Scientific Abstract Presentations
Toolkits and Machine Learning Algorithms

Date:
MON, OCT 2

Time:
10:30 AM – 12:00 PM ET

Location:
Turner Auditorium

Continuing Education:
ASRT-RT | CAMPEP-MPCEC | SIIM IIP-CIIP

A Comprehensive Guide to Preparing Medical Imaging Data for AI: A SIIM Survey
+ Sanaz Vahdati, MD, Postdoctoral Research Fellow, Mayo Clinic
+ Bardia Khosravi, MD, MPH, MHPE; Elham Mahmoudi, MD, MPH; Pouria Rouzrokh, MD, MPH, MHPE; Shahriar Faghani, MD; Mana Moassefi, MD; Aylin Tahmasebi, MD; Katherine Andriole, PhD, FSIIM; Peter Chang, MD; Keyvan Farahani, PhD; Mona G. Flores, MD; Judy W. Gichoya, MD, MS; Sina Houshmand; MD; Bradley J. Erickson, MD, PhD, CIIP, FSIIM

Care to ExplAIn? Differential Impacts of Explanation Types on Physician Trust in AI
+ Drew Prinster, PhD Candidate, Johns Hopkins University
+ Amama Mahmood; Suchi Saria, PhD; Jean Jeudy, MD; Cheng T. Lin, MD; Chien-Ming Huang, PhD; Paul H. Yi, MD, MS

Evaluating the Utility of Self-Configuring Capsule Networks for Brain Image Segmentation
+ Durga Sritharan, Postgraduate Associate, Yale School of Medicine
+ Sanjay Aneja, MD; Arman Avesta, MD, PhD; Rahul D'Souza; Mariam Aboian, MD, PhD; MingDe Lin, PhD

Exploring Interpretation Maps as a Path to Discover Radiogenomics Biomarkers: A Call for Rethinking
+ Shahriar Faghani, MD, Postdoctoral Research Fellow, Mayo Clinic Rochester
+ Mana Moassefi, MD; Gian Marco Conte, MD, PhD; Bradley J. Erickson, MD, PhD, CIIP, FSIIM

From Isolation to Collaboration: Harmonizing Heterogeneous Medical Imaging Datasets with Partial Annotations
+ Pranav Kulkarni, Bioinformatics Software Engineer, University of Maryland School of Medicine
+ Adway Kanhere, MS; Paul H. Yi, MD, MS; Vishwa S. Parekh, PhD

Machine Learning for the Prediction of Osteopenia/osteoporosis Using the Bone Attenuation of Multiple Osseous Sites from Chest Computed Tomography
+ Ronnie Sebro, MD, PhD, Musculoskeletal Radiologist, Mayo Clinic Florida
+ Cynthia De la Garza-Ramos, MD

A Comprehensive Guide to Preparing Medical Imaging Data for AI: A SIIM Survey

Sanaz Vahdati, MD, Postdoctoral Research Fellow, Mayo Clinic; Bardia Khosravi, MD, MPH, MHPE; Elham Mahmoudi, MD, MPH; Pouria Rouzrokh, MD, MPH, MHPE; Shahriar Faghani, MD; Mana Moassefi, MD; Aylin Tahmasebi, MD; Katherine Andriole, PhD, FSIIM; Peter Chang, MD; Keyvan Farahani, PhD; Mona G. Flores, MD; Judy W. Gichoya, MD, MS; Sina Houshmand; MD; Bradley J. Erickson, MD, PhD, CIIP, FSIIM

Introduction

The increasing rate of Artificial Intelligence(AI) model development to address numerous clinical challenges has escalated the need to prepare high-quality clinical imaging data. Optimal data preparation is of paramount importance since it leads to the development of standard, reproducible AI models and alleviates biases. The ideal tool should assist developers and researchers in preparing the data in the fastest and most well-curated manner. To achieve this, one should be aware of existing tools with core features that provide most of the tasks at hand. Becoming familiar with existing tools can be beneficial for not only selecting the best tool for the assigned task (e.g., detection, segmentation, or classification) but also can point out the possible limitations the user might face if starting to work with the wrong tool.

Hypothesis

The current study was designed to collect the most used tools for data curation prior to using them to train AI models.

Methods

A questionnaire including the tool's name, description, and core features was prepared and distributed among 54 active members of the Society of Imaging Informatics(SIIM) from 26 medical informatics centers. Duplicates, general answers not describing a specific tool, and institutional custom (not publicly available)-based applications were excluded, resulting in a total of 30 tools. The tools were investigated based on the core features, cloud features, input data, data curation including de-identification functions and data annotation, workflow, federated learning support, and data storage. We categorize different steps of data preparation with a comprehensive collected list of tools that are designed for each task.

Results

Tools with their extracted core features are described in the attached tables for data identification, data curation, and data annotation and labeling. Image normalization, data conversion, input and output data, labeling and segmentation, and auto-segmentation application are some of the main tasks depicted in our findings. It is noted to mention that data curation is defined as the process of selecting and organizing data from the time it is acquired to the point it is ready for use by AI; Thus, data de-identification and annotation can be considered as part of data curation. As depicted, several tools may provide more than one of the aforementioned tasks.

Conclusion

We propose a comprehensive guideline of stages for data curation for AI applications and present lists of tools in each step for optimizing the decision-making.

Statement of Impact

Providing a list of tools for different steps of medical imaging data preparation for AI model development.

Table1. Tools used for data de-identification.

Deidentifier tool	Cloud-based	Viewer	Input/Output	Data storage
Anonymizer (by John Perry - RSNA)			DCM/DCM+ spreadsheet	✓
Flywheel	✓	✓	DCM/DCM	✓
Glendor PHI Sanitizer	Both		Image folder + text data	
CTP (Clinical Trial Processor)		✓	DCM/DCM	
Horos	Both	✓	DCM+database in txt files with comments or notes on single studies	

Table2. Tools used for data curation with their specific core features.

Tool	De-identification	Viewer	Cloud-based	Input/Output	Conversion	Normalization
MD.ai	✓	✓	✓	Dcm/Dcm, CSV, JSON	✓	
MONAI Label			Both	DCM, NiFTI/DCM, NiFTI	✓	✓
3D slicer	✓	✓	Both	DCM, NiFTI/DCM, NiFTI	✓	✓
ImageJ/FIJI	✓	✓		DCM, NiFTI/DCM, NiFTI	✓	✓
The Medical Imaging Interaction Toolkit (MITK)		✓		DCM, NiFTI/DCM, NiFTI	✓	✓
dicom2nifti				DCM/NiFTI	✓	
DCM2niix	✓			DCM/NiFTI	✓	
MOOSE				DCM/NiFTI	✓	✓
Niffler	✓			DCM/NiFTI	✓	
MANGO		✓		DCM, NiFTI/NiFTI	✓	✓
ITK-SNAP		✓		DCM, NiFTI/NiFTI	✓	✓

Table3. Tools used for data annotation with their specific core features.

Tool	Cloud-base	Input/ Output	Labeling	Segmentation	Active learning	Object detection	3D rendering	Co-registration	classification
MD.ai	✓	Dcm/Dc m, CSV, JSON		✓+ Auto segmentation		✓			✓
MONAI Label	Both	DCM, NiFTI/DC M, NiFTI	✓	✓+ Auto segmentation	✓				✓
3D slicer	Both	DCM, NiFTI/DC M, NiFTI	✓	✓+ Auto segmentation	✓		✓	✓	

ITK-SNAP		DCM, NiFTI/Nif TI		✓+ Auto segmentation			✓	✓	
ImageJ		DCM, NiFTI/DCM, NiFTI	✓	✓			✓	✓	
ImageJ/FIJI		DCM, NiFTI/DCM, NiFTI	✓	✓		✓	✓	✓	✓
The Medical Imaging Interaction Toolkit (MITK)		DCM, NiFTI/DCM, NiFTI		✓+ Auto segmentation			✓	✓	
Ril-Contour		NiFTI/Nif TI		✓+ Auto segmentation			✓		
NCI Imaging Data Commons (IDC)	✓	DCM, NiFTI/DCM, NiFTI	✓						
MOOSE		DCM/Nif TI (3D)		Auto segmentation					
MANGO		DCM, NiFTI/Nif TI	✓				✓		
Prodigy		DCM, NiFTI/DCM, NiFTI (2D)+ text data			✓	✓			✓

Horos	Both	DCM+ database in text files with comment or notes on single studies	✓						
Markit	✓		✓			✓			✓

Keywords
Artificial Intelligence; Data preparation; Data curation; Toolkit

SIIM CMIMI23
OCTOBER 1-2
Johns Hopkins | Baltimore, MD
Conference on Machine Intelligence in Medical Imaging

Care to ExplAIn? Differential Impacts of Explanation Types on Physician Trust in AI

Drew Prinster, PhD Candidate, Johns Hopkins University; Amama Mahmood; Suchi Saria, PhD; Jean Jeudy, MD; Cheng T. Lin, MD; Chien-Ming Huang, PhD; Paul H. Yi, MD, MS

Introduction
Although AI is increasingly used in clinical practice, few studies have evaluated how different AI explanation methods and other design considerations like AI uncertainty communication, might impact physician diagnostic performance or trust in AI advice. We evaluated how AI explanation types, confidence levels, and correctness would impact physician diagnostic performance for chest x-ray (CXR) diagnosis.

Hypothesis
AI explanation types will impact physician diagnostic performance and trust in AI advice.

Methods
We conducted a prospective randomized experiment of 220 physicians (132 radiologists, 88 internal or emergency medicine physicians) who were asked to evaluate 8 CXR cases (Fig.1). We evaluated how AI explanation types, confidence levels, and advice correctness impact diagnostic performance, confidence in final diagnosis, and perception of AI advice. The between-subjects factor was the AI explanation type: either a “local” explanation that localized key CXR image features with a bounding box, or a “global” explanation that compared a given CXR image to a prototypical example of the CXR in question. AI advice correctness and AI confidence were varied as within-subjects factors. We analyzed the data using appropriate generalized linear mixed-effects (GLME) models. Finally, we evaluated if our key findings can be understood through a ‘novel ‘simple trust” mechanism.

Results
Local AI explanations increased physician diagnostic accuracy compared to global AI explanations when AI advice was correct (Fig.2). Furthermore, AI confidence and task expertise modulated the effect of explanation type on diagnostic accuracy. Physicians used local AI explanations more efficiently (time required for diagnosis) than global AI explanations, though explanation type did not impact physician’s subjective perceptions of AI advice or their confidence in their final diagnosis. Lastly, we identified a potential explanation for our results via a novel heuristic for “simple trust”—which can be roughly understood as reliance without verification—that suggests that physicians tend to more quickly align their diagnosis with local AI explanations than global AI explanations, regardless of AI advice correctness.

Conclusion
AI explanation types impact physician diagnostic performance and trust in AI advice more than the physicians themselves are aware of. AI explanation types substantially impact multiple behavioral measures (diagnostic accuracy, efficiency, and simple trust in AI advice). AI developers and clinicians alike should carefully consider the differential impacts of AI explanation types on diagnostic performance and trust when designing and using AI systems.

Statement of Impact
AI Explanation mechanisms impact how physicians perform in CXR diagnosis and their trust in AI advice, which need to be considered during clinical deployment of AI.

Figure 1: Illustration of experimental setup and procedure: (A) Participating physicians first view the X-ray case without AI advice. (B) Once ready, physicians view AI advice including AI explanation and AI confidence, with design conditions applied here. (C) Physicians then decide whether or how to use AI advice, finalize their diagnosis, provide their confidence, and rate AI advice usefulness. (D) Physicians then repeat for 8 X-ray cases.

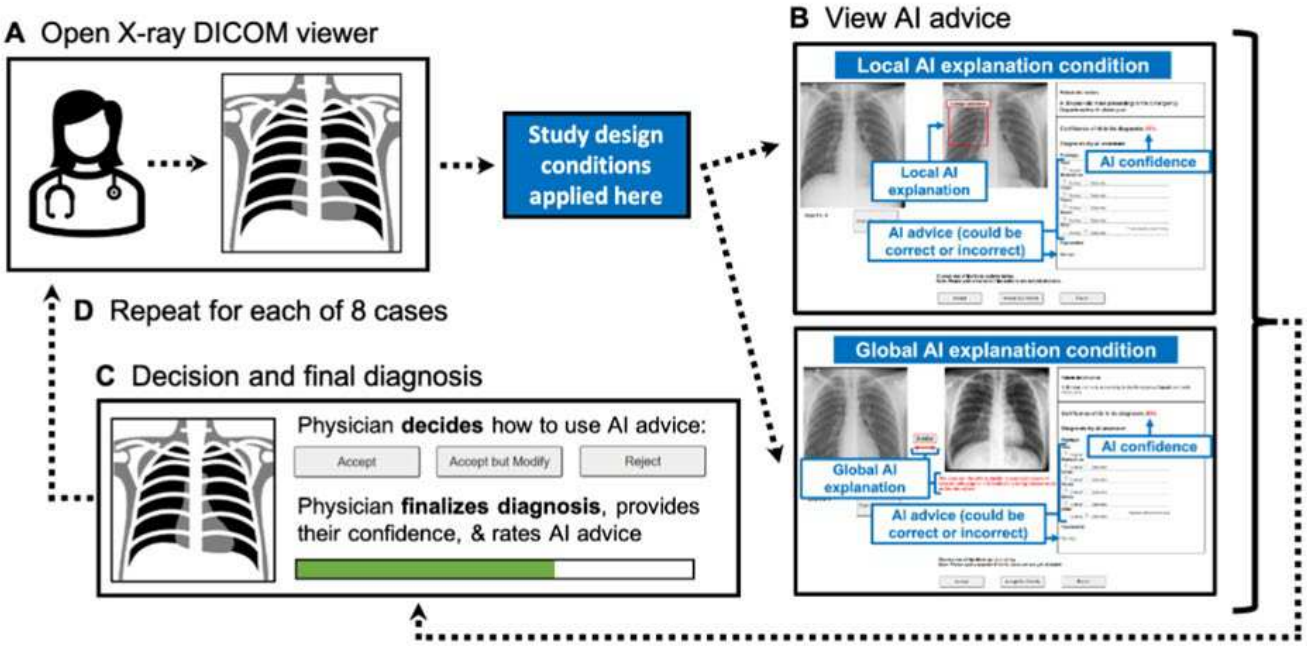
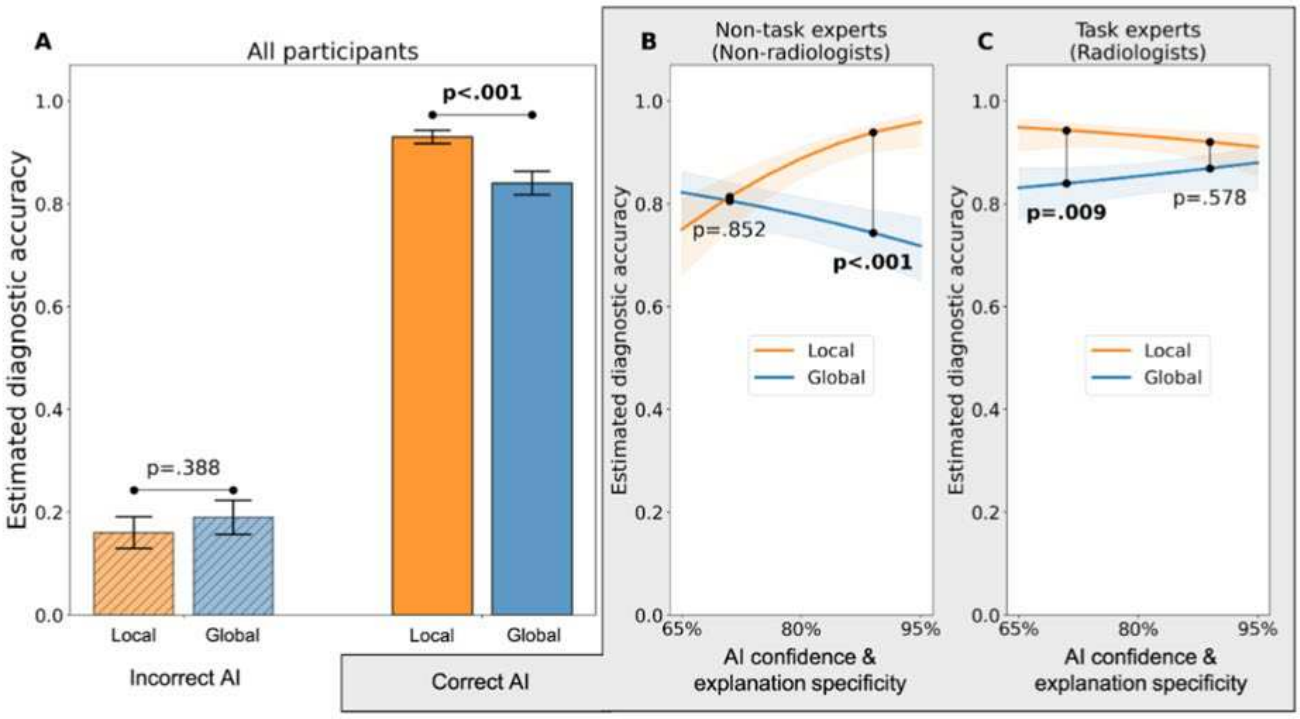


Figure 2: Main results for diagnostic accuracy outcome: Interaction plots for significant interaction effects among experimental variables (addressing primary research questions) for the outcome of marginal-mean estimated diagnostic accuracy. Figure 2A displays the interaction plot for explanation type × AI advice correctness (interaction coefficient β = 1.092, p = .001 [p_adj =]) from the mixed-effects logistic regression model, demonstrating that the impact of AI advice correctness on physician diagnostic accuracy depends on the type of explanation used by the AI: In particular, local AI explanations improve diagnostic accuracy relative to global explanations when AI advice is correct (β = 0.859, p < .001), but we cannot conclude if the explanation type alters the impact of incorrect AI advice on diagnostic accuracy (β = −0.234, p = .388). Only among the correct AI advice condition, figures 2B and 2C display the interaction plots for explanation type × AI advice confidence × physician task expertise (three-way interaction coefficient β = −1.034, p = .012), where radiologists are considered task experts and non-radiologists are non-task experts. In particular, figure 2B illustrates that for non-task experts given correct AI advice, local explanations improve physician diagnostic accuracy relative to global explanations when AI confidence is high (figure 2B, β = 1.598, p < .001), but we do not observe such a difference when AI confidence is low (figure 2B, β = 0.065, p = .852 [p_adj =]). In figure 2C, on the other hand, we see that for task experts given correct AI advice, local explanations improve diagnostic accuracy relative to global explanations when AI confidence is low (figure 2C, β = 1.115, p = .009), but we do not observe such an effect when AI confidence is high (figure 2C, β = 0.578, p > .05).



Keywords
Chest X-ray; Explainability; Human computer interaction; Deep Learning; Trust

Evaluating the Utility of Self-Configuring Capsule Networks for Brain Image Segmentation

Durga Sritharan, Postgraduate Associate, Yale School of Medicine; Sanjay Aneja, MD; Arman Avesta, MD, PhD; Rahul D'Souza; Mariam Aboian, MD, PhD; MingDe Lin, PhD

Introduction

Although a number of deep learning techniques leveraging convolutional neural networks have shown promise for anatomical segmentation, they often require significant amounts of computational memory and training data. Capsule networks represent an alternative and potentially more efficient method for image auto segmentation. We sought to evaluate the utility of self-configuring capsule networks for diagnostic image segmentation.

Hypothesis

We hypothesized that self-configuring capsule networks would be a more computationally efficient method for image segmentation while maintaining high fidelity.

Methods

Using a dataset of 755 MRIs for patients diagnosed with high grade gliomas across multiple facilities within a single healthcare system we trained a self configuring capsule network for tumor identification. Specifically, self-configuring capsule networks were trained to segment tumor enhancing core on the post-T1 contrast sequences. 603 MRIs were used for training, 76 MRIs used for validation, and an additional 76 MRIs were used as a blinded test set. The self-configuring paradigm of the capsule network algorithm included automated adjustments for slice thickness, MRI imaging parameters, and computational resources available for training. Self configuring capsule network performance was compared to traditional convolutional U-NET based auto-segmentation techniques. Dice scores were used to evaluate segmentation performance. Model convergence time, deployment time, and model size (GB) were used to evaluate computational efficiency.

Results

The self-configuring capsule networks showed high fidelity in tumor delineation among gliomas within our dataset. The segmentation accuracy between self-configuring capsule networks was similar to traditional convolutional U-NET based auto-segmentation techniques (89% vs 88%, $p = 0.27$). Self-configuring capsule networks had notably shorter convergence time during training compared to U-NET based models (11 hours vs 38 hours, $p < .001$) and similar deployment time (4 min vs 3.5 min, $p=.15$). Self-configuring capsule networks required significantly less memory when compared to traditional U-NET based segmentation techniques (5 GB vs 31 GB, $p < .001$).

Conclusion

Self configuring capsule networks are a promising computationally efficient method for diagnostic image segmentation tasks with performance that rivals traditional convolutional U-NET based deep learning auto

segmentation techniques. Further studies can help elucidate the potential utility of these novel algorithms within clinical practice.

Statement of Impact

Self-configuring capsule networks are a novel method for image auto segmentation which have the advantage of exquisite computational efficiency compared to alternative deep learning based methods. As the number of algorithms potentially being deployed for image analysis increase, there is an increasing need for more computationally efficient methods which are scalable across all settings.

Keywords

Capsule Networks; Autosegmentation; Deep learning

SIIM CMIMI23
OCTOBER 1-2
Johns Hopkins | Baltimore, MD
Conference on Machine Intelligence in Medical Imaging

Exploring Interpretation Maps as a Path to Discover Radiogenomics Biomarkers: A Call for Rethinking

Shahriar Faghani, MD, Postdoctoral Research Fellow, Mayo Clinic Rochester; Mana Moassefi, MD; Gian Marco Conte, MD, PhD; Bradley J. Erickson, MD, PhD, CIIP, FSIIM

Introduction

Deep learning (DL) has demonstrated promising results in predicting genetic status based on imaging data. However, these models often lack interpretability. Integrated gradients (IG) is a technique that assigns importance scores to pixels, providing insight into the model's decision-making process and the contribution of each pixel to the output. This study explores the role of IG maps in interpreting the classification of the isocitrate dehydrogenase (IDH) gene and O-6-methylguanine-DNA methyltransferase (MGMT) promoter methylation status in glioblastomas using MRI.

Hypothesis

Interpretation maps can facilitate the discovery of new imaging biomarkers for radiogenomics.

Methods

We analyzed the publicly available The University of California San Francisco Preoperative Diffuse Glioma MRI (UCSF-PDGM) dataset, which includes various MRI sequences and corresponding tumor genetic profiles. We trained 3D-Densenet121 models using these sequences independently and in combination to predict IDH mutation and MGMT promoter methylation status. 3D IG maps were utilized to identify the imaging areas that influenced the decision-making process.

Results

The area under the receiver operating curve (AUC) for IDH classification was 0.94, 0.93, and 0.94 for T2, contrast-enhanced T1 (CT1), and T2-CT1 sequences, respectively. The model trained on T2 for MGMT prediction achieved an AUC of 0.65. The model's attention for IDH classification on T2 primarily focused on cerebrospinal fluid, while on CT1, it emphasized contrast-enhanced areas. However, the model primarily highlighted the tumoral area when both CT1 and T2 images were used. Conversely, the model for the MGMT task highlighted all areas without a specific anatomical or functional preference. (Figure 1)

Conclusion

In radiogenomics, the imaging features the model detects, such as IDH classification, depend on the imaging sequence. Using multiparametric MRI directs the model's attention more towards the tumor region. However, no specific imaging areas are emphasized when the model fails to predict the radiogenomics outcome accurately. To draw comprehensive conclusions, employing more interpretation maps, analyzing different combinations of MRI sequences, and considering neuroradiologists' reports on the highlighted areas is necessary. This approach

presents a novel method for identifying imaging biomarkers for radiogenomics through DL.

Statement of Impact

Interpretation maps hold the potential to facilitate the discovery of new imaging biomarkers.

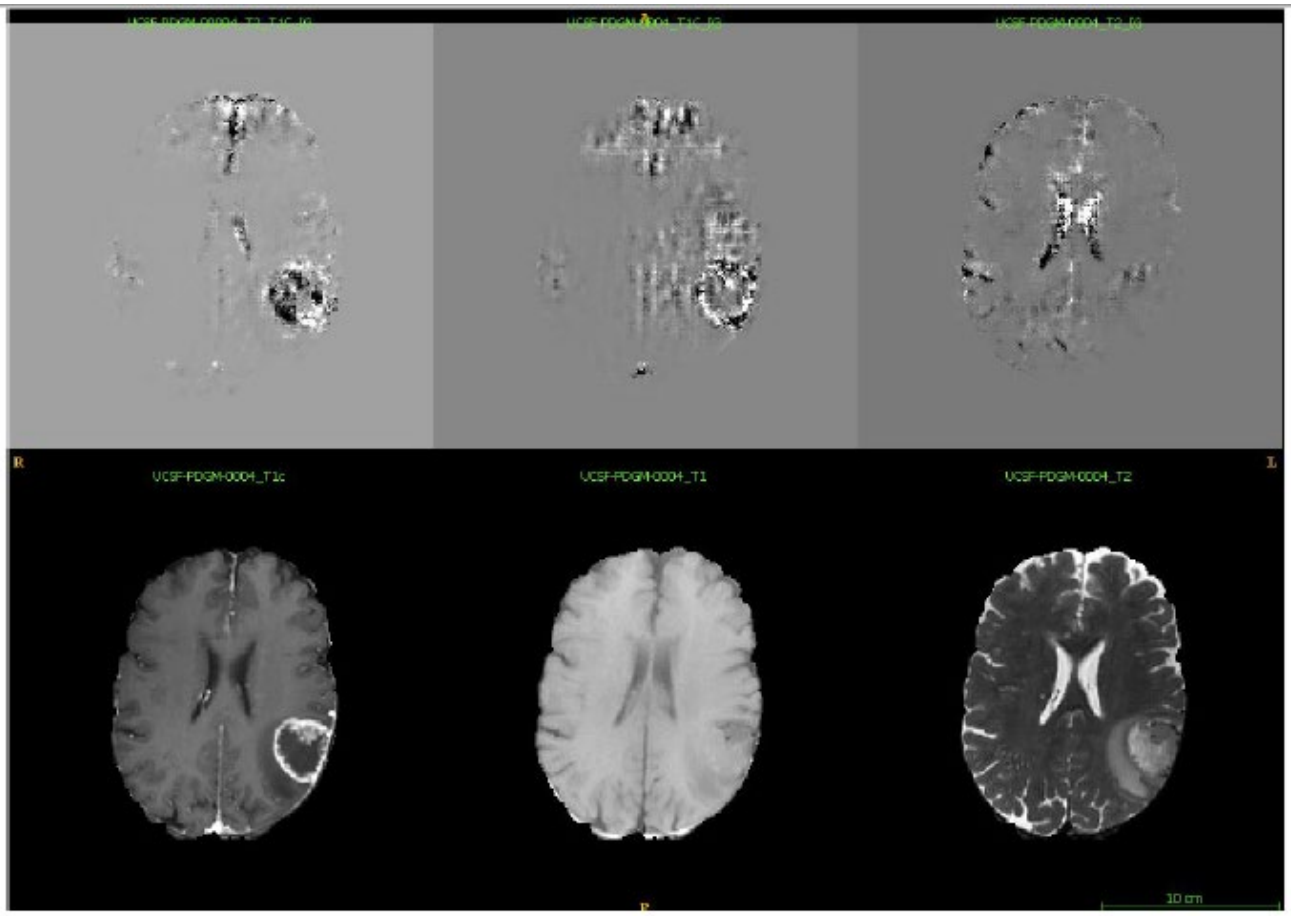


Figure 1 illustrates the brain magnetic resonance imaging (MRI) of a patient with glioma, showcasing T1, T2, and contrast-enhanced T1 (CTE1) images. The top row of the figure displays three distinct integrated gradient maps generated by three different model for predicting the isocitrate dehydrogenase gene status trained on CTE1 and T2 data, CTE1 data alone, and T2 data alone respectively. Notably, all the images within the figure depict the same slice for comparative analysis.

Figure 1 illustrates the brain magnetic resonance imaging (of a patient with glioma, showcasing T 1 T 2 and contrast enhanced T 1 (CTE 1 images The top row of the figure displays three distinct integrated gradient maps generated by three different model for predicting the isocitrate dehydrogenase gene status trained on CTE 1 and T 2 data, CTE 1 data alone, and T 2 data alone respectively Notably, all the images within the figure depict the same slice for comparative analysis

Keywords

Deep learning; Interpretation maps; biomarkers; IDH; MGMT; Glioma

SIIM CMIMI23
OCTOBER 1-2
Johns Hopkins | Baltimore, MD
Conference on Machine Intelligence in Medical Imaging

From Isolation to Collaboration: Harmonizing Heterogeneous Medical Imaging Datasets with Partial Annotations

Pranav Kulkarni, Bioinformatics Software Engineer, University of Maryland School of Medicine;
Adway Kanhere, MS; Paul H. Yi, MD, MS; Vishwa S. Parekh, PhD

Introduction

Although several large-scale chest x-ray datasets have facilitated the development of deep learning (DL) models, heterogeneity in disease labeling schemes limits their inter-operability [Figure 1]. For example, two datasets with different disease labeling conventions cannot be used to directly train a single DL model. Partial annotations – where datasets have non-overlapping disease labels – prevent use of these datasets in aggregate for DL model training. We developed a collaborative learning framework called surgical aggregation to harmonize imaging datasets with heterogeneous label schemes into a single model even in the presence of partial annotations.

Hypothesis

Surgical aggregation will allow for high-performing DL models trained from datasets with partial annotations that outperform conventional approaches.

Methods

Surgical aggregation is a semi-supervised, model-and-task-agnostic framework that selectively aggregates task-specific knowledge from each participating client to train a global model across all observed labels [Figure 2]. Each client can contribute knowledge for its tasks without interacting with any other client or may choose to import knowledge for different tasks via the global model. We evaluate surgical aggregation's ability to harmonize the NIH CXR14 (n=112,120) and CheXpert (n=224,316) datasets using 70%-10%-20% train-validation-test splits (split at patient level). Each dataset has 14 disease labels, of which 7 are common, for a total of 20 unique labels. We used these datasets to train a 20-label classifier, with external testing on the MIMIC-CXR-JPG dataset (n=377,110). We evaluate the surgical aggregation model using mean area under the ROC curve (AUROC) and compare to models trained using conventional methods (baseline, central aggregation, and federated learning) using bootstrapping and paired t-tests; significance was defined as $p< 0.05$.

Results

On the NIH test set (n=22,330), surgical aggregation performed comparably to the NIH baseline with an AUROC of 0.81 ($p=0.06$), while outperforming other conventional approaches (AUROC of 0.67-0.68; $p< 0.001$) and outperformed all conventional approaches with an AUROC of 0.75 on the CheXpert test set (n=45,208) compared to 0.69-0.74 ($p< 0.001$). Similarly, on the external MIMIC test set, surgical aggregation outperformed all approaches with an AUROC of 0.75 compared to 0.71-0.72 ($p< 0.001$) [Figure 3].

Conclusion

Surgical aggregation allows for harmonization of datasets with heterogeneous and non-overlapping disease labeling conventions to train high-performing DL models for CXR diagnosis. This method can scale to any medical imaging use case with heterogeneously labeled datasets.

Statement of Impact

As DL-assisted disease characterization becomes a mainstay in radiology, surgical aggregation provides a framework to leverage heterogeneous medical imaging datasets in aggregate to train large-scale clinically-useful models.

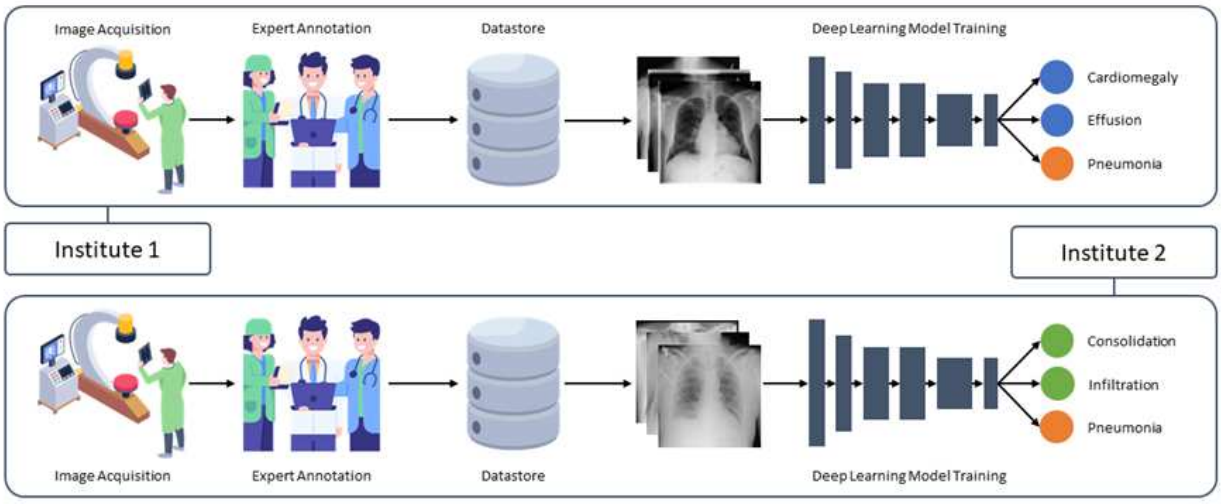


Figure 1. Siloed approach of developing and training models separately on large-scale medical imaging datasets. In this illustration, both institutes curate chest x-ray datasets to train deep learning models on similar tasks. However, due to data and label heterogeneity and patient data privacy, harmonizing and leveraging knowledge from both datasets is difficult.

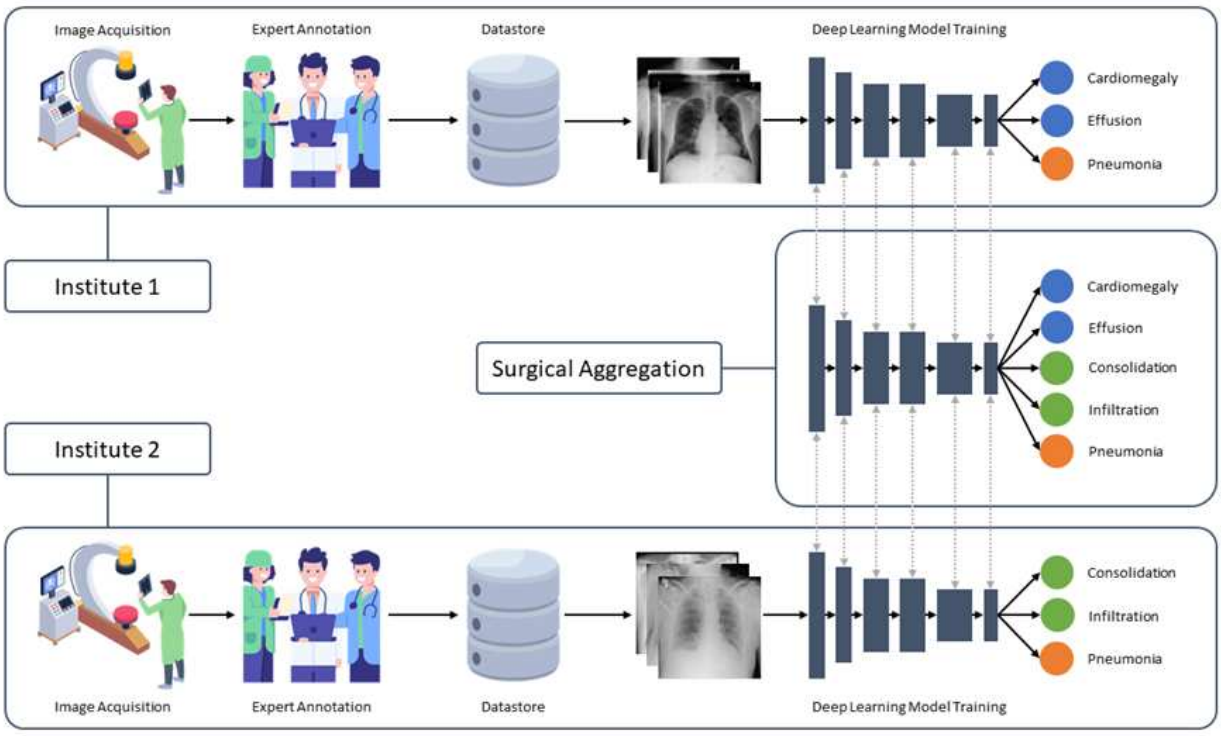


Figure 2. An overview of the surgical aggregation framework. Due to inherent differences in image acquisition, annotation, and curation, large-scale medical imaging datasets are heterogeneous and focus on similar but different disease annotations. Surgical aggregation harmonizes and aggregates knowledge from these heterogeneous datasets into a global deep learning model.

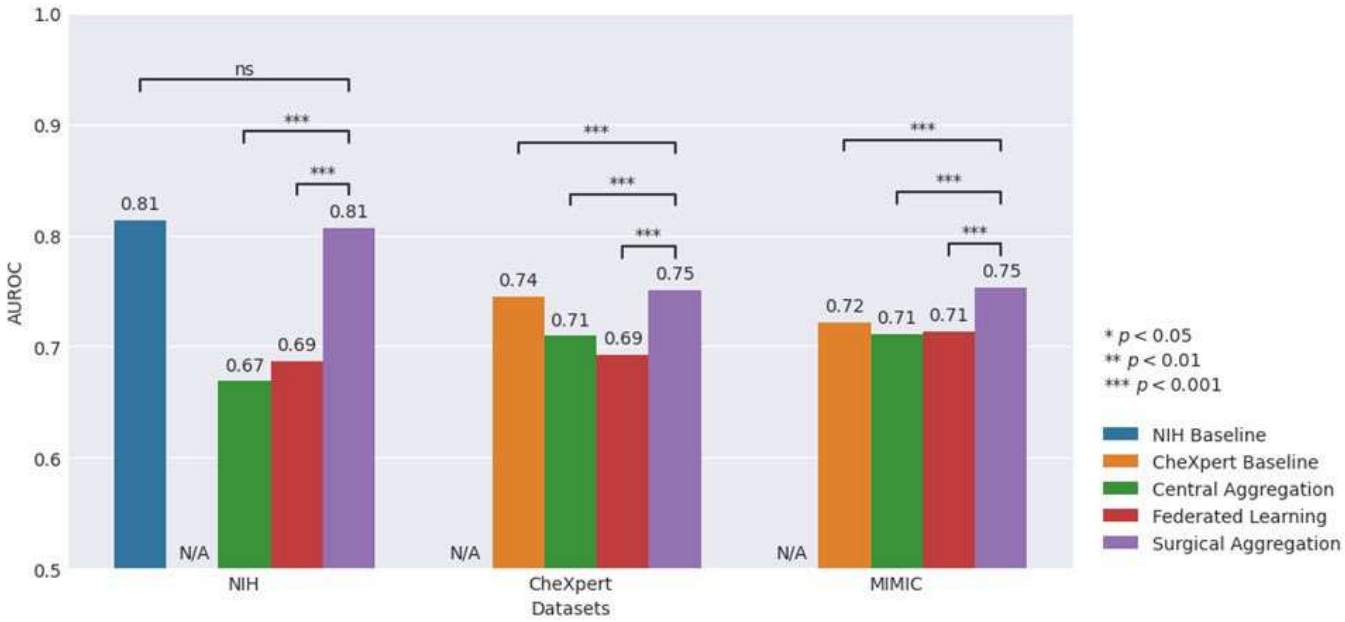


Figure 3. Comparison between the mean AUROC score metrics of surgical aggregation, central aggregation, federated learning, and baseline models on held-out NIH and CheXpert test sets and external MIMIC test set across all 20 disease labels.

Keywords
Data heterogeneity; Data harmonization; Federated learning; Collaborative learning; Chest x-ray; Classification

Machine Learning for the Prediction of Osteopenia/osteoporosis Using the Bone Attenuation of Multiple Osseous Sites from Chest Computed Tomography

Ronnie Sebro, MD, PhD, Professor of Radiology, Mayo Clinic Florida; Cynthia De la Garza-Ramos, MD

Introduction

The attenuation of the lumbar and thoracic spine trabecular bone from CT scans of the abdomen and pelvis have shown utility in predicting BMD measurements. Prior studies have used 2D assessments of the vertebral body trabecular CT attenuation measured on a single slice, however, we hypothesized that volumetric 3D measurement of the vertebral body trabecular CT attenuation would provide a better estimate of the patient's BMD than 2D measurements because the entire trabecular vertebral body is assessed rather than a sample.

Hypothesis

To use machine learning and the 3D CT attenuation of all bones visible on chest CT scans to predict osteopenia/osteoporosis.

Methods

We retrospectively evaluated 364 patients with chest CT and Dual-energy X-ray absorptiometry (DXA) scans within 6 months of each other between 01/01/2015-08/01/2021. Volumetric segmentation of the ribs, thoracic vertebrae, sternum, and clavicle was performed using 3D Slicer to obtain the mean CT attenuation of each bone. The study sample was randomly split into training/validation (80%, n=291) and test (20%, n=73) datasets. Univariate analyses were used to identify the optimal CT attenuation thresholds to diagnose osteopenia/osteoporosis. We used penalized multivariable logistic regression models including Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, and Ridge regression, and Support Vector Machines (SVM) with radial basis functions (RBF) to predict osteopenia/osteoporosis and compared these results to the CT attenuation threshold at T12.

Results

There were positive correlations between the CT attenuation between all bones ($r>0.6$, $P< 0.001$ for all). There were positive correlations between CT attenuation of the bones and the L1-L4 BMD T-score, total hip T-score, and femoral neck T-scores ($r>0.4$, $P< 0.001$ for all). A CT attenuation threshold of 170.2 Hounsfield units (HU) at T12 had an AUC of 0.702, while a threshold of 192.1 HU at T4 had an AUC of 0.757. The SVM with RBF had the highest AUC (AUC=0.864) and was better than the LASSO (P=0.011), Elastic Net (P=0.011), Ridge regression (P=0.011) but was not better than using the CT attenuation at T12 (P=0.060).

Conclusion

The CT attenuation of the ribs, thoracic vertebra, sternum, and clavicle can be used individually and collectively to predict BMD and to predict osteopenia/osteoporosis.

Statement of Impact

Although the CT attenuation of T12 has been historically used to screen for osteopenia/osteoporosis, we found that a T4 CT attenuation threshold of 192.1 HU had a higher AUC than a T12 threshold of 170.2 HU.

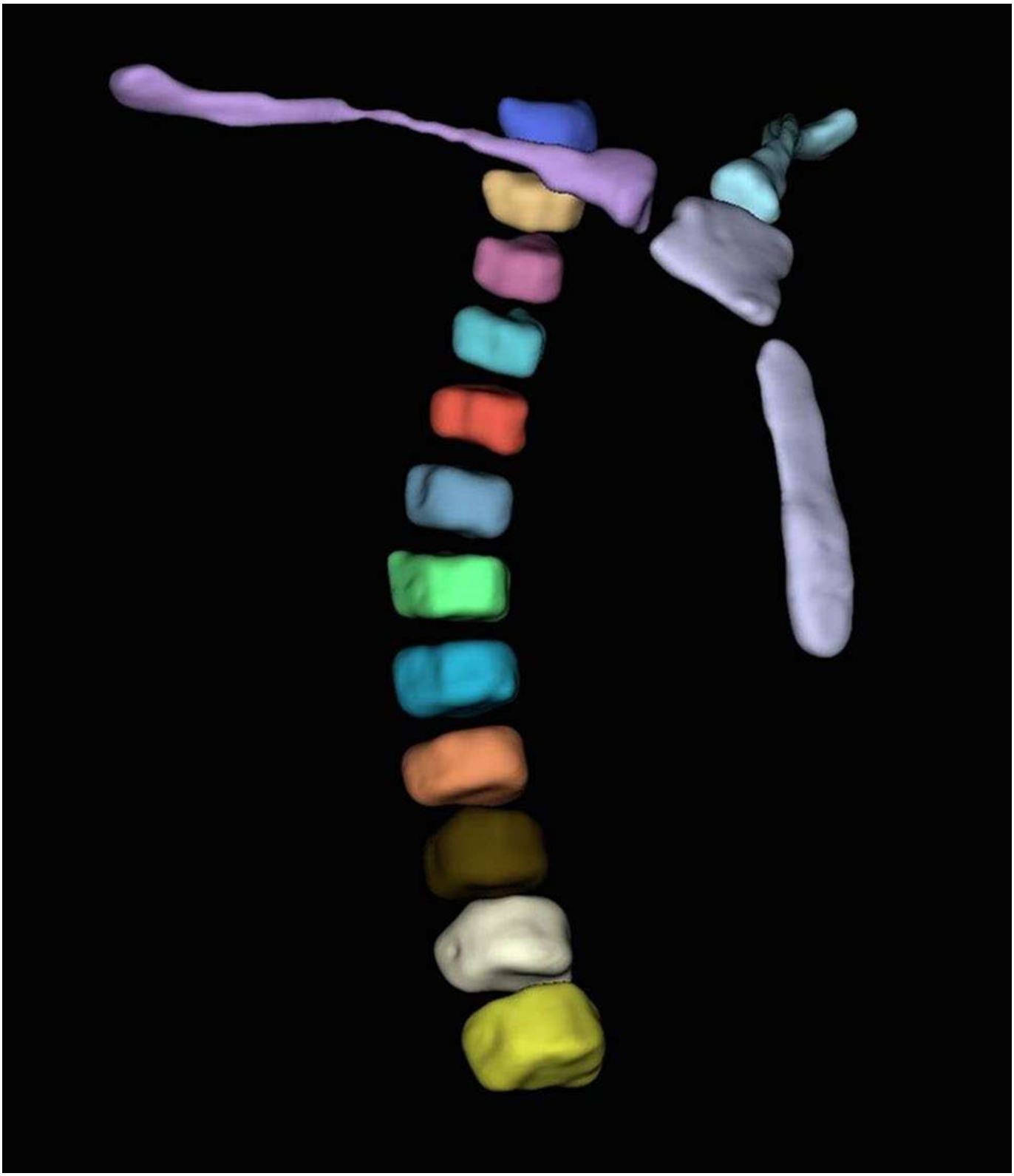


Fig. 1. Three-dimensional volumetric segmentation of the sternum, clavicles, and thoracic vertebrae (T1-T12).

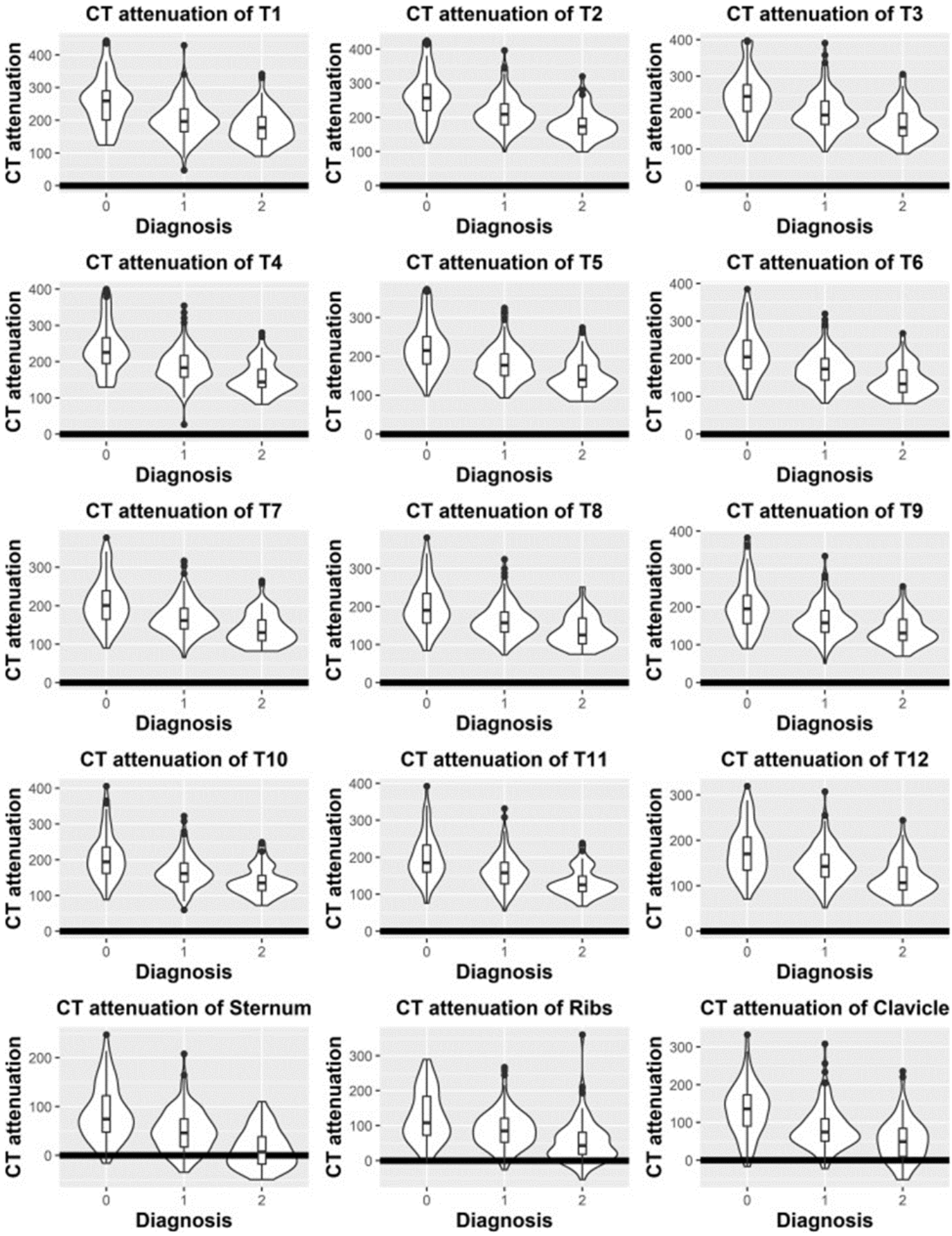

CT attenuation of T1
CT attenuation of T2
CT attenuation of T3
CT attenuation of T4
CT attenuation of T5
CT attenuation of T6
CT attenuation of T7
CT attenuation of T8
CT attenuation of T9
CT attenuation of T10
CT attenuation of T11
CT attenuation of T12
CT attenuation of Sternum
CT attenuation of Ribs
CT attenuation of Clavicle
CT attenuation
Diagnosis
0
1
2

Fig. 2. Violin plots of the mean CT attenuation of each bone by WHO diagnosis. 0 = Normal BMD. 1 = Osteopenia. 2 = Osteoporosis.

Keywords
thoracic spine; computed tomography; osteoporosis; osteopenia; bone mineral density