borish

University of Pittsburgh

Hee-Jeong Yang

National Institute of Allergy and Infectious Diseases

Danielle M. Weiner

National Institute of Allergy and Infectious Diseases

Michelle Sutphin

National Institute of Allergy and Infectious Diseases

Michaela K. Piazza

National Institute of Allergy and Infectious Diseases

Emmanuel Dayao

National Institute of Allergy and Infectious Diseases

Pauline Maiello

University of Pittsburgh

Alexander White

University of Pittsburgh

Charles A. Scanga

University of Pittsburgh

Philana Ling Lin

University of Pittsburgh School of Medicine

Clifton E. III Barry

National Institute of Allergy and Infectious Diseases

JoAnne Flynn

University of Pittsburgh

Anant Madabhushi

Emory University

Yingda Xie

y1x1@njms.rutgers.edu

Rutgers New Jersey Medical School

Article

Keywords: Machine learning, tuberculosis, radiology, diagnostic imaging

Posted Date: March 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-6822856/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Tuberculosis (TB) lesions display structural heterogeneity associated with disease status and progression. We explore the potential of radiomic features from computed tomography (CT) scans to classify TB lesions at a microscopic level.

Treated marmosets and untreated macaques infected with *Mycobacterium tuberculosis* underwent PET-CT imaging after a minimum of 12 weeks. Lesions were classified post-necropsy into necrotizing granulomas, non-necrotizing granulomas, inflammation beyond defined granulomas, and fibrotic granulomas. CT scans were segmented, radiomic features were extracted, and the top ten features were used to develop machine learning (ML) models for classification based on a one-versus-all approach. Top performing models were evaluated on a separate validation dataset.

151 lesions from macaques and 149 lesions from marmosets were identified. Top features included GrayLevelNonUniformity and SmallAreaHighGrayLevelEmphasis for both datasets. Linear discriminate analysis for macaques and adaptive boosting for marmosets averaged AUC-ROCs of 0.66-0.70 for discriminating lesion types, ranging from 0.45-0.52 for necrotizing granulomas to 0.91 for fibrotic granulomas.

Radiomics distinguished fibrotic from non-fibrotic TB granulomas in untreated macaques, potentially relevant for identifying disease progression risk. Classification of other lesion types was modest. Despite small, heterogeneous datasets across primate species, models performed consistently, supporting CT radiomics as a promising and trainable tool for non-invasive lesion phenotyping.

Introduction

Tuberculosis (TB) is characterized by extensive heterogeneity at both the individual and lesion levels, spanning variations in clinical presentation, disease progression, and treatment response. This heterogeneity underpins differences in pharmacokinetics, bacterial metabolism and host immune response [1–4], yet remains poorly captured by conventional diagnostic tools such as sputum culture and molecular testing which offer limited insight into the spatial and biological complexity of in vivo disease [5]. More precise characterization of this heterogeneity could enable earlier diagnosis, individualized risk stratification, and tailored treatment approaches [6, 7].

High resolution imaging modalities such as computed tomography (CT) provide a unique opportunity to characterize diverse lesion features in a non-invasive manner. Beyond simple visual interpretation, radiomics – a high throughput approach to extract quantitative features from radiologic images – enables systematic analysis of lesion texture, shape, and intensity patterns. Radiomic features have been applied in oncology to predict molecular and histopathologic attributes that distinguish benign and malignant lesions [8], stage and prognosticate tumors [9], and predict response to different therapies[10, 11].

Radiomics have been explored less extensively in infectious diseases, including a paucity of applications for characterizing heterogeneous TB lesion pathology that may be clinically informative. Early lesion features on CT and chest-X-rays, such as the presence or absence of fibrosis, have been linked to TB disease progression risk over the following months to years [12, 13]. Additionally, lesion microenvironments influence critical determinates of treatment response [6], such as bacterial physiologic states and persistence [14–16] and drug penetration [17, 18]. A high-resolution radiomic framework could provide a non-invasive biomarker of lesion phenotypes to guide research studies in improving TB treatment regimens and personalized clinical diagnoses.

In this study, we explored whether CT-derived radiomic features could predict TB lesion histopathology. Given the absence of large-scale, lesion-level PET/CT-histopathology mapped datasets in humans, we leveraged two distinct non-human primate cohorts—marmosets and macaques—to represent different spectra of TB pathology and to assess the robustness of radiomic classification across primate species. By training and validating machine learning classifiers separately in each

dataset, we aimed to explore the feasibility and generalizability of CT radiomics as a non-invasive tool for TB lesion phenotyping.

Methods

This project uses two existing databases of *Mycobacterium tuberculosis* (Mtb) infected non-human primate 2-deoxy-2-[¹⁸F] fluorodeoxyglucose positron emission tomography/computed tomography (PET/CT) scans from common marmosets at the NIAID Tuberculosis Research Section ('NIH') and macaques at the Flynn lab at University of Pittsburgh ('Pitt'). All direct interactions with animals occurred in the context of other studies [19–30]; no interactions with animals occurred as part of this study. All necessary permissions were obtained for the use of archived data and biological samples included in this study.

Common marmosets generally exhibit increased susceptibility to Mtb infection compared to macaques, with more rapid disease progression and pronounced pathology that recapitulate heterogeneity and complexity of lesions in human TB disease [31]. On the other hand, macaques, particularly cynomolgus macaques, display a spectrum of disease outcomes from latent to active tuberculosis, closely mirroring the spectrum of human infection patterns [32]. These differences are likely due to species-specific immunological differences [32, 33]. Taken together, these datasets comprise scans across disease stages from early to advanced, lesion types from single granulomas to coalescent and cavitory, and treated and untreated disease. The workflow for the acquisition and segmentation of scans, feature extraction, feature selection and supervised and unsupervised classification analyses are provided in Fig. 1.

Untreated macaques (University of Pittsburgh)

Pitt samples were selected from an existing set of lesions from 29 unvaccinated and untreated macaques (27 cynomolgus and 2 rhesus, mostly males of Chinese origin) for 14 TB immunologic studies conducted between 2010 and 2018 [22–30]. The macaques were previously infected intra-bronchially with 2.7–85 colony forming units (CFU) of Mtb strain Erdman and allowed to progress for 12–86 weeks prior to necropsy. CT scans were obtained within 3–12 days prior to necropsy. Scans were performed using the NeuroLogica CereTom Clinical CT and Siemens MicroPET scanners, with imaging acquisition protocol as detailed in White A et al [34]. Images were exported as raw DICOM files. During necropsy, individual lung lesions were plated and identified using the pre-necropsy CT image. The plated lesions were used to prepare histologic sections consisting of formalin fixed tissues embedded in paraffin, sectioned, and stained with hematoxylin and eosin. In total, 314 histologic sections from the 29 animals in 14 studies mapped to CT were available and scored for histologic classification. All experiments on live vertebrates were approved by the University of Pittsburgh Institutional Animal Care and Use Committee (IACUC protocol numbers 15126588, 11110045, 11090030, 15055811, 15035407, 1105870, 14043492, 12060181, 17091258, 12080653 and 1003622). Experiments were performed and reported in accordance with the ARRIVE guidelines.

Treated marmosets (NIH)

The NIH samples consisted of lesions from 44 common marmosets (*Callithrix jacchus*) exposed to a low dose of aerosolized Mtb strain H37Rv. The disease was allowed to progress for 7 to 8 weeks, and antibiotic treatment was given for 4 or 8 weeks. The animals were treated for one to two months with linezolid, one of five other investigational oxazolidinones, or rifampicin, isoniazid, pyrazinamide or moxifloxacin or a combination regimen of these four drugs to prevent mortality and achieve an infection time of at least 12 weeks so that fibrotic cuffs and other signs of an immunological response could occur. Some of the animals and lesions used in this analysis were also previously analyzed for drug-specific responses to treatment, although in those analyses only features such as lesion mean HU, mean SUV, total lesion glycolysis and the change in these features between time points were explored [21]. To be eligible for inclusion in the dataset, the animals had a pre-necropsy scan within 48 hours of euthanasia, labeled histological samples preserved, and CFU/lesion determined [31]. Scans were also performed using the NeuroLogica CereTom Clinical CT and Siemens MicroPET scanners aligned using a

common bed as with the Pitt scanner, with the same imaging acquisition protocol as Pitt detailed in White et al. [34] with the following exceptions: the NIH CT data was collected using 120 Kv, 5 mA/s, and the FDG PET scan began collection at 60 minutes post injection of the tracer as only two fields of view were necessary to cover the lung region of the marmoset in the Siemens Focus 220. After necropsy, there were 278 fixed hematoxylin and eosin-stained slides of classifiable TB lesions from the 44 animals studied. All live vertebrate procedures were performed in accordance with the recommendations of the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the NIAID Animal Care and Use Committee in Protocol LCIM-9 (Permit issued to NIH as A-4149-01). Both male and female marmosets ages 2 to 6 years old were included in the study.

Histological classification of TB lesions

All available TB lesion histologic sections were classified using a classification protocol for histologic scoring developed by a veterinary pathologist to capture TB lesion phenotypes shared across human and nonhuman primates [35] (Fig. 2, Supplementary materials appendix S1). The histological classes include necrotizing granulomas (type 1), non-necrotizing granulomas (type 2), architecturally unstructured granulomatous inflammation outside of defined granuloma structure formation (type 3), and fibrotic granulomas (type 4).

Archived fixed or digital slides were reviewed by pathologists who scored lesions according to one of four categories in the manual. Slides which contain multiple lesion types were scored according to the predominant lesion type. Slides that could not be scored due to missing lesions or unclear labeling were marked as indeterminate.

Segmentation of TB lesions from CT

TB lesions were manually segmented from CT scans using Amira (Thermo Fisher Scientific version 6.5) to create binary mask region of interest (ROI) for each TB lesion (See supplementary materials for protocol). Segmentation was performed by a trained individual guided by two individuals performing quality control. Individuals performing segmentation were blinded to lesion classification. DICOM files of scan images from the NIH marmosets and Pitt macaques were imported into Amira and segmentation of the ROIs were performed as follows: (i) 3D ROIs were created by wrapping or interpolating between manually segmented cross sections, (ii) ROIs were refined by removing non-lesion tissue (e.g. large vasculature/airways, mediastinal structures, chest wall), (iii) Any material with a Hounsfield (HU) lower than - 500 attributed to aerated, normal lung tissue was removed from the ROI. This segmentation process was performed for each lesion following a standard operating procedure (see Supplementary materials appendix S2). The segmentation was reviewed for quality by a separate reader to ensure protocol adherence, and inclusion of only TB lesions within the ROIs.

Separation of validation set

Data from both sites were randomly split into training and validation sets (D1 and D2 for Pitt, D3 and D4 for NIH). The split was performed at the primate level to prevent data leakage, ensuring that lesions from the same primate did not appear in both training and validation sets. The number of lesions and primates in each dataset (D1–D4) is detailed in Fig. 3.

Radiomic feature extraction

A total of 74 first and second-order radiomic features (Supplementary table S1) were extracted from the lesion data using PyRadiomics, an open-source radiomic extraction python package [36]. The 74 3D features included: (i) six first-order statistics features to describe the basic distribution of voxel intensities within the identified lesion ROIs [36], (ii) 22 Haralick feature [37] from a gray-level co-occurrence matrix that captures textural patterns that could encode variation in the different presentations of lesions and (iii) 46 additional texture features from the gray-level run length[38, 39] size zone [40]. PET images had larger voxel sizes compared to CT images (~ 10mm for PET vs. ~2mm for CT), resulting in imprecise alignment

with the CT-based ROI. Consequently, feature extraction from PET was limited to SUVmax because SUVmax is less affected by the partial volume effect, where PET voxels extend beyond the smaller CT-based lesion ROIs.

Lesion size exclusion criteria

The two groups of primates in the study allowed us to observe biological variability in lesion representation on CT images. The lesion sizes differed significantly between the two groups. For each lesion, we identified the axial slice with the largest region of interest (ROI). The median value ROI area at this slice (TB lesion area) is 220 pixels in marmosets and 20 pixels in macaques.

Since our study employed handcrafted features, one of the standard methods for quantifying image textures involves the use of a kernel with a size of 5×5 pixels [36]. In contrast, Gray Level Co-occurrence Matrix (GLCM) features rely on the statistical distribution of pixel intensities within the region of interest (ROI). Therefore, ROIs that were less than 5 pixels were excluded in the primary analysis due to concern of poor reliability and representation of the computed features [41].

Feature selection

To enhance model generalizability and minimize the risk of overfitting, we performed radiomic feature selection as a pre-processing step. Feature selection was carried out using the Minimum Redundancy Maximum Relevance (mRMR) algorithm to ensure that the selected features had the maximum relevance to the target variable while minimizing redundancy. This approach facilitates the identification of a subset of features that contribute the most to prediction accuracy, while simultaneously reducing the complexity of the model[42]. We evaluated the performance of the resulting models for the top 5 and the top 10 features.

Unsupervised analysis

t-Distributed Stochastic Neighbor Embeddings (tSNE) of normalized training data using the top 10 selected features was used to visualize the topology of the N-dimensional clustering in a three-dimensional space [43]. This was accomplished by constructing probability distributions such that similar high-dimensional samples have higher probabilities, and dissimilar samples have lower probabilities. These probabilities were then mapped onto a three-dimensional space for visualization. The resulting plots demonstrate the effectiveness of the selected feature representation in capturing the structure of the data in a reduced-dimensional space. tSNE clustering was performed in R, version 4.1.1, using the Rtsne package [44].

Supervised analysis

Supervised analysis was used to assess how well the model could generalize to new, unseen data and its ability to correctly identify the histologic class of a TB lesion. We used two different approaches to distinguish between multiple classes. The first approach employs a multi-class (MC) classification strategy [45], where a single model is trained to classify the input data into multiple classes. In this strategy, the classifier distinguishes quantitative features into distinct histological types, with each type encoded by a unique class label (e.g., histological type 1 is encoded as class 1, type 2 as class 2, etc.).

The second approach follows the "one-versus-all" (OvA) strategy [46]. Unlike the multi-class method, this strategy involves designing an individual classification model for each histological type. For instance, one model is dedicated to classifying histological type 1—where histological type 1 is encoded as class 1 and all other types as class 0—and this process is repeated for every histological type. Here we focus on the results from the OvA approach. Results from the MC approach are presented in the supplementary materials.

Two training strategies were employed: Leave-One-Out (LOO) cross-validation and standard training with a held-out validation set. The LOO method evaluates model performance and robustness by iteratively using every sample as a test case once, while the remaining data is used to train the model. This approach provides an unbiased accuracy estimate and maximizes

data utility—particularly valuable for smaller datasets. Only the training datasets (D1 for Pittsburgh data and D3 for NIH) were used in LOO to maintain consistency with the validation split. LOO cross-validation was implemented as follows:

1. In each iteration, one sample was held out as test data, while the rest trained the model.
2. The mRMR algorithm dynamically selected the top 5 and top 10 most discriminative features per iteration, ensuring adaptive feature filtering.
3. This process was repeated across all samples, systematically assessing which texture features best differentiated lesion types.

By isolating test data in every iteration, LOO mitigates overfitting and offers a rigorous evaluation of feature generalizability.

Following the Leave-One-Out (LOO) cross-validation performed on the training datasets (D1 for Pitt and D3 for NIH), we proceeded to train final diagnostic models using the complete D1 and D3 datasets. These models incorporated the most discriminative features identified during the LOO analysis. The trained models were then evaluated on the separate validation datasets (D2 for Pitt and D4 for NIH) to assess their diagnostic performance. This approach ensured that the validation was performed on entirely independent data, preventing any information leakage and providing a reliable measure of the models' generalization capability.

Statistical analysis

For both the supervised and unsupervised analysis methods, the following algorithms were used for model training: random forests (RF), support vector machine (SVM), adaptive boosting (Ada-Boost), and linear discriminant analysis (LDA). Results were quantified by calculating overall accuracy, weighted accuracy, and the area under the receiving operating characteristic curve (ROC-AUC) for each trained classifier, calculating these metrics on an OvA basis.

The methodology described above, including LOO cross-validation and training with validation, was implemented and examined for two feature selection setups: selecting the top 5 and the top 10 best features. The top 10 features were chosen for the primary analysis because they gave the best performance for the training dataset in terms of AUC for the supervised models. Results of the analysis using the top 5 features are presented in the supplementary materials.

Results

Among 556 segmented lesions from 29 Pitt macaques and 488 segmented lesions from 44 NIH marmosets annotated on PET/CT, 151 lesions from 29 macaques and 149 lesions from 41 marmosets matched with associated histopathology class labels and met the lesion ROI size criteria (Fig. 3). Among these, 65 lesions from 15 marmosets and 27 lesions from 5 macaques were randomly set aside in an independent dataset (D2 and D4 respectively). The remaining eligible dataset was assigned to datasets D1 and D3.

Pitt dataset

The 150 untreated macaque lesions in the Pitt dataset used in the analysis were compromised of 76 (50.7%) type 1 lesions, 29 (19.3%) type 2 lesions, and 45 (30.0%) type 4 lesions. A single type 3 lesion representing granulomatous inflammation outside of the defined granuloma structure formation was removed from the analysis due to the paucity of type 3 lesions leading to severe class imbalance. Class imbalance in machine learning occurs when the distribution of samples across different classes is highly uneven, meaning some classes have significantly fewer examples than others [47]. This imbalance can harm model performance by causing the algorithm to become biased toward the majority class [48], leading to poor predictive accuracy for the underrepresented (but often critical) minority class, such as rare diseases in medical diagnosis or fraud cases in financial transactions. Therefore, no judgments can be made about separability of type 3 lesions in the Pitt dataset.

Selected features from the training/test set are shown in Table S2. Experiments were done with both the OvA and the MC classification approach, with similar results achieved using both approaches. Unsupervised tSNE clustering results of the training dataset are shown (Fig. 4a). The LDA model was selected as the best-performing model for the leave-one-out approach, with an average AUC of 0.66 and a balanced accuracy of 52.76% (Fig. 4b).

Applied to the independent validation set, the LDA model yielded AUCs of 0.52 (95% CI = 0.26–0.77) for type 1 lesions, 0.69 (95% CI = 0.48–0.88) for type 2 lesions, and 0.91 (95% CI = 0.77–1.00) for type 4 lesions.

NIH dataset

The 149 treated marmoset lesions in the NIH dataset were comprised of 31 (20.8%) type 1 lesions, 28 (18.8%) type 2 lesions, 60 (40.3%) type 3 lesions, and 30 (20.1%) type 4 lesions. Selected features are shown in Table S2. Like the analysis of the Pitt dataset, the OvA and MC approaches yielded comparable results. Unsupervised tSNE clustering results of the training dataset are shown (Fig. 5a). Supervised classification results demonstrate a higher balanced accuracy but lower AUC compared to the results from the Pitt dataset (Fig. 5b and 5c). The Ada-boost model was selected as the best-performing model for the training set, with an average AUC of 0.70 and a balanced accuracy of 38.8% (Fig. 5b). Applied to the validation set, the Ada-boost model yielded AUCs of 0.47 (95% CI = 0.29–0.65) for type 1 lesions, 0.58 (95% CI = 0.38–0.78) for type 2 lesions, 0.70 (95% CI = 0.56–0.83) for type 3 lesions, and 0.65 (95% CI = 0.50–0.79) for type 4 lesions.

Discussion

We applied radiomic feature extraction to identify shape- and texture-based characteristics that differentiate TB lesion histopathologic characteristics across two primate species, with and without treatment. Radiomic models have been widely used in cancer research and more recently in TB diagnosis, including to distinguish TB from other respiratory conditions [49–58] and predict drug resistance [59, 60], but have not looked at the resolution of histopathology. To our knowledge, this is the first application of radiomics to discriminate TB lesion pathology at a microscopic level. Using these extracted features and existing experimental data from a limited cohort of untreated macaques and treated marmosets, we trained machine learning models that exhibited modest performance in classifying TB lesion histology types among a small and biologically diverse non-human primate dataset, supporting the feasibility and trainability of this approach with further refinement on existing datasets. Notably, the models distinguished fibrotic from non-fibrotic granulomas with high fidelity in untreated macaques, highlighting the potential for radiomics to resolve biologically meaningful lesion types that may relate to clinical risk.

Compared to macaques, marmosets exhibited larger and more extensive lesions. Model performance metrics were generally consistent across species, feature sets, and classifier types, suggesting that the approach does not require extensive fine-tuning to yield stable results. However, performance varied by lesion type in the independent validation sets. Specifically, necrotizing granulomas (Type I lesions) were the most difficult to classify, with AUC ROC values close to 0.5 in both datasets, and were often misclassified as Type IV lesions—or as Type III lesions in the NIH dataset. In contrast, Type IV lesion predictions were more accurate in untreated macaques, with an AUC ROC of 0.9 [95% CI 0.77–1.00], than in treated marmosets, where the AUC ROC dropped to 0.47 [95% CI 0.29–0.65]. This discrepancy may reflect treatment-induced changes in lesion phenotype among marmosets, which could obscure distinctions present in untreated disease. Treated animals may also have shown a greater prevalence of mixed histopathologic features, which were categorized based on the predominant phenotype. In clinical imaging surveys of chest X-ray and PET-CT among individuals with risks for TB exposure, non-fibrotic (“active”) TB-like lesions were several times more likely to progress to active TB than fibrotic (“inactive”) lesions [12, 13]. Therefore, the ability to distinguish fibrotic (Type IV) lesions from non-fibrotic types in untreated individuals may become a valuable biomarker for predicting and characterizing early TB before individuals develop illness or transmit to others. Notably, recent and ongoing studies are now applying CT and chest-X-ray imaging in high-risk human populations to identify early TB, offering an opportunity to test and refine this classifier while informing training of new clinical models of TB progression risk.

Interestingly, SUVmax was not among the top-selected features in either dataset, indicating that classification was driven by CT-based radiomics rather than PET metabolic activity. The contribution of SUVmax may have also been limited by the considerably lower PET resolution compared to the CT-based ROI. Rather, features associated with structural organization and gray-level intensity variations were more predictive, reflecting the structural complexity of lesions. Features such as Small Area High Gray Level Emphasis were higher in inflammation beyond granulomas (Type 3; marmosets), potentially reflecting greater fragmentation and high-intensity structures with dense cellular infiltration, while fibrotic granulomas had lower values potentially reflecting more uniform collagen deposition. In both datasets, necrotizing granulomas (Type 1) and fibrotic granulomas (Type 4) also tended to have higher Contrast values, potentially reflecting sharp intensity transitions in necrotic cores and fibrotic lesions, whereas inflammation beyond granulomas (Type 3; marmosets) tended to have lower contrast, potentially due to more diffuse inflammatory involvement. These findings suggest that radiomic features, particularly those related to small-scale heterogeneity and intensity transitions, may be relevant in distinguishing TB pathology across species and treatment conditions.

Early data from large autopsy studies have identified granulomas with culturable Mtb among individuals who died from non-TB causes, suggesting a prevalence and range of clinically-occult pathology that may range from self-resolution to requiring various preventive interventions [13]. The ability to phenotype lesions in-vivo would provide an opportunity to characterize the range of minimal TB disease in contacts or high-risk individuals that herald a likelihood of progression to bacteriologically positive disease. Additionally, in-vivo lesion phenotyping may also provide a useful tool to personalize treatment approaches and design host-directed or anti-microbial drug regimens effective across the range of lesion microenvironments [3]. This high-resolution imaging-based treatment response tool could apply to sputum paucibacillary and drug-resistant TB where treatment monitoring methods are limited or critical.

There were several limitations of the study. Our decision not to analyze a combined NIH and Pitt dataset due to the presence of batch effects highlights an important limitation: the differences in acquisition related factors can have a substantial impact on the corresponding radiomic features, even in CT scans[61, 62]. Therefore, models were separately generated for each dataset. Additionally, the study used existing datasets collected from different facilities, primate species, and infection/treatment protocols for different primary objectives, conditions leading to substantial heterogeneity and reduced sample sizes for each group in part because not all observed lesions in the PET/CT scans were captured in a corresponding histological section. As a result of the small sample sizes, the analysis was restricted to traditional classification methods. Deep learning approaches such as convolutional neural networks or vision transformers were not explored in this study. Acquiring larger datasets may allow for the use of deep learning models to validate our results. Future work to explore methods of normalizing the extracted features across imaging acquisition sites such as ComBat batch effect adjustment [63] are also relevant to verify these findings. Nevertheless, these limitations may be viewed as a representation of the diversity of human pathology from early to later disease stages and supports the ability of radiomic classifiers to profile TB lesions to microscopic resolutions among vastly different disease states. Finally, a substantial proportion of the macaque lesions were small (< 5 voxels in the largest slice) which did not meet the run length of several radiomic features and were thus excluded. Therefore, in very small lesions (e.g., early lesions or certain animal models), there may be a limit to what the radiomic models can reliably measure.

Altogether, these findings highlight the potential of radiomics to non-invasively classify histologic TB lesion types, with consistent performance for fibrotic lesions and modest results for others. With expanding access to annotated experimental and clinical datasets, model refinement may improve classification of other lesion types and advance lesion-level phenotyping tools for clinical and research applications in TB pathogenesis, treatment strategies, and detection of early TB and TB across the disease spectrum.

Declarations

Data Availability

All extracted radiomic feature data generated and/or analyzed during this study are included in this published article (and its Supplementary Materials files). PET/CT scans are available from the corresponding author on reasonable request.

Acknowledgements

Funding was provided in part by the Bill and Melinda Gates Foundation through OPP053284 (PLL), OPP1162695 and OPP1024021 (CEB), OPP1034408 and INV020435 (JF), Grand Challenges – Annual Meeting “Call to Action” (LX, JF), and in part by the Division of Intermural Research, NIAID, NIH (CEB and LEV). We thank the Comparative Medicine Branch of NIAID, NIH for clinical care of the marmosets and appreciate the technical expertise of Emmanuel Dayao DVM, and Becky Sloan, and of NIAID.

Author Contributions

EK, LEV, HY performed lesion analysis. LEV, DMW, MS, MP, ED, HJB, AW, PM, CS, PLL collected the PET/CT studies, labeled scans, catalogued lesions, and performed necropsies. DC and VD performed the machine learning analysis under the supervision of AM. ZK performed segmentations, data cleaning, data analysis and manuscript writing. RR performed data analysis and manuscript writing. JF and CEB supervised the non-human primate studies and helped conceptualize the original study design. YLX conceptualized and supervised the study design, analyses, and manuscript writing.

Competing Interests

The authors declare no competing interests.

References

1. Cadena, A.M., S.M. Fortune, and J.L. Flynn, *Heterogeneity in tuberculosis*. Nat Rev Immunol, 2017. **17**(11): p. 691-702.
2. Dhar, N., J. McKinney, and G. Manina, *Phenotypic Heterogeneity in Mycobacterium tuberculosis*. Microbiol Spectr, 2016. **4**(6).
3. Lenaerts, A., C.E. Barry Iii, and V. Dartois, *Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses*. Immunological Reviews, 2015. **264**(1): p. 288-307.
4. Lin, P.L., et al., *Radiologic Responses in Cynomolgus Macaques for Assessing Tuberculosis Chemotherapy Regimens*. Antimicrob Agents Chemother, 2013. **57**(9): p. 4237-4244.
5. Chen, X. and T.Y. Hu, *Strategies for advanced personalized tuberculosis diagnosis: Current technologies and clinical approaches*. Precis Clin Med, 2021. **4**(1): p. 35-44.
6. Xie, Y.L., et al., *Fourteen-day PET/CT imaging to monitor drug combination activity in treated individuals with tuberculosis*. Sci Transl Med, 2021. **13**(579).
7. Walter, N.D., et al., *Lung microenvironments harbor Mycobacterium tuberculosis phenotypes with distinct treatment responses*. Antimicrob Agents Chemother, 2023. **67**(9): p. e0028423.
8. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data*. Radiology, 2016. **278**(2): p. 563-77.
9. Aerts, H.J.W.L., et al., *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach*. Nature Communications, 2014. **5**(1): p. 4006.
10. Jain, P., et al., *Novel Non-Invasive Radiomic Signature on CT Scans Predicts Response to Platinum-Based Chemotherapy and Is Prognostic of Overall Survival in Small Cell Lung Cancer*. Front Oncol, 2021. **11**: p. 744724.
11. Bera, K., et al., *Predicting cancer outcomes with radiomics and artificial intelligence in radiology*. Nat Rev Clin Oncol, 2022. **19**(2): p. 132-146.
12. Esmail, H., et al., *High resolution imaging and five-year tuberculosis contact outcomes*. medRxiv, 2023.

13. Sossen, B., et al., *The natural history of untreated pulmonary tuberculosis in adults: a systematic review and meta-analysis*. The Lancet Respiratory Medicine, 2023. **11**(4): p. 367-379.
14. Lavin, R.C. and S. Tan, *Spatial relationships of intra-lesion heterogeneity in Mycobacterium tuberculosis microenvironment, replication status, and drug efficacy*. PLoS Pathog, 2022. **18**(3): p. e1010459.
15. Gold, B. and C. Nathan, *Targeting Phenotypically Tolerant Mycobacterium tuberculosis*. Microbiol Spectr, 2017. **5**(1).
16. Barry, C.E., et al., *The spectrum of latent tuberculosis: rethinking the biology and intervention strategies*. Nature Reviews Microbiology, 2009. **7**(12): p. 845-855.
17. Prideaux, B., et al., *The association between sterilizing activity and drug distribution into tuberculosis lesions*. Nature Medicine, 2015. **21**(10): p. 1223-1227.
18. Dartois, V., *The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells*. Nature Reviews Microbiology, 2014. **12**(3): p. 159-167.
19. Budak, M., et al., *A systematic efficacy analysis of tuberculosis treatment with BPAL-containing regimens using a multiscale modeling approach*. CPT Pharmacometrics Syst Pharmacol, 2024. **13**(4): p. 673-685.
20. Boshoff, H.I.M., et al., *Mtb-Selective 5-Aminomethyl Oxazolidinone Prodrugs: Robust Potency and Potential Liabilities*. ACS Infectious Diseases, 2024. **10**(5): p. 1679-1695.
21. Greenstein T, V.L., Moraes MP, Weiner DM, et al., *PET/CT multivariate tuberculosis treatment response profiles in marmosets unify disparate preclinical biomarkers*. Sci Transl Med, 2025.
22. Maiello, P., et al., *Rhesus Macaques Are More Susceptible to Progressive Tuberculosis than Cynomolgus Macaques: a Quantitative Comparison*. Infect Immun, 2018. **86**(2).
23. Ganchua, S.K.C., et al., *Lymph nodes are sites of prolonged bacterial persistence during Mycobacterium tuberculosis infection in macaques*. PLoS Pathog, 2018. **14**(11): p. e1007337.
24. Lin, P.L., et al., *PET CT Identifies Reactivation Risk in Cynomolgus Macaques with Latent M. tuberculosis*. PLoS Pathog, 2016. **12**(7): p. e1005739.
25. Medrano, J.M., et al., *Characterizing the Spectrum of Latent Mycobacterium tuberculosis in the Cynomolgus Macaque Model: Clinical, Immunologic, and Imaging Features of Evolution*. J Infect Dis, 2023. **227**(4): p. 592-601.
26. Coleman, M.T., et al., *Early Changes by (18)Fluorodeoxyglucose positron emission tomography coregistered with computed tomography predict outcome after Mycobacterium tuberculosis infection in cynomolgus macaques*. Infect Immun, 2014. **82**(6): p. 2400-4.
27. Gideon, H.P., et al., *Variability in tuberculosis granuloma T cell responses exists, but a balance of pro- and anti-inflammatory cytokines is associated with sterilization*. PLoS Pathog, 2015. **11**(1): p. e1004603.
28. Rodgers, M.A., et al., *Preexisting Simian Immunodeficiency Virus Infection Increases Susceptibility to Tuberculosis in Mauritian Cynomolgus Macaques*. Infect Immun, 2018. **86**(12).
29. Lin, P.L., et al., *Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing*. Nat Med, 2014. **20**(1): p. 75-9.
30. Darrah, P.A., et al., *Prevention of tuberculosis in macaques after intravenous BCG immunization*. Nature, 2020. **577**(7788): p. 95-102.
31. Via Laura, E., et al., *Differential Virulence and Disease Progression following Mycobacterium tuberculosis Complex Infection of the Common Marmoset (Callithrix jacchus)*. Infection and Immunity, 2013. **81**(8): p. 2909-2919.
32. Scanga, C.A. and J.L. Flynn, *Modeling tuberculosis in nonhuman primates*. Cold Spring Harb Perspect Med, 2014. **4**(12): p. a018564.
33. Peña Juliet, C. and W.-Z. Ho, *Non-Human Primate Models of Tuberculosis*. Microbiology Spectrum, 2016. **4**(4): p. 10.1128/microbiolspec.tb2-0007-2016.
34. White, A.G., et al., *Analysis of 18FDG PET/CT Imaging as a Tool for Studying Mycobacterium tuberculosis Infection and Treatment in Non-human Primates*. J Vis Exp, 2017(127).

35. Leong, F.J., Dartois, V., & Dick, T. (Eds.), *A Color Atlas of Comparative Pathology of Pulmonary Tuberculosis (1st ed.)*. 2010: CRC Press.
36. van Griethuysen, J.J.M., et al., *Computational Radiomics System to Decode the Radiographic Phenotype*. Cancer Research, 2017. **77**(21): p. e104-e107.
37. Haralick, R.M., K. Shanmugam, and I. Dinstein, *Textural Features for Image Classification*. IEEE Transactions on Systems, Man, and Cybernetics, 1973. **SMC-3**(6): p. 610-621.
38. Galloway, M.M., *Texture analysis using gray level run lengths*. Computer Graphics and Image Processing, 1975. **4**(2): p. 172-179.
39. Chu, A., C.M. Sehgal, and J.F. Greenleaf, *Use of gray value distribution of run lengths for texture analysis*. Pattern Recognition Letters, 1990. **11**(6): p. 415-419.
40. Thibault, G., et al., *Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification*, in *10th International Conference on Pattern Recognition and Information Processing*. 2009.
41. Jensen, L.J., et al. *Stability of Radiomic Features across Different Region of Interest Sizes—A CT and MR Phantom Study*. Tomography, 2021. **7**, 238-252 DOI: 10.3390/tomography7020022.
42. van Timmeren, J.E., et al., *Feature selection methodology for longitudinal cone-beam CT radiomics*. Acta Oncol, 2017. **56**(11): p. 1537-1543.
43. Van der Maaten, L. and G. Hinton, *Visualizing data using t-SNE*. Journal of machine learning research, 2008. **9**(11).
44. Krijthe, J.H., *Rtsne: T-distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation*. 2015.
45. Aly, M., *Survey on multiclass classification methods*. Neural Netw, 2005. **19**(1-9): p. 2.
46. Rifkin, R. and A. Klautau, *In defense of one-vs-all classification*. Journal of machine learning research, 2004. **5**(Jan): p. 101-141.
47. Guo, X., et al. *On the Class Imbalance Problem*. in *2008 Fourth International Conference on Natural Computation*. 2008.
48. Japkowicz, N. and S. Stephen, *The class imbalance problem: A systematic study*. Intell. Data Anal., 2002. **6**(5): p. 429–449.
49. Li, P., et al., *A CT-based radiomics predictive nomogram to identify pulmonary tuberculosis from community-acquired pneumonia: a multicenter cohort study*. Frontiers in Cellular and Infection Microbiology, 2024. **14**.
50. Zhang, X., et al., *Deep learning PET/CT-based radiomics integrates clinical data: A feasibility study to distinguish between tuberculosis nodules and lung cancer*. Thorac Cancer, 2023. **14**(19): p. 1802-1811.
51. Li, Y., et al., *Machine learning-based radiomics to distinguish pulmonary nodules between lung adenocarcinoma and tuberculosis*. Thoracic Cancer, 2024. **15**(6): p. 466-476.
52. Thattaamuriyil Padmakumari, L., et al., *The Role of Chest CT Radiomics in Diagnosis of Lung Cancer or Tuberculosis: A Pilot Study*. Diagnostics, 2022. **12**(3): p. 739.
53. Hu, Y., et al., *Lung CT-based multi-lesion radiomic model to differentiate between nontuberculous mycobacteria and Mycobacterium tuberculosis*. Med Phys, 2025. **52**(2): p. 1086-1095.
54. Li, H.L., et al., *Multimodal machine learning-based model for differentiating nontuberculous mycobacteria from mycobacterium tuberculosis*. Front Public Health, 2025. **13**: p. 1470072.
55. Li, P., et al., *A CT-based radiomics predictive nomogram to identify pulmonary tuberculosis from community-acquired pneumonia: a multicenter cohort study*. Front Cell Infect Microbiol, 2024. **14**: p. 1388991.
56. Zhou, L., et al., *A retrospective study differentiating nontuberculous mycobacterial pulmonary disease from pulmonary tuberculosis on computed tomography using radiomics and machine learning algorithms*. Ann Med, 2024. **56**(1): p. 2401613.
57. Yan, Q., et al., *CT-based radiomics analysis of consolidation characteristics in differentiating pulmonary disease of non-tuberculous mycobacterium from pulmonary tuberculosis*. Exp Ther Med, 2024. **27**(3): p. 112.

58. Wei, S., et al., *Differentiating mass-like tuberculosis from lung cancer based on radiomics and CT features*. Translational Cancer Research, 2021. **10**(10): p. 4454-4463.

59. Jiang, F., et al., *A CT-based radiomics analyses for differentiating drug-resistant and drug-sensitive pulmonary tuberculosis*. BMC Med Imaging, 2024. **24**(1): p. 307.

60. Li, Y., et al., *Radiomics analysis of lung CT for multidrug resistance prediction in active tuberculosis: a multicentre study*. Eur Radiol, 2023. **33**(9): p. 6308-6317.

61. Patrick, L., et al., *Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images*. Journal of Medical Imaging, 2016. **3**(4): p. 047502.

62. Cherezov, D., et al., *Rank acquisition impact on radiomics estimation (AcquIRE) in chest CT imaging: A retrospective multi-site, multi-use-case study*. Comput Methods Programs Biomed, 2024. **244**: p. 107990.

63. Zhang, Y., G. Parmigiani, and W.E. Johnson, *ComBat-seq: batch effect adjustment for RNA-seq count data*. NAR Genomics and Bioinformatics, 2020. **2**(3): p. lqaa078.

Figures

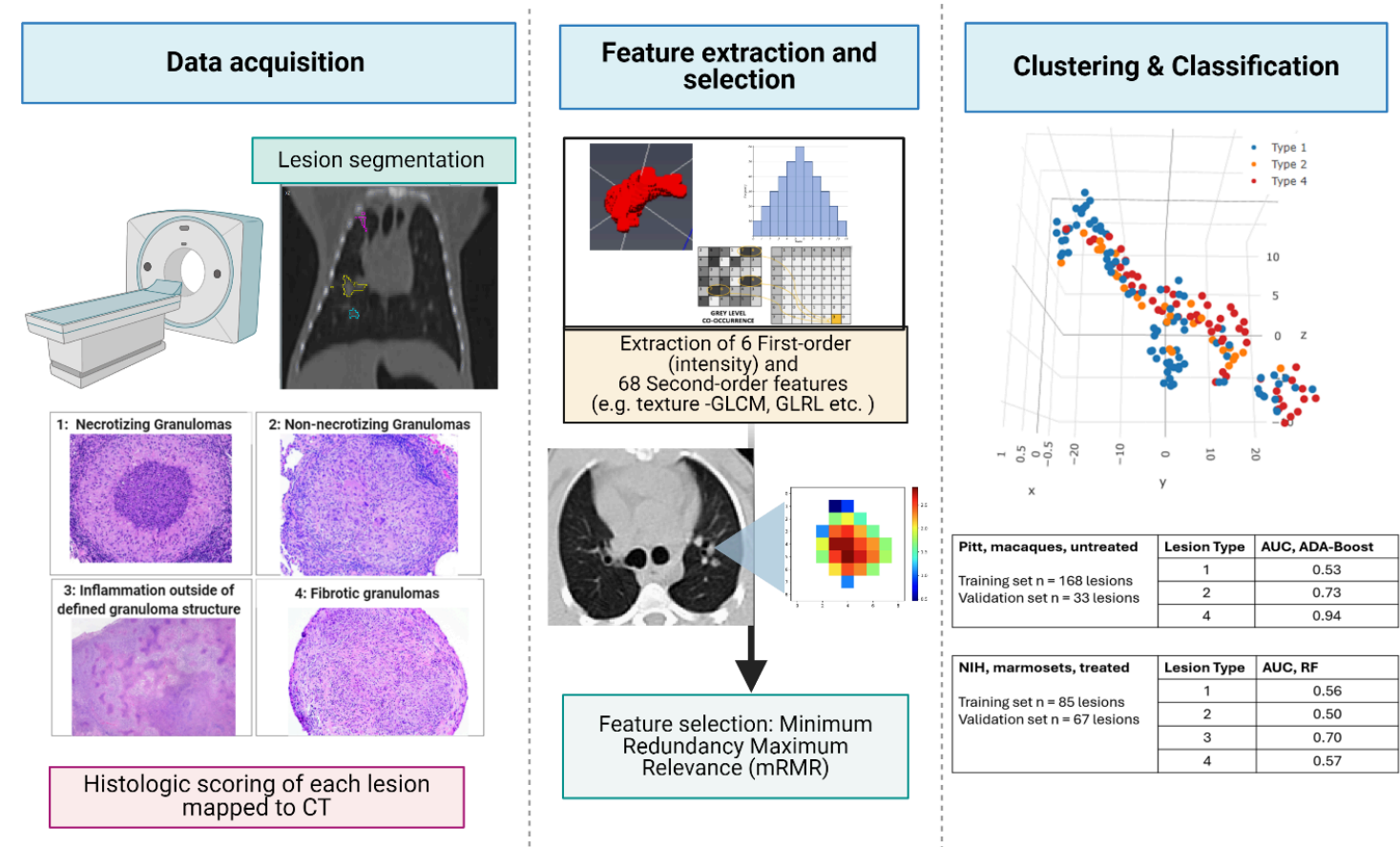


Figure 1

Workflow of image analysis, feature extraction and selection, and application of machine learning algorithms for radiomic prediction of TB lesion histopathology in marmosets and macaques. Performance of hierarchical classifier methods using train-test split for predicting the 4 TB lesion histopathologic classes shown in bottom right.

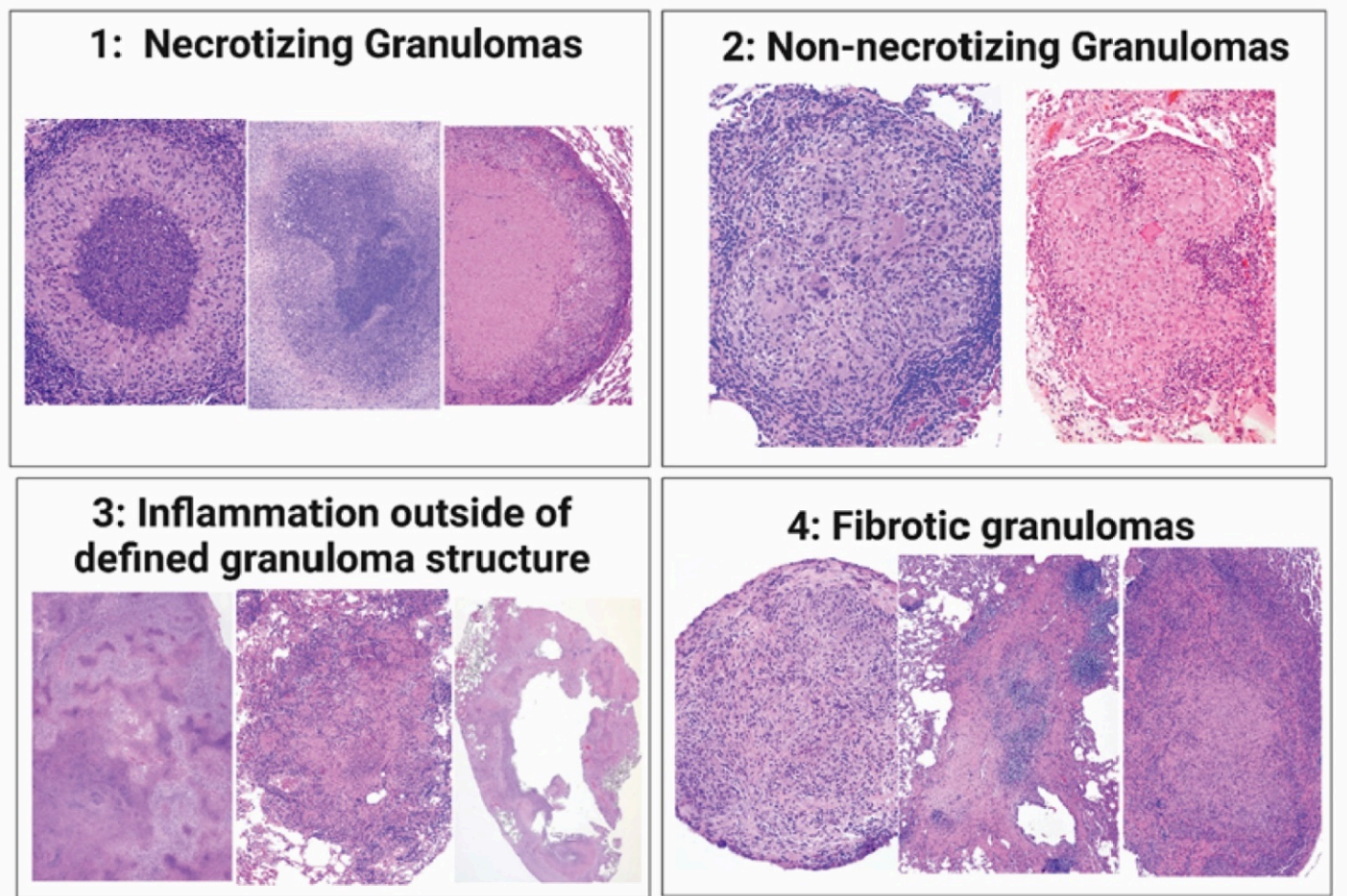


Figure 2

Histology and descriptions of the four TB lesion classes: (1) Type 1: Necrotizing granulomas: These lesions demonstrate central necrosis with eosinophilic, amorphous acellular debris (caseum) +/- occasional stippled mineralization, concentrically surrounded by a mantle of epithelioid macrophages/multinucleated giant cells, and further peripherally margined by lymphoplasmacytic inflammatory infiltrates. (2) Type 2: Non-necrotizing granulomas: Circumscribed aggregates of inflammatory cells (generally histiocytic). Demonstrate an architectural organization, although without a necrotic component and the various zones and cellular types associated therein, their architecture is generally less complex. (3) Type 3: Granulomatous inflammation outside of defined granuloma structure formation: Large area of coalescing or confluent necrotizing (caseous) granulomas or lesions in which histiocytic inflammation without frank granuloma structure extends in a direct, seemingly unimpeded fashion from alveolus-to-alveolus. This type of lesion is sometimes designated as tuberculous pneumonia. (4) Type 4: Fibrotic granulomas. Granulomatous architecture has been largely replaced by evolving dense fibrous connective tissue.

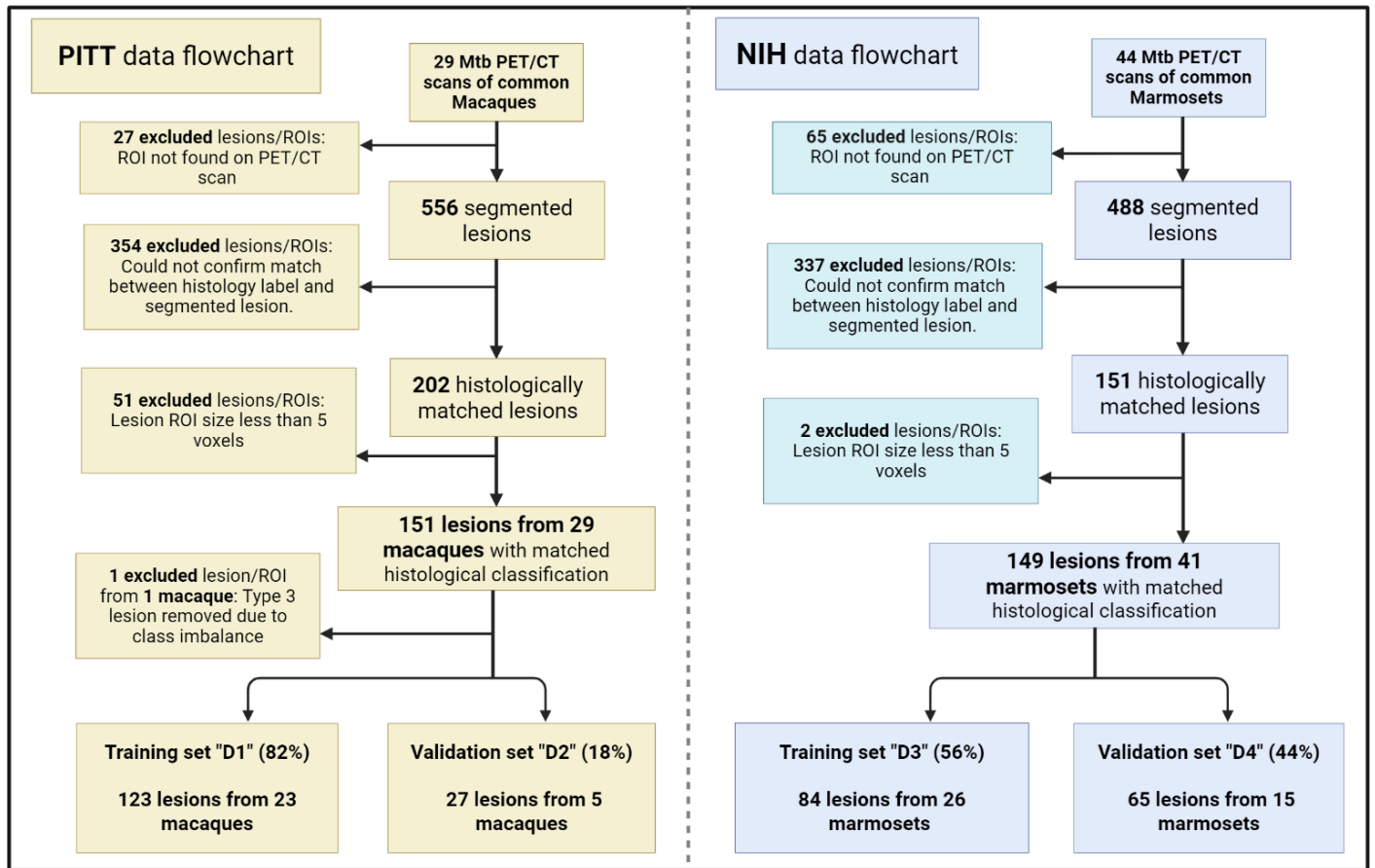


Figure 3

Study flowchart.

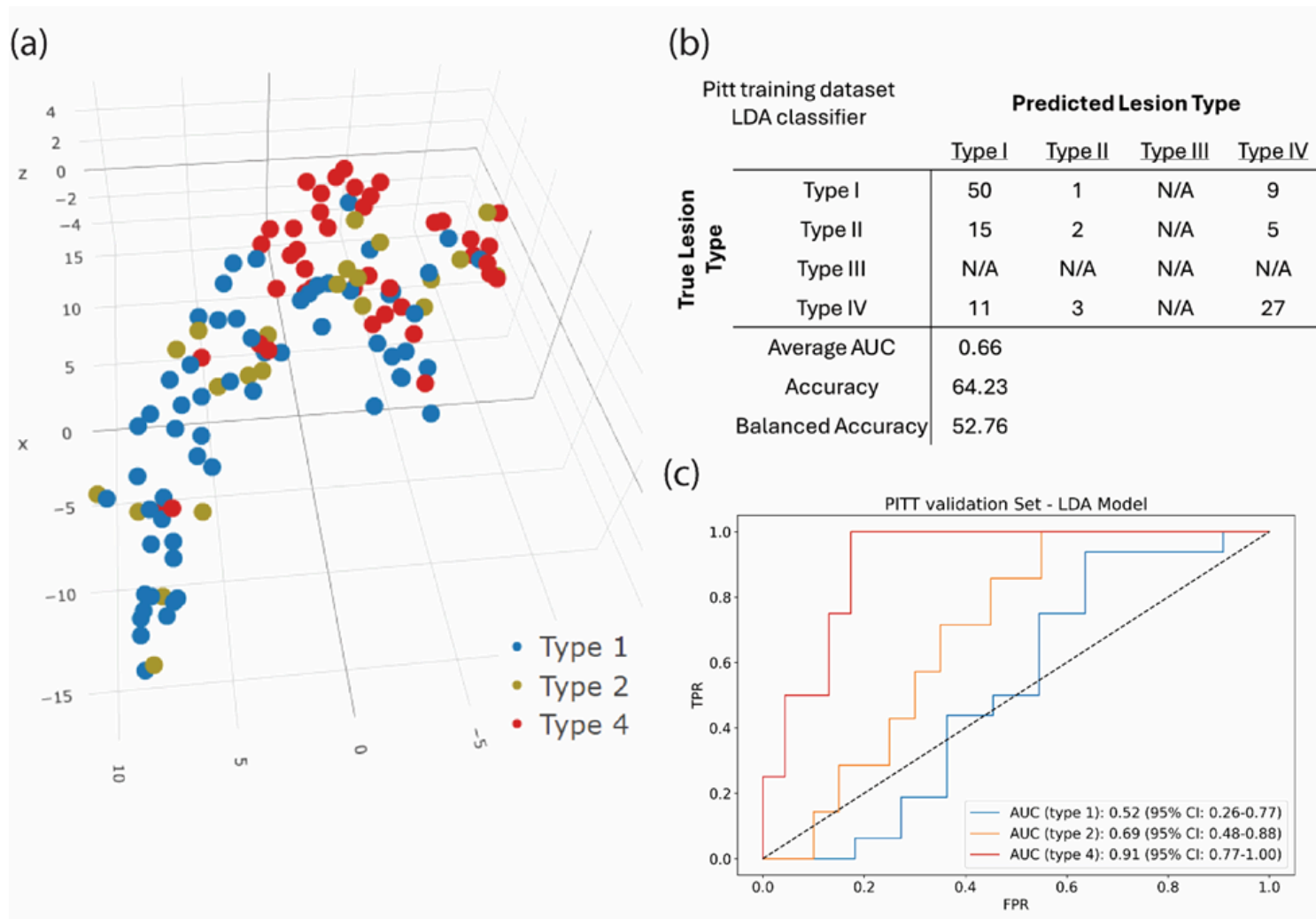


Figure 4

Unsupervised and supervised classification results for the Pitt dataset with (a) three-dimensional tSNE projection, (b) performance metrics for the LDA classifier based on LOO cross-validation of the training dataset D1 and (c) ROC curves and AUC values for the LDA classifier based on the validation dataset D2.

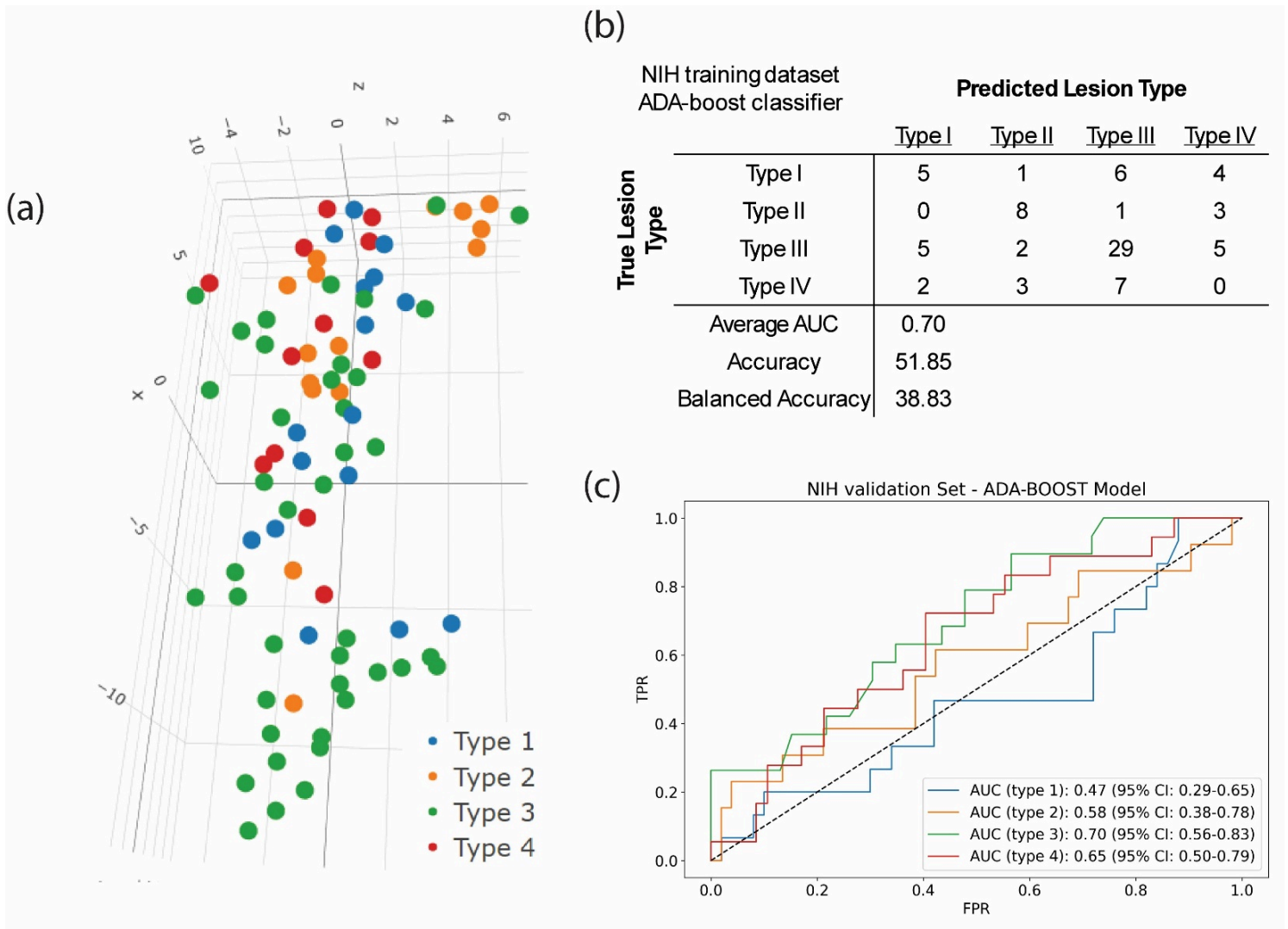


Figure 5

Unsupervised and supervised classification results for the NIH dataset with (a) three-dimensional tSNE projection, (b) performance metrics for the ADA-boost classifier based on LOO cross-validation of the training dataset D3 and (c) ROC curves and AUC values for the ADA-boost classifier based on the validation dataset D4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialsTBradiomicmsv4.213May2025.docx](#)
- [SupplementaryMaterialsappendixS1IllustrationsandFurtherClarificationforRadiomicsGranulomaClassificationSystem.pdf](#)
- [SupplementaryMaterialsappendixS2SOPforsegmentation14May2025.docx](#)
- [SupplementaryMaterialsappendixS3RadiologicFeaturesDataset.csv](#)
- [SupplementarymaterialsappendixS4Primatelists.xlsx](#)