

# Biomedical Informatics 260 (BIOMEDIN260, CS235, RAD260, BMP260)

## Introduction

### Lecture 1

Mirabela Rusu, PhD  
Spring 2025

# Outline for Today

- About the course
- Expectations
- Logistics
- Resources
- Definitions
- Project Description and Summary

# About the course

# Objectives

- Understand medical imaging modalities
- Learn about how to get computers to “understand” images
- See applications for computerized image analysis
- Get hands on experience!

# Audience

- Graduate students
- Medical Students
- Medical Residents / Fellows
- Undergraduates

# Principle Instructor

**Mirabela Rusu, PhD**



- Assistant Professor
  - Radiology
  - Biomedical Data Science (courtesy)
  - Urology (courtesy)
- Industry experience
  - Medical Image Analysis Scientist
  - Lead Engineer
- Research:
  - Medical Image Analysis
  - Disease Detection
  - Disease Subtyping
  - Multimodal Data Integration
  - Foundation Models

## Other Instructors

Stanford | PROFILES



### Bruno Passebon Soares, MD

ASSOCIATE PROFESSOR OF RADIOLOGY (PEDIATRIC)

Radiology - Pediatric Radiology

Practices at Stanford Health Care, Stanford Medicine Children's Health

Web page: <http://web.stanford.edu/people/bruno.soares>

# Other Instructors



Research

Focus areas ▾

Blog

Publications

Careers

About ▾

[Home](#)  
↳ [People](#)

Work location

IBM Research - Almaden San Jose,  
CA USA

Contact

Topics

Data Management

Impact Science

Computer Science

Title  
IBM Fellow, Chief Scientist, Multimodal Bioinspired AI

# Tanveer Syeda-Mahmood



## Other Instructors

Stanford | PROFILES



**Roxana Daneshjou, MD, PhD**

ASSISTANT PROFESSOR OF BIOMEDICAL DATA SCIENCE AND OF  
DERMATOLOGY

Department of Biomedical Data Science

Practices at Stanford Health Care, Stanford Medicine Children's Health

## Other Instructors

**Stanford | PROFILES**



**Nicole Martinez-Martin**  
ASSISTANT PROFESSOR (RESEARCH) OF PEDIATRICS (BIOMEDICAL ETHICS) AND, BY COURTESY, OF PSYCHIATRY AND BEHAVIORAL SCIENCES (CHILD AND ADOLESCENT PSYCHIATRY AND CHILD DEVELOPMENT)  
Pediatrics - Center for Biomedical Ethics

## Other Instructors

**Stanford | PROFILES**



**Ehsan Adeli**

ASSISTANT PROFESSOR (RESEARCH) OF PSYCHIATRY AND BEHAVIORAL  
SCIENCES (PUBLIC MENTAL HEALTH AND POPULATIONS SCIENCES)

## Other Instructors



**Jeong Hoon Lee**  
POSTDOCTORAL SCHOLAR, RADIOLOGY



**Cynthia Li**  
PH.D. STUDENT IN COMPUTATIONAL AND MATHEMATICAL ENGINEERING,  
ADMITTED SPRING 2022

# Teaching Assistants

**Alejandro Lozano Garcia**



**Elana Simon**

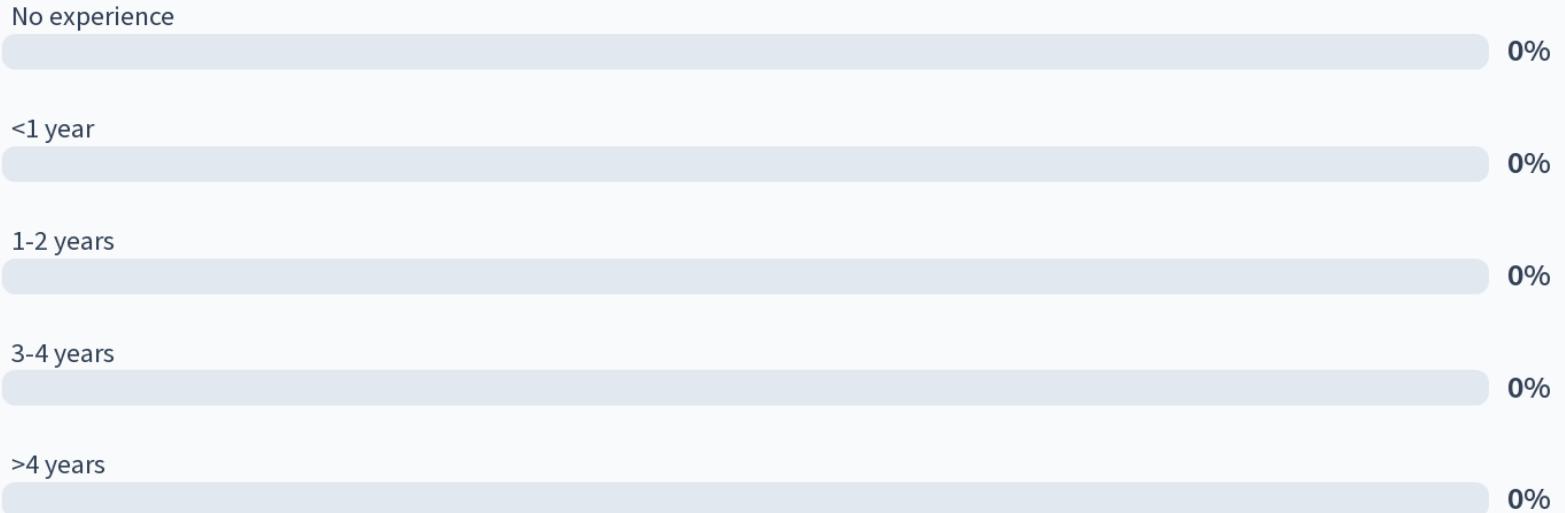


**Ben Viggiano**

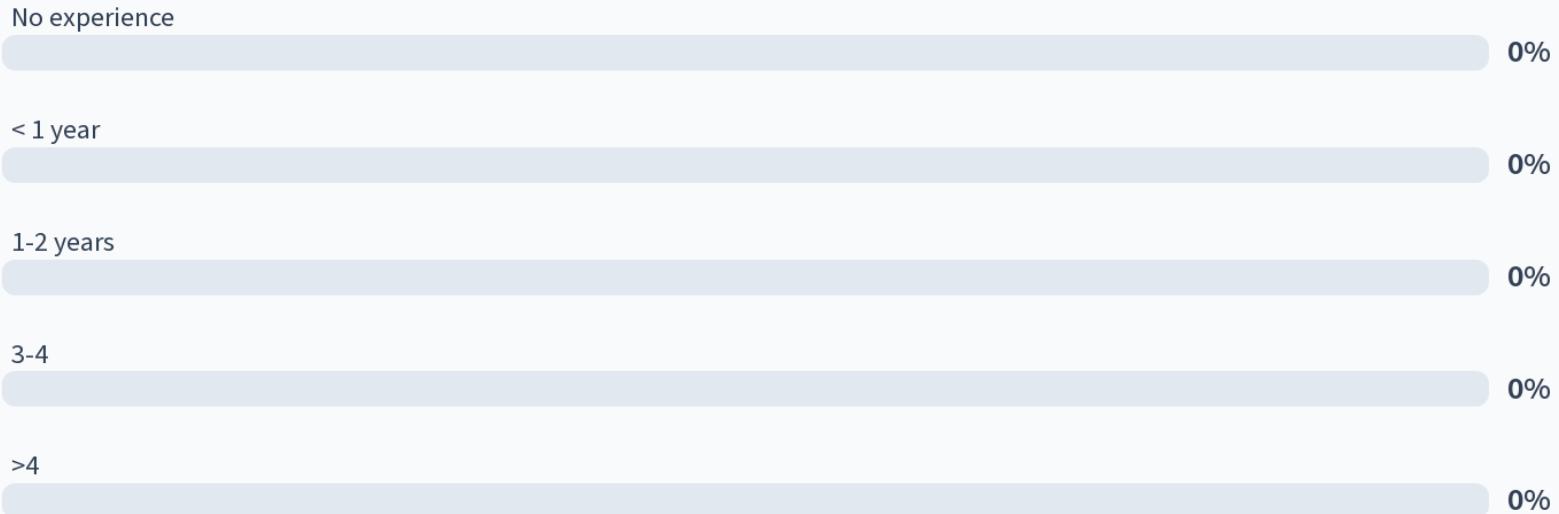


# Tell us about yourself

## Do you have programming experience? (CS106A not included)



## Do you have experience with biomedical data?





## What do you hope to learn in this class?

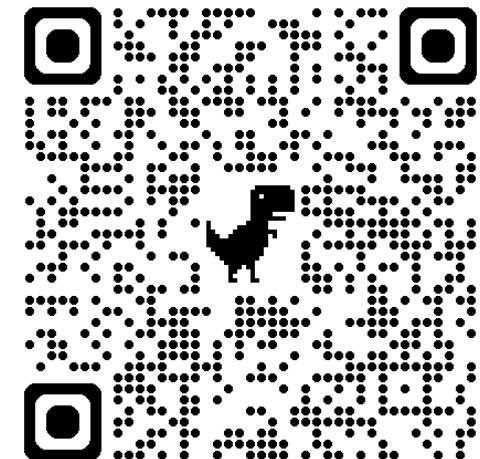
Nobody has responded yet.

Hang tight! Responses are coming in.

# Expectations

# Attendance

- This is an in-person class
  - Exceptions apply for SCPD/CGOE students and for medical reasons
  - Students have up to 3 unexcused absences, need to attend 15 lectures
- Attendance is recorded for every lecture
- Every lecture, there will be a code word provided during the lecture
- Record attendance using the form:
  - <https://forms.gle/kjPwLkS1BnZdLMnf6>
  - Submit attendance form during class
- SCPD/CGOE Students have until Sunday to record attendance



# BIOMEDIN 260 Attendance Form

Please fill out the attendance form by the end of class (unless you received prior approval to make up attendance by end of week) - the code word will be announced during lecture. CGOP students: please fill out the attendance form by Sunday night for both the Monday and Wednesday lectures of the prior week.

mrusu@stanford.edu [Switch account](#)



\* Indicates required question

Email \*

Your email

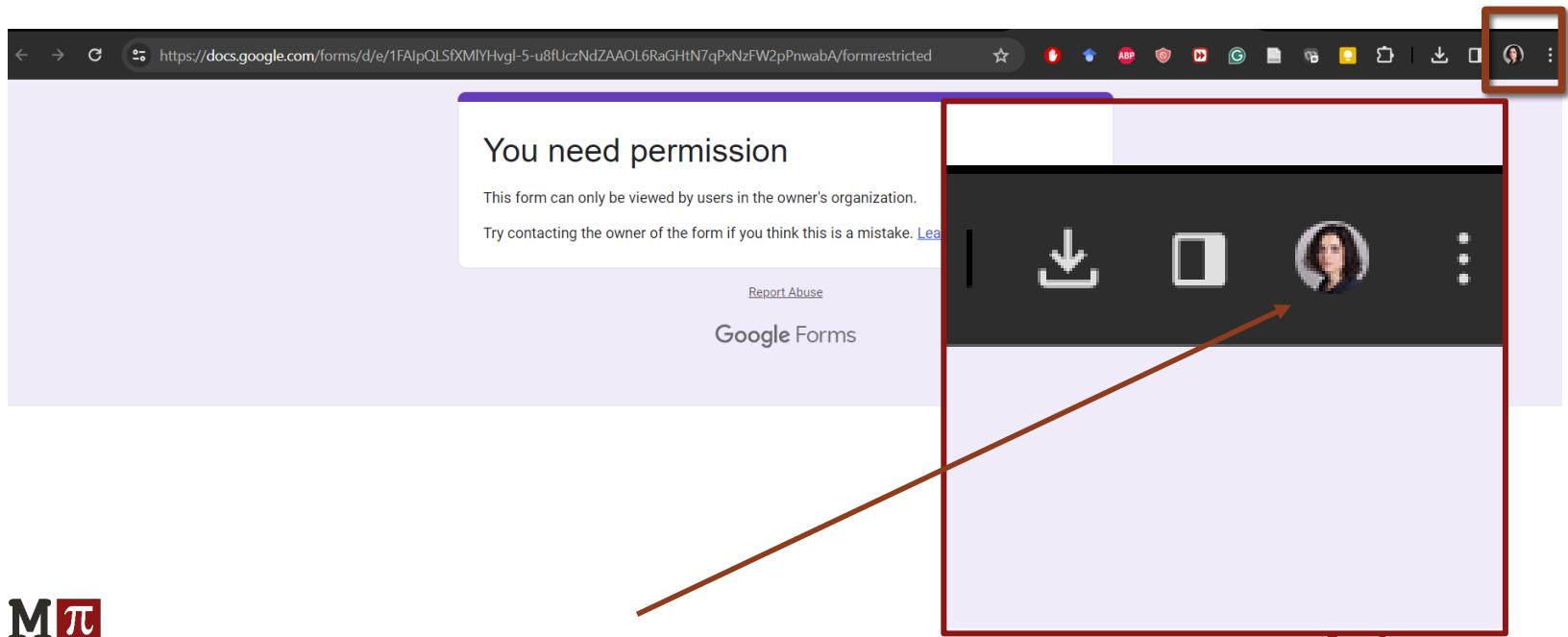
Name (Last, First) \*

Your answer

SUNET ID (sunet@stanford.edu) \*

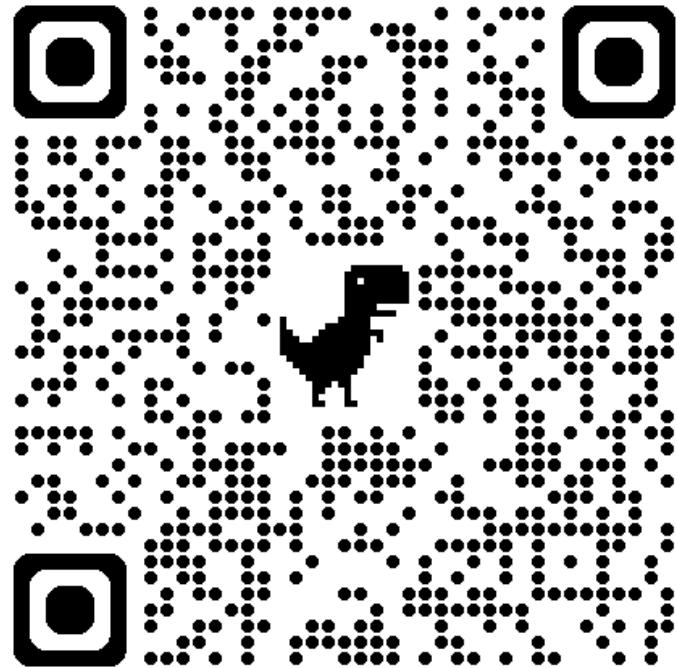
Your answer

# If you get this error message – use your Stanford account



Today's Attendance Code:

Medical Image Analysis



<https://forms.gle/kjPwLkS1BnZdLMnf6>

# Pre-requisites

- What you absolutely need to know
  - Programming ability (CS 106A or similar)
  - Basic statistics helpful
- What would be really nice to know
  - Familiarity with Google Colab  
(<https://colab.research.google.com/>)
  - Jupyter Notebooks

# Readings

- **Articles**
  - Two recommended articles per class
  - Provided with each lecture; posted on Canvas
- **Books**
  - Not required
  - Supplement required readings
- (see Canvas, <https://canvas.stanford.edu/courses/206247> )

# Coursework

- **Assignments (N=3)**
  - Involves programming (Python)
  - Released on Mondays, DUE on Sunday  
(see the syllabus on Canvas for dates)
  - Up to groups of 2
- **Project (5 requirements)**
  - 3 writeups or reports (Proposal, Milestone, Final)
  - 2 presentations
    - Project Milestone Presentation (2 classes over one week)
    - Final project presentations (During final exam slot)
  - Submissions on Assignments section on Canvas

# Final Project

- A substantive programming project that covers element(s) of both ***quantitative*** and ***qualitative*** aspects of imaging
- Group projects: 2-3 students (N students, N quantitative approaches, N={2,3})
  - Project report (due Sun 4/20)
  - Milestone presentation (5/05 or 5/07)
  - Milestone report (due Sun 5/11)
  - Final project report (Due Fri 5/30 by 11:59 PM)
  - Project Final Presentation (***During final exam slot,***

M<sub>E</sub>D<sub>T</sub> **Fri, Jun 6, 3:30-6:30pm)**

Stanford University

# Logistics

# Schedule

- **Lectures:** Mon / Wed 10:30-11:50pm
  - Zoom on during lectures
  - Classes are also recorded and posted on canvas
  - Beckman Center B060
  - No class on 5/26 and 6/4
- **Section/Office hours**
  - Tuesday 8-9am (Elana, zoom)
  - Wednesday 5-6pm (Ben, zoom)
  - Friday 4-5pm (Ale, Room: Packard 340)
  - First TA by Ale: Python, setting up jupyter notebooks and colab

# Course Outline

- Overview of imaging modalities
- Visualization
- Image processing (Detection, Segmentation, Quantification, Registration, Radiomics)
- Machine Learning and Deep Learning for Images
- Diffusion Models
- Foundation models
- Clinical applications

|    |           | Mon 10:30 AM - 11:50 PM  |                           | Wed 10:30 AM - 11:50 PM   |
|----|-----------|--|---------------------------|---|
| 1  | 3/31/2025 | Course Framework, Introduction, Project Introduction (Rubrics)   | 4/1/2025                  | Imaging Modalities and Image Interpretation (Bruno Suarez)  |
|    |           | Image Visualization and Segmentation (3D rendering, image fusion, perception, augmented reality, 3D printing, thresholding and binary images, clustering) Image Registration (Interpolation, transformations, optimization for registration) | 4/9/2025                  | Radiomics (Feature Extraction, Feature Reduction, Classification) CNN (architectures, activation functions, initialization, loss functions) |
| 2  | 4/7/2025  | Image Processing, Image Analysis, Visualization/ Registration (sensitivity vs. specificity, ROC curves, segmentation eval)   | 4/16/2025                 | Multimodal Deep Learning & Radioprecise / Approaches for multi-modality integration, data complementarity, data correlation,                |
| 3  | 4/13/2025 | Guest Lecture: Mirabela Rusu (Date TBD)  | 4/17/2025                 | Guest Lecture: Bruno Soares (Date TBD)  |
| 4  | 4/21/2025 | Foundation Models 1 (Tanveer Syeda-Mahmood)  | 4/23/2025                 | Foundation Models   |
| 5  | 4/28/2025 | Foundation Models 2 (Tanveer Syeda-Mahmood)  | 4/30/2025                 | Generative Models (Basics, Medical Examples, Diffusion Models)  |
| 6  | 5/5/2025  | MIDTERM - Project Presentation   | 5/7/2025                  | MIDTERM - Project Presentation  |
| 7  | 5/12/2025 | Image Content (Jed Hwang, Postdoc in Rusu Lab)   | 5/14/2025                 | Natural Language Processing and Vision Language Models  |
| 8  | 5/19/2025 | Guest Lecture: Ersan Adali (Date TBD)  | 5/21/2025                 | Guest Lecture: Roxana Daneshiou (Date TBD)  |
| 9  | 5/26/2025 | Memorial Day - No Class  | 5/28/2025                 | Guest Lecture: Nicole Martinez - Bioethics  |
| 10 | 6/2/2025  | Guest Lecture: Mirabela Rusu: Quantitative imaging and deep learning<br>Bruno Soares: Clinician<br>Tanveer Syeda-Mahmood<br>Alejandro Lozano<br>Guest lectures: Applications   | 6/4/2025                  | No Class  |
|    |           |  | Fri 6/06/2025 3:30-6:30pm | Final Presentations   |

# Syllabus

## Advanced Topics

### Image Processing, Image Analysis, Visualization/ Registration

### Image Content

### Applications

# Grading

- **Grade Breakdown**
  - 3 Assignments: 30% total (10% each)
  - Final project 60%
    - Proposal Report 20%
    - Milestone Report 10%
    - Milestone Presentation 10%
    - Project report 10%
    - Final Project Presentation 10%
- **Class Attendance (5%)**
- **Class participation (5%)**
  - Asking/answering questions on Ed Discussion forum;
  - Participating in TA Sections

# Late Submission Policy

- Assignments are due at or before 11:59 PM on the due date
- **You have 4 free late days total**
  - Not valid for final project assignments
- After that, **10% off your grading score per day late** (late time within a day used is counted as a full late day; i.e., it rounds up)

# Resources

# Course Support Resources

- Course Support
  - Canvas for files, assignments  
<https://canvas.stanford.edu/courses/206247>
  - Ed Discussion  
<https://edstem.org/us/courses/76292/discussion>
  - Assignments  
<https://canvas.stanford.edu/courses/206247/assignments>

<https://canvas.stanford.edu/courses/206247>



Sp25-RAD-260-01... > Syllabus

Spring 2025

Home

Lane Library  
Resources

Syllabus

Assignments

Quizzes

Discussions

Files

People

Collaborations

Zoom

Panopto Course  
Videos

Ed Discussion

## Computational Methods for Biomedical Image Analysis and Interpretation

### Overview:

Biomedical imaging is a rapidly advancing field, revolutionizing how we visualize and understand the human body. Modern imaging technologies now produce incredibly detailed views of tissue morphology, reveal dynamic physiological processes, and even capture molecular events such as gene expression. Imaging is central to medical practice—nearly every patient undergoes some form of imaging during their care, and a single study can generate thousands of images. Just as the surge in genomic data propelled the rise of bioinformatics, the exponential growth in digital imaging is driving the need for specialized techniques in imaging informatics.

**Imaging Informatics** is the interdisciplinary science dedicated to the acquisition, storage, analysis, retrieval, and interpretation of imaging data in biomedicine. This field not only optimizes the management of vast imaging datasets but also integrates them with molecular and clinical information to enhance diagnosis and treatment. Imaging informatics spans multiple disciplines, including engineering, computer science, statistics, radiology, pathology, genomics and medicine.

This course provides a comprehensive introduction to imaging informatics, covering foundational techniques for image processing, analysis, and interpretation via computational approaches such as Artificial Intelligence. By mastering these methods, students will gain the skills needed to harness imaging data for scientific discovery and real-world applications in medicine.

Topics covered in this course:

- Types of imaging methods and how they are used in biomedicine
- Image processing, enhancement, and visualization
- Computer-assisted AI detection, diagnosis, and decision support
- Access and utility of publicly available image data sources
- Linking imaging data to clinical data and phenotypes
- Computer reasoning with images
- Foundation models for medical image analysis
- New questions in biomedicine using imaging informatics. Case studies.

View Course Stream

View Course Calendar

View Course Notifications

### To Do

- Problem Set 1 X
- Computational Methods  
for Biomedical Image  
Analysis and  
Interpretation  
100 points |  
Apr 13 at 11:30pm

ersity

<https://canvas.stanford.edu/courses/206247>

## ***Course work and syllabus (bottom of home page)***

### **Coursework:**

|               | Description                                  | Out Date  | Due Date    | Percent |
|---------------|--|-----------|-------------|---------|
| Problem Set 1 | Lung Field Segmentation                      | 3/31/2025 | 4/13/2025   | 10%     |
| Project       | Proposal                                     |           | 4/20/2025   | 20%     |
| Problem Set 2 | Brain Tumor Segmentation                     | 4/14/2025 | 4/27/2025   | 10%     |
| Project       | Milestone Presentation                       |           | 5/5 and 5/7 | 10%     |
| Project       | Milestone Report                             |           | 5/11/2025   | 10%     |
| Problem Set 3 | Multimodal LLMs for Radiology                | 5/12/2025 | 5/25/2025   | 10%     |
| Project       | Final Report                                 |           | 5/30/2025   | 10%     |
| Project       | Final Presentation                           |           | 6/6/2025    | 10%     |
| Attendance    | In class                                     |           |             | 5%      |
| Participation | In class, during OH, and/or on Ed Discussion | -         | -           | 5%      |

<https://canvas.stanford.edu/courses/206247>

≡ Sp25-RAD-260-01... > Assignments

Spring 2025

Search...

SHOW BY DATE

SHOW BY TYPE

Home

Lane Library  
Resources

Syllabus

Assignments

Quizzes

Discussions

Files

People

Collaborations

Zoom

Panopto Course  
Videos

Ed Discussion

▼ Upcoming Assignments

Problem Set 1

Not available until Mar 31 at 9am | Due Apr 13 at 11:30pm | -/100 pts

# Discussion forum



People



Pages



Syllabus



Rubrics



Quizzes



Modules



Collaborations



Chat



Roster Photos



Panopto Course  
Videos



New Analytics



Zoom

Ed Discussion



Gradescope

## Computational Methods for Biomedical Image Analysis and Interpretation



Edit

[Who can view my syllabus?](#)

## Welcome to BIOMEDIN 260/RAD 260/CS 235!

### Overview:

Biomedical imaging is an exploding field. The technologies for visualizing the body (the imaging modalities) are becoming very powerful, providing exquisite images of tissue morphology, revealing tissue function, and even beginning to see molecular events such as gene expression. Imaging is at the core of medical practice; nearly all patients have imaging of some sort during care, and many studies produce thousands of images. Just as the genetic data explosion has fueled the field of bioinformatics, the growth in digital imaging is necessitating techniques in imaging informatics.

Imaging Informatics is the science of analytic, storage, retrieval, and interpretive methods to optimally use the increasingly voluminous imaging data in biomedicine, and integrate and

# Discussion forum

ed BMI 260 – Ed Discussion

[New Thread](#)

COURSES +  
BMI 260

CATEGORIES  
General  
Lectures  
Sections  
Problem Sets  
Assignments  
Social

Search [Filter](#)

Last Week

Google Cloud Coupon ✓  
General Robert Igboekwe :) 11h 1

Welcome! ❤ 4  
General Elana Pearl Simon STAFF 1d

This screenshot shows a discussion forum interface. At the top, a purple header bar displays the text "ed BMI 260 – Ed Discussion". Below the header, there's a blue button labeled "New Thread". To the left, a sidebar lists "COURSES" with "BMI 260" selected, and "CATEGORIES" including General, Lectures, Sections, Problem Sets, Assignments, and Social. The main area features a search bar with a "Filter" dropdown. A section titled "Last Week" shows two threads: "Google Cloud Coupon" by Robert Igboekwe, posted 11 hours ago, and "Welcome!" by Elana Pearl Simon, a staff member, posted 1 day ago. The "Welcome!" thread has 4 likes.

# Use of Large Language Models

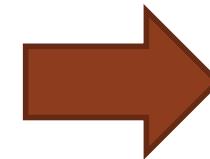
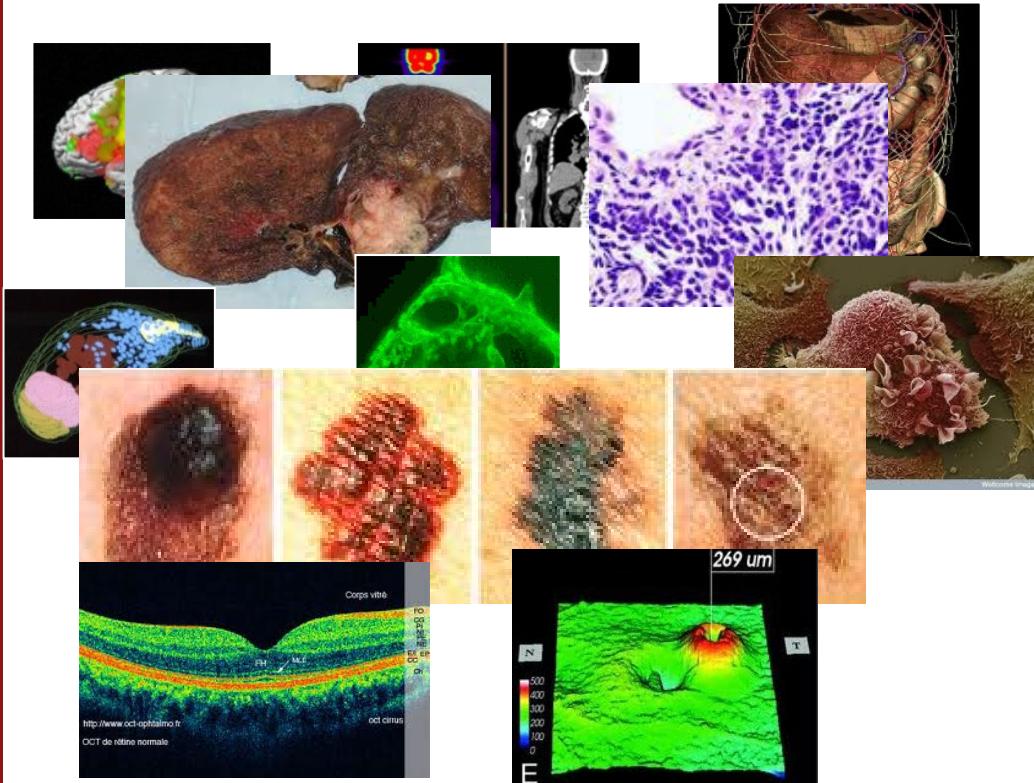
Large Language Models are very good at getting you started with a topic in computational Methods for biomedical image analysis

- Provide high level medical information
- Provide high level information on Computational methods
- Help with coding, debugging, running

Rules on using: Large Language Models

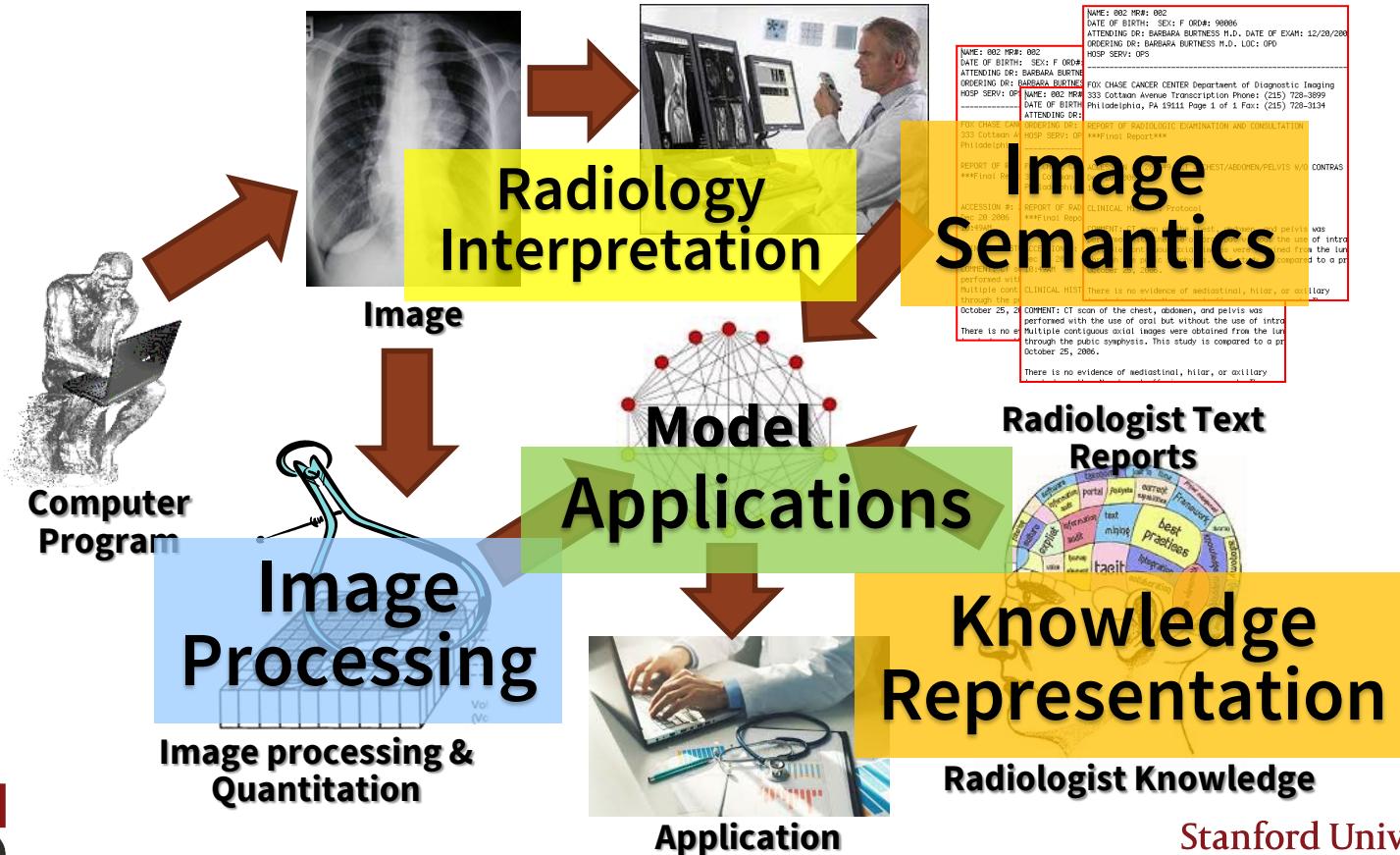
1. You are responsible of the content you are submitting: homework doesn't compile, reports are incomplete, etc - you don't get the points
2. You need to acknowledge its use: you use it for coding in your homework, you have to mention it at the end of your assignment. You used it during the report writing, you need to mention how you use it, e.g. grammar editing, vs content creation.
3. If you use them and not mention it, you may not receive full points
4. Check all the information it provides (it does hallucinate, e.g., referencing non existing papers)
5. Can NOT be used to create figures, results.

# Our goal is to go from images to understanding...



Diagnosis  
Biomarker  
Disease Progression  
Biological Models

# From medical images to clinical understanding



# Definitions

# Computational Methods for Biomedical Image Analysis

# General terms

Data: basic values, quantities, characters, or symbols on which operations are performed by a computer

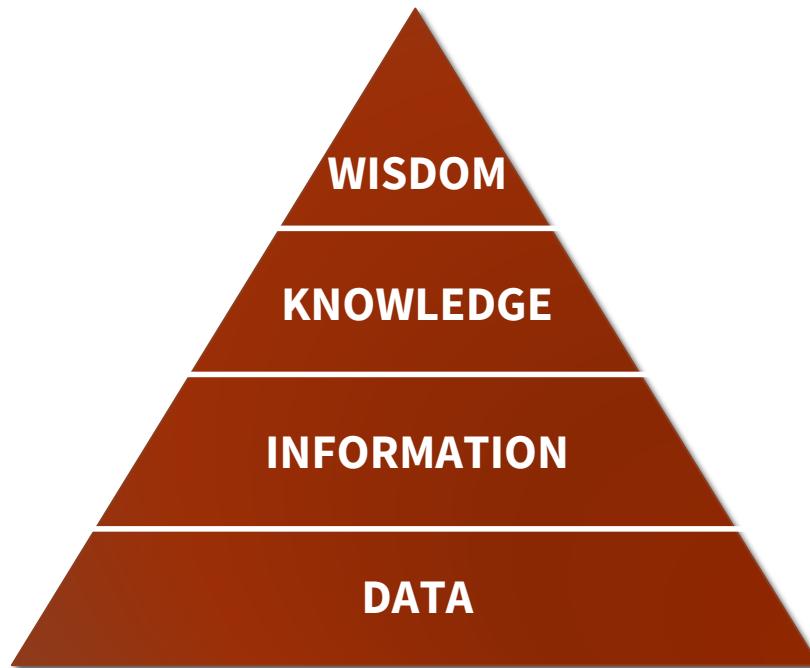
- Referred to as Raw Data
- Specifics: the numbers, e.g., 5,

Information: Data with a meaning or purpose: 5mm  
“Information is inferred from data”

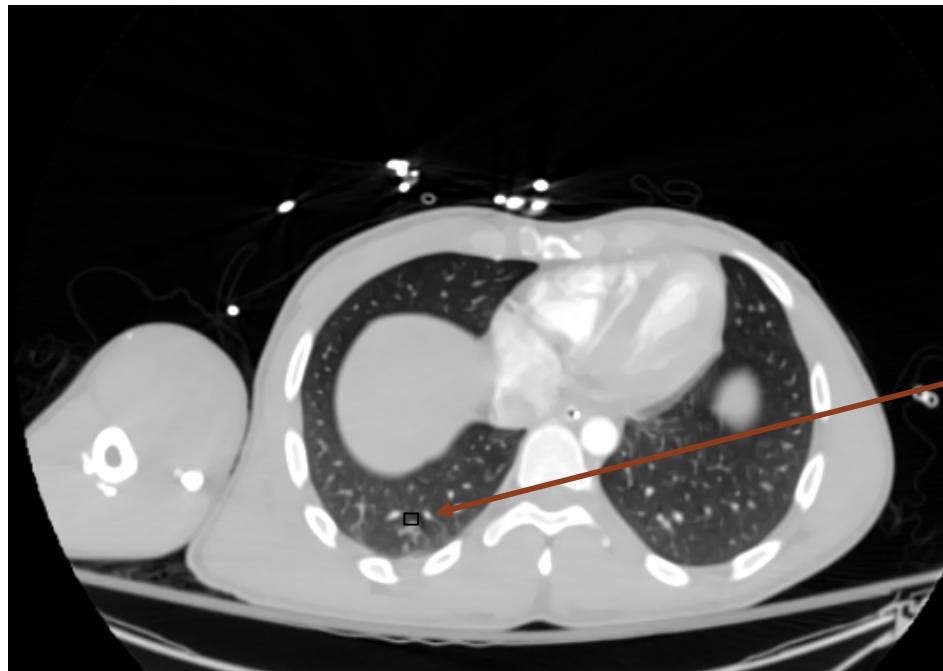
Knowledge: information that has been processed organized or structured,  
- Specifics: Knowledge is put into action

Wisdom: integrated knowledge, information make super-useful: ability to increase effectiveness

# Pyramid

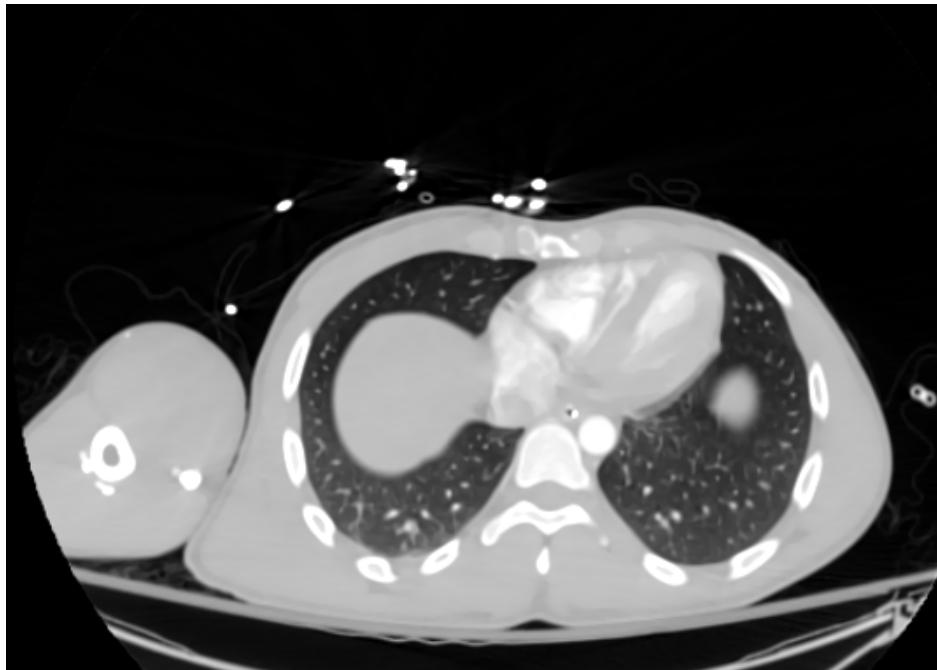


# Chest CT – Data are the Voxel values



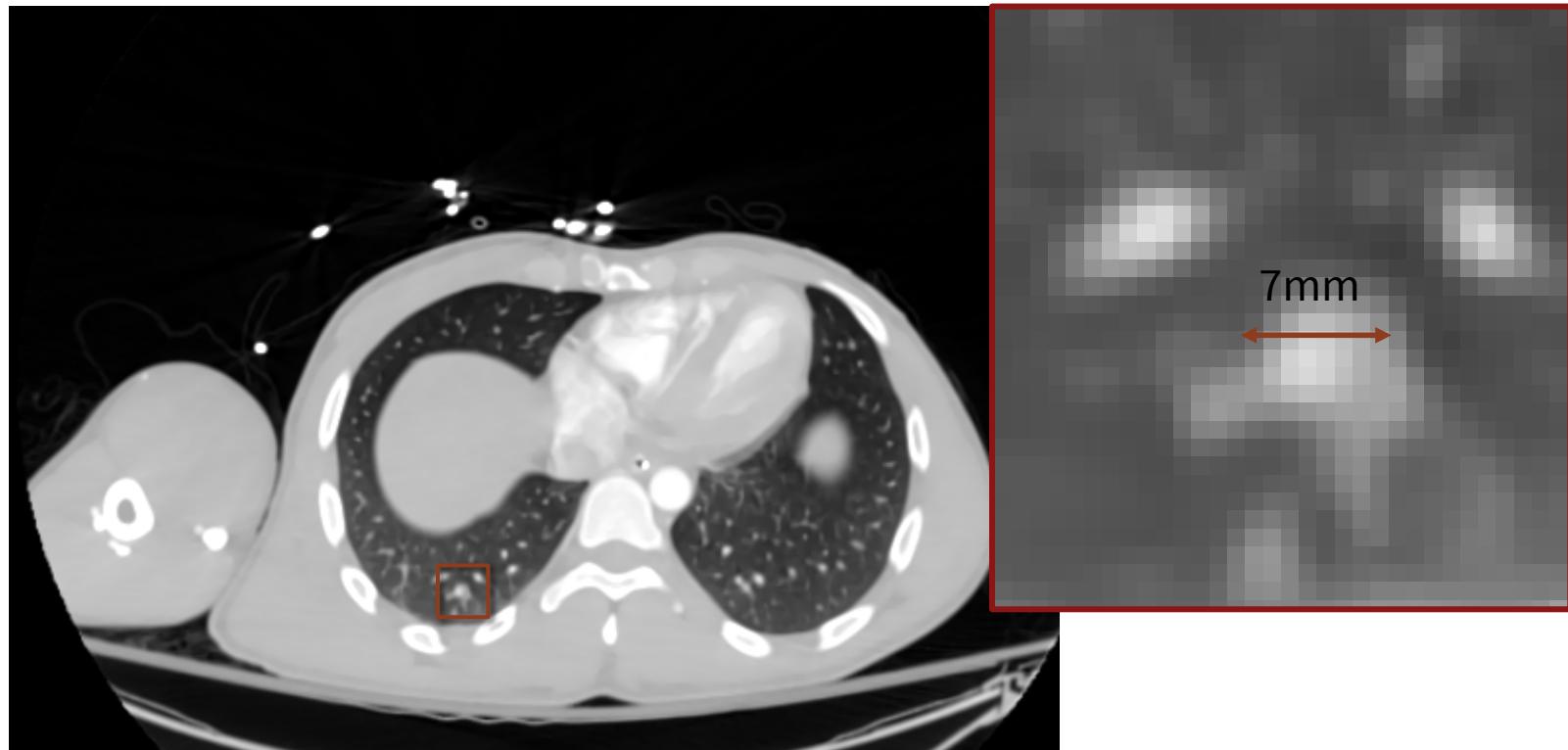
|     |     |     |
|-----|-----|-----|
| 250 | 129 | 200 |
| 128 | 100 | 50  |
| 200 | 10  | 75  |

# Chest CT – Information: Adding Hounsfield units to voxel values (data)



|        |        |        |
|--------|--------|--------|
| 250 HU | 129 HU | 200 HU |
| 128 HU | 100 HU | 50 HU  |
| 200    | 10 HU  | 75 HU  |

# Chest CT – Knowledge: this is a lung nodule of 7 mm



# Wisdom

Ground glass nodules have good outcomes when resected

Lung cancer has a 5-year survival rate of 26.6%

# Computational Methods for Biomedical Image Analysis

# Image Analysis

Image Analysis: computer-based process to extract quantitative information from images

Requires: Digital Images

Example:

- Reading code bars
- Reading handwritten zip codes



# Image Analysis – Other examples



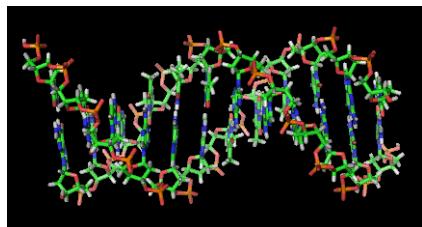
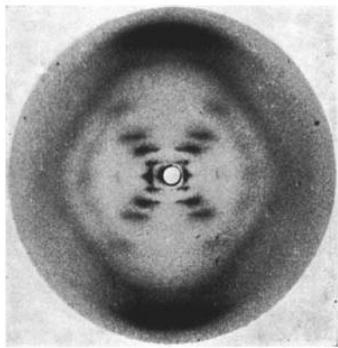
Vs.



# Computational Methods for Biomedical Image Analysis

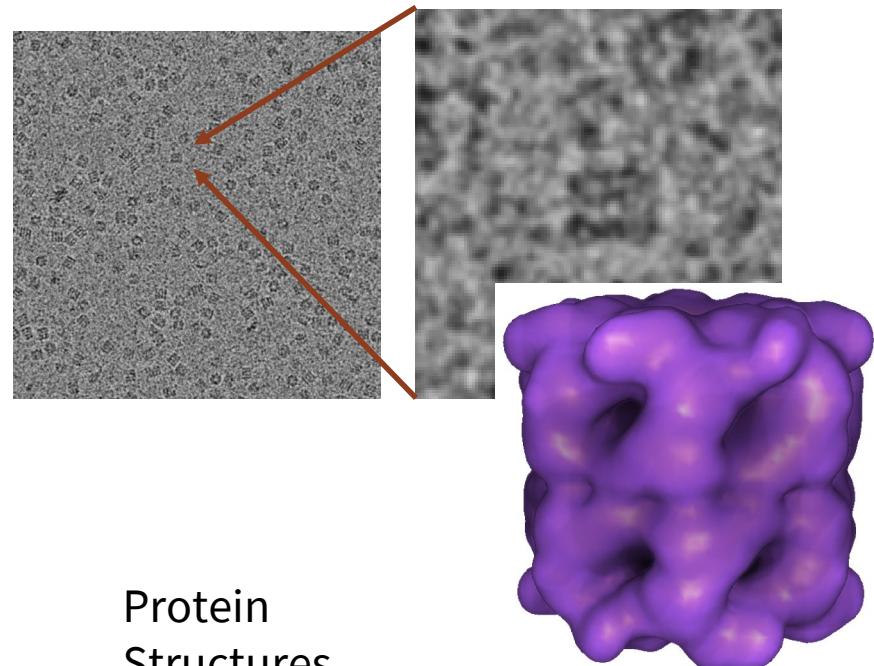
# Biomedical Images –Molecules

Wide range of scales

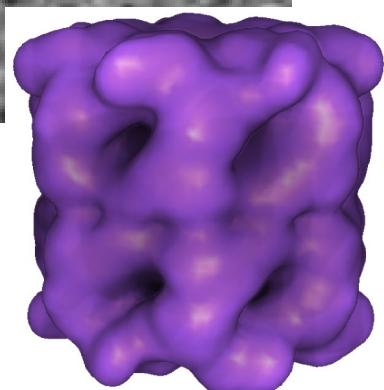


DNA structure

M $\pi$   
ED

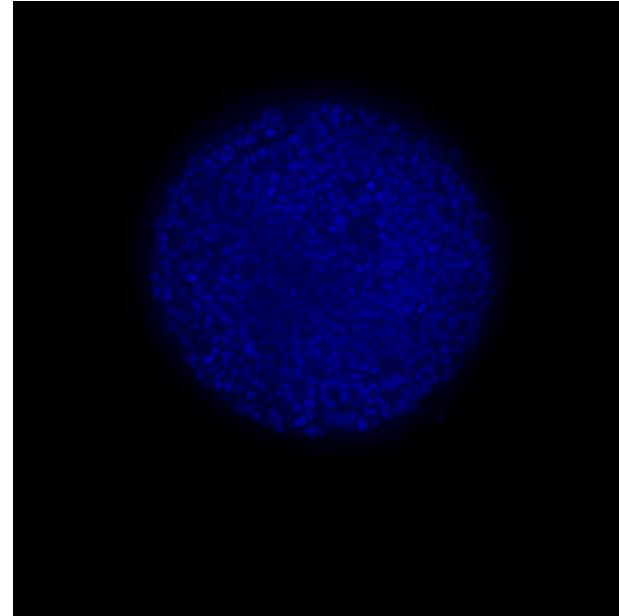
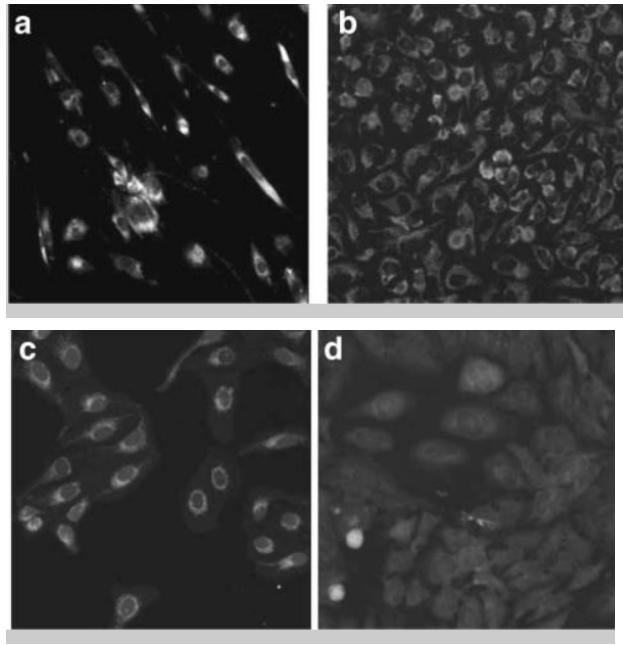


Protein  
Structures



Stanford University

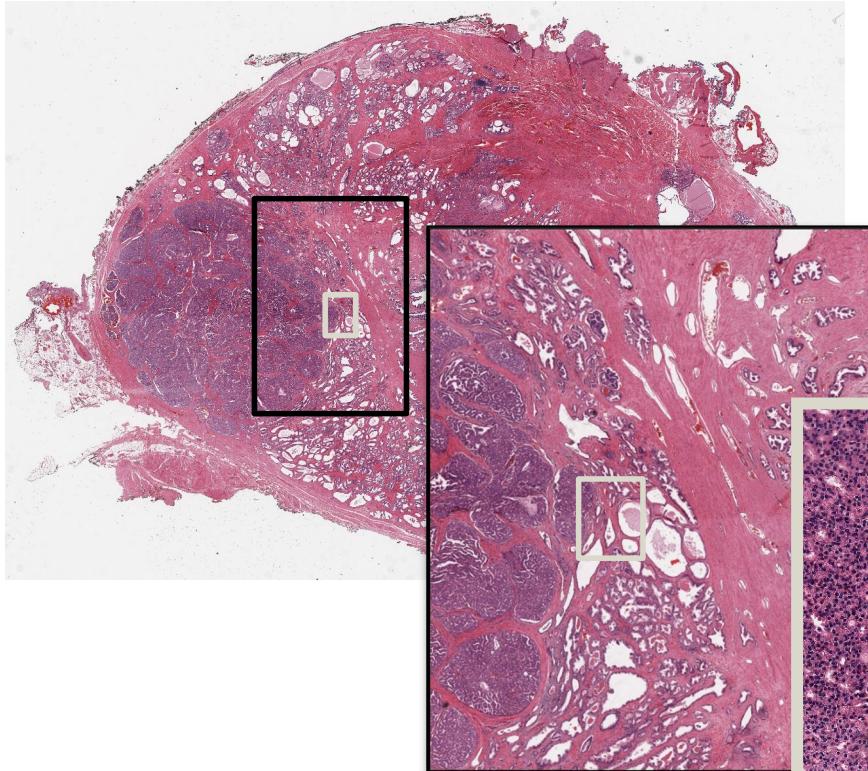
# Biomedical Images – drug assay – cell cultures



Al-Kofahi et al, 2018

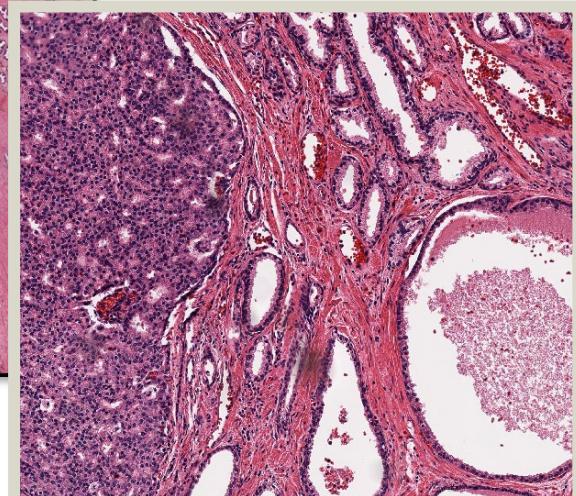
<https://vids.moleculardevices.com/watch/Uzi1MZHJUPGNRTZWht6q2?>

# Medical Images – histopathology of excise tissue

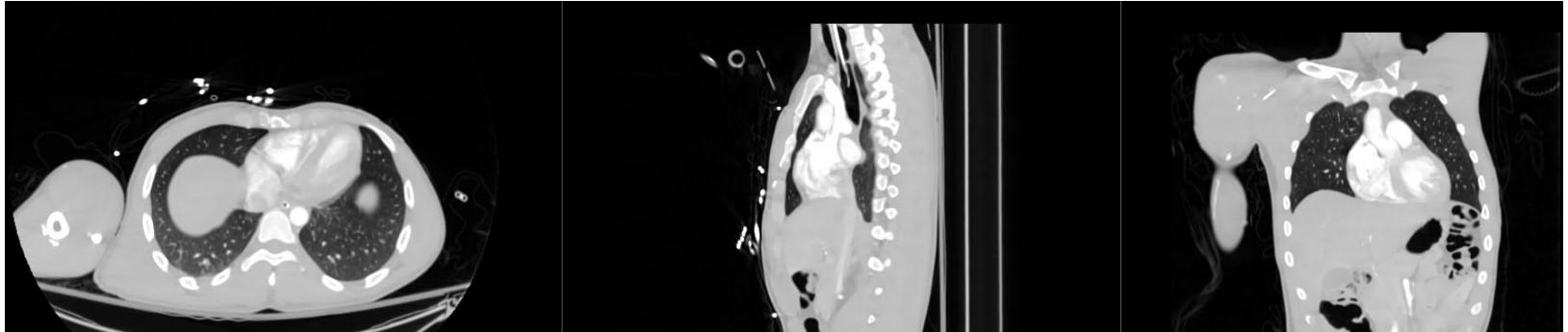


Hematoxylin and eosin stain

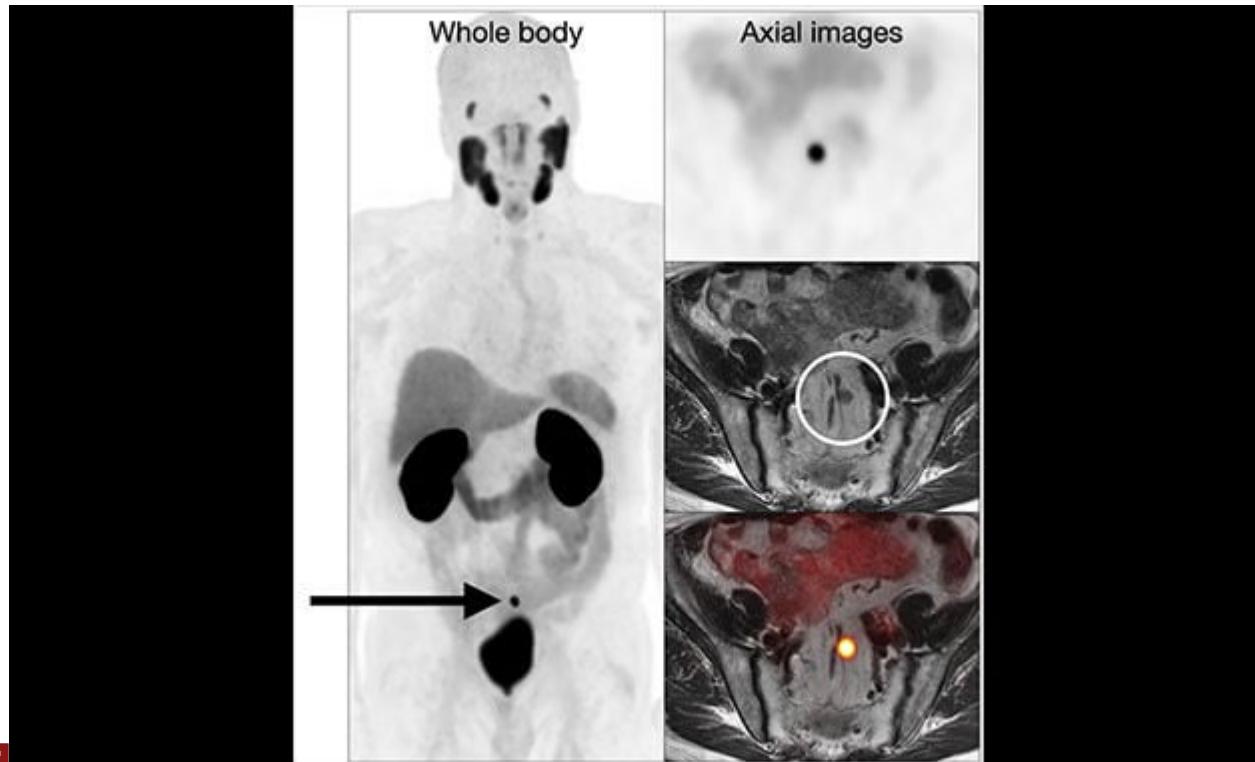
- hematoxylin stains cell nuclei blue
- eosin stains the extracellular matrix and cytoplasm pink



# Medical Images – Chest CT – Localize lung nodules



# Medical Images – Whole Body PET Images - Metastasis

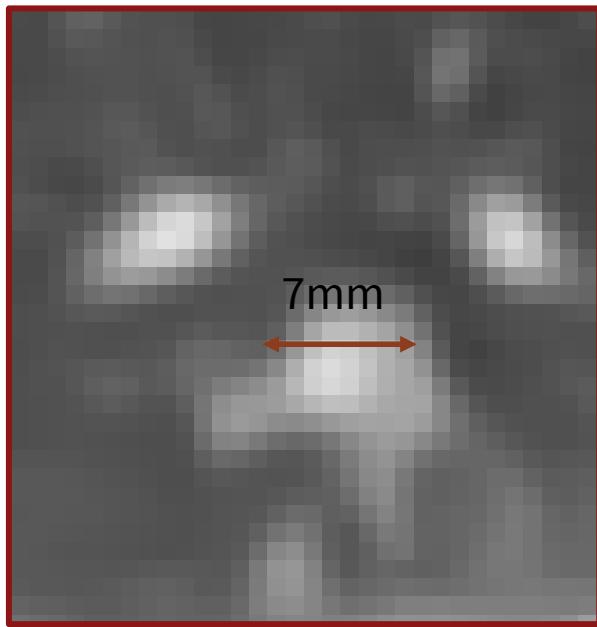


# Computational Methods for Biomedical Image Analysis

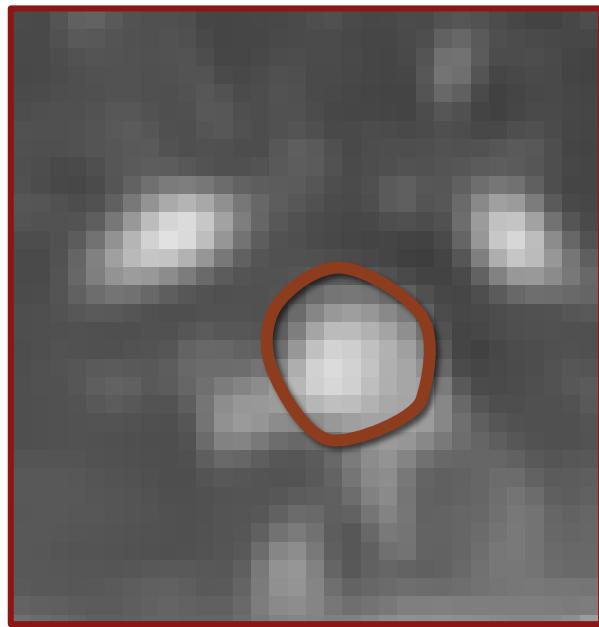
# Manual Assessment

vs

# Computational Methods



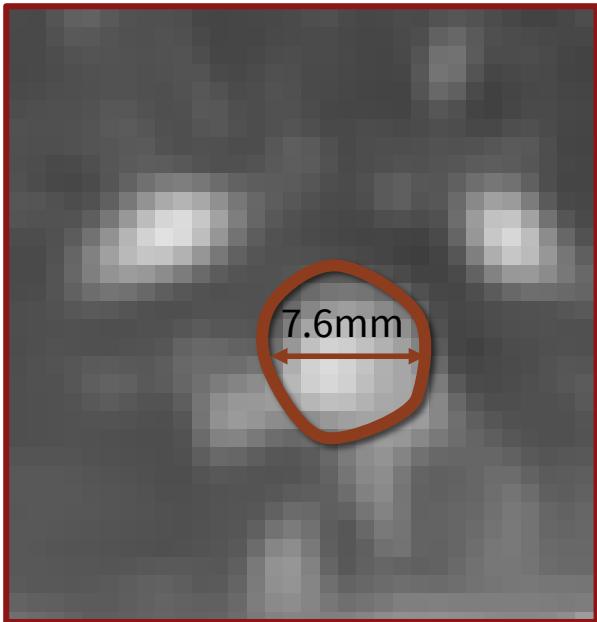
Malignant Nodule



# Manual Assessment

vs

# Computational Methods



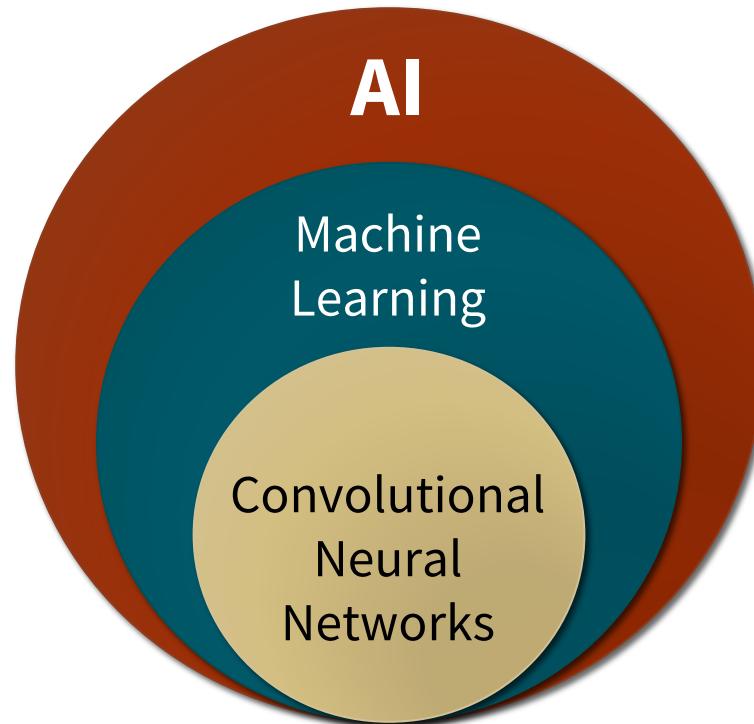
1. Classification: Suspicious Nodule
2. Classification: Benign vs Malignant
3. Segmentation: Red outline  
Area:  $181 \text{ mm}^2$   
Diameter = 7.61mm

# Introduction to Artificial Intelligence (AI)

AI: Development of computer systems to perform task that normally require human intelligence

Machine Learning: learn and adapt without explicit instructions

Convolutional Neural Networks:  
class of artificial neural networks



# AI Tasks

2012 AlexNet (CNN on GPU) best on ImageNet  
1000 categories (Dogs vs. Cats)



vs.



Muffins vs Chihuahua



@TeenyBiscuit

Stanford University

# Challenges of Medical AI

## Limited Medical Data (vs. Natural Images)

- Low Patient Volume
- Protected Health Information

## Limited Labels

- Labels are tedious to generate
- “Ground truth”: pathology, but not always available
- Radiologist Labels: Interpretation of radiology images (possibly with pathology confirmation)
  - Problem: same conditions are hard to see on radiology images

# Project Expectation

# Motivation

- The field of Biomedical Image Analysis moves very quickly
  - Image Modalities are being upgraded often
  - Scientific problems are updated often
  - Computational approaches used to solve these problems
- Consistent elements
  - Scientific approach: experimental design
  - Rigor: Model development, validation and testing
  - Statistical Analysis
  - The clinical relevance of the methods

# Goals

- Goal Class:  
Introduce you to a wide range of methods (not exhaustive list)
- Goals Project:  
Get comfortable applying the methods, solving biomedical problems
- In the space of Biomedical Image Analysis:  
Everything is a short-term project

# Example Research Project

<https://link.springer.com/article/10.1007/s00330-023-10540-3>

Almeida et al. European Radiology \_#####\_  
<https://doi.org/10.1007/s00330-023-10540-3>



## IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

# Prediction of disease severity in COPD: a deep learning approach for anomaly-based quantitative assessment of chest CT



Silvia D. Almeida<sup>1,2,3,4\*</sup>, Tobias Norajitra<sup>1,2</sup>, Carsten T. Lüth<sup>5,6</sup>, Tassilo Wald<sup>1,6</sup>, Vivienn Weru<sup>7</sup>, Marco Nolden<sup>1,8</sup>, Paul F. Jäger<sup>5,6</sup>, Oyunbileg von Stackelberg<sup>8,9</sup>, Claus Peter Heußel<sup>2,9,10</sup>, Oliver Weinheimer<sup>2,9</sup>, Jürgen Biederer<sup>2,9,11,12</sup>, Hans-Ulrich Kauczor<sup>2,9</sup> and Klaus Maier-Hein<sup>1,2,4,6,8\*</sup>

## IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

# Prediction of disease severity in COPD: a deep learning approach for anomaly-based quantitative assessment of chest CT



Silvia D. Almeida<sup>1,2,3,4\*</sup>, Tobias Norajitra<sup>1,2</sup>, Carsten T. Lüth<sup>5,6</sup>, Tassilo Wald<sup>1,6</sup>, Vivieni Weru<sup>7</sup>, Marco Nolden<sup>1,8</sup>, Paul F. Jäger<sup>5,6</sup>, Oyunbileg von Stackelberg<sup>8,9</sup>, Claus Peter Heußel<sup>9,10</sup>, Oliver Weinheimer<sup>2,9</sup>, Jürgen Biederer<sup>2,9,11,12</sup>, Hans-Ulrich Kauczor<sup>2,9</sup> and Klaus Maier-Hein<sup>1,2,4,6,8\*</sup>

### Abstract

**Objectives** To quantify regional manifestations related to COPD as anomalies from a modeled distribution of normal-appearing lung on chest CT using a deep learning (DL) approach, and to assess its potential to predict disease severity.

**Materials and methods** Paired inspiratory/expiratory CT and clinical data from COPDGene and COSYCONET cohort studies were included. COPDGene data served as training/validation/test data sets ( $N = 3144/786/1310$ ) and COSYCONET as external test set ( $N = 446$ ). To differentiate low-risk (healthy/minimal disease, [GOLD 0]) from COPD patients (GOLD 1–4), the self-supervised DL model learned semantic information from  $50 \times 50 \times 50$  voxel samples from segmented intact lungs. An anomaly detection approach was trained to quantify lung abnormalities related to COPD, as regional deviations. Four supervised DL models were run for comparison. The clinical and radiological predictive power of the proposed anomaly score was assessed using linear mixed effects models (LMM).

**Results** The proposed approach achieved an area under the curve of  $84.3 \pm 0.3$  ( $p < 0.001$ ) for COPDGene and  $76.3 \pm 0.6$  ( $p < 0.001$ ) for COSYCONET, outperforming supervised models even when including only inspiratory CT. Anomaly scores significantly improved fitting of LMM for predicting lung function, health status, and quantitative CT features (emphysema/air trapping;  $p < 0.001$ ). Higher anomaly scores were significantly associated with exacerbations for both cohorts ( $p < 0.001$ ) and greater dyspnea scores for COPDGene ( $p < 0.001$ ).

**Conclusion** Quantifying heterogeneous COPD manifestations as anomaly offers advantages over supervised methods and was found to be predictive for lung function impairment and morphology deterioration.

**Clinical relevance statement** Using deep learning, lung manifestations of COPD can be identified as deviations from normal-appearing chest CT and attributed an anomaly score which is consistent with decreased pulmonary function, emphysema, and air trapping.

\*Correspondence:  
Silvia D. Almeida  
silvia.almeida@dfz-heidelberg.de  
Klaus Maier-Hein  
kmaier-hein@dfz-heidelberg.de  
Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### Key Points

- A self-supervised DL anomaly detection method discriminated low-risk individuals and COPD subjects, outperforming classic DL methods on two datasets (COPDGene AUC = 84.3%, COSYCONET AUC = 76.3%).
- Our contrastive task exhibits robust performance even without the inclusion of expiratory images, while voxel-based methods demonstrate significant performance enhancement when incorporating expiratory images, in the COPDGene dataset.
- Anomaly scores improved the fitting of linear mixed effects models in predicting clinical parameters and imaging alterations ( $p < 0.001$ ) and were directly associated with clinical outcomes ( $p < 0.001$ ).

**Keywords** Chronic obstructive pulmonary disease, Deep learning, Artificial intelligence, Computed tomography

### Introduction

Chronic obstructive pulmonary disease (COPD) affects approximately 10.3% of the global population [1]. However, a significant portion of individuals remain undiagnosed [2], mainly due to the limitations of clinical tests and spirometry [3], in capturing the heterogeneous manifestations of the disease. These can range from predominant involvement of the airways to predominant damage and loss of the lung parenchyma. Quantitative CT imaging has emerged as a potential diagnostic tool [4], offering valuable insights into COPD manifestation [5, 6]. Nevertheless, the analysis and interpretation of CT images in these patients is challenging due to the variability and inhomogeneous distribution of findings related to COPD.

In recent years, supervised deep learning (DL) methods have been proposed to assist physicians in studying the various imaging characteristics of COPD [7–10]. Unfortunately, supervised approaches face limitations in capturing the full spectrum of COPD manifestations and representing them on the training data. In reality, this appears difficult, and typically results in poor generalizability. Additionally, these methods rely on obtaining local and global labels, which can be difficult and subjective to acquire. Moreover, previous DL methods have primarily focused on utilizing a single CT image during full inspiration, neglecting the potential benefits of incorporating full expiration CT scans as surrogate markers for small-airway inflammation [11, 12]. While large-scale cohort studies include expiratory CT scans for this purpose, their value for DL methods in COPD diagnosis has not yet been explored.

To address the limitations of existing supervised methods, this study used a self-supervised contrastive pretext model deep learning (DL) approach. Unlike supervised learning, self-supervised learning does not require labeling of pre-defined features. Instead, it leverages inherent patterns or relationships within the data, such as similarity between images, to create its own training labels and learn useful representations. Our hypothesis is that by leveraging the inherent similarities within normal-appearing lung regions and identifying deviations from

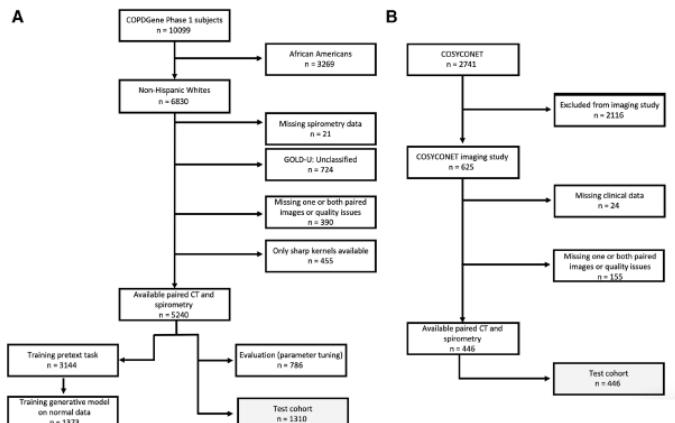
these characteristics, COPD regions can be detected without explicitly learning all possible image features. By this approach, the spectrum of diseased lung parenchyma is implicitly captured as anomalies from what is found in healthy subjects (never-smoker controls) or patients with minimal disease (GOLD 0, no airflow limitation and no/minimal emphysema) [13]. The presented study compares the proposed self-supervised anomaly detection approach with state-of-the-art supervised methods, explores the potential added value of incorporating expiratory CT scans under different input configurations, and contributes to a deeper understanding of the clinical implications in two nationwide cohorts.

### Materials and methods

#### Study cohorts

Two multicenter cohorts were retrospectively used: COPDGene (Genetic Epidemiology of COPD) and COSYCONET (COPD and SYstemic consequences-COrmorbidities NETwork). The COPDGene study (ClinicalTrials.gov Identifier: NCT00608764) [14] recruited current and former self-reported non-Hispanic whites and African Americans smokers ( $\geq 10$  pack-years), aged 45–80 years, between 2008 and 2011. The COSYCONET imaging sub-study (ClinicalTrials.gov Identifier: NCT02629432) [15] recruited individuals with a diagnosis of COPD, according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [16], or chronic bronchitis, aged 40 years or older, between 2010 and 2013. Both studies collected paired chest CT in inspiration (Insp.) and expiration (Exp.), pulmonary function tests (PFT), and questionnaires. Reconstruction protocols were comparable, but the maximum dose level for the CT acquisition differed (3.5 mSv for COSYCONET and 10 mSv for COPDGene, respectively). For additional details on the CT protocol, see Supplementary S1.

To account for disparities in health outcomes and quality of life between African Americans and non-Hispanic whites with COPD [17–19], a self-reported non-Hispanic white population was conveniently selected from COPDGene to match the ethnicity of COSYCONET.



**Fig. 1** Study design flowcharts for data selection. **A** Data selection from the COPDGene cohort. Data from 3144 participants were used to train the pretext task, from which 1373 normal never-smoker control and GOLD 0 subjects were used to train the generative model. The evaluation task (hyperparameter tuning) was performed on 786 subjects. The final testing was performed only once on the test set (1310 subjects). **B** Data selection from the COSYCONET cohort. This was entirely used as an external test set (n = 446)

Exclusion criteria and study design are shown in Fig. 1A and B.

The “control” class was defined as individuals without airflow obstruction, as indicated by forced expiratory volume in 1 s (FEV<sub>1</sub>) to forced vital capacity (FVC) ratio equal to or greater than 70%. This category encompassed never-smoker controls and GOLD 0 individuals, none of whom met the full criteria for a COPD diagnosis. As defined by the inclusion criteria to the cohort studies, these individuals did not present other lung diseases.

The “diseased” class was defined by individuals with clinical manifestations of COPD and airflow obstruction (FEV<sub>1</sub>/FVC < 70%), corresponding to GOLD 1–4.

Written consent was obtained, and the study protocol was approved by each clinical center’s review board.

#### Pre-processing

The input for the DL model were 3D ROIs (patches) extracted from single Insp scans or from dual-channel Insp and registered Exp (Expr) CT scans, respectively, covering > 70% of the segmented lung parenchyma volume of each individual. The chosen patch size of 50<sup>3</sup> voxels (50 × 50 × 50 voxels) was meant to cover the typical

size of a secondary pulmonary lobule, the basic unit of lung structure [20]. Two patch-overlapping strategies were implemented (0% and 20%) and applied to Insp CT (1 channel) and Insp + Expr (2 channels), resulting in four different configurations of input patches to be tested.

To ensure that patches extracted from the “control” class individuals were representative of healthy lung regions, only those with less than 1% emphysema were included when acquiring the representative distribution model of normal-appearing lung. Further pre-processing details are found in Supplementary S2 and Supplementary Figure 1.

#### COPD Classification as Out-Of-Distribution anomaly detection (cOoD)

A sequence of  $B$  3D lung patches ( $x_i$ ) $_{i=1}^B$  is taken per patient  $i$ , from a single or paired CT scan  $X$ . A latent representation is obtained per each patch, for a maximum of 100 patches per subject selected at random, using a trained self-supervised contrastive encoder  $z_i = f(x_i)$ . The maximum of 100 patches per subject was defined based on previous experiments, showing that using all patches available per subject only introduces redundancy while

increasing computational costs. Then, the distribution of normal representations from all patches of control individuals is learned through a generative model  $p(z)$  and anomalies (assumed as stemming from COPD) are identified and given an anomaly score defined as the negative log likelihood  $s(x_i) = -\log(p(f(x_i)))$ . A patient-level score  $S(X)$  is obtained by aggregating all patch-level scores, which is then used to predict the binary class (“control” vs “diseased” classes). Final aggregation strategy was chosen based on the highest area under receiver operator curve (AUC) on three runs on the validation set, for all the input configurations (Insp 0%, Insp 20%, Insp Expr 0%, Insp Expr 20%). These steps are detailed in [13] and Supplementary S3–5 and presented in Fig. 2.

#### Compared methods

Four established supervised deep learning methods, detailed in Supplementary S-6, were compared to cOoD: three voxel-based (end-to-end) Patch Classifier with a recurrent neural network [PatchClass + RNN], a multiple instance learning [MIL] with RNN as aggregation [MIL + RNN], an attention-based MIL [MIL + Att] and one representation-based [ReContrastive].

#### Statistical analysis and clinical prediction

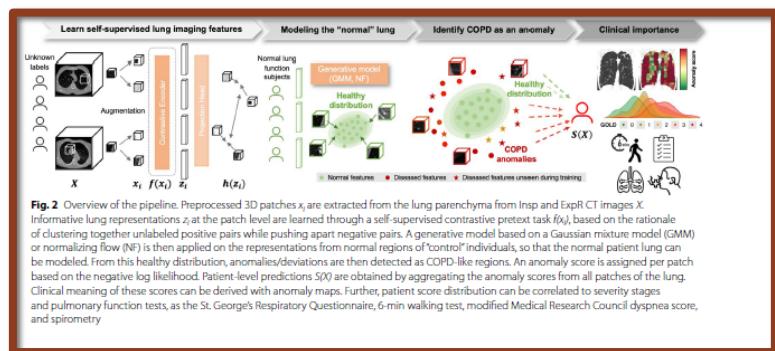
The main evaluation metric for the COPD binary classification was the AUC (more details in Supplementary S-7).

For exploratory purposes, a two-way ANOVA was performed independently per dataset to evaluate the effect of the method, input configuration, and their interaction on the main performance metric. A post hoc Tukey test was then performed to assess the multiple

pairwise comparisons in between each method and input configuration.

Linear mixed effects models (LMM) were used to predict clinical parameters and radiological features based on the cOoD anomaly score ( $S(X)$ ), adjusted for age, gender, body mass index (BMI), smoking status (0: never-smoker control, 1: former smoker, and 2: current smoker), smoking duration, and a random term for the study site. Clinical parameters included the following: St. George’s Respiratory Questionnaire (SGRQ), 6-min walking test (6MWT), FEV<sub>1</sub>, or FEV<sub>1</sub>/FVC. Radiological features were the %Emphysema or the %Air Trapping, measured by Vida Diagnostics (Coralville, IA) (for COPDGene) and by YACTA [21–23] (for COSYCONET). As these radiological features were skewed, a log transformation was employed. We report the overall conditional coefficient of determination ( $R^2$ ), adjusted for the number of regressors, and the individual  $R^2$  decomposition of the explained variation. Forest plots present the standardized beta coefficients (estimates) of the fixed terms, per model. 95% confidence intervals (CIs) and p-values were computed using a Wald t distribution approximation. Added value of the anomaly score, compared to nested baseline methods, was assessed through the likelihood ratio test of nested models. Additionally, to test whether the anomaly score provides additional information beyond morphological lung changes, models were also adjusted for %Emphysema and %Air Trapping, for the prediction of PFT.

COPD outcomes prediction by the anomaly score was also analyzed, namely reported severe exacerbations in the past year (Wilcoxon rank-sum test) and dyspnea, by



**Table 1** Demographic data, functional parameters, and low-attenuation (LAA) percentages, for the pretext self-supervised contrastive task, for the generative model on the “control” class from the COPDGene dataset, and for the internal (COPDGene) and external test set (COSYCONET). Note: Attenuation percentages were measured by different methods: VIDA Diagnostics for COPDGene; YACTA for COSYCONET

| Characteristic                                 | Training pretext task<br>(COPDGene)<br>“normal” and<br>“diseased” class) | Training generative model<br>(COPDGene)<br>only “normal” class) | Internal test set<br>(COPDGene)<br>“normal” and<br>“diseased” classes) | External test set<br>(COSYCONET)<br>“normal” and<br>“diseased”<br>classes) |
|--|--|---|--|--|
| Demographic data                               |  |   |  |  |
| N of patients [N]                              | 3144   | 13/3  | 1310   | 446  |
| M [N]  | 1699   | 694   | 692  | 272  |
| F [N]  | 1445   | 679   | 618  | 174  |
| Age (y) [mean, (IQR)]                          | 63 (55–69)   | 60 (52–66)  | 63 (56–69)   | 63 (58–69)   |
| BMI [mean, (SD)] <sup>a</sup>                  | 28.3 (5.7)   | 29.1 (5.6)  | 28.1 (5.7)   | 26.6 (4.6)   |
| Smoking habits                                 |  |   |  |  |
| Never-smoker [N, (%)]                          | 63 (2.0%)  | 63 (4.6%)   | 29 (2.2%)  | 34 (7.6%)  |
| Former smoker [N, (%)]                         | 1862 (59.2%)   | 755 (54.9%)   | 817 (62.4%)  | 298 (66.7%)  |
| Current smokers [N, (%)]                       | 1219 (38.8%)   | 555 (40.4%)   | 464 (35.4%)  | 114 (25.6%)  |
| Smoking duration (y) [mean, (SD)] <sup>a</sup> | 36 (12)  | 31 (13)   | 36 (12)  | 33 (15)  |
| Spirometry                                     |  |   |  |  |
| FEV <sub>1</sub> %_pred [mean, (SD)]           | 75.4 (26.8)  | 97.2 (11.3)   | 74.2 (27.1)  | 57.6 (19.1)  |
| FEV <sub>1</sub> /FVC [mean, (SD)]             | 0.6 (0.2)  | 0.8 (0.1)   | 0.6 (0.2)  | 0.7 (0.2)  |
| Non-smoker control [N]                         | 63   | 63  | 29   | 0  |
| GOLD 0 [N]                                     | 1310   | 1310  | 538  | 23   |
| GOLD 1 [N]                                     | 350  | 0   | 128  | 30   |
| GOLD 2 [N]                                     | 764  | 0   | 315  | 215  |
| GOLD 3 [N]                                     | 423  | 0   | 195  | 146  |
| GOLD 4 [N]                                     | 234  | 0   | 105  | 32   |
| Severe exacerbations [N (%)]                   | 318 (10.1%)  | 40 (3%)   | 137 (10.5%)  | 224 (50.3%)  |
| 6MWT (ft) [mean, (SD)] <sup>a,b</sup>          | 433.8 (120.7)  | 485.4 (98.6)  | 427.1 (124.8)  | 453.3 (99.8)   |
| SGRQ [mean, (SD)] <sup>a</sup>                 | 25.6 (22.5)  | 13.3 (15.5)   | 25.9 (22.9)  | 40.3 (18.8)  |
| mMRC dyspnea score 0 [N] <sup>a,c</sup>        | 1502   | 963   | 588  | 274  |
| mMRC dyspnea score 1 [N] <sup>a,c</sup>        | 460  | 193   | 221  | 110  |
| mMRC dyspnea score 2 [N] <sup>a,c</sup>        | 364  | 104   | 158  | 60   |
| mMRC dyspnea score 3 [N] <sup>a,c</sup>        | 549  | 98  | 225  | 1  |
| mMRC dyspnea score 4 [N] <sup>a,c</sup>        | 264  | 15  | 115  | 0  |
| Imaging  |  |   |  |  |
| Imaging (Insp and Exp)                         |  |   |  |  |
| LAA-950%* [mean, (SD)] <sup>a,d</sup>          | 7.8 (10.4)   | 2.5 (3.0)   | 8.1 (10.6)   | 17.0 (13.6)  |
| LAA-856%* [mean, (SD)] <sup>a,d</sup>          | 25.4 (20.4)  | 11.7 (9.9)  | 26.1 (21.1)  | 45.2 (20.5)  |

COPDGene: Genetic Epidemiology of COPD; COSYCONET: COPD and Systemic consequences-Comorbidities NETwork; N, number; SD, standard deviation; y, years; BMI, body mass index; FEV<sub>1</sub>, forced expiratory volume in 1 s; FEV<sub>1</sub>/FVC, FEV<sub>1</sub>-to-forced vital capacity ratio; GOLD, Global Initiative for Chronic Obstructive Lung Disease; 6MWTT, 6-min walking test; SGRQ, St. George's Respiratory Questionnaire; LAA-950%, percentage of LAA under - 950 HU; LAA-856%, percentage of LAA under - 856 HU.

For some cases, data was not available:

\* (48 for 6MWTT, 5 for mMRC dyspnea score, 7 for LAA-950%, 163 for LAA-856%);

<sup>a</sup> (7 for 6MWTT, 3 for LAA-950%, 65 for LAA-856%);

<sup>b</sup> (7 for 6MWTT, 3 for mMRC dyspnea score, 3 for LAA-950%, 76 for LAA-856%);

<sup>c</sup> (1 for BMI, 3 smoking duration, 22 for 6MWTT, 2 for SGRQ, 1 for mMRC dyspnea score)

the modified Medical Research Council (mMRC) dyspnea scale (Jonckheere-Terpstra test).

Statistical analyses were performed with R (version 4.2.3; R Foundation for Statistical Computing). A p-value of < .05 was considered statistically significant and was adjusted for multiple comparisons, using the Holm-Bonferroni method.

The applied code is available on a public repository on GitHub (<https://github.com/MIC-DKFZ/cOoPd>).

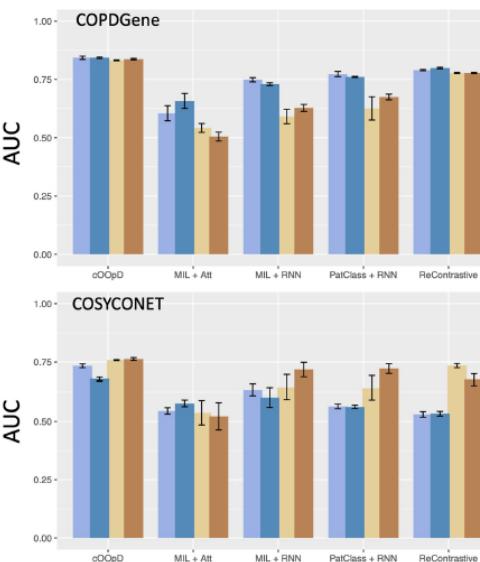
## Results

### Clinical characteristics

Table 1 presents demographic data, functional parameters, and radiological measures for the training and test sets. The self-supervised pretext task training cohort consisted

of 3144 COPDGene participants, of which 63 never-smokers and 1310 GOLD 0 met the criteria for the “control” class, as defined above. Healthy-appearing samples from this group were used to acquire the representative distribution model of the “control” lung. The average percentages of emphysema and air trapping in the entire lungs of these subjects were 2.5% and 11.7%, respectively. This was considered to be consistent with low-risk and control groups from previously published studies [24–26].

The internal testing cohort consisted of 1310 COPDGene participants (692 men, 618 women), with mean age 63 years (interquartile range [IQR] 56–69). The external testing cohort consisted of 446 COSYCONET participants (272 men, 174 women), with mean age 63 years (IQR 58–69).



**Fig. 3** Performance assessment through the area under the receiver operating curve (AUC) for the internal (COPDGene) and external (COSYCONET) test sets, for four different input configurations (0% and 20% patch-overlapping applied to Insp CT (1 channel) and Insp + ExpR (2 channels)), for the anomaly detection method (cOoPd), and for four supervised deep learning methods: end-to-end Patch Classifier with a recurrent neural network [PatClass + RNN], a multiple instance learning [MIL] with RNN as aggregation [MIL + RNN], an attention-based MIL [MIL + Att], and one representation-based [ReContrastive]

### COPD binary classification and effect of including the expiratory CT

The average performances of all models and all input configurations on the internal and external test sets (Fig. 3) were examined. On COPDGene, the best average performance was achieved by our proposed model (cOOp), reaching an AUC of  $83.2 \pm 0.2$ ,  $83.7 \pm 0.3$ ,  $84.3 \pm 0.7$ , and  $84.3 \pm 0.3$  for Insp-0%, Insp-20%, Insp + Expr-0%, and Insp + Expr-20% respectively. All methods showed a slightly lower performance on the external test set (COSYCONET), but the proposed method remained the most performant, achieving an AUC of  $75.8 \pm 0.2$ ,  $76.3 \pm 0.6$ ,  $73.4 \pm 0.8$ , and  $67.9 \pm 0.7$  for Insp-0%, Insp-20%, Insp + Expr-0%, and Insp + Expr-20%, respectively (Supplementary Table 3).

For exploratory purposes, the effect of method, input configuration, and their interaction for overall performance was evaluated (Table 2). For both datasets (COPDGene and COSYCONET), we found a statistically significant difference in the average AUC by the method, the input configuration, and their interaction.

The Tukey post hoc test revealed that, for both test sets, our proposed method yielded, on average, higher AUC scores than all other methods ( $p < 0.001$ ), regardless of the input configuration.

The effect of adding the Expr, on the other hand, only showed, on average, statistically significant improvements

for the voxel-wise supervised methods (MIL + Att, MIL + RNN, PatClass + RNN) for COPDGene ( $p < 0.001$ ). When applied to the data from COSYCONET, this effect was no longer observed.

The subsequent analyses were conducted for the configuration that achieved higher mean AUC and lower standard deviation (InspExpr-20%).

### Visualization of anomaly maps

Figure 4 shows representative coronal CT views with an overlay of the patch-level anomaly scores obtained for InspExpr-20% and Insp-20% (as reference), which illustrate how much certain lung regions differentiate from normal-appearing ones. Min-max normalization was applied for visualization purposes, corresponding to the 5th and 95th percentiles of the corresponding test set.

### Clinical and radiological predictive value: COPDGene

Adding the anomaly score statistically improved the fitting of all LMM to predict clinical and radiological measures ( $p < 0.001$  each) (Table 3). Similar models adjusted for %Emphysema, or adjusted for the %Air Trapping, to predict PFT, also benefit from adding the anomaly score as predictor (6MWT, FEV<sub>1</sub>, FEV<sub>1</sub>/FVC,  $p < 0.001$  for each). This indicates that the anomaly score comprises more than the morphological information provided by %Emphysema or %Air Trapping.

Furthermore, the explanatory power ( $R^2$ ) of all models increased by adding the anomaly score, with the most significant improvement observed in predicting radiological features. Compared to other predictors, the anomaly score had the highest individual explained variance for all models (Supplementary Table 4). In detail, it explained 19% (95%CI 16, 25), 12% (95%CI 8, 15), 39% (95%CI 35, 44), 41% (95%CI 38, 45), 28% (95%CI 23, 32), and 40% (95%CI 37, 45) of the variance of the SGQR, 6MWT, FEV<sub>1</sub>, FEV<sub>1</sub>/FVC, %Emphysema, and %Air Trapping, respectively.

As shown in Fig. 5, BMI, smoking duration, and the anomaly score were positively correlated with predicting SGQR. An increase in one standard deviation (SD) of the anomaly score resulted in a 0.44 increase of the SD of SGQR. For the 6MWT, the lower the BMI, age, smoking duration, and anomaly score, the higher the distance a patient can walk. This distance was lower for a female patient than for a male patient. Here, for an increase in one SD of the anomaly score, the SD of the distance a patient

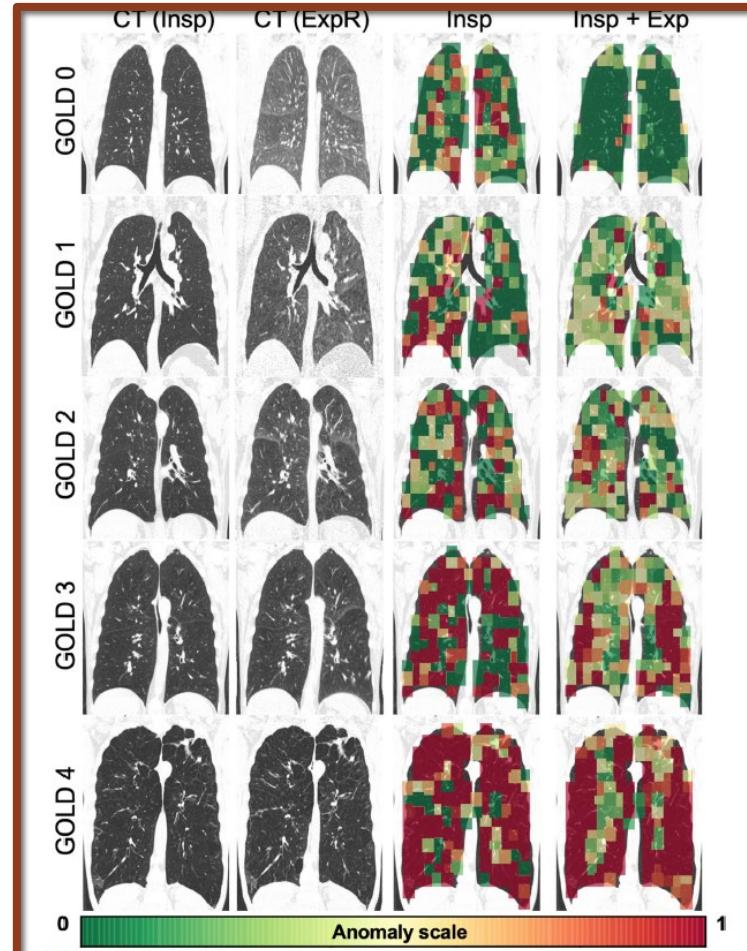
**Table 2** Two-way analysis of variance (ANOVA) per dataset (COPDGene and COSYCONET) to evaluate the effect of the method, input configuration, and their interaction on the main performance metric (AUC). The second column presents the  $F$  statistics, where the number in parentheses corresponds to the degrees of freedom ( $N - 1$ )

| Source of variation                         | F statistics    | P-value     |
|---|-----------------|-------------|
| COPDGene                                    |                 |             |
| Method                                      | $F(4) = 393.94$ | $p < 0.001$ |
| Input configuration                         | $F(3) = 94.43$  | $p < 0.001$ |
| Method <input/> configuration (interaction) | $F(12) = 14.36$ | $p < 0.001$ |
| COSYCONET                                   |                 |             |
| Method                                      | $F(4) = 50.43$  | $p < 0.001$ |
| Input configuration                         | $F(3) = 24.00$  | $p < 0.001$ |
| Method <input/> configuration (interaction) | $F(12) = 7.46$  | $p < 0.001$ |

COPDGene, Genetic Epidemiology of COPD; COSYCONET, COPD and Systemic consequences—Comorbidities NETwork; N, number of variables

(See figure on next page.)

**Fig. 4** Representative coronal views of the cOOp score map using the inspiratory image alone (3rd column) and using the Insp and the Expr image (4th column) on five subjects from COPDGene with different degrees of severity, for a 20% patch-overlapping strategy. CT images are scaled from  $-1300$  to  $50$  HU for visualization purposes. Color maps show the normalized patch anomaly score (negative log likelihood), normalized by the min-max normalization corresponding to the 5th and 95th percentiles of the dataset. Red symbolizes higher degrees of severity



**Fig. 4** (See legend on previous page.)

**Table 3** Linear mixed effects model to predict several clinical (SGRQ, 6MWT, FEV<sub>1</sub>, FEV<sub>1</sub>/FVC) and radiological (%Emphysema, %Air Trapping) dependent variables using the produced anomaly score as a predictor, for the COPDGene test cohort ( $n = 1310$ ). Note: Models are adjusted for age, gender, BMI, smoking status, smoking duration, and a random term for the study site. This model was compared with a baseline model that omits the anomaly score. Conditional  $R^2$  is adjusted for the number of regressors added. Bold values indicate a greater  $R^2$  per dependent variable.  $p$ -values are reported per model and for the comparison between them and are corrected for multiple comparisons. \*%Emphysema and %Air Trapping were skewed, so a log transformation was applied

| Dependent variable    | Predictor   | Adjusted conditional $R^2$ | $p$ -value |
|-----------------------|---|----------------------------|------------|
| SGRQ                  | Age, gender, BMI, smoking status, smoking duration, (center)                        | 0.24                       | $p < .001$ |
|                       | Age, gender, BMI, smoking status, smoking duration, <u>anomaly score</u> , (center) | <b>0.36</b>                | $p < .001$ |
| 6MWT                  | Age, gender, BMI, smoking status, smoking duration, (center)                        | 0.41                       | $p < .001$ |
|                       | Age, gender, BMI, smoking status, smoking duration, <u>anomaly score</u> , (center) | <b>0.45</b>                | $p < .001$ |
| FEV <sub>1</sub>      | Age, gender, BMI, smoking status, smoking duration, (center)                        | 0.22                       | $p < .001$ |
|                       | Age, gender, BMI, smoking status, smoking duration, <u>anomaly score</u> , (center) | <b>0.49</b>                | $p < .001$ |
| FEV <sub>1</sub> /FVC | Age, gender, BMI, smoking status, smoking duration, (center)                        | 0.26                       | $p < .001$ |
|                       | Age, gender, BMI, smoking status, smoking duration, <u>anomaly score</u> , (center) | <b>0.54</b>                | $p < .001$ |
| Emphysema % *         | Age, gender, BMI, smoking status, smoking duration, (center)                        | 0.38                       | $p < .001$ |
|                       | Age, gender, BMI, smoking status, smoking duration, <u>anomaly score</u> , (center) | <b>0.54</b>                | $p < .001$ |
| Air Trapping % *      | Age, gender, BMI, smoking status, smoking duration, (center)                        | 0.33                       | $p < .001$ |
|                       | Age, gender, BMI, smoking status, smoking duration, <u>anomaly score</u> , (center) | <b>0.58</b>                | $p < .001$ |

$R^2$ , overall conditional coefficient of determination; BMI, body mass index; SGRQ, St. George's Respiratory Questionnaire; 6MWT, 6-min walking test; FEV<sub>1</sub>, forced expiratory volume in 1 s; FEV<sub>1</sub>/FVC, FEV<sub>1</sub>-to-forced vital capacity ratio

can walk in 6 min decreased 0.31 times. For FEV<sub>1</sub>, both the effect of longer smoking duration and higher anomaly scores were significantly correlated with a low FEV<sub>1</sub>. No statistically significant differences were found for gender or BMI. For FEV<sub>1</sub>/FVC, the trends were similar, except for BMI, which was positively correlated. For both sputometry measures, an increase in one SD of the anomaly score was associated with a decrease of 0.60 of the SD of FEV<sub>1</sub> or FEV<sub>1</sub>/FVC, almost four times more than for a unit decrease in the SD of smoking duration.

For radiological features, lower BMI and higher anomaly scores were significantly correlated with higher %Emphysema and %Air Trapping. Both measures were significantly lower for female patients than for male patients. An increase in one SD of the anomaly score was associated with an increase of 0.27 SD for %Emphysema and %Air Trapping. This relationship is further highlighted in Fig. 6A and B, where the distributions of the anomaly score and the %Emphysema and %Air Trapping are shown. The anomaly score density plots (top) colored by the GOLD stage show a distinction between the GOLD classes, especially for GOLD 0 and 4. This distinction is no longer clear for the %Emphysema density plots, as the GOLD classes are highly overlapped.

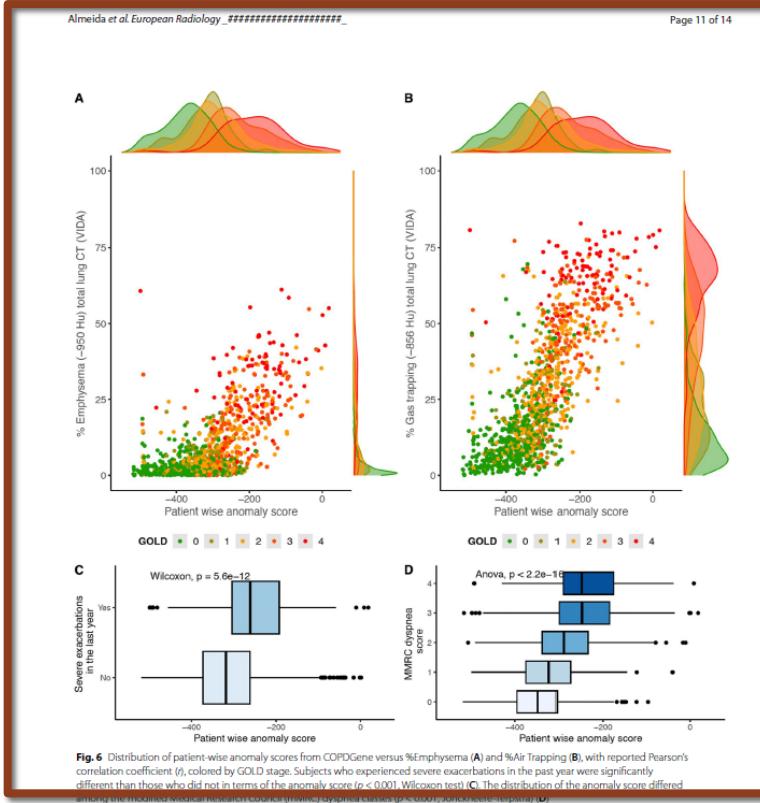
Finally, the anomaly score was associated with worsening of COPD symptoms. Patients who reported severe exacerbations in the past year had a higher anomaly score ( $p < 0.001$ ) (Fig. 6C). Increases in dyspnea scores were also associated with higher anomaly scores ( $p < 0.001$ ) (Fig. 6D).

#### Clinical and radiological predictive value: COSYCONET

Adding the anomaly score statistically improved the fitting of all LMM models ( $p < 0.001$  each), except for SGRQ. The explanatory power followed the same trends as in COPDGene, although less strong (Supplementary Table 5). Individual explained variance of the predictors (Supplementary Table 6) and standardized beta forest plots (Supplementary Figure 2) were consistent with the results reported before: higher anomaly scores were associated with greater SGRQ, more severe emphysema and air trapping (Supplementary Figure 3A, B, and D) and lower walking distances and lung function decline. Again, patients who reported severe exacerbations in the past year had higher anomaly scores ( $p < 0.001$ ) but increases in dyspnea scores were no longer associated with higher anomaly scores (Supplementary Figure 3C, D).

#### Discussion

In this study, we reformulate COPD binary classification into an anomaly detection task. We leverage the heterogeneity of COPD by modeling characteristics of "normal" lung tissue from low-risk "control" individuals and identifying deviations as anomalies indicative of impaired regions. Our method, based on self-supervised contrastive deep learning, outperformed supervised classification models in two cohorts: COPDGene (AUC 84.3 ± 0.3,  $p < 0.001$ ) and COSYCONET (AUC 76.3 ± 0.6,  $p < 0.001$ ). The anomaly detection task was trained solely on differentiating COPD patients from healthy individuals and



**Fig. 6** Distribution of patient-wise anomaly scores from COPDGene versus %Emphysema (A) and %Air Trapping (B), with reported Pearson's correlation coefficient ( $r$ ), colored by GOLD stage. Subjects who experienced severe exacerbations in the past year were significantly different than those who did not in terms of the anomaly score ( $p < 0.001$ , Wilcoxon test) (C). The distribution of the anomaly score differed among the modified median research council (mMRC) dyspnea classes ( $p < 0.001$ , Jonckheere-Terpstra) (D)

voxel-based methods, which heavily rely on labeled data and showed a significant improvement for COPDGene when expiratory images were incorporated. As a possible explanation for the robustness of our contrastive task model even when limited to inspiratory images alone instead of using both, in- and expiratory images, one could have discussed an already high prevalence

of small airway disease and consecutive findings on expiratory CT images (i.e., air trapping) in the minimal risk group (defined as "control" for the purposes of this study). This could have reduced the difference to the "diseased" group, in particular for features related to expiratory CT images. However, this appears unlikely, since in the above defined "control" class, which we employed to

features as seen in the COPDGene test set. High class imbalance, different cohort inclusion criteria, and differences in CT protocols, namely dose differences, may be the origin of this variation. Still, anomaly scores showed the same associations to clinical and radiological features as seen in the COPDGene test set, which may be considered suggestive for the generalizability of this approach.

Some limitations of our study need to be discussed: Although we circumvent the need to label all the data, our method still requires defining a population of normal individuals, which in our study consisted of healthy never-smoker individuals and individuals with minimal disease and minimal emphysema and without airflow limitation (GOLD 0). As a measure to reduce potential bias from this, we only included patches with less than 1% emphysema to exclude regions with a certain degree of lung damage. Furthermore, it is worth noting that assembling a truly healthy population, comparable in size to the diseased cases in the COPDGene training dataset, poses significant challenges. This is particularly true in the context of our study, which necessitated paired inspiratory and expiratory CT scans. Such scans involve a radiation burden that may not be ethically justifiable for healthy individuals, further complicating the recruitment of an ideal control group. Finally, the study was focused on COPD; therefore, the performance of the approach in other lung and vascular diseases cannot be predicted. Further research on the applicability for other disease entities is warranted.

In conclusion, we demonstrated the feasibility of identifying COPD as a deviation from normality. The produced region anomaly scores provide visual representations of local deviations and can potentially serve as surrogate markers to disentangle COPD phenotypes and early identification of individuals at risk. The subject-wise anomaly score provides an interpretable metric that explains common clinical and radiological manifestations. Future work will focus on generalizability to other datasets and lung diseases, as well as longitudinal analysis.

## Abbreviations

|                  |  |
|------------------|--|
| COPD             | Chronic obstructive pulmonary disease  |
| COPDGene         | Genetic Epidemiology of COPD   |
| COSYCONET        | Systematic Sources-CMorbilities NETwork  |
| Epi              | Inspiratory CT images  |
| ExpR             | Registered expiratory CT images  |
| FEV <sub>1</sub> | Forced expiratory volume in 1 s  |
| FVC              | Forced vital capacity  |
| GOLD             | Global Initiative for Chronic Obstructive Lung Disease                         |
| Insp             | Inspiratory CT images  |
| LAA-856          | Percentage of lung voxels with CT attenuation less than -856 HU on inspiration |
| LAA-950          | Percentage of lung voxels with CT attenuation less than -950 HU on inspiration |
| SGRQ             | St. Georges Respiratory Questionnaire  |

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s00330-023-10540-3>.

Below is the link to the electronic supplementary material Supplementary file 1 (PDF 749 KB)

## Funding

Open Access funding enabled and organized by Projekt DEAL. This research was funded by the State Ministry of Baden-Württemberg for Sciences, Research and Arts, Germany, grant number 30.4000/SB/3; and by Helmholtz Imaging (H4), a platform of the Helmholtz Institute on Information and Data Sciences.

The COPDGene study was funded by National Institutes of Health grants U01HL089856 and U01 HL089897 and also supported by the COPD Foundation through contributions made by an Industry Advisory Board comprised of Pfizer, Astrazeneca, Boehringer Ingelheim, Novartis, and Sunovion. The COSYCONET study is supported by the German Center for Lung Research (DZL), grant number B2DZL005A2 (COSYCONET), and the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) grant number 01GQ0881, and is furthermore supported by unrestricted grants from Astrazeneca GmbH, Boehringer Ingelheim Pharma GmbH & Co. KG, GlaxoSmithKline GmbH&Co. KG, Grifols Deutschland GmbH, and Novartis Deutschland GmbH.

## Declarations

### Guarantor

The scientific guarantor of this publication is Prof. Dr. Klaus Maier-Hein.

### Conflict of interest

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

### Statistics and biometry

One of the authors has significant statistical expertise: Vivien Weru, Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany.

### Informed consent

Written informed consent was obtained from all subjects (patients) in this study.

### Ethical approval

Institutional Review Board approval was obtained.

### Study subjects or cohorts overview

The test cohorts (COPDGene and COSYCONET) are open access, therefore have been reported in several previous studies, including our own: Almeida S.D. et al. (2022). cOpD: reformulating COPD classification on chest CT scans as anomaly detection using contrastive representations. In: Greenspan, H., et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2023. MICCAI 2023. Lecture Notes in Computer Science, vol 14242. Springer, Cham. [https://doi.org/10.1007/978-3-031-43904-9\\_4](https://doi.org/10.1007/978-3-031-43904-9_4).

### Methodology

- retrospective
- observational
- multicenter study

### Author details

- 1Division of Medical Image Computing, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. <sup>2</sup>Translational Lung Research Center Heidelberg, Department of General Thoracic Research Center (DZL), Heidelberg, Germany. <sup>3</sup>Medical Faculty Heidelberg, University Heidelberg, Germany. <sup>4</sup>National Center for Tumor Diseases (NCT), NCT Heidelberg, a partnership between DKFZ and Heidelberg University Medical Center, Heidelberg, Germany. <sup>5</sup>Interactive Machine Learning Group
- Karch A, Vogelmeier C, Witte T et al (2016) The German COPD cohort COSYCONET: aims, methods and descriptive analysis of the study population at baseline. *Respir Med* 110:247–257
- GOLD (2020) GOLD 2020 Global Strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. <https://goldcopd.org/wp-content/uploads/2019/11/GOLD-2020-REPORT-ver1.0wms.pdf>

- 6ML, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>7</sup>Helmholtz Imaging, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>8</sup>Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>9</sup>Pattern Analysis and Learning Group, Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany. <sup>10</sup>Department of Diagnostic and Interventional Radiology, Heidelberg University Hospital, Heidelberg, Germany. <sup>11</sup>Diagnostic and Interventional Radiology with Nuclear Medicine, Thoraxklinik at University Hospital, Heidelberg, Germany. <sup>12</sup>Faculty of Medicine, University of Latvia, Raina Bulvaris 19, Riga LV-1366, Latvia. <sup>13</sup>Faculty of Medicine, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany.
- Received: 5 October 2023 Revised: 13 November 2023 Accepted: 11 December 2023  
Published online: 27 December 2023

## References

- 1 Adelom, D, Song P, Zhu Y, Campbell H, Sheikh A, Rudan I (2022) Global, regional, and national burden of disease and risk factors for chronic obstructive pulmonary disease (COPD): a systematic review and modeling analysis. *Lancet Reg Reg Health Public Health* 37:447–458
- 2 Martinez CH, Mannino DM, Jaimes F et al (2015) Undiagnosed obstructive lung disease in the United States: Associated factors and long-term mortality. *Am J Thorac Soc* 12:1788–1795
- 3 Andreava E, Pokharelina M, Lebedev A, Moiseeva I, Kurnosova O, Degryse JM (2017) Spirometry is not enough to diagnose COPD in epidemiological studies: a follow-up study. *NPJ Prim Care Respi Med* 27:62
- 4 Lowe KE, Regan EA, Anzueto A et al (2019) COPDGene 2019: redefining the diagnosis of chronic obstructive pulmonary disease. *Chronic Obstr Pulm Dis* 13:1–10
- 5 Jo H, Lee SM, Seo JB et al (2019) Prediction of pulmonary function in patients with chronic obstructive pulmonary disease: correlation with quantitative CT parameters. *Korean J Radiol* 20:2638–2645
- 6 Lynch DA, Moore CM, Wilson C et al (2018) CT-based visual classification of emphysema: association with mortality in the COPDGene Study. *Radiology* 288:859–866
- 7 González C, Ash SJ, Vega-Sánchez F, Merello G et al (2018) Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 197:193–203
- 8 Tang LYW, Corcoran C, Lam S, Leipisch J, Tam RC, Sin DD (2020) Towards a more accurate, faster, and less invasive imaging and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. *Lancet Digit Health* 2:e259–e267
- 9 Singla S, Gong M, Riley C, Scultra B, Battaglinelli K (2021) Improving clinical disease subtyping and future events prediction through a chest CT-based deep learning approach. *Med Phys* 48:1168–1181
- 10 Sun J, Liao X, Yan Y et al (2022) Detection and staging of chronic obstructive pulmonary disease using a computed tomography-based weakly supervised deep learning approach. *Eur Radiol* 32:5319–5329
- 11 Ganesh A, Muthukumaraswamy S, Scheffel H et al (2018) Time to add additional layers of respiratory chest CT in chronic obstructive pulmonary disease. *Can Respir J* 2018:1–9
- 12 Cao X, Guo Y, Xu N et al (2021) Potential value of expiratory CT in quantitative assessment of pulmonary vessels in COPD. *Front Med* 8:761804
- 13 Almeida SD, Lüth CT, Norajitra T et al (2023) cOpD: reformulating COPD classification on chest CT scans as anomaly detection using contrastive representations. In: Greenspan, H., et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2023 (eds. Greenspan H et al) vol 14242. Springer, Cham. <https://doi.org/10.1007/978-3-031-2301-0216>
- 14 Heussel CP, Herth FJ, Kappler J et al (2009) Fully automatic quantitative assessment of emphysema in computed tomography: comparison with pulmonary function testing and normal values. *Eur Radiol* 19:2391–2402
- 15 Park H, Yun J, Lee SM et al (2023) Deep learning-based approach to predict pulmonary function at chest CT. *Radiology* 307:e221488
- 16 Li F, Choi J, Zou C et al (2021) Latent traits of lung tissue patterns in former smokers derived by dual channel deep learning in computed tomography images. *Sci Rep* 11:9196
- 17 Lüth CT, Zimmerer D, Koehler G et al (2023) CRADL: Contrastive Representations for Unsupervised Anomaly Detection and Localization. <https://doi.org/10.48550/ARXIV/2301.02126>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Goal

- Project aims to provide the opportunity to develop computational methods for biomedical image analysis
- Write about computational methods
- Present computational methods
  
- Make teams of 2-3 members
- Find a problem that interests you
- Understand the biomedical question
- Identify if you have the right data to answer the question
- Perform the experiments, write about your approach and results, present them

[Home](#)[Lane Library  
Resources](#)[Syllabus](#)[Assignments](#)[Quizzes](#)[Discussions](#)[Files](#)[People](#)[Collaborations](#)[Zoom](#)[Panopto Course  
Videos](#)[Ed Discussion](#)

## Course assignments

During the quarter, the students will undertake developing and implementing four substantial imaging informatics applications (3 problem sets and 1 project), increasing in difficulty during the quarter. **Students will turn in problem sets as \*.pdf conversions of their Jupyter notebooks.** Grades for these assignments will be released on Canvas and weighted as in the table above; midterms and final project **may** be graded on a curve.

### Due dates and Late Day Policy

Assignments are due at or before 11:59 PM on the dates indicated above, and new assignments are released on Mondays. You may have up to **4 late days** in the quarter that you can use, except for final project writeup and final project presentation. **No extra credit will be given for unused late days.**

Once you are out of late days, you will lose **10% of your grade** on late Problem Sets each day past the deadline (each "day" rounds up - so if you are one hour past the deadline, that is a day).

You will lose **10% of your grade** for each day that the Project Proposal or Milestone is turned in late but the Final Write-up and Presentation **will not be accepted** past the deadline.

Please contact the TAs for special accommodation if applicable.

## Project

There is no midterm or final exam; however, you must complete a project with five deliverables to pass this class. The deliverables are:

### Project Proposal:

- Project Proposal Writeup

### Project Milestone:

- Project Milestone Presentation
- Project Milestone Writeup

### Final Project Presentations

- Project Presentation
- Project Writeup

**Project Rubric:** [Project Rubric Link](#) ↗



## Other Considerations

This course will use Python 3 for all programming-related assignments. We recommend using the [Anaconda distribution](#) as it comes with many useful scientific computing packages. It also allows you to use Jupyter notebooks, which will be used for submitting assignments.

'ssh' and 'cardinal'

# Project Details and Rubric

[https://docs.google.com/document/d/1Qx-9gpwYAA0fj4yAsbuvVnm\\_Qc7A6CIQCVUCG5luk7E/edit?tab=t.0](https://docs.google.com/document/d/1Qx-9gpwYAA0fj4yAsbuvVnm_Qc7A6CIQCVUCG5luk7E/edit?tab=t.0)

## BIOMEDIN 260/RAD 260/CS 235/BMP260 Project Information

For group-specific questions regarding projects, please create a post on Ed. Please first have a look through the frequently asked questions (located at the bottom).

One of BIOMEDIN 260's main goals is to provide hands-on experience on biomedical imaging analysis. The final project is intended to be an introduction towards this goal.

---

### Team Formation

Form your teams early! The Project can be done in teams of two to three students. The ideal team size is 3. Single person projects are not recommended. CGOE students are especially encouraged to start thinking about team formation early. Even if contributions are different, you will need to assure us that every teammate's contribution is significant.

Note: Only one group member is supposed to submit the assignment, and tag the rest of the group members (do not all submit separately, or on the flip side forget to tag your teammates if you are the group's designated submitter). If you do not do this, then you can submit a regrade request and we will fix it.

---

### Project Topics

Your first task is to pick a project topic. If you're looking for project ideas, please come to office hours, and we'd be happy to brainstorm and suggest some project ideas.

# Choosing your project

1. If you already have a research subject
  - Feel free to use it for this class
  - BUT: apply new methods that you are learning in this class
  - Make sure there is no overlap with the existing analysis
2. If you don't have a subject, yet:
  1. Find the problem
  2. Find the data

# Data Sources

- [MICCAI](#)
- [GrandChallenges](#)
- [Kaggle](#)
- [DDSM: Digital Database for Screening Mammography](#)
- [Imaging Data commons](#)
- [The Cancer Imaging Archive](#)
- [The Cancer Genome Atlas \(TCGA\)](#)
- [MicroscopyU](#)
- [Multimedia Database of Interstitial Lung Diseases](#)
- [Medical Segmentation Decathlon](#)
- [Stanford AIMI center datasets](#)
- [github repos that compile sources](#)

# Finding Data for Final Project

Grand challenges (many types):

<https://grand-challenge.org/challenges/>

The Cancer Imaging Archive:

<http://cancerimagingarchive.net>

The Imaging Data Commons:

<https://datacommons.cancer.gov/repository/imaging-data-commons>

Stanford AIMI datasets:

<https://aimi.stanford.edu/research/public-datasets>

Medical Imaging and Data Resource Center (focus on COVID-19 images):

<https://www.midrc.org/>

# Finding Data for Final Project

<https://sites.google.com/site/aacruzr/image-datasets>

*(link also on Canvas)*

Angel Cruz-  
Roa

Welcome

Angel's Blog

Biography

Deep learning

Doctoral Thesis

Funding sources

GECCO - Grupo de Estudio  
en Ciencias de la  
Computación

Image Datasets

Interesting links

Interesting papers

- Current state of the art of most used computer vision datasets: Who is the best at X? [http://rodrigob.github.io/are\\_we\\_there\\_yet/build/](http://rodrigob.github.io/are_we_there_yet/build/)
- Grand Challenges in Medical Image Analysis [http://www.grand-challenge.org/index.php/Main\\_Page](http://www.grand-challenge.org/index.php/Main_Page)
- Multimodal databases
  - Center for Invivo Microscopy (CIVM), Embrionic and Neonatal Mouse (H&E, MR) <http://www.civm.duhs.duke.edu/devatlas/>  
<http://www.civm.duhs.duke.edu/devatlas/UserGuide.pdf>
  - LONI image data archive <https://ida.loni.usc.edu/services/Menu/IdaData.jsp?project=>
- Radiology (Ultrasound, Mammograms, X-Ray, CT, MRI, fMRI, etc.)
  - COllaborative Informatics and Neuroimaging Suite (COINS) <https://portal.mrn.org/micis/index.php?subsite=dx>
  - The Cancer Imaging Archive (TCIA) <http://www.cancerimagingarchive.net/> (Collections)
  - Alzheimer's Disease Neuroimaging Initiative (ADNI) <http://adni.loni.ucla.edu/>
  - The Open Access Series of Imaging Studies (OASIS) <http://www.oasis-brains.org/>
  - Breast Cancer Digital Repository <http://bcdr.inegi.up.pt/>
  - DDSM: Digital Database for Screening Mammography <http://marathon.csee.usf.edu/Mammography/Database.html>
  - The Mammographic Image Analysis Society (MIAS) mini-database <http://peipa.essex.ac.uk/info/mias.html>
  - Mammography Image Databases 100 or more images of mammograms with ground truth. Additional images available by request, and links to several other mammography databases are provided <http://marathon.csee.usf.edu/Mammography/Database.html>
- NLM HyperDoc Visible Human Project color, CAT and MRI image samples - over 30 images [http://www.nlm.nih.gov/research/visible/visible\\_human.html](http://www.nlm.nih.gov/research/visible/visible_human.html)
- Histology and Histopathology (H&E, IHQ ...)
  - The International Society for Digital Imaging of the Skin (ISDIS): <https://isdis.net/home>

# Thank you!

Image Modalities, Interpretation and Semantic Features,  
By Dr. Bruno Suarez, Practicing Radiologist

What are your thoughts after the first lecture?

Nobody has responded yet.

Hang tight! Responses are coming in.

# Thank you!