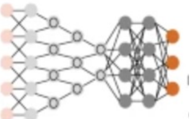
FOUNDATION MODELS FOR IMAGING & RT APPLICATIONS

Lei Xing, PhD
Jacob Haimson & Sarah S. Donaldson Professor & Director of Medical Physics Division

Departments of Radiation Oncology & Electrical Engineering (by courtesy),
ICME & Molecular Imaging Program (MIPS)
Stanford University, Stanford, California

Stanford University
Department of Radiation Oncology
School of Medicine

ARTIFICIAL INTELLIGENCE IN 2024



2014 - Deep learning

2016-AlphaGo

2021-AlphaFold

2023-Foundation models

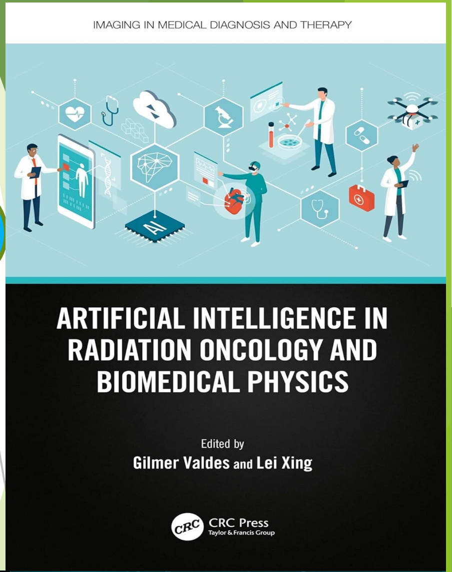
data-driven medicine

Sam Altman @sama · Dec 21

it's been a crazy year.

i'm grateful that we put a tool out in the world that people really love and get so much benefit from.

more than that, i am glad that 2023 was the year the world started taking AI seriously.

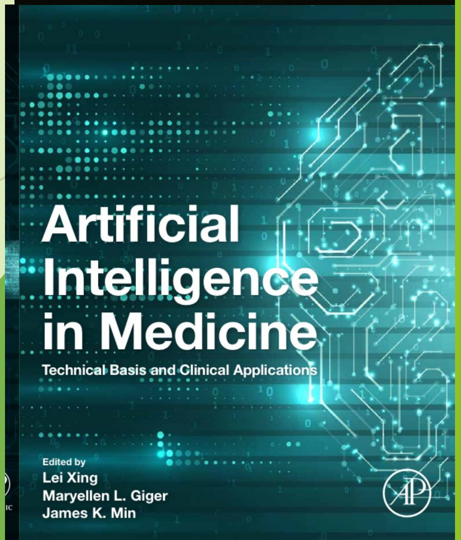


IMAGING IN MEDICAL DIAGNOSIS AND THERAPY

ARTIFICIAL INTELLIGENCE IN RADIATION ONCOLOGY AND BIOMEDICAL PHYSICS

Edited by
Gilmer Valdes and Lei Xing

CRC Press
Taylor & Francis Group



Artificial Intelligence in Medicine
Technical Basis and Clinical Applications

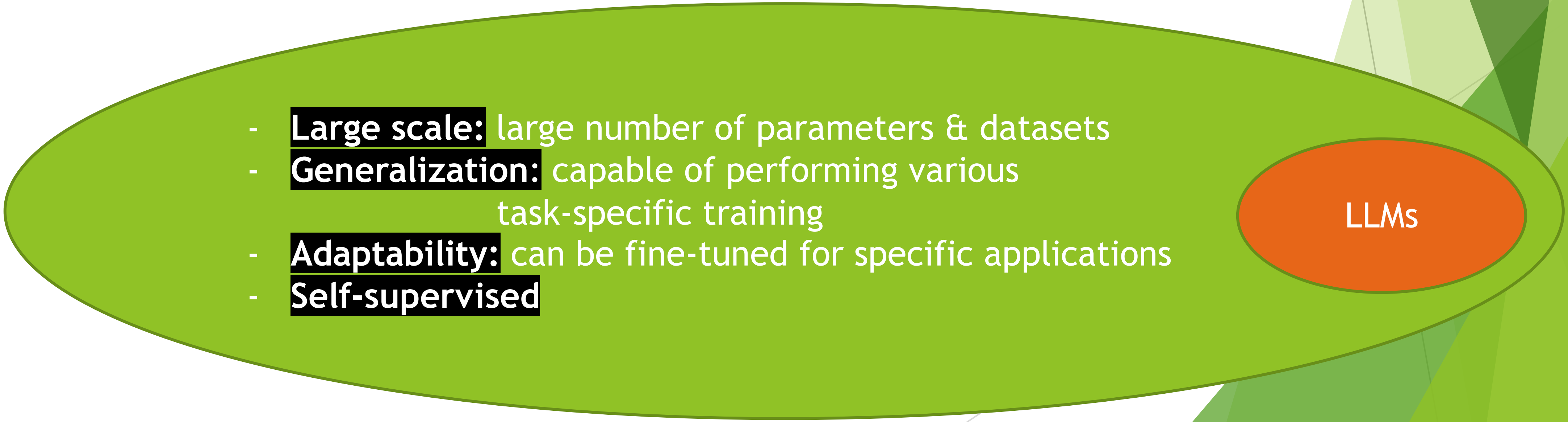
Edited by
Lei Xing
Maryellen L. Giger
James K. Min

AI FOUNDATION MODELS

- **Definition**:

Foundation models are large-scale machine learning models trained on vast amounts of data. They are designed to be adaptable to a wide range of tasks.

Characteristics:



- **Large scale:** large number of parameters & datasets
- **Generalization:** capable of performing various task-specific training
- **Adaptability:** can be fine-tuned for specific applications
- **Self-supervised**

LLMs

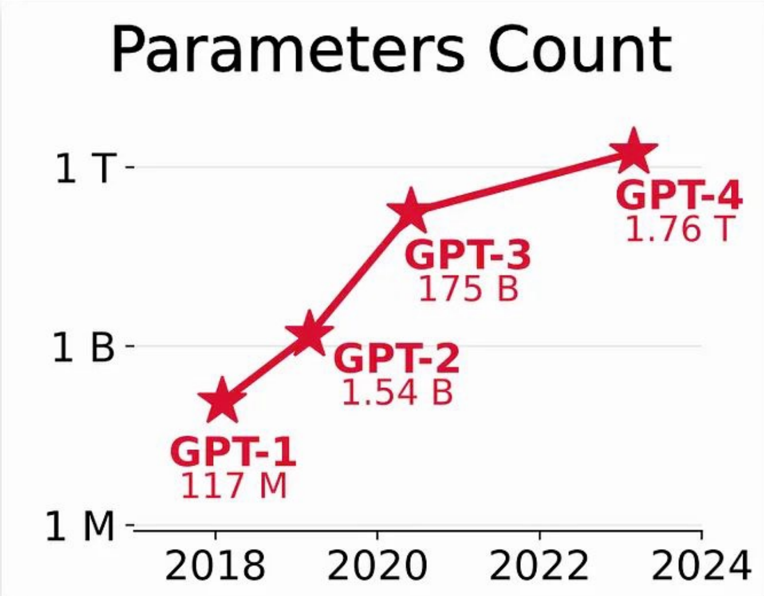
FOUNDATION MODELS

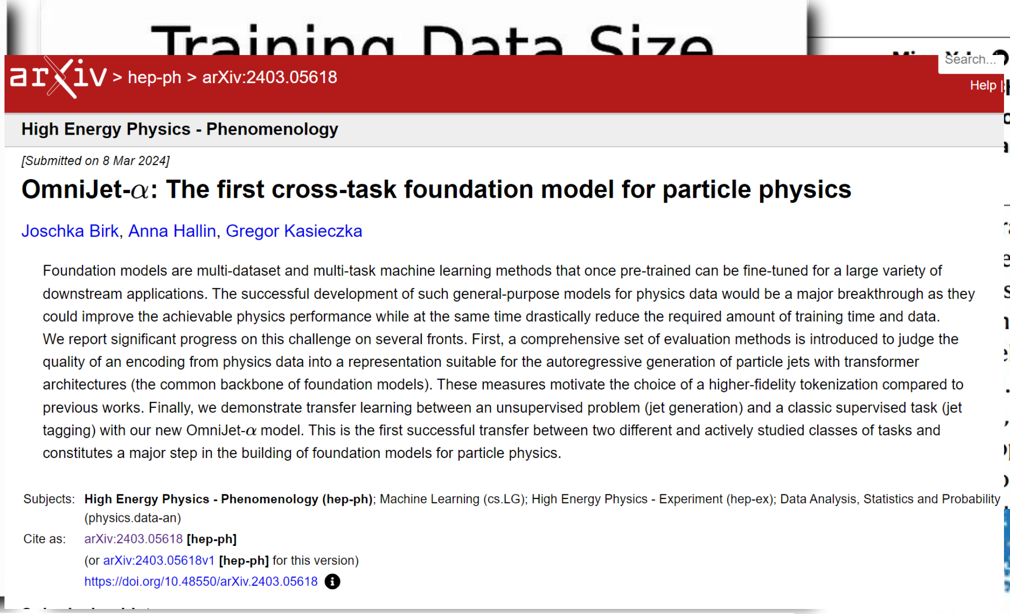
nature medicine

Article

https://doi.org/10.1038/s41591-024-02856-4

A visual-language foundation model for computational pathology





Training Data Size

High Energy Physics - Phenomenology

[Submitted on 8 Mar 2024]

OmniJet-α: The first cross-task foundation model for particle physics

Joschka Birk, Anna Hallin, Gregor Kasieczka

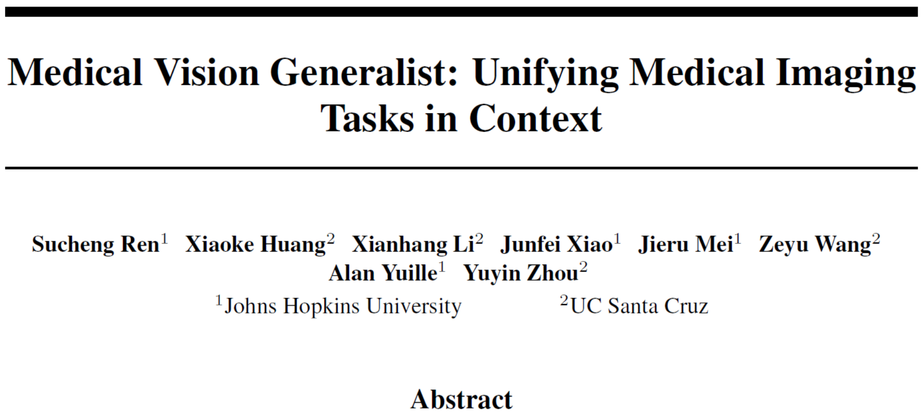
Foundation models are multi-dataset and multi-task machine learning methods that once pre-trained can be fine-tuned for a large variety of downstream applications. The successful development of such general-purpose models for physics data would be a major breakthrough as they could improve the achievable physics performance while at the same time drastically reduce the required amount of training time and data. We report significant progress on this challenge on several fronts. First, a comprehensive set of evaluation methods is introduced to judge the quality of an encoding from physics data into a representation suitable for the autoregressive generation of particle jets with transformer architectures (the common backbone of foundation models). These measures motivate the choice of a higher-fidelity tokenization compared to previous works. Finally, we demonstrate transfer learning between an unsupervised problem (jet generation) and a classic supervised task (jet tagging) with our new OmniJet-α model. This is the first successful transfer between two different and actively studied classes of tasks and constitutes a major step in the building of foundation models for particle physics.

Subjects: **High Energy Physics - Phenomenology (hep-ph)**; Machine Learning (cs.LG); High Energy Physics - Experiment (hep-ex); Data Analysis, Statistics and Probability (physics.data-an)

Cite as: arXiv:2403.05618 **[hep-ph]**
(or arXiv:2403.05618v1 **[hep-ph]** for this version)
https://doi.org/10.48550/arXiv.2403.05618

[1,2,3,4,5,11], Bowen Chen[1,2,11], Drew F. K. Williamson[1,2,3,11],
hen[1,2,3,4,6], Ivy Liang[1,7], Tong Ding[1,7], Guillaume Jaume[1,2,3,4],
ov[1], Long Phi Le[2], Georg Gerber[1], Anil V. Parwani[8],
ang[1,2,3,4,9] & Faisal Mahmood[1,2,3,4,10]

ated adoption of digital pathology and advances in deep learning
ed the development of robust models for various pathology
s a diverse array of diseases and patient cohorts. However,
ing is often difficult due to label scarcity in the medical domain,
l's usage is limited by the specific task and disease for which
. Additionally, most models in histopathology leverage only
, a stark contrast to how humans teach each other and reason
pathologic entities. We introduce CONtrastive learning from
r Histopathology (CONCH), a visual-language foundation model

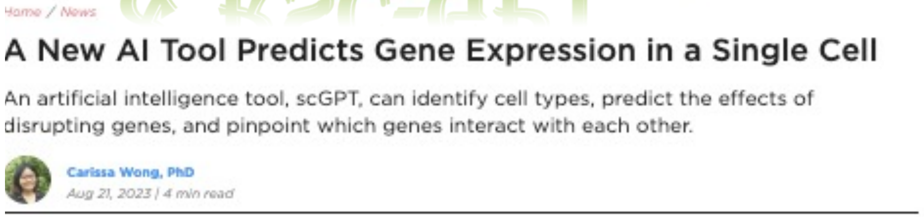


Medical Vision Generalist: Unifying Medical Imaging Tasks in Context

Sucheng Ren[1] Xiaoke Huang[2] Xianhang Li[2] Junfei Xiao[1] Jieru Mei[1] Zeyu Wang[2]
Alan Yuille[1] Yuyin Zhou[2]
[1]Johns Hopkins University [2]UC Santa Cruz

Abstract

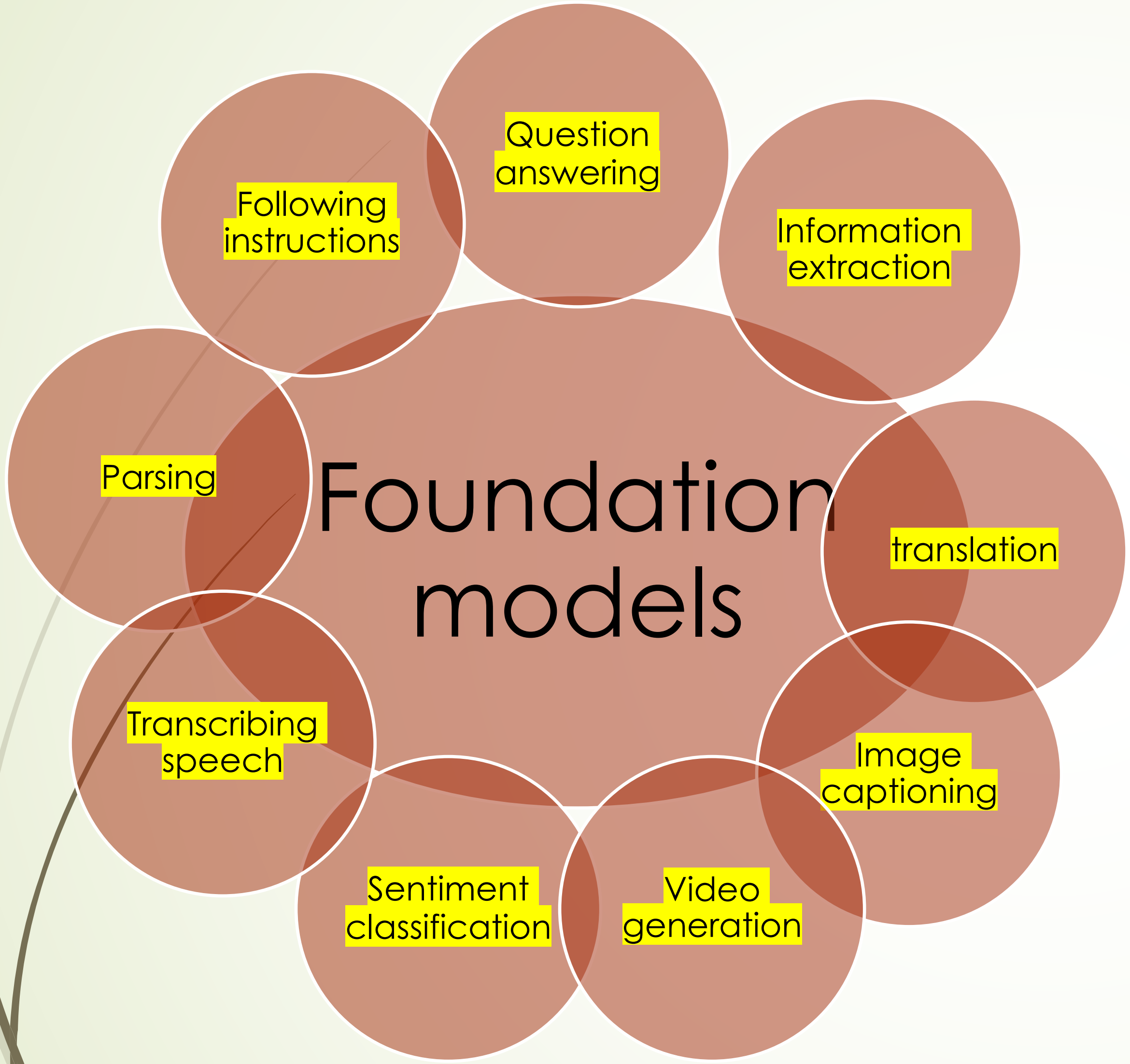
GENEFORMER & RSC-GPT



A New AI Tool Predicts Gene Expression in a Single Cell

An artificial intelligence tool, scGPT, can identify cell types, predict the effects of disrupting genes, and pinpoint which genes interact with each other.

Carissa Wong, PhD
Aug 21, 2023 | 4 min read



Question answering

Following instructions

Information extraction

Parsing

Foundation models

translation

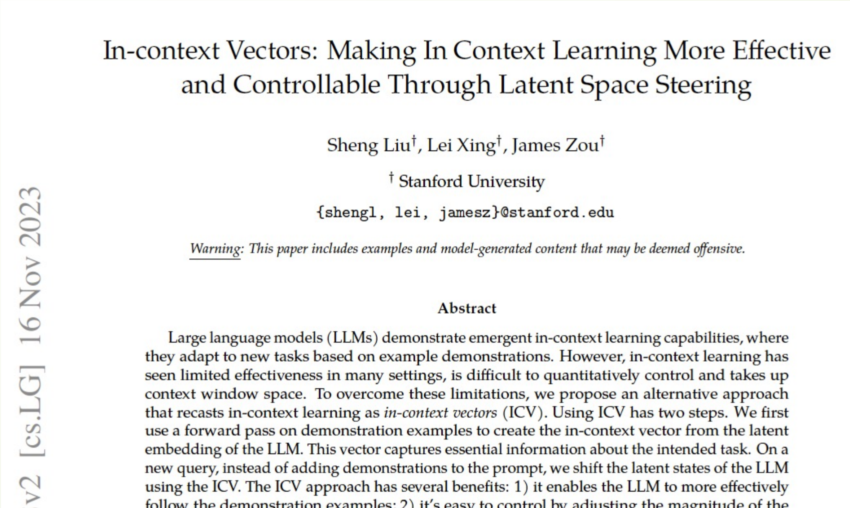
Transcribing speech

Image captioning

Sentiment classification

Video generation

- Fine-tuning
- prompting
- In-context learning



v2 [cs.LG] 16 Nov 2023

In-context Vectors: Making In Context Learning More Effective and Controllable Through Latent Space Steering

Sheng Liu[†], Lei Xing[†], James Zou[†]

[†] Stanford University

{shengl, lei, jamesz}@stanford.edu

Warning: This paper includes examples and model-generated content that may be deemed offensive.

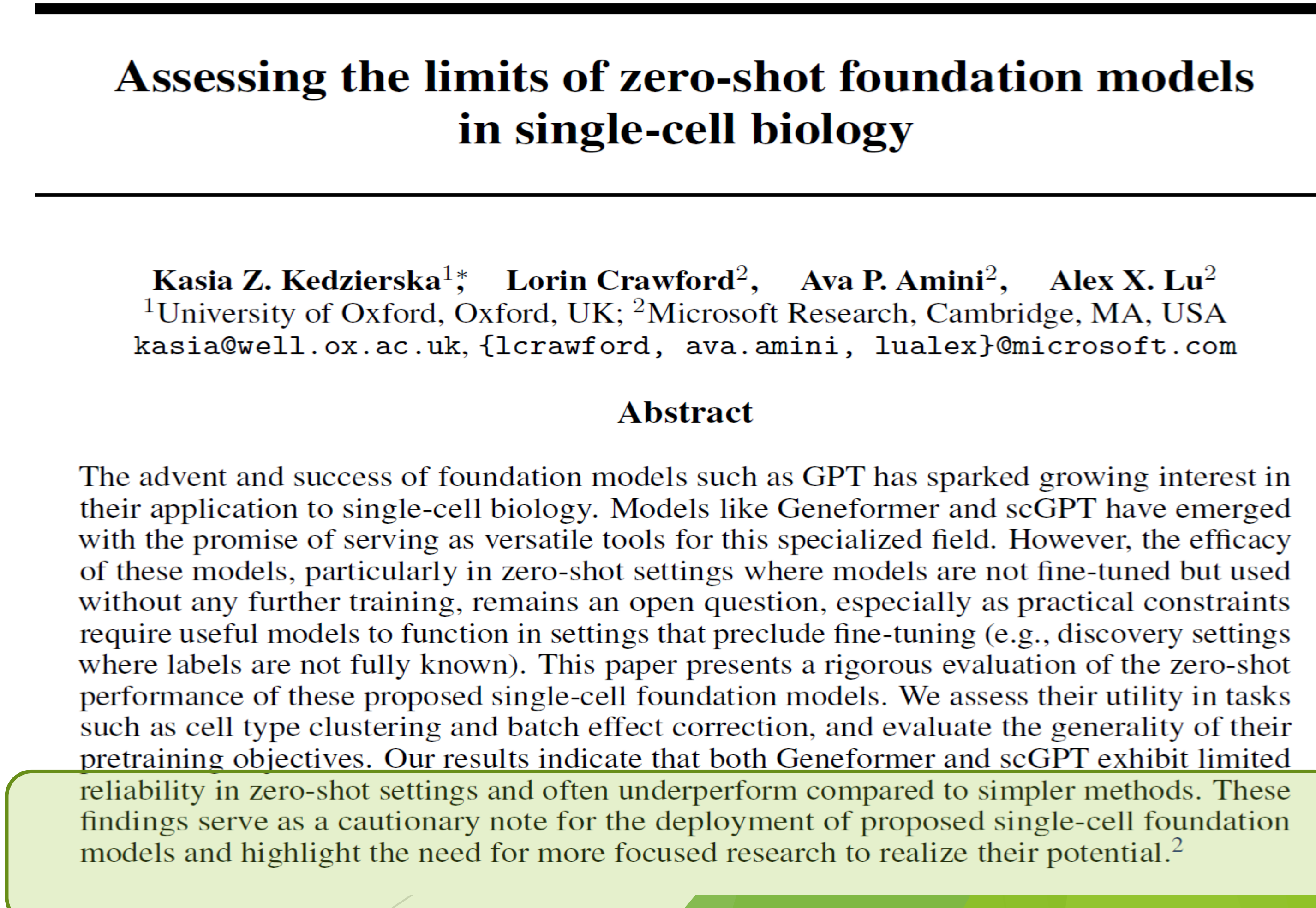
Abstract

Large language models (LLMs) demonstrate emergent in-context learning capabilities, where they adapt to new tasks based on example demonstrations. However, in-context learning has seen limited effectiveness in many settings, is difficult to quantitatively control and takes up context window space. To overcome these limitations, we propose an alternative approach that recasts in-context learning as *in-context vectors* (ICV). Using ICV has two steps. We first use a forward pass on demonstration examples to create the in-context vector from the latent embedding of the LLM. This vector captures essential information about the intended task. On a new query, instead of adding demonstrations to the prompt, we shift the latent states of the LLM using the ICV. The ICV approach has several benefits: 1) it enables the LLM to more effectively follow the demonstration examples; 2) it's easy to control by adjusting the magnitude of the

AI Foundation Models

- **Versatility**: One model can be adapted for multiple tasks (e.g., language translation, summarization, question answering).
- **Efficiency**: Reduces the need for developing and training models from scratch for different tasks.

The deployment of FMs need to be cautious - more focused research is necessary to realize the potential of FMs.

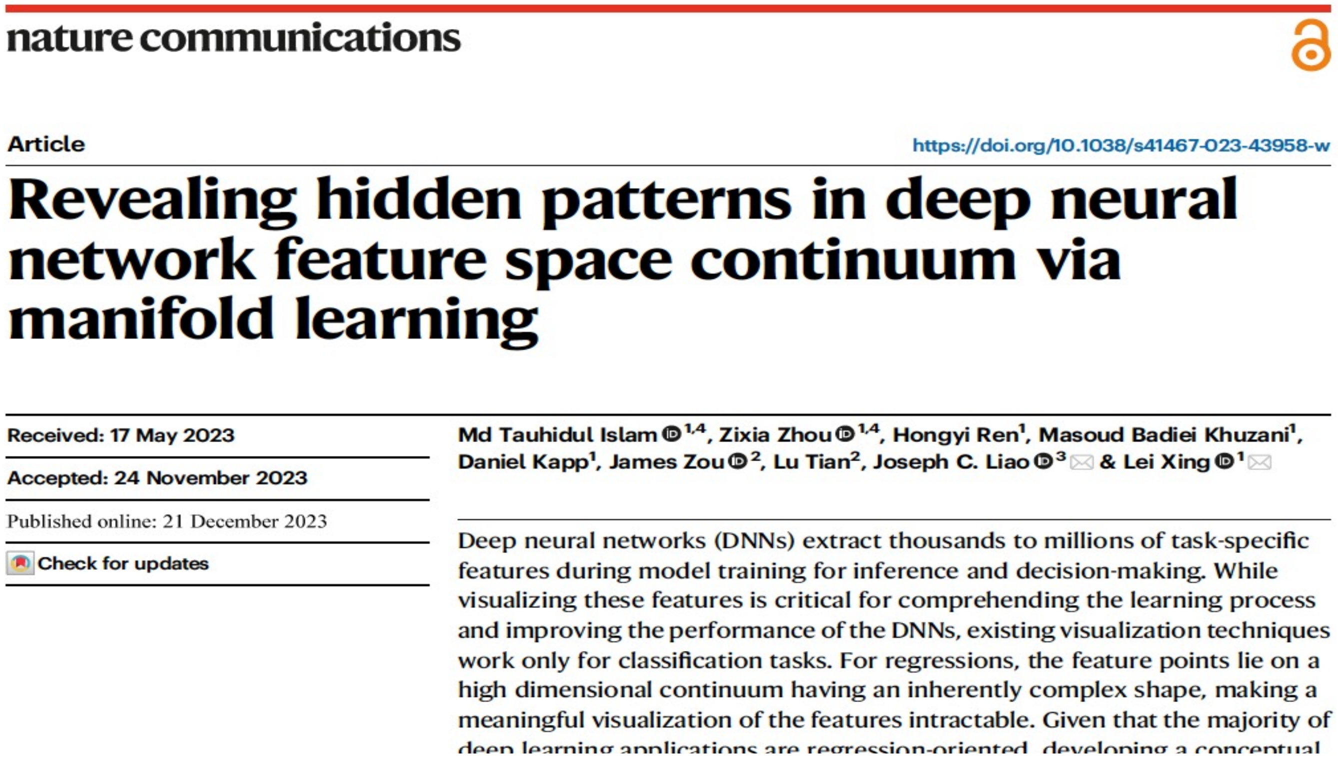


Assessing the limits of zero-shot foundation models in single-cell biology

Kasia Z. Kedzierska[1]*, **Lorin Crawford**[2], **Ava P. Amini**[2], **Alex X. Lu**[2]
[1]University of Oxford, Oxford, UK; [2]Microsoft Research, Cambridge, MA, USA
`kasia@well.ox.ac.uk`, `{lcrawford, ava.amini, lualex}@microsoft.com`

Abstract

The advent and success of foundation models such as GPT has sparked growing interest in their application to single-cell biology. Models like Geneformer and scGPT have emerged with the promise of serving as versatile tools for this specialized field. However, the efficacy of these models, particularly in zero-shot settings where models are not fine-tuned but used without any further training, remains an open question, especially as practical constraints require useful models to function in settings that preclude fine-tuning (e.g., discovery settings where labels are not fully known). This paper presents a rigorous evaluation of the zero-shot performance of these proposed single-cell foundation models. We assess their utility in tasks such as cell type clustering and batch effect correction, and evaluate the generality of their pretraining objectives. Our results indicate that both Geneformer and scGPT exhibit limited reliability in zero-shot settings and often underperform compared to simpler methods. These findings serve as a cautionary note for the deployment of proposed single-cell foundation models and highlight the need for more focused research to realize their potential.[2]



nature communications

Article

https://doi.org/10.1038/s41467-023-43958-w

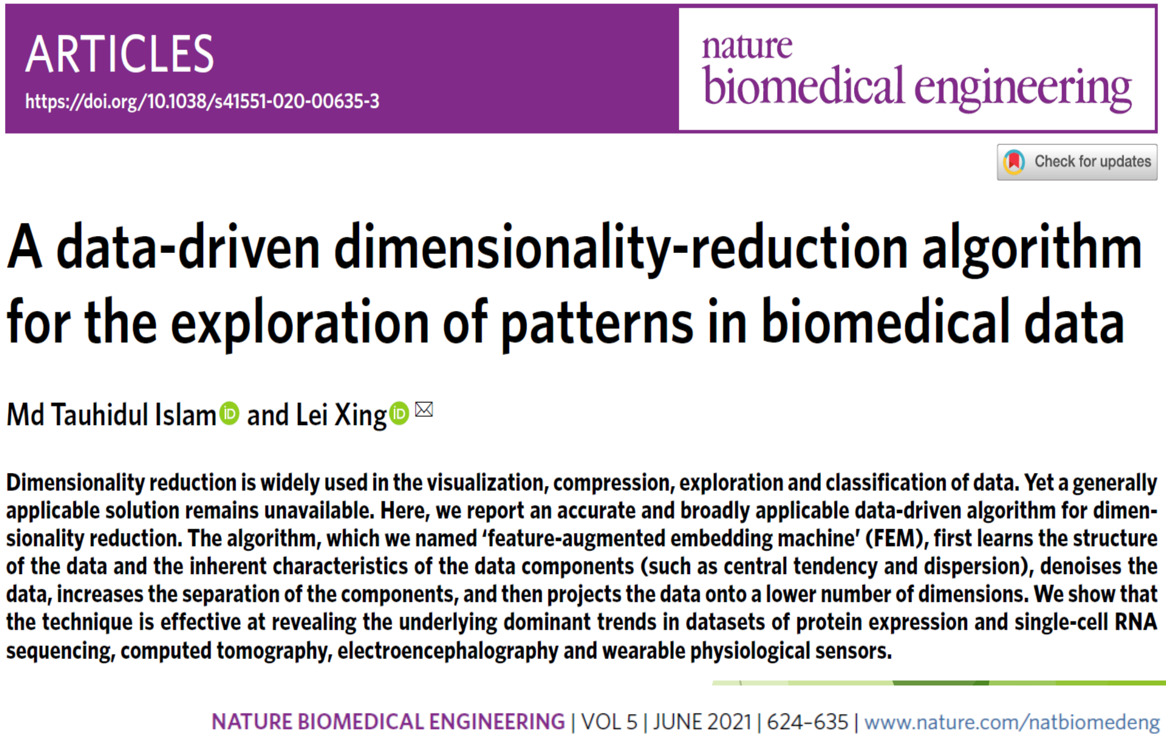
Revealing hidden patterns in deep neural network feature space continuum via manifold learning

Received: 17 May 2023

Accepted: 24 November 2023

Md Tauhidul Islam[1,4], Zixia Zhou[1,4], Hongyi Ren[1], Masoud Badiei Khuzani[1], Daniel Kapp[1], James Zou[2], Lu Tian[2], Joseph C. Liao[3] & Lei Xing[1]

Deep neural networks (DNNs) extract thousands to millions of task-specific features during model training for inference and decision-making. While visualizing these features is critical for comprehending the learning process and improving the performance of the DNNs, existing visualization techniques work only for classification tasks. For regressions, the feature points lie on a high dimensional continuum having an inherently complex shape, making a meaningful visualization of the features intractable. Given that the majority of deep learning applications are regression-oriented, developing a conceptual



ARTICLES

https://doi.org/10.1038/s41551-020-00635-3

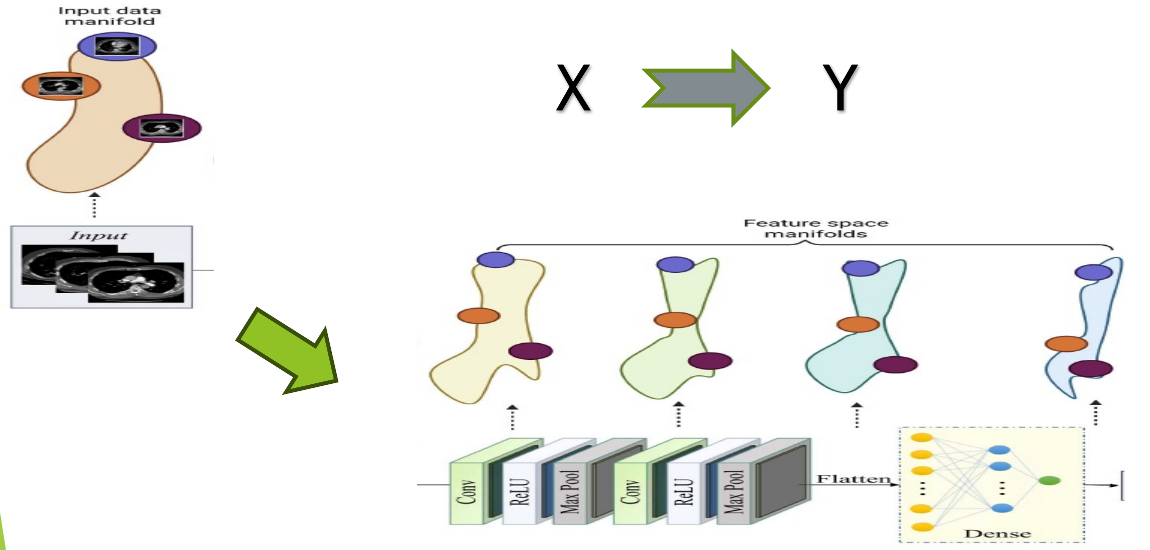
nature biomedical engineering

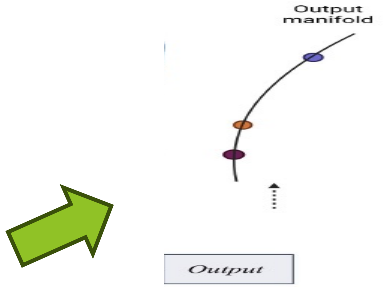
A data-driven dimensionality-reduction algorithm for the exploration of patterns in biomedical data

Md Tauhidul Islam and Lei Xing

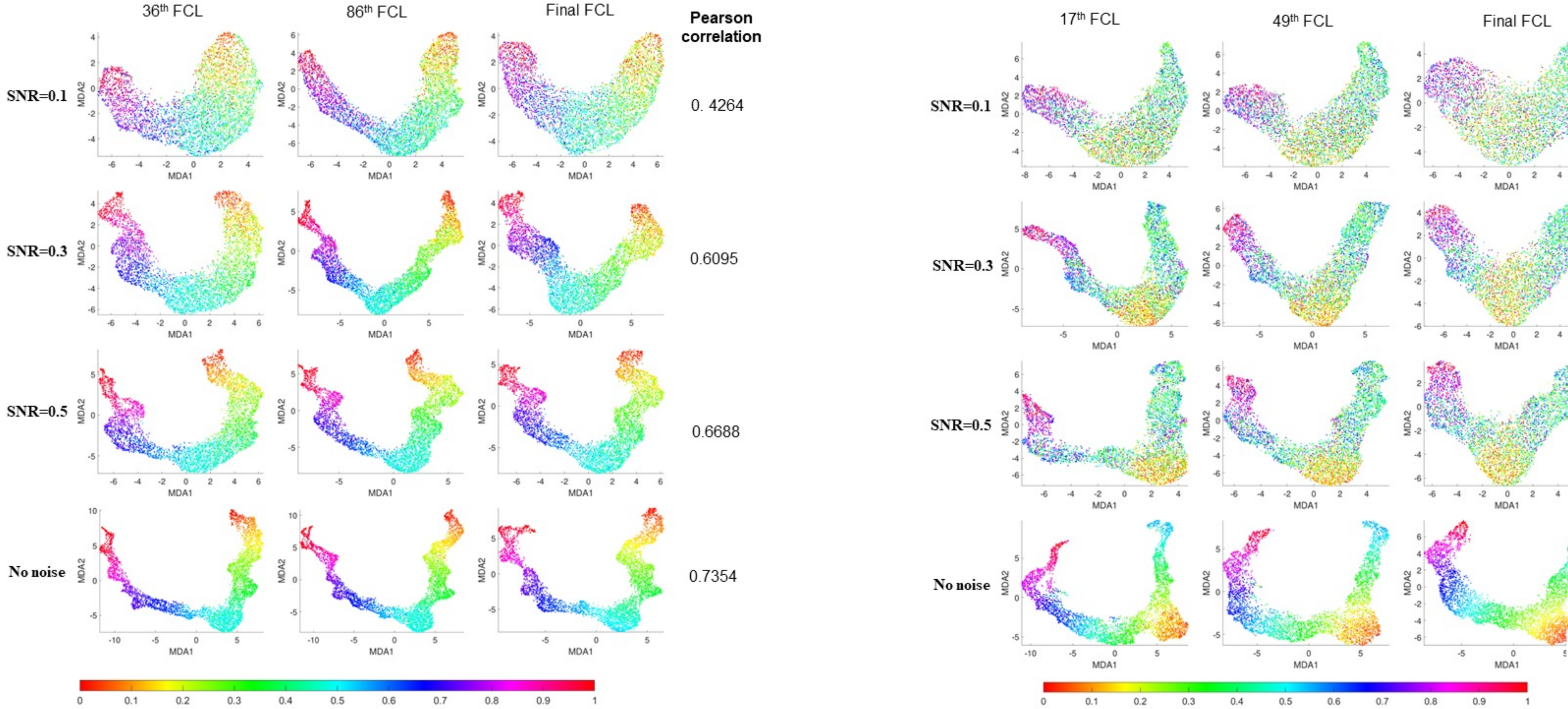
Dimensionality reduction is widely used in the visualization, compression, exploration and classification of data. Yet a generally applicable solution remains unavailable. Here, we report an accurate and broadly applicable data-driven algorithm for dimensionality reduction. The algorithm, which we named 'feature-augmented embedding machine' (FEM), first learns the structure of the data and the inherent characteristics of the data components (such as central tendency and dispersion), denoises the data, increases the separation of the components, and then projects the data onto a lower number of dimensions. We show that the technique is effective at revealing the underlying dominant trends in datasets of protein expression and single-cell RNA sequencing, computed tomography, electroencephalography and wearable physiological sensors.

NATURE BIOMEDICAL ENGINEERING | VOL 5 | JUNE 2021 | 624–635 | www.nature.com/natbiomedeng



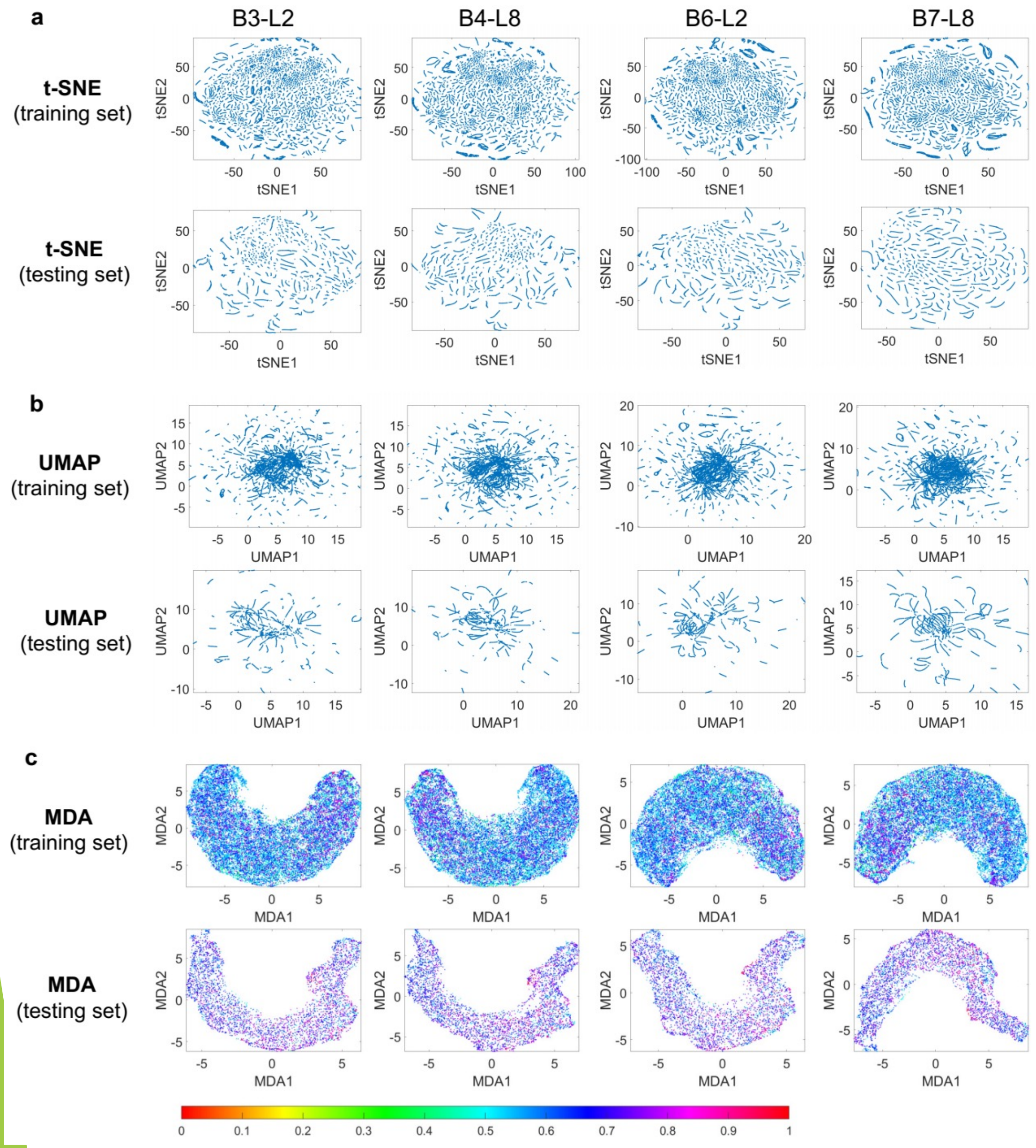


Nature Communications 14, 8506, 2023

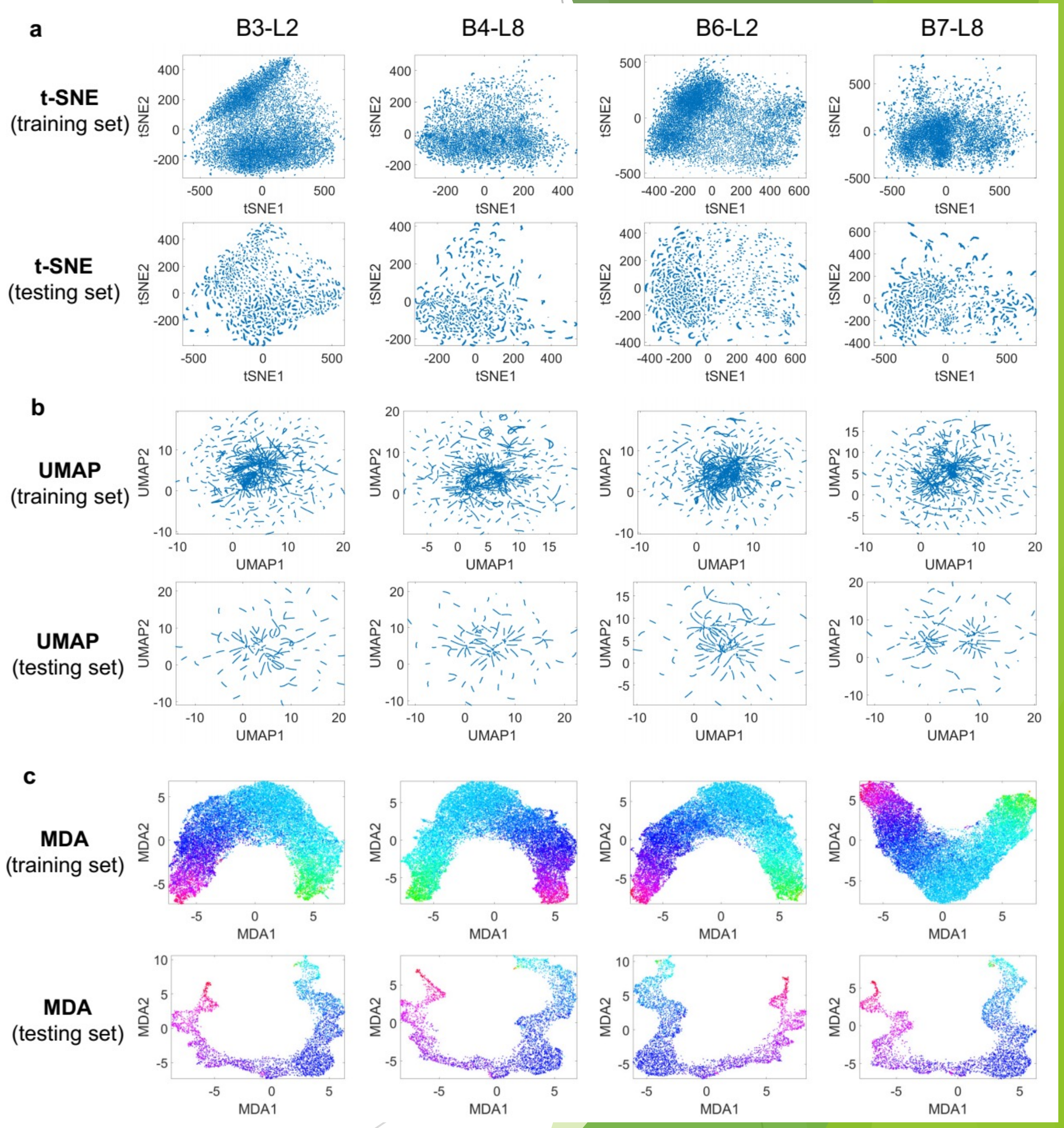


MDA visualization of GPT-2 and BERT finetuning features at 3 different layers at different SNRs. Red to violet color denotes the survival da 10,000 days. Colorbar denotes the normalized survival days.

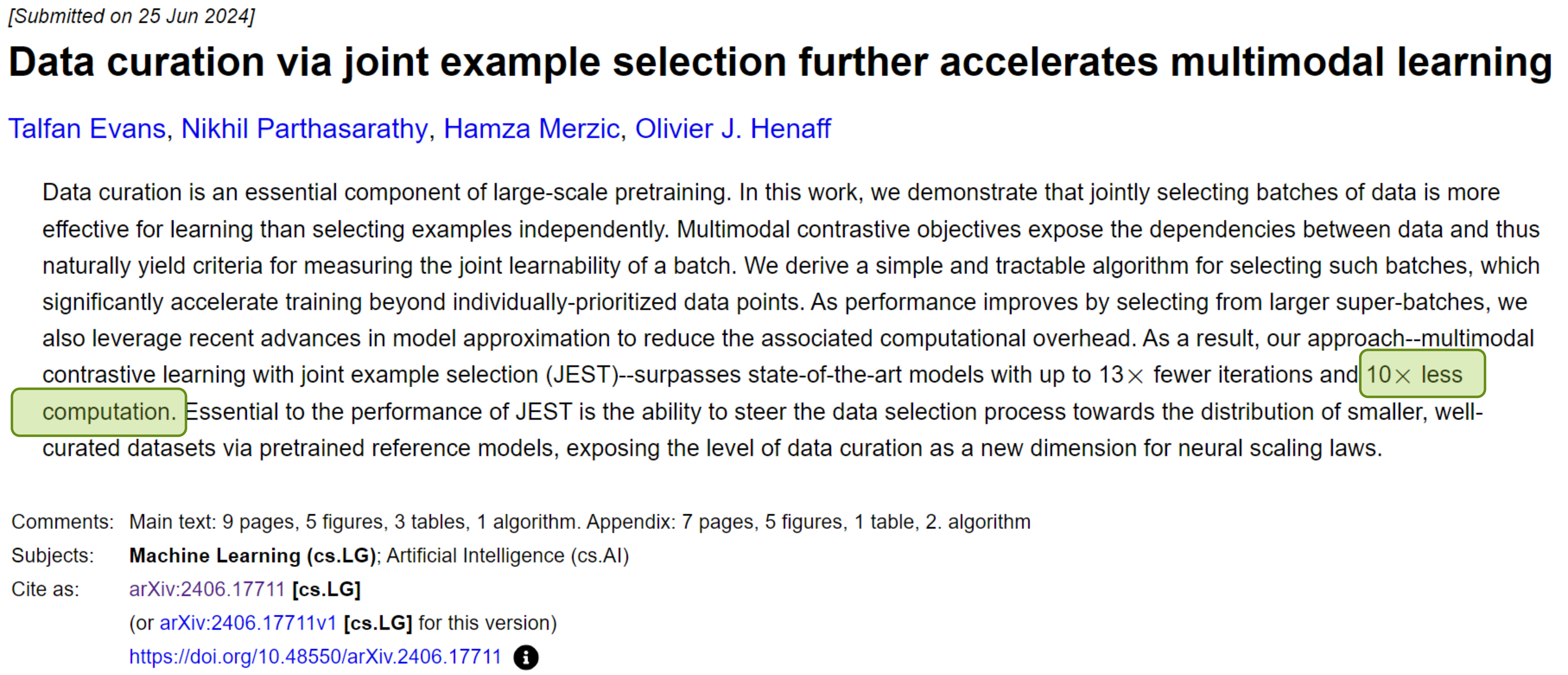
Before training



After training



Brain tumor segmentation in MRI- Dense-UNet



[Submitted on 25 Jun 2024]

Data curation via joint example selection further accelerates multimodal learning

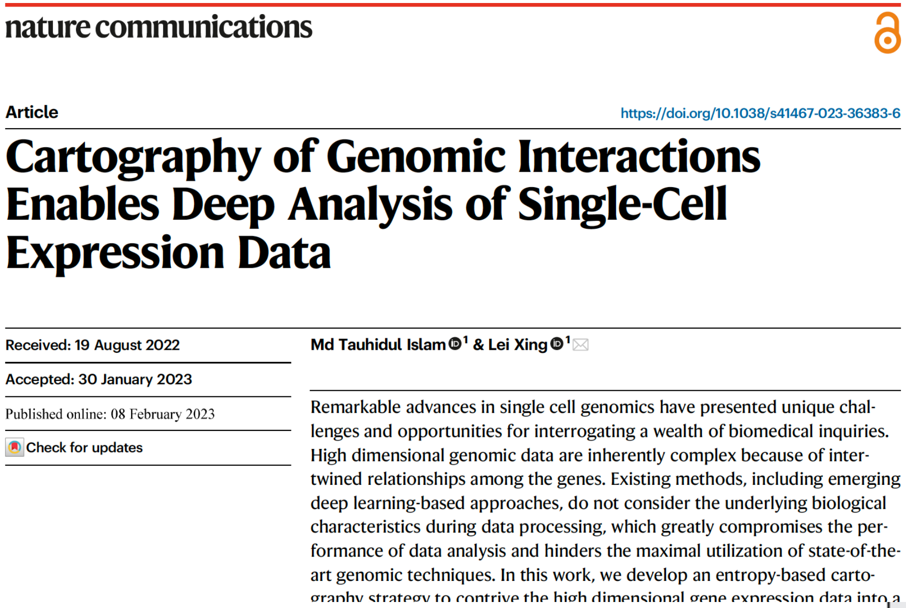
Talfan Evans, Nikhil Parthasarathy, Hamza Merzic, Olivier J. Henaff

Data curation is an essential component of large-scale pretraining. In this work, we demonstrate that jointly selecting batches of data is more effective for learning than selecting examples independently. Multimodal contrastive objectives expose the dependencies between data and thus naturally yield criteria for measuring the joint learnability of a batch. We derive a simple and tractable algorithm for selecting such batches, which significantly accelerate training beyond individually-prioritized data points. As performance improves by selecting from larger super-batches, we also leverage recent advances in model approximation to reduce the associated computational overhead. As a result, our approach--multimodal contrastive learning with joint example selection (JEST)--surpasses state-of-the-art models with up to 13× fewer iterations and 10× less computation. Essential to the performance of JEST is the ability to steer the data selection process towards the distribution of smaller, well-curated datasets via pretrained reference models, exposing the level of data curation as a new dimension for neural scaling laws.

Comments: Main text: 9 pages, 5 figures, 3 tables, 1 algorithm. Appendix: 7 pages, 5 figures, 1 table, 2. algorithm
Subjects: **Machine Learning (cs.LG)**; Artificial Intelligence (cs.AI)
Cite as: arXiv:2406.17711 **[cs.LG]**
(or arXiv:2406.17711v1 **[cs.LG]** for this version)
https://doi.org/10.48550/arXiv.2406.17711

AI FOR OMICS/RADIOMICS DATA PROCESSING

www.AnalyXus.com



nature communications

Article https://doi.org/10.1038/s41467-023-36383-6

Cartography of Genomic Interactions Enables Deep Analysis of Single-Cell Expression Data

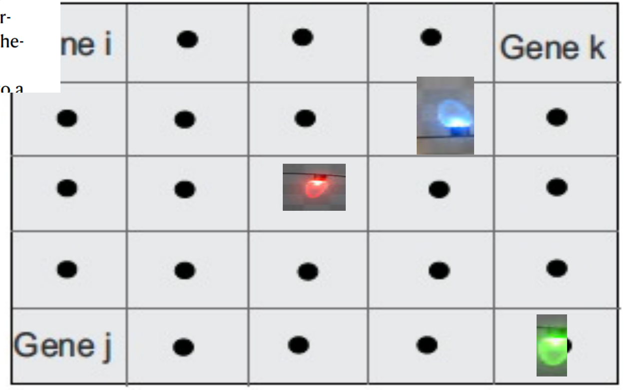
Received: 19 August 2022

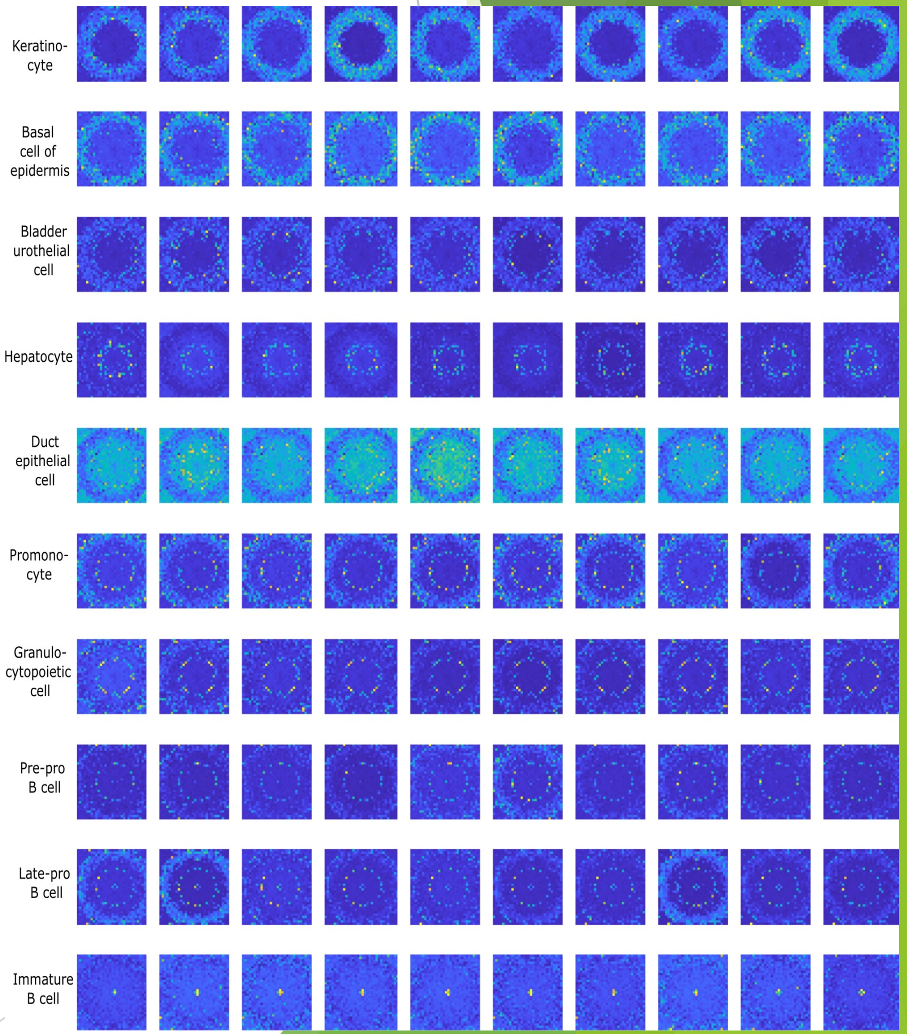
Accepted: 30 January 2023

Md Tauhidul Islam 1 & Lei Xing 1

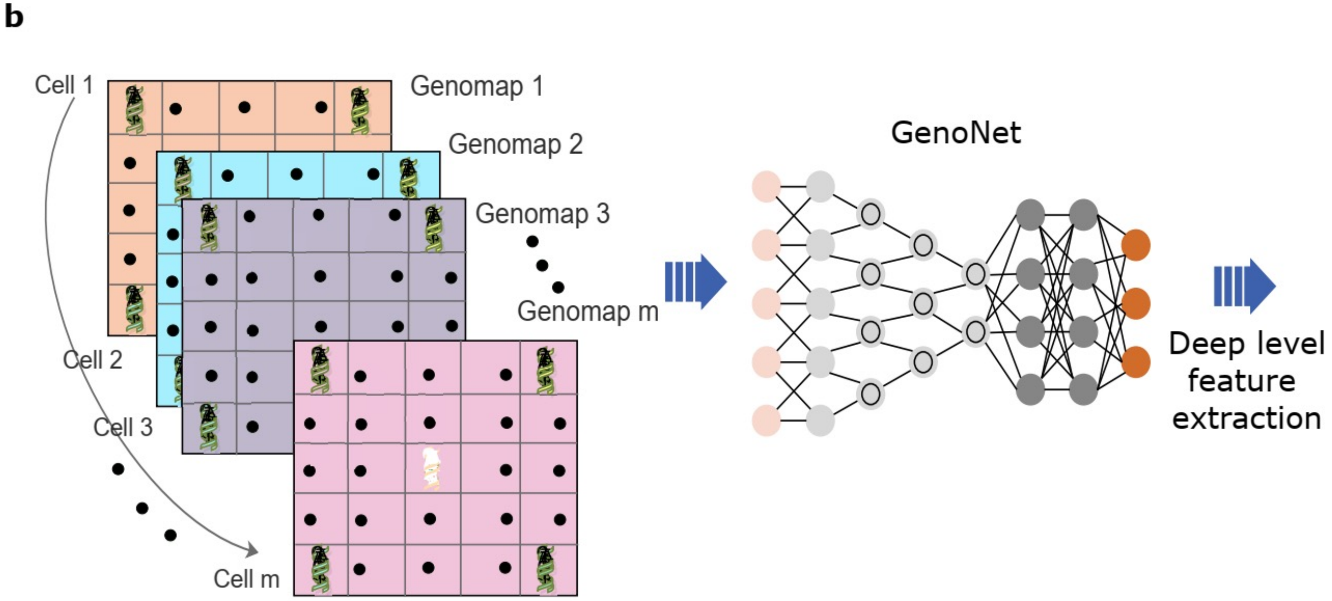
Remarkable advances in single cell genomics have presented unique challenges and opportunities for interrogating a wealth of biomedical inquiries. High dimensional genomic data are inherently complex because of intertwined relationships among the genes. Existing methods, including emerging deep learning-based approaches, do not consider the underlying biological characteristics during data processing, which greatly compromises the performance of data analysis and hinders the maximal utilization of state-of-the-art genomic techniques. In this work, we develop an entropy-based cartography strategy to contrive the high dimensional gene expression data into a

https://www.nature.com/articles/s41467-023-36383-6

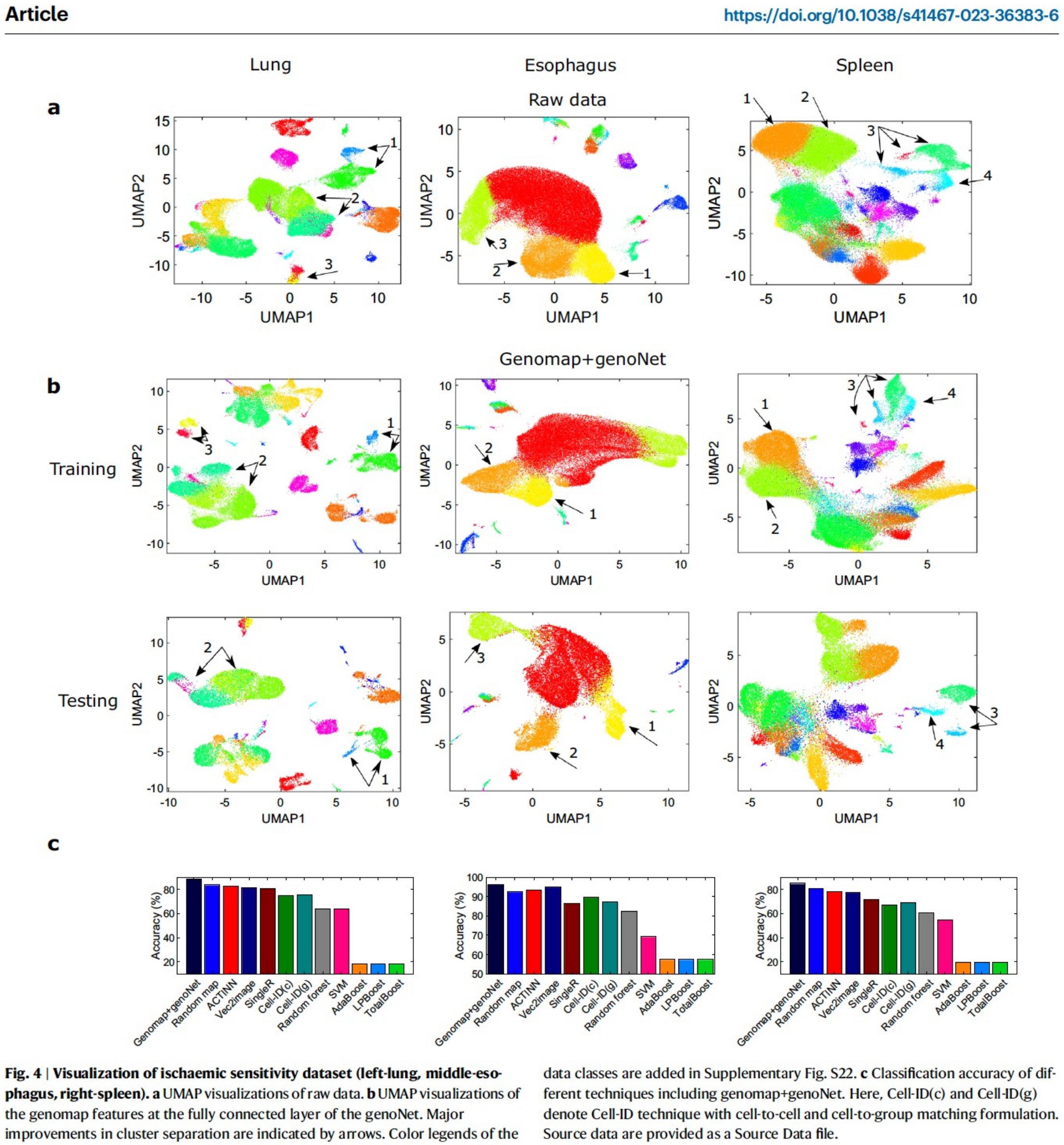




Deep analysis of RNA-seq data



Deep level feature extraction


Article

https://doi.org/10.1038/s41467-023-36383-6

Fig. 4 | Visualization of ischaemic sensitivity dataset (left-lung, middle-esophagus, right-spleen). a UMAP visualizations of raw data. **b** UMAP visualizations of the genomap features at the fully connected layer of the genoNet. Major improvements in cluster separation are indicated by arrows. Color legends of the data classes are added in Supplementary Fig. S22. **c** Classification accuracy of different techniques including genomap+genoNet. Here, Cell-ID(c) and Cell-ID(g) denote Cell-ID technique with cell-to-cell and cell-to-group matching formulation. Source data are provided as a Source Data file.

www.AnalyXus.com

How can FMs help AI Radiation Therapy?

Introduce additional information to enhance the solution.

- Tumor volume delineation
- Treatment planning
- Clinical decision making
- ……

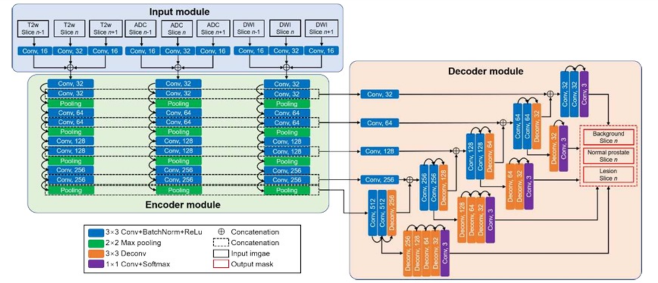
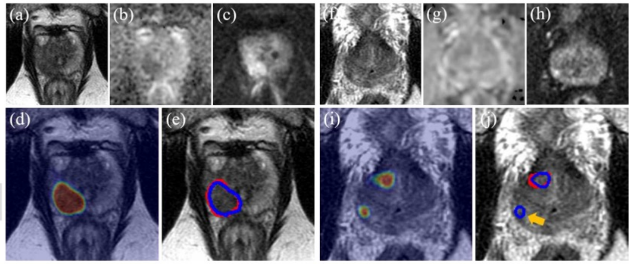
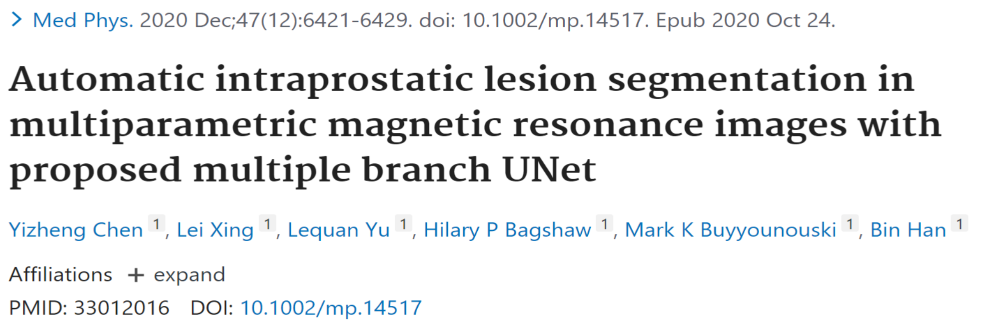


Figure 1. Architecture of the proposed MB-UNet.





> Med Phys. 2020 Dec;47(12):6421-6429. doi: 10.1002/mp.14517. Epub 2020 Oct 24.

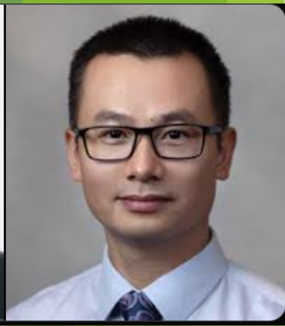
Automatic intraprostatic lesion segmentation in multiparametric magnetic resonance images with proposed multiple branch UNet

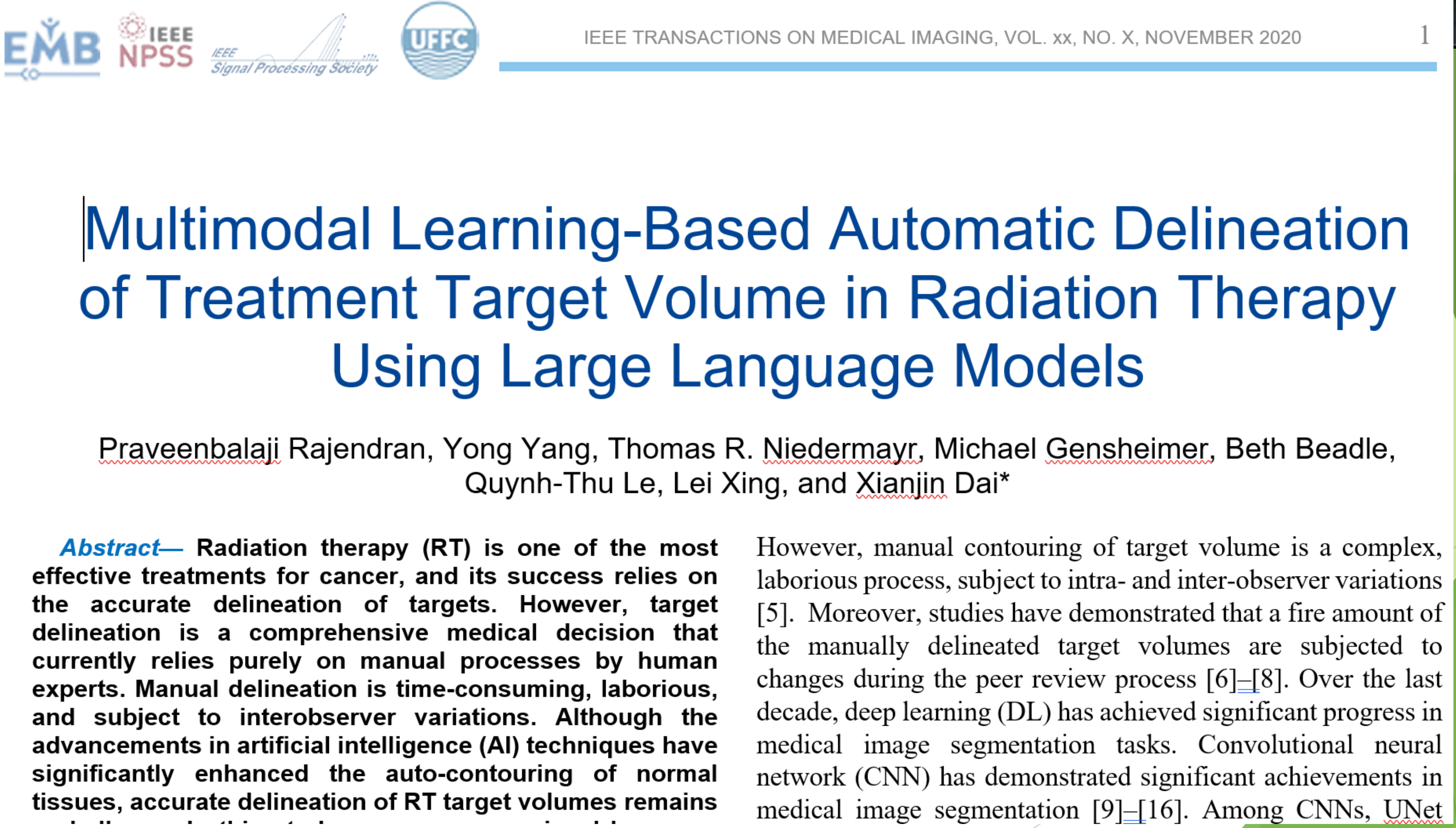
Yizheng Chen [1], Lei Xing [1], Lequan Yu [1], Hilary P Bagshaw [1], Mark K Buyyounouski [1], Bin Han [1]

Affiliations + expand

PMID: 33012016 DOI: 10.1002/mp.14517

Radformer





IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. xx, NO. X, NOVEMBER 2020 1

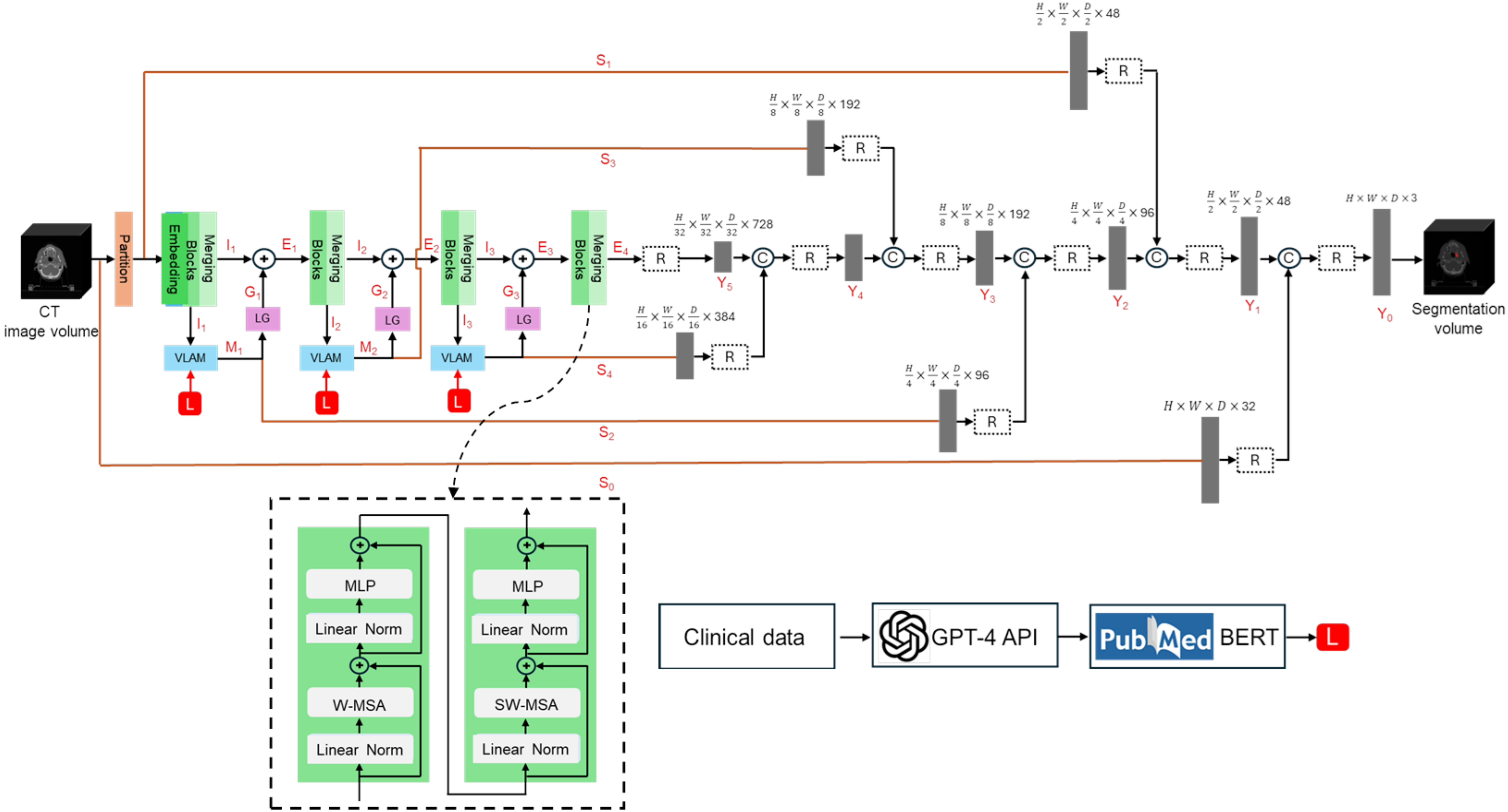
Multimodal Learning-Based Automatic Delineation of Treatment Target Volume in Radiation Therapy Using Large Language Models

Praveenbalaji Rajendran, Yong Yang, Thomas R. Niedermayr, Michael Gensheimer, Beth Beadle,
Quynh-Thu Le, Lei Xing, and Xianjin Dai*

***Abstract*— Radiation therapy (RT) is one of the most effective treatments for cancer, and its success relies on the accurate delineation of targets. However, target delineation is a comprehensive medical decision that currently relies purely on manual processes by human experts. Manual delineation is time-consuming, laborious, and subject to interobserver variations. Although the advancements in artificial intelligence (AI) techniques have significantly enhanced the auto-contouring of normal tissues, accurate delineation of RT target volumes remains**

However, manual contouring of target volume is a complex, laborious process, subject to intra- and inter-observer variations [5]. Moreover, studies have demonstrated that a fire amount of the manually delineated target volumes are subjected to changes during the peer review process [6]–[8]. Over the last decade, deep learning (DL) has achieved significant progress in medical image segmentation tasks. Convolutional neural network (CNN) has demonstrated significant achievements in medical image segmentation [9]–[16]. Among CNNs, UNet

Radformer


CT
image volume
Partition
Embedding
Blocks
Merging
I1
E1
G1
LG
VLAM
M1
L
Blocks
Merging
I2
E2
G2
M2
I3
E3
G3
E4
R
Y5
Y4
Y3
Y2
Y1
Y0
C
S1
S2
S3
S4
S0
H/2 × W/2 × D/2 × 48
H/8 × W/8 × D/8 × 192
H/32 × W/32 × D/32 × 728
H/16 × W/16 × D/16 × 384
H/8 × W/8 × D/8 × 192
H/4 × W/4 × D/4 × 96
H/2 × W/2 × D/2 × 48
H × W × D × 3
H/4 × W/4 × D/4 × 96
H × W × D × 32
Segmentation
volume
MLP
Linear Norm
W-MSA
Linear Norm
MLP
Linear Norm
SW-MSA
Linear Norm
Clinical data
GPT-4 API
PubMed BERT
L

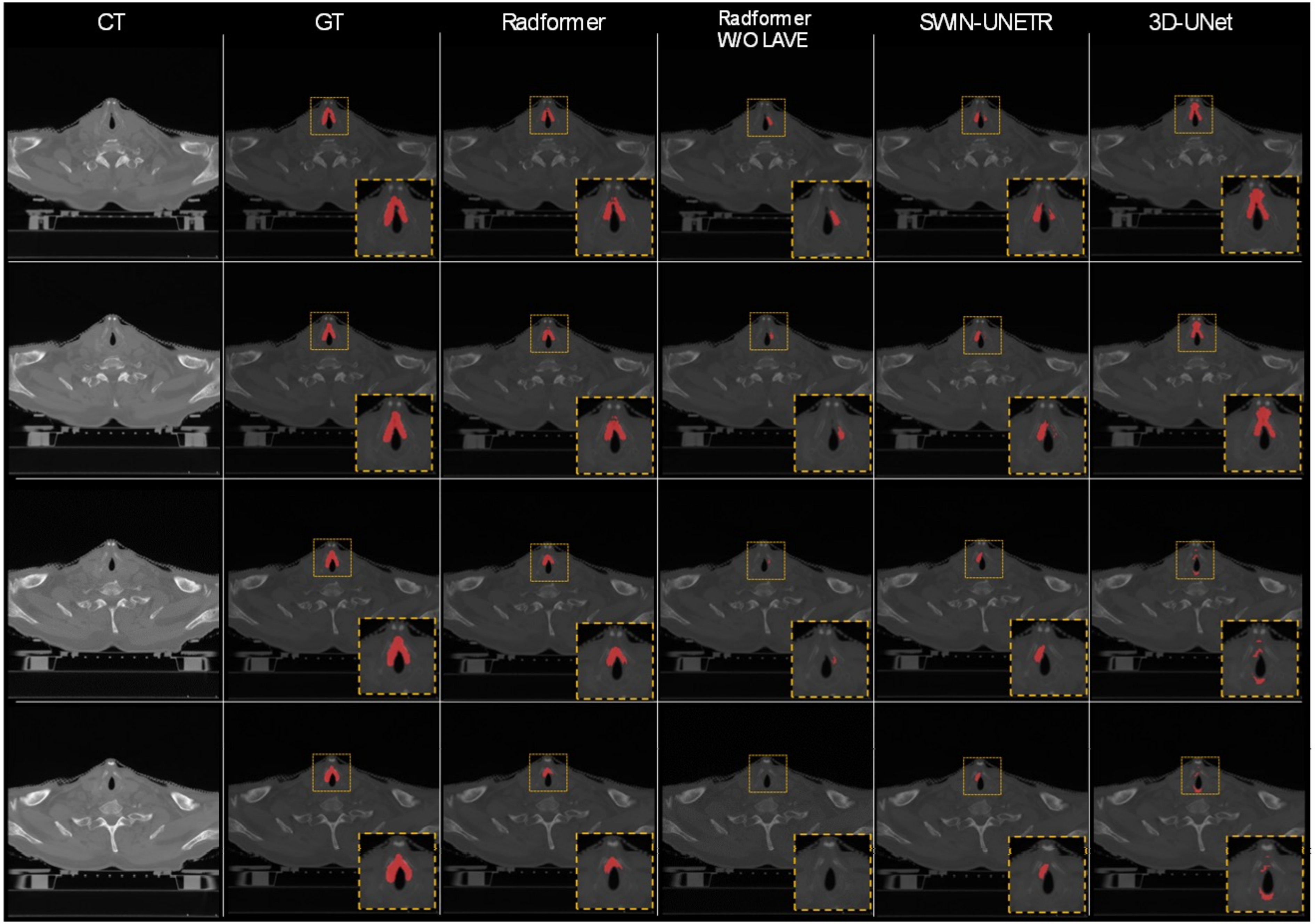


Fig. 4. Illustrative example of the GTV segmentation for a patient with early-stage laryngeal cancer

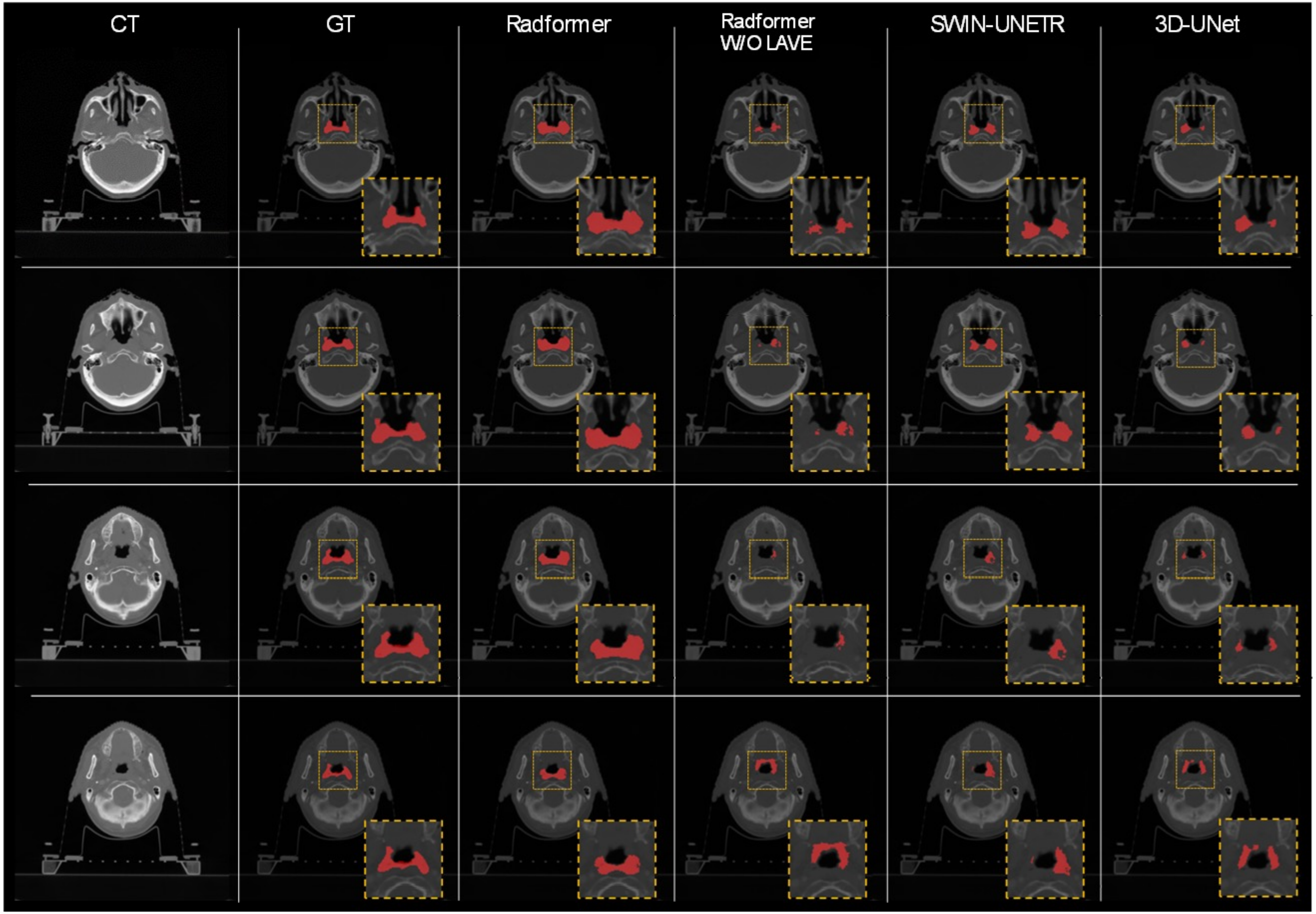


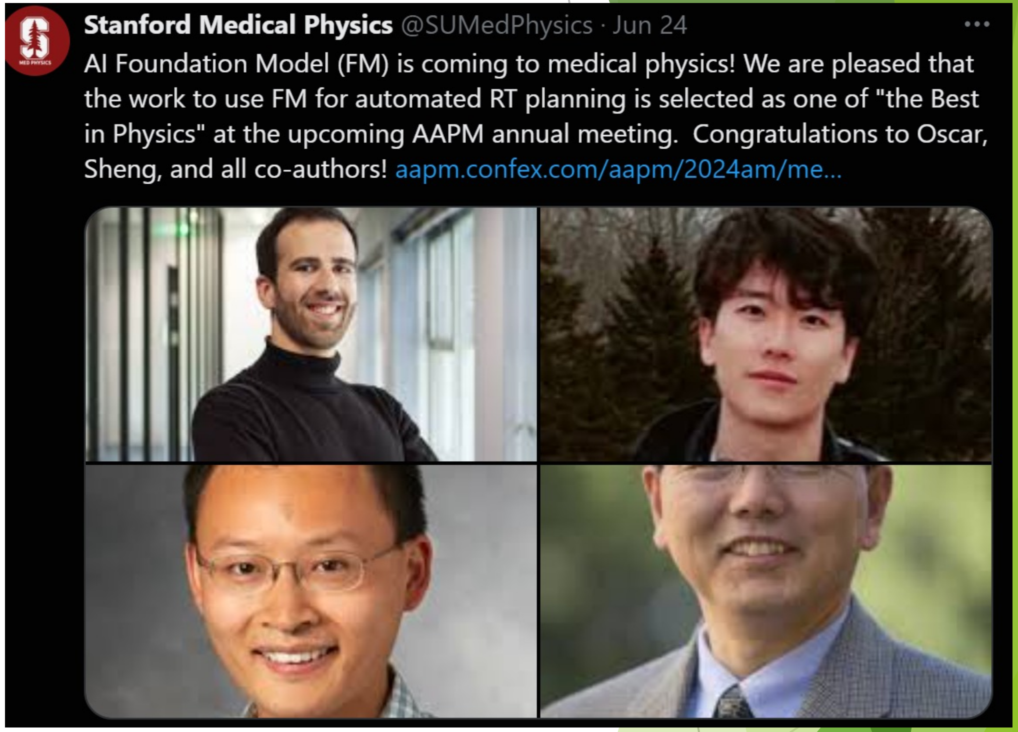
Fig. 5. Illustrative example of the GTV segmentation for a patient stage III nasopharyngeal cancer

Automated radiotherapy treatment planning guided by GPT-4Vision

Sheng Liu[1,2]*, Oscar Pastor-Serrano[1]*, Yizheng Chen[1], Matthew Gopaulchan[1],
Weixing Liang[3], Mark Buyyounouski[1], Erqi Pollom[1], Quynh-Thu Le[1],
Michael Gensheimer[1], Peng Dong[1], Yong Yang[1], James Zou[2,3]† and Lei Xing[1]†
[1]Department of Radiation Oncology, Stanford University, Stanford, CA, USA
[2]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
[3]Department of Computer Science, Stanford University, Stanford, CA, USA

Abstract

Radiotherapy treatment planning is a time-consuming and potentially subjective process that requires the iterative adjustment of model parameters to balance multiple conflicting objectives. Recent advancements in frontier AI models offer promising avenues for addressing the challenges in planning and clinical decision-making. This study introduces GPT-RadPlan, an automated treatment planning framework that integrates radiation oncology knowledge with the reasoning capabilities of large multi-modal models, such as GPT-4Vision (GPT-4V) from OpenAI. Via in-context learning, we incorporate clinical protocols for various disease sites to enable GPT-4V to acquire treatment planning domain knowledge. The resulting GPT-RadPlan agent is integrated into our in-house inverse treatment planning system through an API. For a given patient, GPT-RadPlan acts as both plan evaluator and planner, first assessing dose distributions and dose-volume histograms (DVHs), and then providing "textual feedback" on how to improve the plan. In this manner, the agent iteratively refines the plan by adjusting planning parameters, such as weights and objective doses, based on its suggestions. The efficacy of the automated planning system is showcased across multiple prostate and head & neck cancer cases, where we compared GPT-RadPlan results to clinical plans produced by human experts. In all cases, GPT-RadPlan either outperformed or matched the clinical plans, demonstrating superior target coverage and organ-at-risk sparing. Consistently satisfying the dosimetric objectives in the clinical protocol, GPT-RadPlan represents the first multimodal large language model agent that mimics the behaviors of human planners in radiation oncology clinics, achieving promising results in automating the treatment planning process without the need for additional training.



Stanford Medical Physics @SUMedPhysics · Jun 24

AI Foundation Model (FM) is coming to medical physics! We are pleased that the work to use FM for automated RT planning is selected as one of "the Best in Physics" at the upcoming AAPM annual meeting. Congratulations to Oscar, Sheng, and all co-authors! aapm.confex.com/aapm/2024am/me...

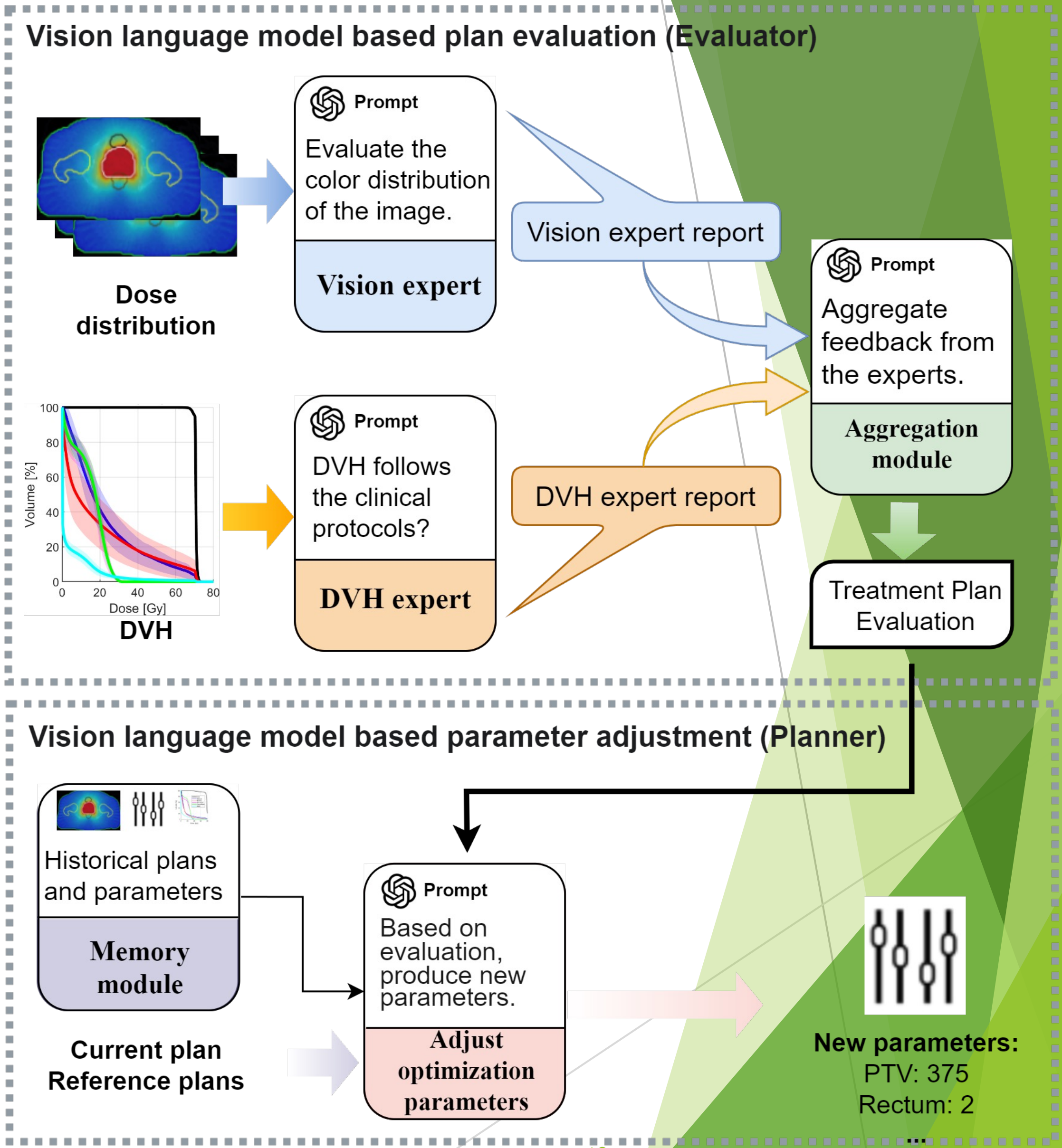
Liu S, Pastor-Serrano O, ….., arXiv preprint arXiv:2406.15609, 2024

GPT-RadPlan

- Based on GPT4-Vision

- Evaluation module:
  - DVH Expert – compares metrics with protocol/intent
  - Image expert – evaluates dose distribution and presence of hot/cold spots
  - Aggregation module – improvement suggestions

- Planner module: suggest new parameters based on
  - Stored information from previous iterations
  - The current plan
  - 3 reference approved plans from the same disease site



Vision language model based plan evaluation (Evaluator)

Dose distribution

Prompt Evaluate the color distribution of the image.

Vision expert

Vision expert report

DVH

Prompt DVH follows the clinical protocols?

DVH expert

DVH expert report

Prompt Aggregate feedback from the experts.

Aggregation module

Treatment Plan Evaluation

Vision language model based parameter adjustment (Planner)

Historical plans and parameters

Memory module

Current plan
Reference plans

Prompt Based on evaluation, produce new parameters.

Adjust optimization parameters

New parameters:
PTV: 375
Rectum: 2
...

DVH comparison

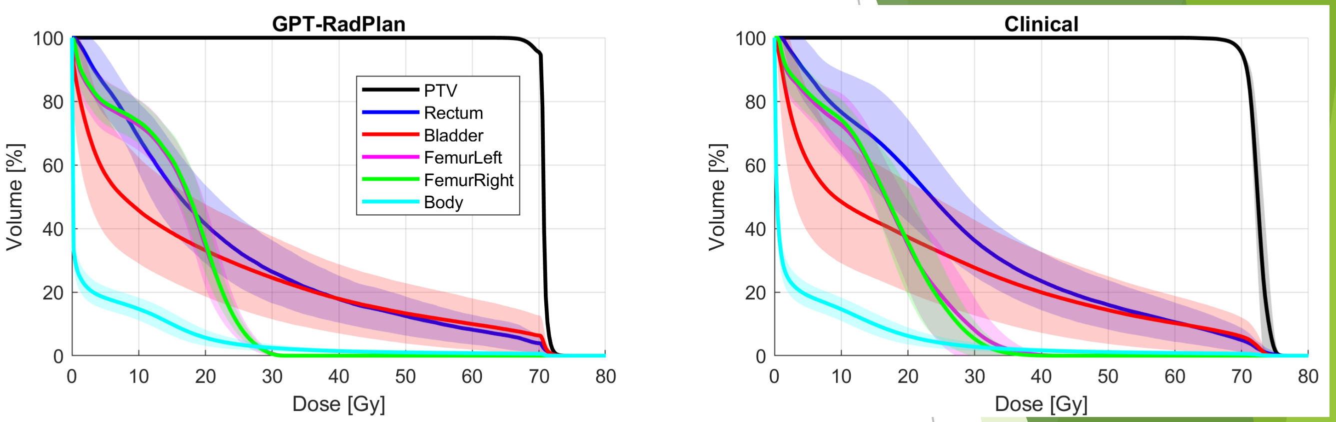
GPT-RadPlan consistently meets the protocols
Overall, 15% reduction in prostate mean dose

Prostate:

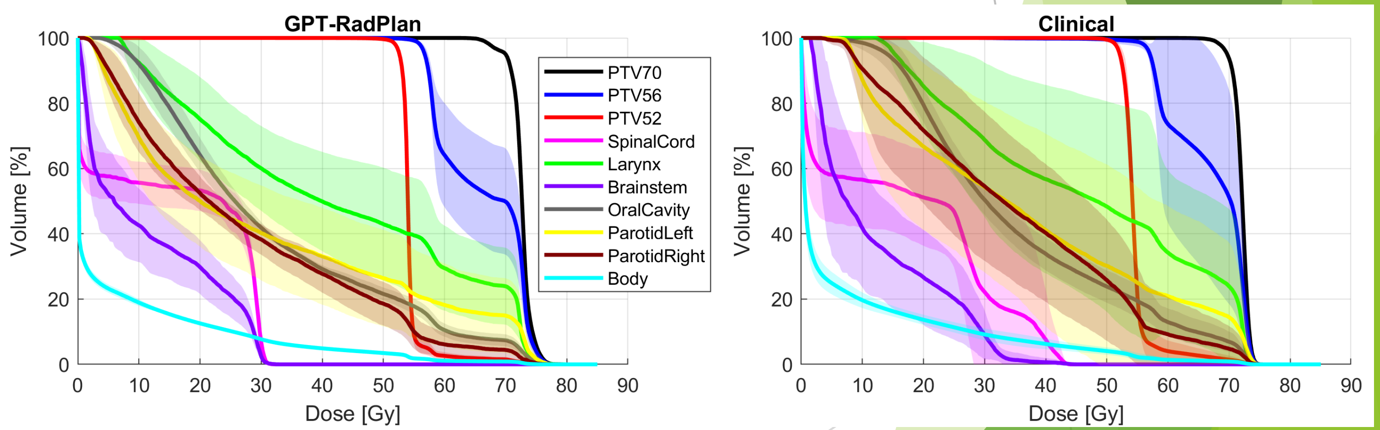
- Better PTV coverage (HI 1.96 vs 5.43, CI 0.92 vs 0.88)
- Better rectum sparing
- Slightly better bladder sparing
- Similar femoral head sparing, avoids higher doses

Head and neck:

- Similar PTV coverage, more conformal (CI 0.96 vs 0.84)
- Better larynx, oral cavity, parotid sparing
- Similar brainstem and spinal cord sparing



Prostate cases

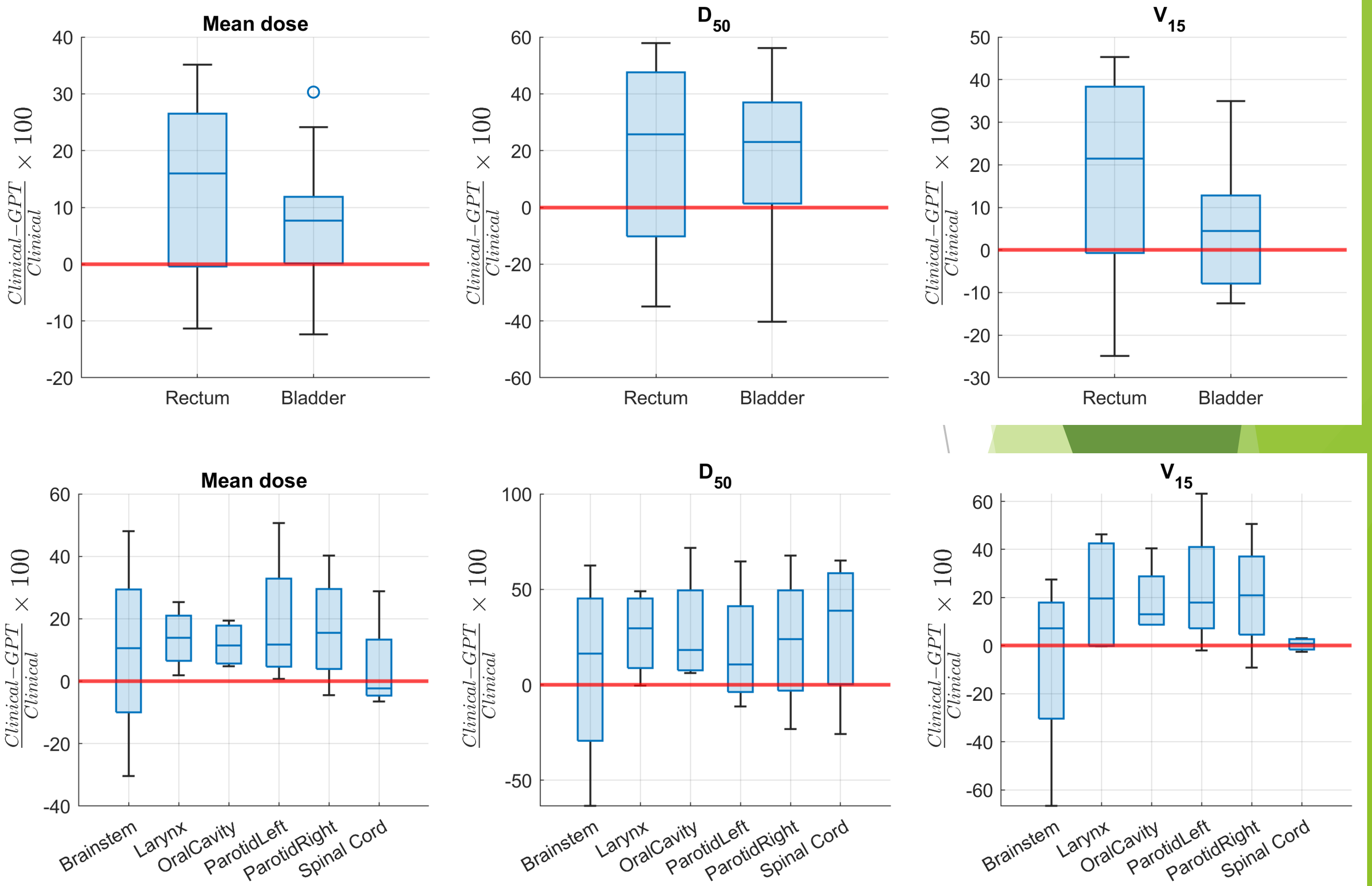


Head-and-neck cases

Liu S, Pastor-Serrano O, ….., arXiv preprint arXiv:2406.15609, 2024

Clinical plans vs GPT-RadPlan

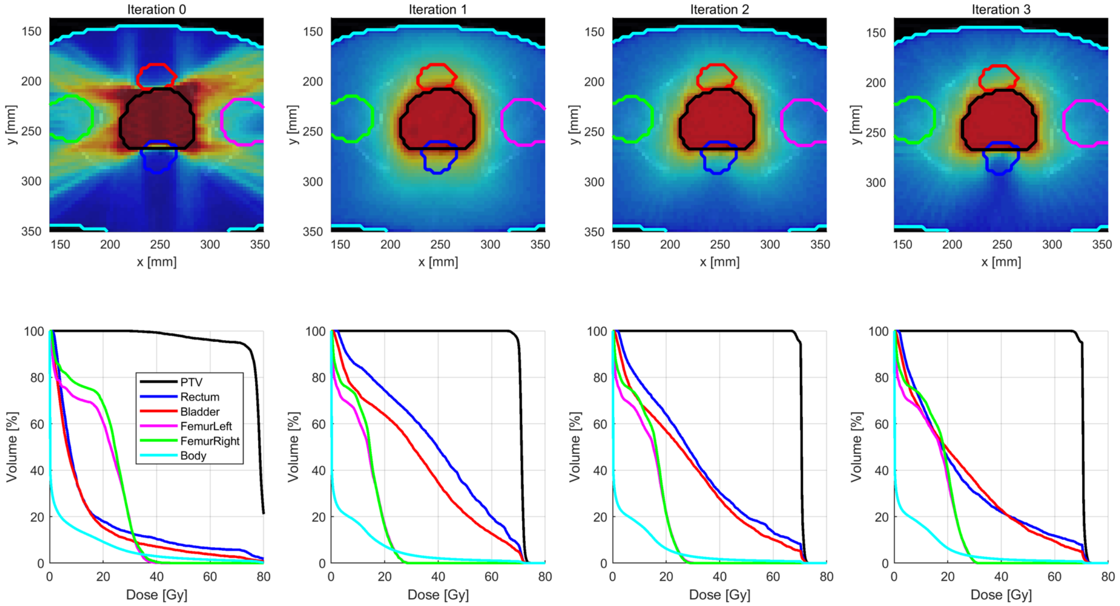
- Relative error between metrics from clinical plans and GPT-RadPlan
- Compares plans **case-by-case**
- Positive values → lower GPT-RadPlan metrics, better sparing

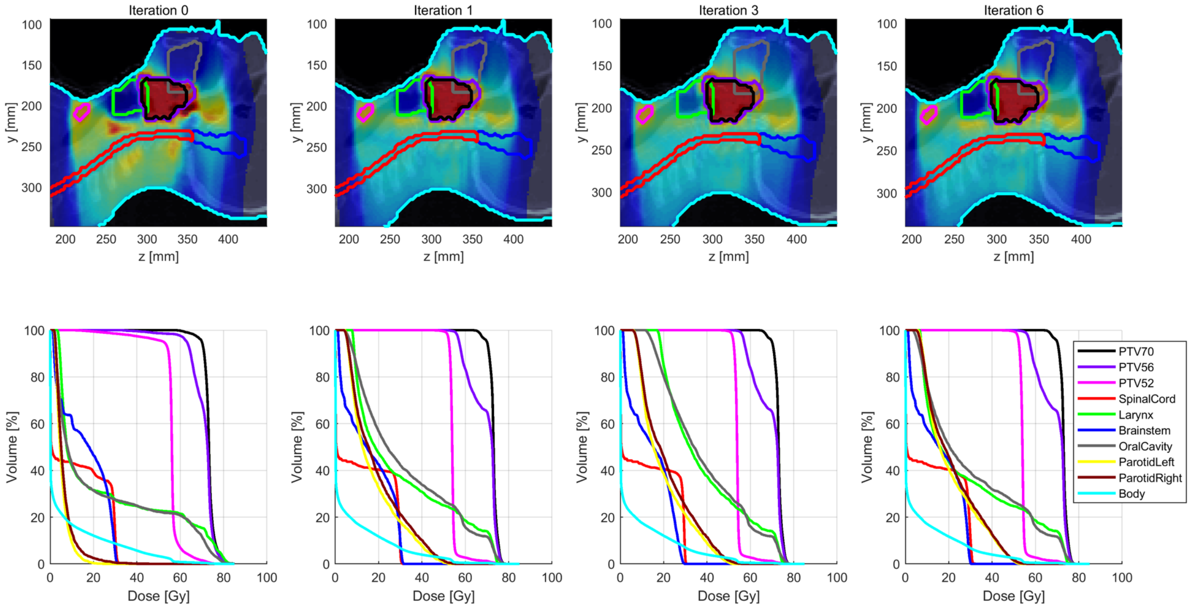


Liu S, Pastor-Serrano O, ….., arXiv preprint arXiv:2406.15609, 2024

Planning trajectories

1. Initialize plan based on average parameters
2. Ensure PTV homogeneity and conformity
3. Spare OARs while maintaining PTV coverage



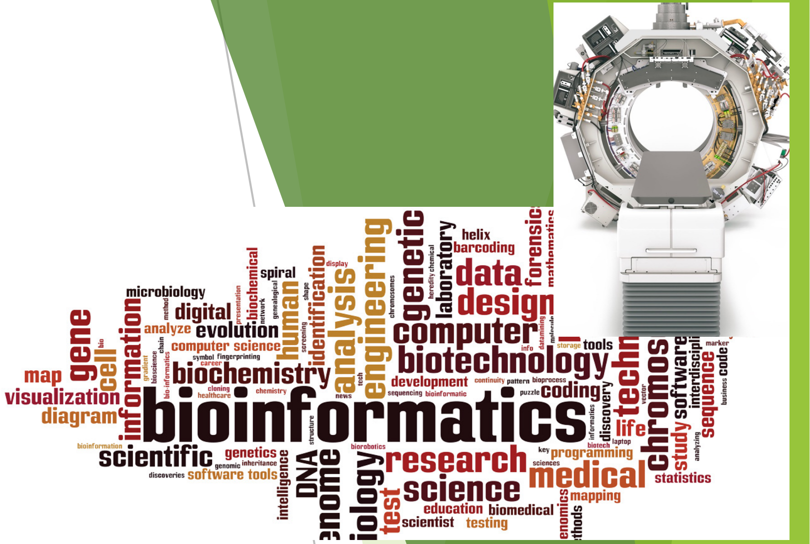


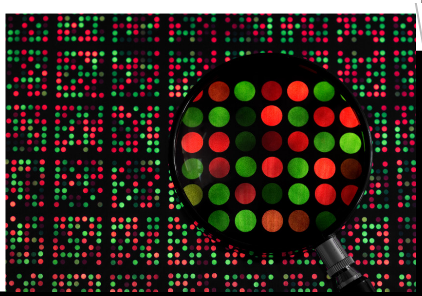
Stanford MEDICINE | School of Medicine

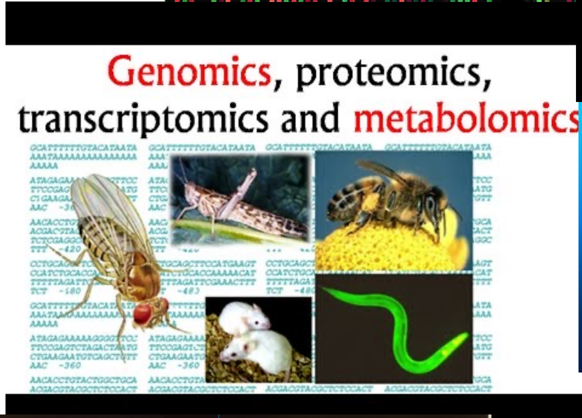
Liu S, Pastor-Serrano O, ….., arXiv preprint arXiv:2406.15609, 2024

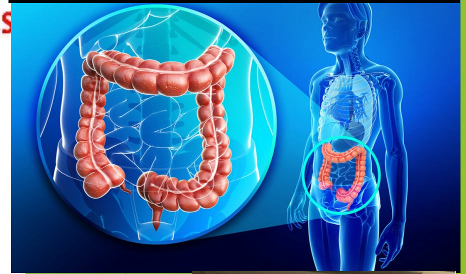
MULTIMODAL FOUNDATION MODELS

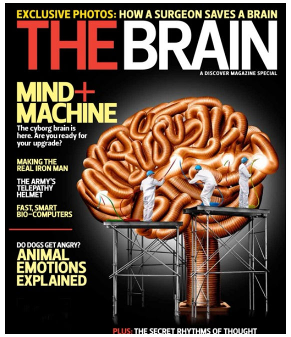
- Huge number of diseases
- Different types of models
- Various data formats
- Noise, small sample size, missing data, …

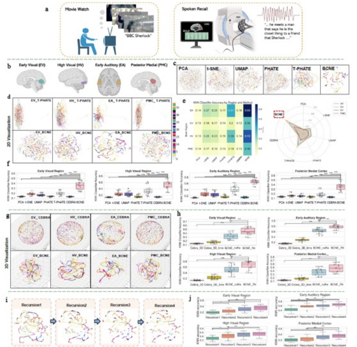


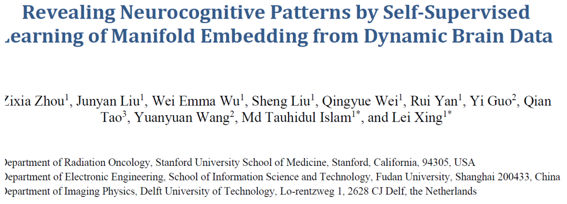








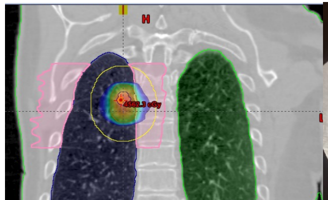


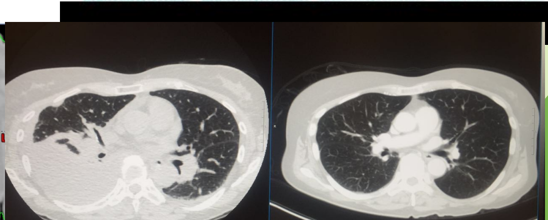


Revealing Neurocognitive Patterns by Self-Supervised Learning of Manifold Embedding from Dynamic Brain Data

Zixia Zhou[1], Junyan Liu[1], Wei Emma Wu[1], Sheng Liu[1], Qingyue Wei[1], Rui Yan[1], Yi Guo[2], Qian Tao[3], Yuanyuan Wang[2], Md Tauhidul Islam[1*], and Lei Xing[1*]

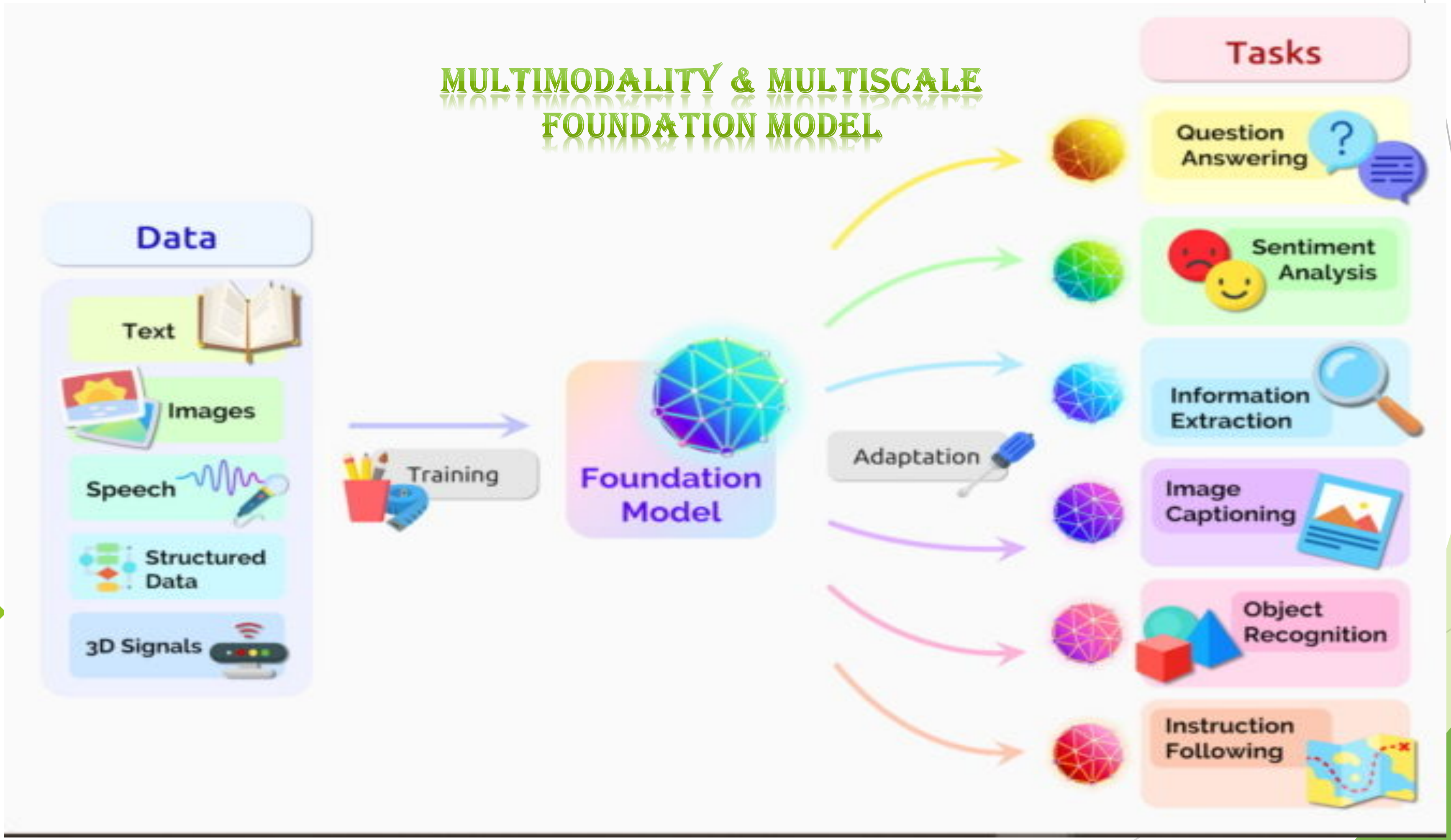
Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California, 94305, USA
Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai 200433, China
Department of Imaging Physics, Delft University of Technology, Lo-rentzweg 1, 2628 CJ Delf, the Netherlands



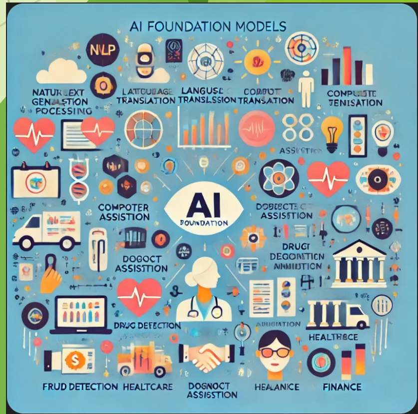


AI & FM FOR FUTURE HEALTHCARE

MULTIMODALITY & MULTISCALE FOUNDATION MODEL



https://blogs.nvidia.com/blog/2023/03/13/what-are-foundation-models/



Summary

AI Foundation Models for RT

Improving safety, quality, efficiency, and cost-effectiveness

1

Foundation models

- Characteristics
- Feature extraction & feature visualization
- Novel deep learning strategies & FMs

2

Challenges in AI FMs

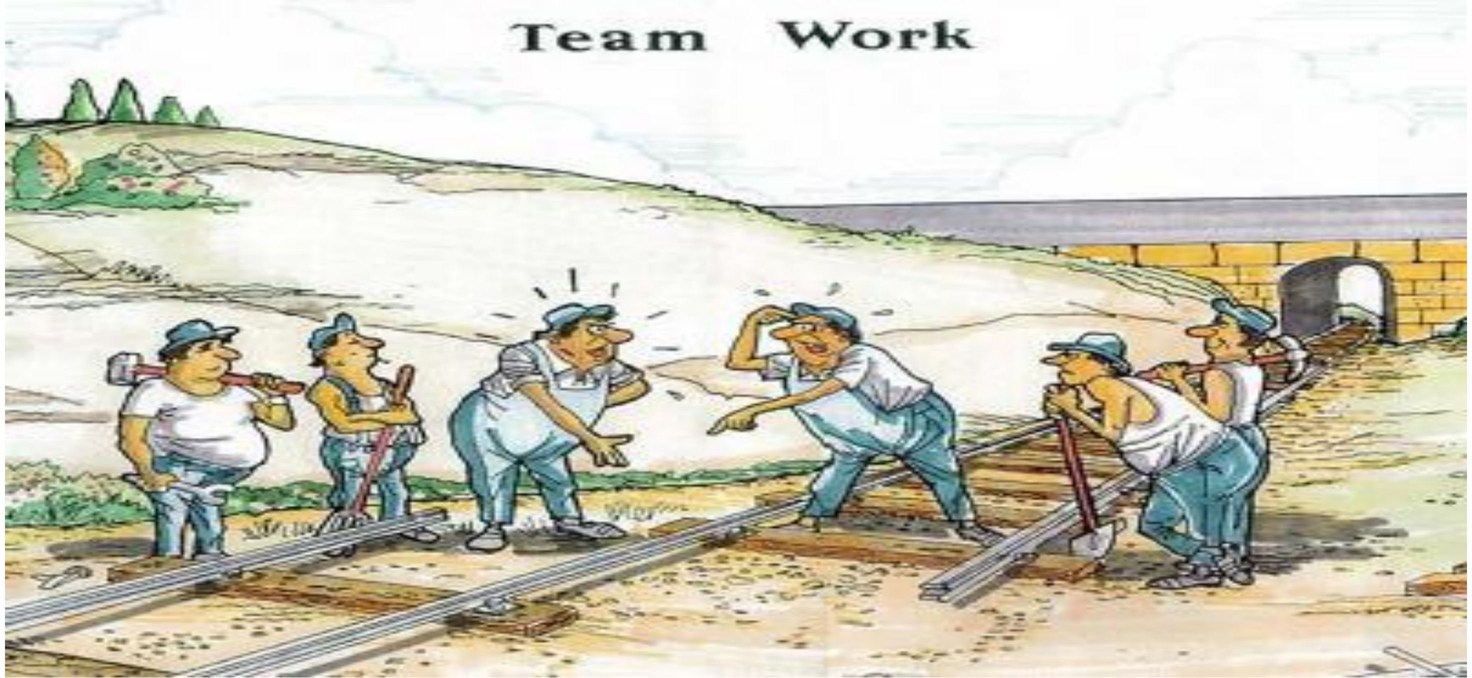
- Data requirements & computational resources
- Deployment – scalability, maintenance & update
- Ethical concerns – bias privacy, responsible use
- Multimodality integration

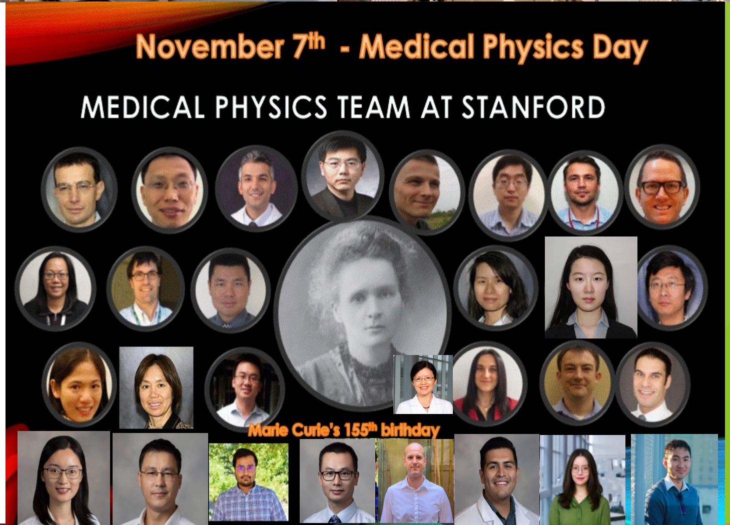
3

Applications of FMs in imaging & RT

- In-context learning
- Image analysis
- RT treatment plan evaluation & planning
- Multimodal FMs for healthcare

Acknowledgements




November 7th - Medical Physics Day

MEDICAL PHYSICS TEAM AT STANFORD

Marie Curie's 155th birthday

- **NIH: 1R01CA223667, 1R01CA227713, and 1R01CA256890**
- **Google Inc.**
- **Varian Medical systems, Mevion Medical Systems**