Scientific Abstract Presentations: Clinical Applications

Date:
SUN, OCT 1

Time:
9:00 AM – 10:30 AM ET

Location:
Turner Auditorium

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

A Comparison of Convolutional Neural Network Architectures in Auto-Segmenting Primary Oropharyngeal Cancers from Contrast-Enhanced CT Scans
+ Onur Sahin, PhD, Medical Student, McGovern Medical School
+ Kareem A. Wahid, PhD; Abdallah S. Mohamed, MD, PhD; Clifton D. Fuller, MD, PhD; Mohamed A. Naser, PhD

Acute Respiratory Distress Syndrome (ARDS) Detection in the Pediatric Intensive Care Unit (PICU) setting Demonstrates High Performance with Transfer Deep Learning
+ Vahid Khalkhali, PhD, Research Scientist, Children's Hospital of Philadelphia
+ Michael Welsh, DO; Patricia P. Rafful, MD, PhD; Dana S. Alkhulaifat, MD; Adarsh Ghosh, MD; Saurav Bose, MD; Nadir Yehya, MD; Susan T. Sotardi, MD

An Improved UNet++ Architecture for Deep Learning based Segmentation of Kidneys and Cysts in Autosomal Dominant Polycystic Kidney Disease (ADPKD)
+ Chetana Krishnan, Graduate Student, University of Alabama at Birmingham
+ Emma Schmidt; Ezinwanne Onuoha, MS; Michal Mrug, MD; Carlos E Cardenas, PhD; Harrison Kim, PhD

Augmenting the MIDRC Dataset using Deep Learning-Based Quantification of Abdominal Aortic Calcification: Proof-of-Concept for Population-Level Disease Screening
+ Devina Chatterjee, Medical Student, University of Maryland School of Medicine;
+ Adway Kanhere, MS; Annie Trang, MS; Vishwa S. Parekh, PhD; Paul H. Yi, MD, MS

Deep Multiclass Multiple-instance Learning For DSA Classification
+ Reza Moein Taghavi, Medical Student, UC Davis School of Medicine
+ Roger Goldman, MD, PhD

Evaluation of an Artificial Intelligence Chatbot for Delivery of Interventional Radiology Patient Education Material
+ Colin J. McCarthy, MD, Interventional Radiologist, Beth Israel Deaconess Medical Center, Harvard Medical School
+ Seth Berkowitz, MD; Vijay Ramalingam, MD; Muneeb Ahmed, MD

A Comparison of Convolutional Neural Network Architectures in Auto-Segmenting Primary Oropharyngeal Cancers from Contrast-Enhanced CT Scans

Onur Sahin, PhD, Medical Student, McGovern Medical School; Kareem A. Wahid, PhD; Abdallah S. Mohamed, MD, PhD; Clifton D. Fuller, MD, PhD; Mohamed A. Naser, PhD

Introduction

Contouring tumors during the development of radiotherapy plans is a labor-intensive process with a high degree of variation when manually performed. Deep learning (DL) models have been developed to automate the segmentation of tumors from anatomical imaging modalities. Previously, our group developed an auto-segmentation model for oropharyngeal cancers (OPCs) from MRI scans. While multiple studies focused on auto-segmentation models from other imaging modalities, there has been no work assessing the development of models using diagnostic-quality contrast-enhanced computed tomography (CE-CT) scans for OPCs , even though CE-CTs can be more easily acquired. In this study, we develop two models using Swin and Resunet convolutional neural networks to auto-segment OPCs from CE-CT scans.

Hypothesis

We hypothesize that DL convolution neural networks can be trained on CE-CT scans to segment OPCs.

Methods

Pre-surgical diagnostic quality CE-CT scans were collected from 474 patients with OPCs. Ground-truth segmentations of OPCs were manually performed and images were randomly assigned into a training (n=380) and test set (n=94) using a 4:1 split. 5-fold cross validation was performed on the training set to train 5 separate models using either a Resunet or Swin network from the MONAI framework. Final model predictions were generated by a consensus mask using either an averaging or simultaneous truth and performance level estimation (STAPLE) method and assessed on the test set. The Dice similarity coefficient (DSC) was used to evaluate model performance. Statistical comparisons were performed using Wilcoxon signed-ranks test.

Results

In the training/test sets, 16.8%/17% of OPCs were T1, 42.1%/42.5% were T2, 22.3%/22.3% were T3, and 18.6%/18.1% were T4. The overall average DSC performance across cross-validation folds was 0.61/0.62 for the Swin/Resunet models, 0.71/0.7 consensus masks developing via an averaging method on the test set, and 0.7/0.69 for the STAPLE method. The Wilcoxon signed-ranks test between the Swin and Resunet showed a $p < 0.05$ for only the consensus masking generated via the averaging method.

Conclusion

We developed the first DL models, to our knowledge, trained on CE-CT scans to segment OPCs, and showed that

they achieved similar performance to DL models trained on MRI images. Further work can focus on improving segmentation performance by utilizing larger patient cohort sizes.

Statement of Impact

In this study we show that more CE-CT scans can be used to develop OPC auto-segmentation models with reasonable performance. With further optimization, the proposed models could be integrated into the workflow of radiation oncologists to increase the efficiency of developing radiation treatment plans.

Keywords

Auto-segmentation; Oropharyngeal Cancer; Convolutional Neural Networks

Acute Respiratory Distress Syndrome (ARDS) Detection in the Pediatric Intensive Care Unit (PICU) setting Demonstrates High Performance with Transfer Deep Learning

Vahid Khalkhali, PhD, Research Scientist, Children's Hospital of Philadelphia; Michael Welsh, DO; Patricia P. Rafful, MD, PhD; Dana S. Alkhulaifat, MD; Adarsh Ghosh, MD; Saurav Bose, MD; Nadir Yehya, MD; Susan T. Sotardi, MD

Introduction

Acute respiratory distress syndrome (ARDS) is a significant cause of morbidity and mortality in the pediatric intensive care unit (PICU). ARDS diagnosis involves chest X-ray (CXR) criteria combined with clinical and laboratory parameters. Machine learning models have demonstrated utility in the detection of ARDS on chest radiographs.

Hypothesis

Machine learning models for radiology CXR can perform better than humans. We evaluated the performance of deep learning (DL) models to diagnose ARDS based on CXR exams. We performed statistical evaluation performance between the models and two experienced radiologists, and also statistical agreement evaluation between the two radiologists as a gold-standard correlation for comparison.

Methods

In this retrospective, IRB-approved study, we identified 368 children admitted to the PICU with a diagnosis of ARDS, at a large pediatric academic center from 2014 to 2019. A single random radiograph from all patients admitted to the PICU without a diagnosis of ARDS during 2018 was used as the control cohort, (n=1127). The train-validation-test ratio was 60/20/20. Using transfer learning, we utilized pretrained DL structures to diagnose ARDS (PyTorch, version 1.2). The area under the receiver operating characteristic (AUROC) was the main performance metric. Two pediatric radiologists independently assigned labels of ARDS/No ARDS and interrater reliability was calculated. Correlations were calculated using the Pearson correlation coefficient and Cohen’s Kappa. All statistical analyses used Type-I error of 5% and power of 80%.

Results

The interrater reliability between the radiologists was 94.5% (Cohen’s Kappa of 85.8%) for the training cohort. The ARDS diagnostic performance of two radiologists yielded an AUROC (balanced accuracy) of 72.5%, while the DenseNet161 model achieved 86.0% (AUROC of 92.5%) and an ensemble of models reach 83.7% (AUROC of 93.5%). Radiologist diagnoses were only 81% correlated (Cohen’s Kappa 51.6%) with the DenseNet161 and 83% (Cohen’s Kappa 60.5%) with the ensemble. While the Pearson correlation between the two radiologists was high (>90%) on the test set, the difference between the detection of different models was statistically significant ($p < 0.01$). Attention maps show that models are able to capture the regions of interest (Figure 1).

Conclusion

Using transfer DL, we trained models to reliably detect ARDS in the PICU and compared their performance with the diagnostic rates of two experienced radiologists. DL can automatically detect ARDS on chest radiographs, with a performance that parallels those of radiologists.

Statement of Impact

Deep learning detection of ARDS could improve the triage of patients in the intensive care unit before the availability of dedicated pediatric radiologist reads.

Figure 1. Attention maps for CNN: left and right examples are non-ARDS and ARDS, respectively.

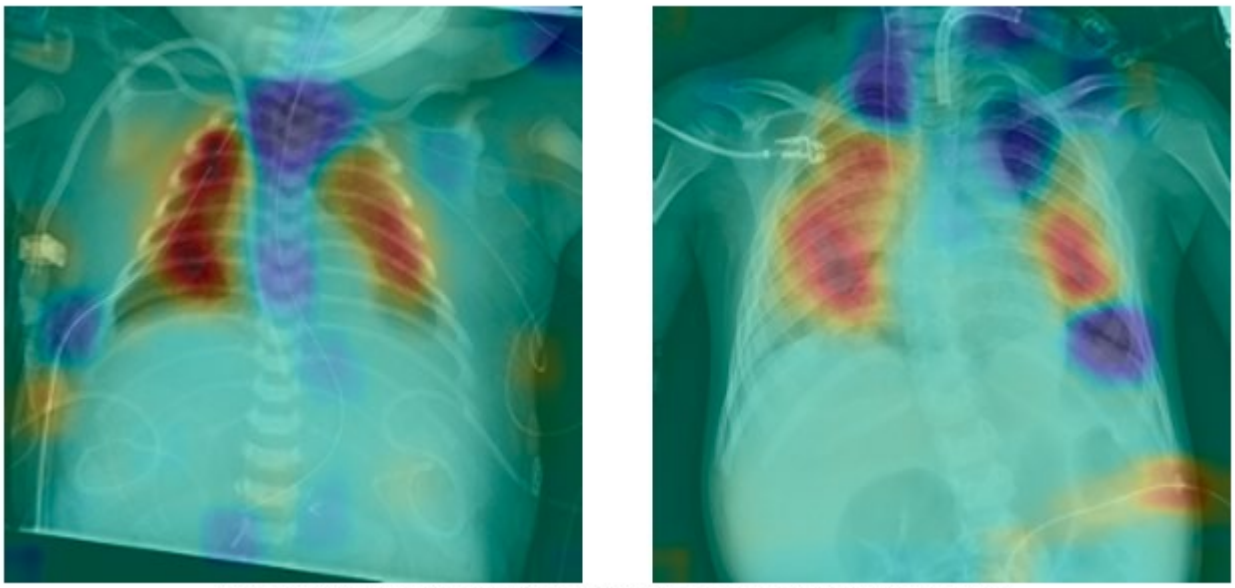


Figure 1. Attention maps for CNN: left and right examples are non-ARDS and ARDS, respectively.

Keywords

Acute Respiratory Distress Syndrome (ARDS)**;** Transfer Deep Learning**;** Chest X-Ray (CXR)

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

An Improved UNet++ Architecture for Deep Learning based Segmentation of Kidneys and Cysts in Autosomal Dominant Polycystic Kidney Disease (ADPKD)

Chetana Krishnan, Graduate Student, University of Alabama at Birmingham; Emma Schmidt; Ezinwanne Onuoha, MS; Michal Mrug, MD; Carlos E Cardenas, PhD; Harrison Kim, PhD

Introduction

UNet++, an enhanced version of UNet, incorporates nested skip connections (NSC), batch normalization (BN), and deep supervision to improve feature extraction and accuracy. However, the UNet++ encoder lacks sufficient feature extraction in small regions of interest (ROIs), leading to inadequate feature fusion during up-sampling and reduced accuracy due to the absence of proper feature normalization. We propose a new architecture called sUNet++, replacing BN with switching normalization (SN) to avoid batch effects and integrate residual blocks in the encoder and decoder. NSCs are replaced with concatenated skip connections (CSC). We compared this architecture with gUNet++, employing group normalization.

Hypothesis

Our hypothesis posits that SN identifies suitable normalization techniques for each layer based on their importance, while residual blocks propagate essential features through skip connections. CSCs improve gradient flow and information propagation. This mitigates performance degradation, allowing the network to learn residual mappings and retain information across layers, resulting in improved segmentation.

Methods

To validate our hypothesis, we trained UNet++, sUNet++, and gUNet++ on T2-weighted MRI images of 95 ADPKD patients, utilizing a total of 756 3D kidney images (604 for training, 76 for validation, and 76 for testing). Preprocessing, cropping, and slicing techniques were applied to generate 2D training samples, resulting in approximately 69,000 samples. The task involved segmenting kidneys and cysts. The models were trained for 50 epochs using a patch-wise approach. Data augmentation techniques were employed to increase the training samples. Leaky ReLU was the activation function. Performance was evaluated using the Dice similarity coefficient (DSC), Hausdorff distance (HD), and Intersection over Union score (IoU).

Results

As summarized in Table 1, sUNet++ achieved higher accuracy than UNet++ for both kidney and cyst segmentation. Moreover, sUNet++ exhibited the highest minimum dice score, indicating superior individual dice performance and success in cases where UNet++ failed. Although gUNet++ showed better performance for kidneys, it was less suitable for cyst segmentation. Notably, sUNet++ required fewer model parameters, converged faster, and demanded less training and inference time. Figure 1 shows the representative test kidney and cyst boundaries determined by sUNet++, gUNet++, and UNet++. Figure 2 illustrates the importance layers of sUNet++ and UNet++.

Conclusion

sUNet++ addresses the challenge of learning normalization in deep learning by dynamically selecting normalizations and statistics for each layer, offering architectural flexibility, and adaptability to varying batch sizes, and eliminating the reliance on sensitive hyperparameters.

Statement of Impact

sUNet++ can improve the diagnosis, treatment planning, and monitoring of ADPKD by detecting early changes in cysts.

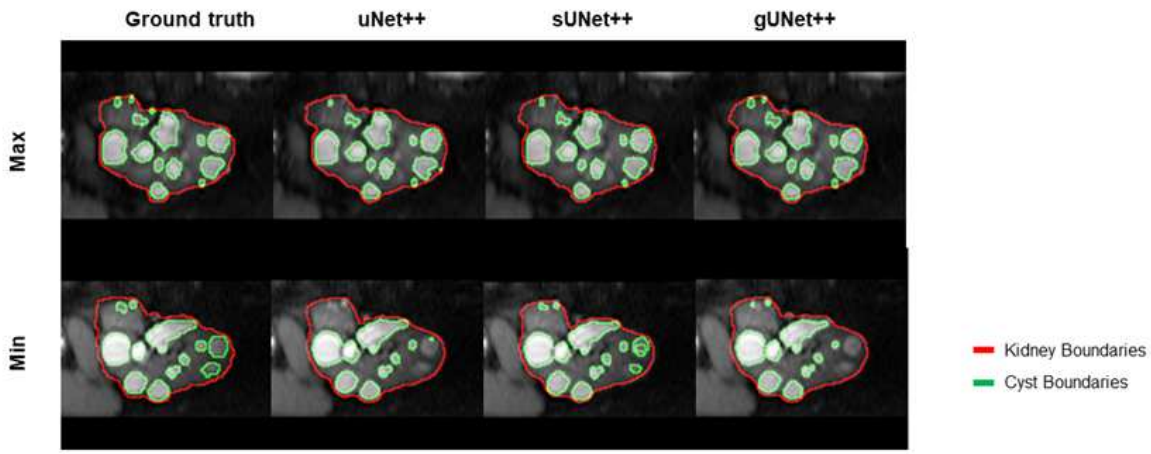


Figure 1. Ground truth Vs predicted kidney and cyst segmentation. Two representative images showing best (max) and moderate (min) performance on the test set, respectively, with kidney (red line) and cyst (green line) boundaries determined by our semi-automatic method (ground truth, first column) and all three models (second to fourth columns) sUNet++ demonstrates improved boundary, edge localization, and enhanced accuracy in capturing collided cyst edges. The replacement of concatenated skip connections enables the preservation of high-resolution edge information by establishing direct paths for information flow between corresponding encoder and decoder layers. This is advantageous for transferring learning to a different task.

Figure 1. Ground truth Vs predicted kidney and cyst segmentation. Two representative images showing best (max) and moderate (min) performance on the test set, respectively, with kidney (red line) and cyst (green line) boundaries determined by our semi-automatic method (ground truth, first column) and all three models (second to fourth columns) sUNet++ demonstrates improved boundary, edge localization, and enhanced accuracy in capturing collided cyst edges. The replacement of concatenated skip connections enables the preservation of high-resolution edge information by establishing direct paths for information flow between corresponding encoder and decoder layers. This is advantageous for transferring learning to a different task.

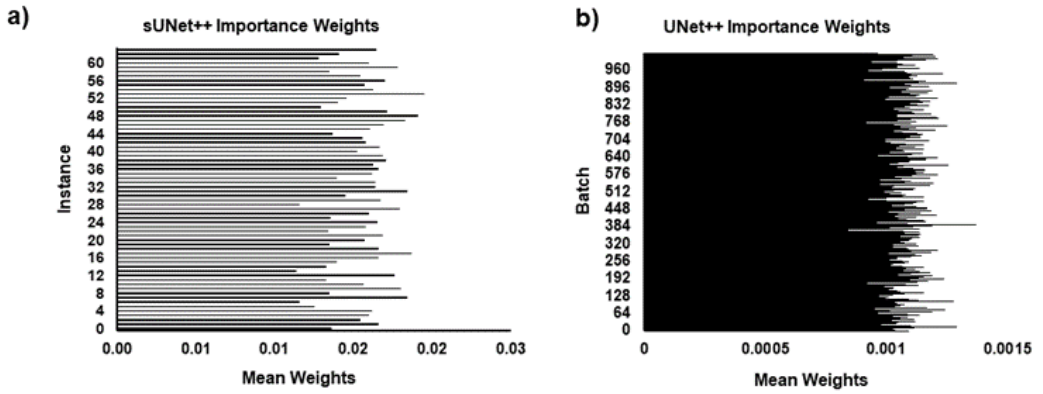


Figure 2. Importance weights. a) The model weight distribution per instance for sUNet++. b) The model weight distribution per batch for UNet++. sUNet++ has fewer model parameters and the concatenated skip connections combined all the layers with equal importance to one single layer to avoid complexity. This reduces redundancy and helps in faster convergence and training. UNet++ with batch normalization considers all the batches irrespective of importance thus leading to re-learning the same features causing overfitting on training. The mean weights are higher in sUNet++ for fewer parameters compared to UNet++ thus having improved representative learning thus removing batch effects.

Figure 2. Importance weights. a) The model weight distribution per instance for sUNet++. b) The model weight distribution per batch for UNet++. sUNet++ has fewer model parameters and the concatenated skip connections combined all the layers with equal importance to one single layer to avoid complexity. This reduces redundancy and helps in faster convergence and training. UNet++ with batch normalization considers all the batches irrespective of importance thus leading to re-learning the same features causing overfitting on training. The mean weights are higher in sUNet++ for fewer parameters compared to UNet++ thus having improved representative learning thus removing batch effects.

Model	ROI	IoU Score	HD (mm)	Test DSC	Min Test DSC	Max Test DSC	Training Time / Epoch (s)	Inference Time (min ± sec)
UNet++	Kidney	0.88±0.47	1.35±0.95	0.93±0.35	0.70	0.97	217±5	05±23
	Cyst	0.77±0.43	1.52±0.78	0.86±0.42	0.71	0.92	220±7	07±14
sUNet++	Kidney	0.90±0.44	1.38±0.94	0.94±0.35	0.84	0.98	110±3	04±12
	Cyst	0.77±0.47	1.30±0.90	0.87±0.42	0.76	0.93	116±6	04±34
gUNet++	Kidney	0.88±0.46	1.43±0.91	0.93±0.34	0.74	0.97	203±4	05±09
	Cyst	0.76±0.44	1.53±0.79	0.86±0.43	0.65	0.93	218±9	05±16

Table 1. Performance Metrics. Intersection over union (IoU) score, Haudsdorff distance (HD), and test Dice similarity score (DSC) of UNet++ and proposed models based kidney and cyst segmentation, together with its training time per epoch, and inference time to predict the boundaries of all test images.

Table 1. Performance Metrics. Intersection over union (IoU) score, Haudsdorff distance (HD), and test Dice similarity score (DSC) of UNet++ and proposed models based kidney and cyst segmentation, together with its training time per epoch, and inference time to predict the boundaries of all test images.

Keywords
Skip connections; Normalization; Image segmentation; Polycystic kidney disease; UNet++; Residual staging

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

Augmenting the MIDRC Dataset using Deep Learning-Based Quantification of Abdominal Aortic Calcification: Proof-of-Concept for Population-Level Disease Screening

Devina Chatterjee, Medical Student, University of Maryland School of Medicine; Adway Kanhere, MS; Annie Trang, MS; Vishwa S. Parekh, PhD; Paul H. Yi, MD, MS

Introduction

Large public imaging datasets like the MIDRC dataset of 20,000+ CT scans have facilitated rapid development of machine learning tools to fight diseases like COVID-19. However, these datasets seldom have disease labels beyond their primary use case (e.g., COVID-19 status for MIDRC), limiting their use for other prediction tasks. We evaluated the feasibility of deep learning-based quantification of abdominal aortic calcification to augment the MIDRC dataset with potential biomarkers for population-level cardiovascular risk assessment.

Hypothesis

Deep learning-based abdominal aortic calcium quantification will allow for augmentation of the MIDRC dataset with population-level assessments of cardiovascular disease risk.

Methods

We first validated the state-of-the-art TotalSegmentator deep learning multi-organ CT segmentation model on two datasets of CT scans of the abdomen/pelvis for segmentation of the abdominal aorta and imaged portions of the thoracic aorta using Dice scores: 1) subset of the MIDRC dataset (COVID-19-NY-SBU [N=1285]; primary dataset) and 2) AMOS dataset (N=250; secondary dataset). Aortic calcifications were segmented on the COVID-19-NY-SBU dataset using Hounsfield Unit voxel thresholding (>250 HU); automated calcium segmentations were validated using 100 manually segmented scans. The aortic calcifications were localized to thoracic vs. abdominal aorta using an automated bodypart regressor model. Agatston score for the abdominal aorta was calculated by multiplying area of calcium by a factor related to maximum plaque attenuation.

Results

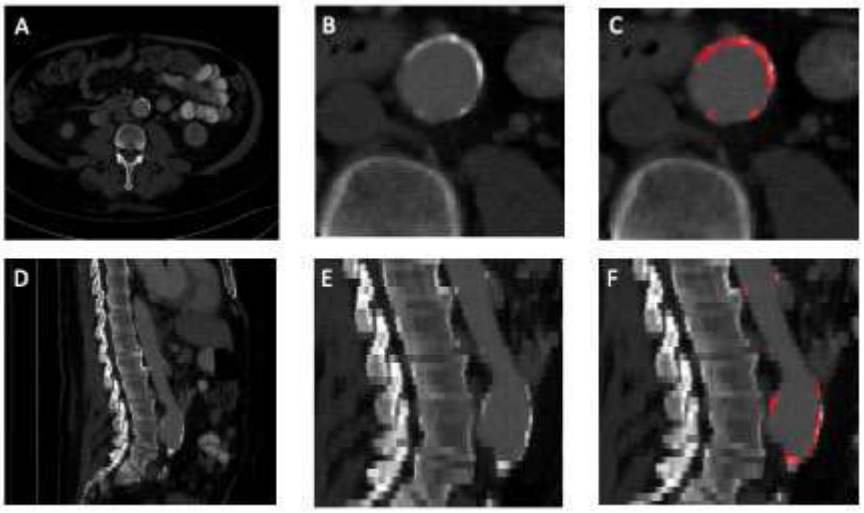
TotalSegmentator had high performance for aortic segmentation on both datasets with mean Dice scores of 0.88 and 0.91 on the MIDRC and AMOS datasets, respectively (Figure 1). Automated aortic calcification similarly had Dice score of 0.88 on our primary dataset (MIDRC) with high correlation with manual segmentations (R2 of 0.9575; Figure 2). In the MIDRC dataset, 75% of patients had aortic calcification; of these, 73% had abdominal aortic calcifications and 77% had thoracic aortic calcifications. For patients with calcifications [N=963], the mean Agatston score in the abdominal aorta was 15,235; 792 patients (62%) had scores >1000 (previously-validated cardiovascular disease risk threshold).

Conclusion

Deep learning-based quantification of abdominal aortic calcifications can augment large public datasets like MIDRC, allowing for population-level assessments of cardiovascular disease risk. Our proof-of-concept paves the way for augmenting other large datasets, including the remainder of the MIDRC dataset, which our group is actively working on next.

Statement of Impact

Augmentation of large imaging datasets for disease risk assessment with deep learning models paves the way for population-level screening for chronic conditions like cardiovascular disease.



Examples of deep learning-based aortic calcification segmentation in the abdominal aorta. Panels (A) and (D) show single axial and sagittal slices, respectively. Panels (C) and (G) demonstrate a portion of the aorta with calcium. Panels (D) and (H) highlight in red the aortic calcification automatically detected by the deep learning model. Note that our models segmented calcium in the entire 3D CT volume, from which the Agatston scores were calculated.

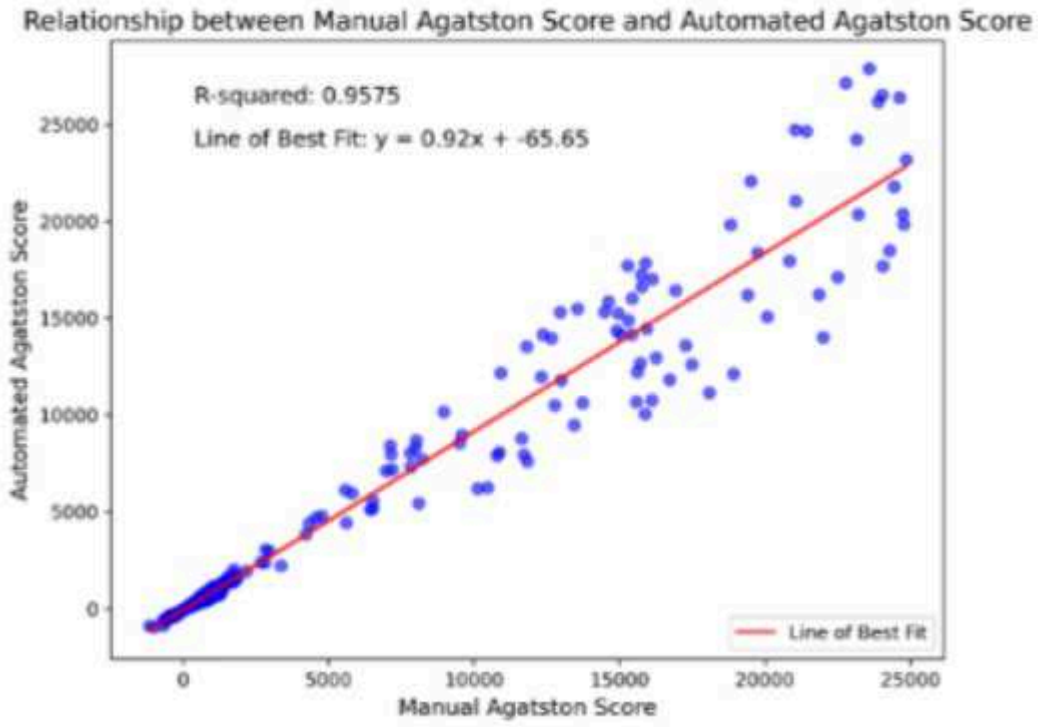


Figure 2: Deep learning-based aortic calcification segmentation and Agatston score calculation correlates well with manual segmentation and score calculations with correlation coefficient of 0.96.

Keywords

MIDRC; Datasets; Aortic Calcification; Opportunistic Screening; Deep Learning

Deep Multiclass Multiple-instance Learning For DSA Classification

Reza Moein Taghavi, Medical Student, UC Davis School of Medicine; Roger Goldman, MD, PhD

Introduction
Anatomic localization is a critical requirement for interpretation of angiography. Many images in digital subtraction angiography (DSA) sequences lack information for localization due to insufficient radiopaque contrast within the vessels, presenting a significant challenge in automated interpretation. The purpose of our study was to evaluate a deep multiclass Multiple Instance Learning (MIL) algorithm for anatomic localization in DSA sequences.

Hypothesis
We hypothesize that MIL model can accurately identify standard anatomic locations in abdominopelvic angiographic sequences.

Methods
We performed a retrospective review of the institutional PACS to identify 689 DSA sequences performed with contrast administration via the aorta, left external iliac artery, right external iliac artery, celiac artery, superior mesenteric artery, and inferior mesenteric artery. Individual images within each DSA sequence were designated as “key” if contrast opacified the identified artery at the location of contrast administration and a first order downstream vessel. Data were divided into 482 sequences for training and validation and 207 sequences for testing. A deep multiclass MIL model was developed using the MONAI Python library to classify DSA sequences from these anatomical locations. The model was trained with inputs of 50 images from each angiographic sequence. To ensure uniformity, all sequences were padded to 50 images with copies of the final image. Classification performance was quantified using accuracy, precision, recall, and F1. MIL model assigned attention weights to each image reflecting contribution to the final classification. Images corresponding to the algorithmically generated top five attention weights were compared with manually-labeled “key” images for overlap. The overlap was quantified as the ratio of the number of images in common to the number of key images.

Results
The deep multiclass MIL algorithm achieved an accuracy of 92.75% (95% CI: 89.22 - 96.28), Precision: 93.99% ± 3.24, Recall: 92.75% ± 3.53, and F1 of 88% ± 4.43 on the held-out test data. The algorithm performance for each anatomical location is provided in table 1. Figure 1 depicts an example of a DSA image sequence, manually-labeled “key” images, and the attention weights. We found an average overlap of 54.8%. In 93.24% of cases, at least one algorithm-chosen image matched an image in the manually-labeled “key” selections.

Conclusion
Deep multiclass MIL is feasible for accurate anatomic localization in DSA imaging.

Statement of Impact
This study demonstrates the potential of deep learning with attention mechanisms for automatically classifying the anatomical locations in time-series DSA data, a critical task in the interpretation of imaging during and after image-guided endovascular procedures.

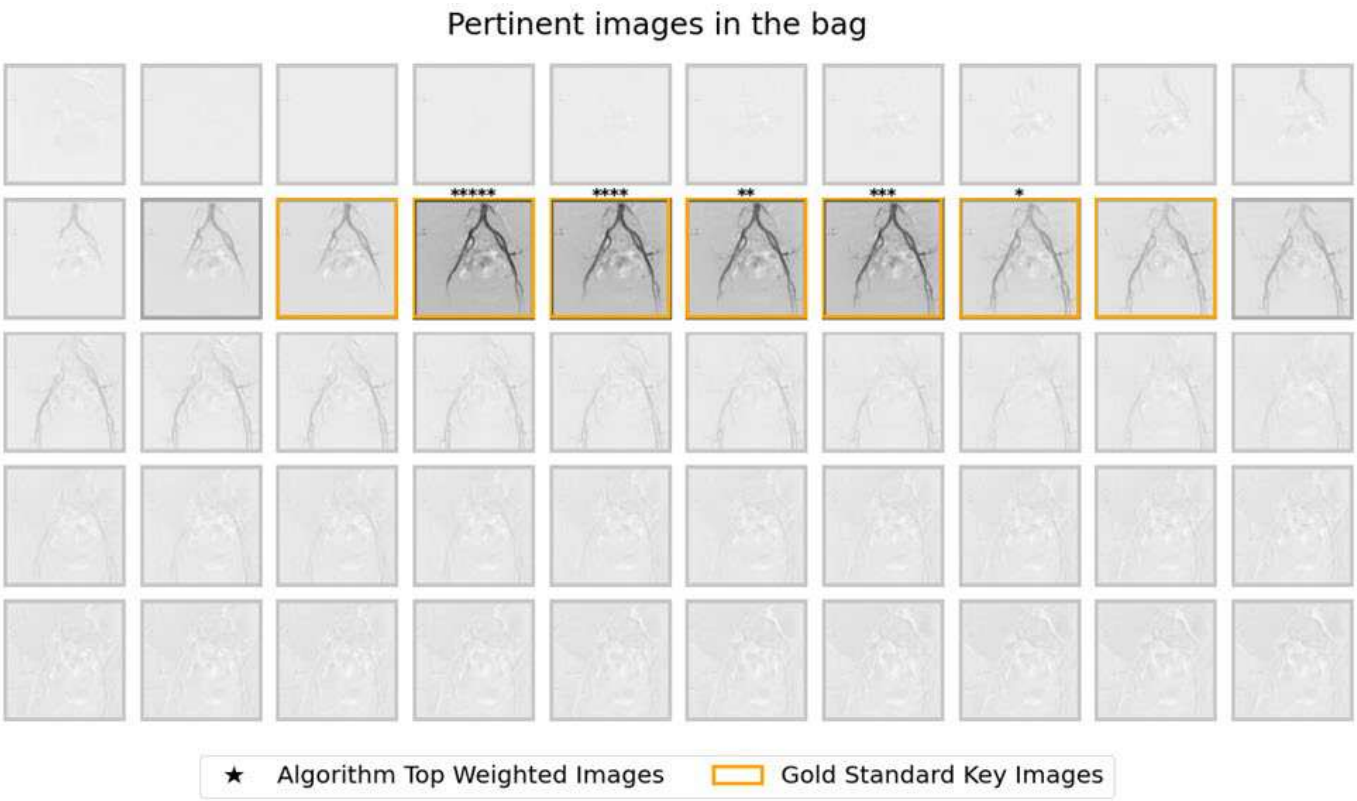


Figure 1: Sample DSA sequence from the testing dataset with an overlay of algorithm-selected images and manually labeled diagnostic “key” images are shown. Images with stars represent the top five weighted images by the algorithm, with the star count correlating to the assigned weight. Similarly, the opacity of the images correlates to algorithm generated weights: the greater the opacity, the larger the weight. Conversely, more transparent images have smaller weights. Images outlined in yellow rectangles represent the diagnostic “key” images.
HAPYGCNX-1580149-2-ANY(4).docx

Anatomic location	Number of samples	Precision	Sensitivity (Recall)	F1
Aorta	27	100	96.3	98.12
CA	66	95.59	98.48	97.01
EIA/R	68	88.16	98.53	93.06
EIA/L	13	100	69.23	81.82
SMA	28	88	78.57	83.02
IMA	5	100	60	75

Table 1 : Performance of the algorithm on the test dataset, stratified by anatomical location: aorta (AO), left external iliac artery (LEIA), right external iliac artery (REIA), celiac artery (CA), superior mesenteric artery (SMA), and inferior mesenteric (IMA).

Keywords
MIL; Deep Learning; DSA

Evaluation of an Artificial Intelligence Chatbot for Delivery of Interventional Radiology Patient Education Material

Colin J. McCarthy, MD, Interventional Radiologist, Beth Israel Deaconess Medical Center, Harvard Medical School; Seth Berkowitz, MD; Vijay Ramalingam, MD; Muneeb Ahmed, MD

Introduction

To assess the potential role for ChatGPT for the delivery of medical information to patients.

Hypothesis

To assess the accuracy, completeness, and readability of patient educational material produced by a machine-learning model and compare the output to that provided by a Societal patient education website.

Methods

Content from the Society of Interventional Radiology (SIR) Patient Center website was retrieved, categorized and organized into discrete questions. These questions were entered into the ChatGPT platform, and the output was analyzed for word and sentence count, readability using multiple validated scales, factual correctness and suitability for patient education using the PEMAT-P instrument.

Results

21,154 words were analyzed, including 7,917 words from the website and 13,377 words representing the total output of the ChatGPT platform across twenty-two text passages. Compared to the Societal website, output from the ChatGPT platform was longer and more difficult to read on 4 of 5 readability scales. The ChatGPT output was incorrect for 12 of 104 (11.5%) questions. When reviewed using the PEMAT-P tool, the ChatGPT content scored lower than the website material. Content from both the website and ChatGPT were significantly above the recommended 5th or 6th grade-level for patient education, with mean Flesch Kincaid Grade Level of 11.1 (+/- 1.3) for the website and 11.9 (+/- 1.6) for the ChatGPT content.

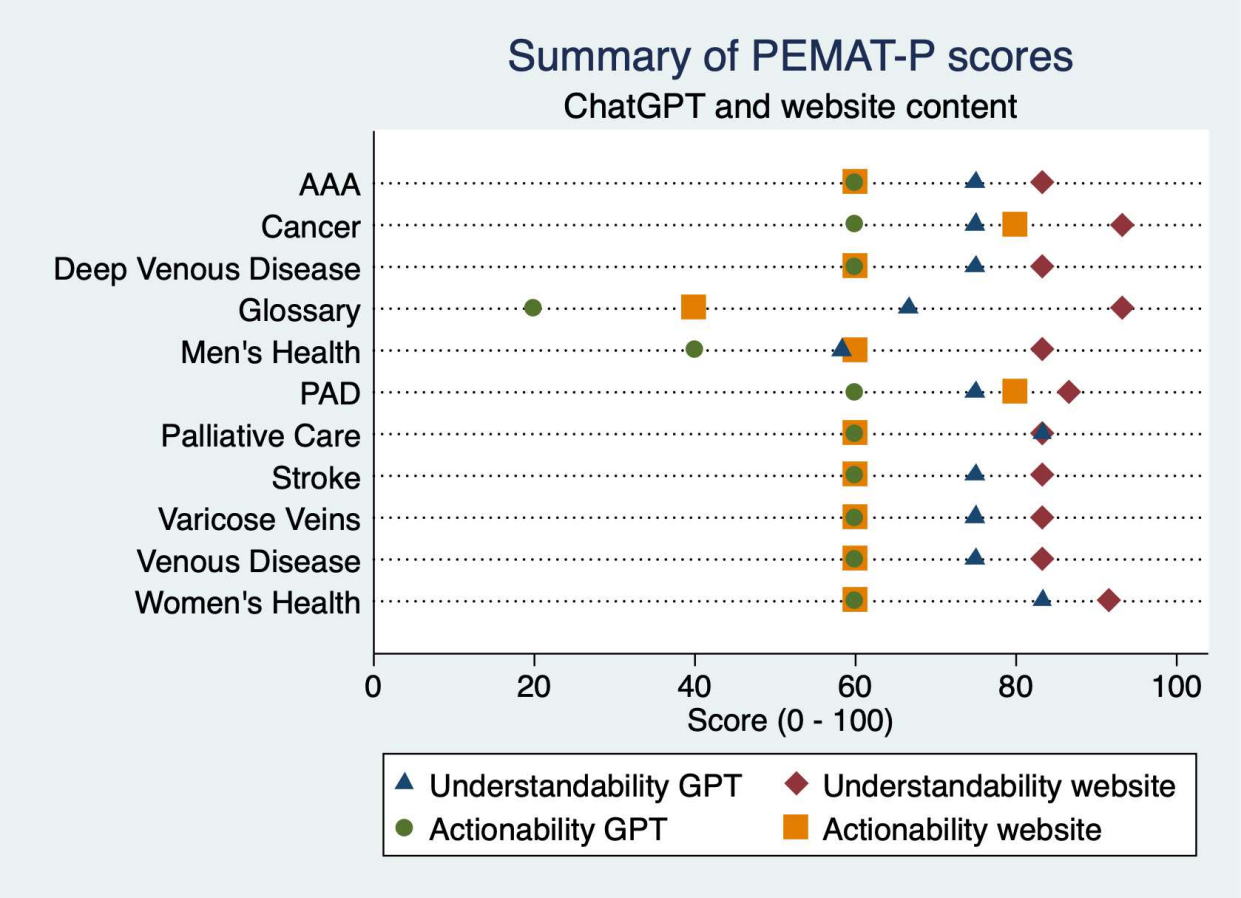
Conclusion

The ChatGPT platform may produce incomplete or inaccurate patient educational content, and providers should be familiar with the limitations of the system in its current form. Opportunities may exist to fine-tune existing large language models, which could be optimized for the delivery of patient educational content.

Statement of Impact

1. Evaluation of this large language Chatbot model as a patient education tool showed that incomplete or incorrect content was provided in 12 of 104 (11.5%) questions. 2. The absence of visual aids in the generated output results in material that is not fully optimized for patient educational purposes. 3. Compared to the Societal website, content from the ChatGPT platform was longer, contained more difficult words and longer sentences, and was more difficult

to read when assessed using several validated reading scales. Content from both the ChatGPT model and Societal website was written at a grade level higher than that recommended for patient education.



Performance of the ChatGPT model against the SIR website, using the PEMAT-P instrument.

Table 3 – overview of analyzed text passages.

	Word Count		Sentence Count		Long Sentence Count (i)		Reading Time (mm:ss)		Dale-Chall Difficult Words (ii)		Automated Readability Index (iii)		Passive Voice Count	
Topic	ChatGPT	SIR	ChatGPT	SIR	ChatGPT	SIR	ChatGPT	SIR	ChatGPT	SIR	ChatGPT	SIR	ChatGPT	SIR
Abdominal Aortic Aneurysm (AAA)	487	426	25	23	21	18	2:09	1:53	191	147	12.36	11.06	4	3
Cancer	873	465	39	28	32	19	3:52	2:04	339	185	13.94	11.01	12	4
Deep venous disease	979	856	52	47	42	32	4:21	3:48	356	309	10.61	10.03	9	9
Glossary of IR Terms	3801	1783	189	77	179	71	16:53	7:55	1340	680	11.31	13.77	97	28
Men's Health	988	738	46	38	39	26	4:23	3:16	421	308	13.93	12.88	15	5
Peripheral artery disease (PAD)	1275	871	68	43	52	29	5:40	3:52	438	288	10.65	12.35	17	11
Palliative care	1199	400	59	24	46	14	5:19	1:46	471	149	12.74	10.92	15	4
Stroke	1026	421	49	23	41	14	4:33	1:52	338	158	11.45	10.89	9	7
Varicose veins	845	469	53	31	42	16	3:45	2:05	356	193	10.61	10.04	21	5
Venous Disease	261	234	17	14	14	7	1:09	1:02	93	87	8.77	9.31	3	0
Women's Health	1603	1254	76	70	59	49	7:07	5:34	676	508	13.37	11.77	20	16
Grand Totals	13337	7917	673	418	567	295	59:11	35:07	5019	3012	129.74	124.03	222	92
Mean	1212.5	719.7	61.2	38	51.5	26.8	5:22	3:11	456.3	273.8	11.8	11.3	20.2	8.4

i) Long sentence count defined as sentences containing > 20 syllables. ii) Dale-Chall Difficult Words are those not found on a list of about 3,000 common words, which 80% or more of US 4th grade students are familiar with. iii) The calculation for the Automated Readability Index (ARI) is outlined in Table XXX, and assesses the approximate U.S. grade level required to read a piece of text.

Keywords

ChatGPT; Patient information; Chatbot; AI; Interventional Radiology