Furthering Machine Learning Applications in COVID-19 Research with Diverse Aggregated Datasets and Generalized Ensemble Model

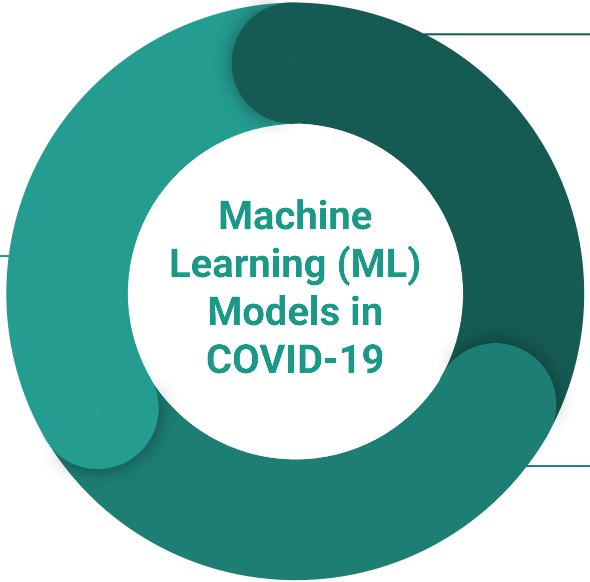
Group 2

Abhijit Gupta, Katerina Kargioti, Megha Joshi, Shayna Sragovicz

Introduction and Significance Katerina

Artificial Intelligence Enhanced Medical Imaging

- Data from sensors
- Different types of networks
- Focus on different aspects of the medical imaging
- Reduces patient-physician interaction
- Standardized
- Reduces radiation exposure



Existing Medical Image Workflow

- Requires physician assistance
- Involves patient positioning, Image Taking, and Data Interpretation

Issues of Current ML Models

- Limited datashets
  ↳ Models cannot be trained and tested properly
- Existing models specialize in certain topics

Purpose	Model	Dataset
Patient positioning	FAST Integrated Workflow [7]	
	Automatic Patient Centering for MDCT [8]	63 patients (36 men, 27 women, mean age: 51, age range: 22-83), chest CT: 18, abdominal: 45)
Body estimation	DARWIN: Deformable Patient Avatar Representation With Deep Image Network [4]	1063 patients from 3 different hospitals
	Patient MoCap [9]	180000 video frames
	Hierarchical Kinematic Human Mesh Recovery [5]	
Infectiousness assessment	Risk-Aware Identification of Highly Suspected COVID-19 Cases [12]	Hospital data of infected patients
Disease qualification	Deep learning-based detection for COVID-19 from chest CT [13]	Training: 499 CT volumes, testing: 131 CT volumes
	Deep Learning-based Quantitative CT Pipeline [14]	Training: CT images from 10 patients
	Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images [15]	50 patient images, 50 healthy individual images
	Automatic detection from X-Ray images utilizing Transfer Learning [16]	1427 X-Ray images (Covid-19: 224, common pneumonia: 700, normal conditions: 504)

Rationale Megha

- It is important to streamline the workflow required to treat and diagnose COVID-19 to minimize threats that frontline healthcare workers are facing
- Choosing the highest preform networks will enable standardization of the methodology
    - Consider a diverse set of data
- Robust and accurate dataset will facilitate the workflow between network-practitioner relationships

Specific Aim 1 Shayna

- Develop a **robust** and **diverse** dataset
- A **robust** dataset: large, variable in data
    - will allow for more accurate model predictions
- A **diverse** dataset: disease severity and demographics
    - ensure that all patients are able to be examined by networks, regardless of their identity.

Specific Aim 2 Abhijit

- Develop an ensemble model that is generalized and scalable
    - Takes pre-existing models and applies standardization of inputs and outputs, plus a final neural network layer
    - Should scale as additional models are introduced
    - Should perform well on a range of input characteristics

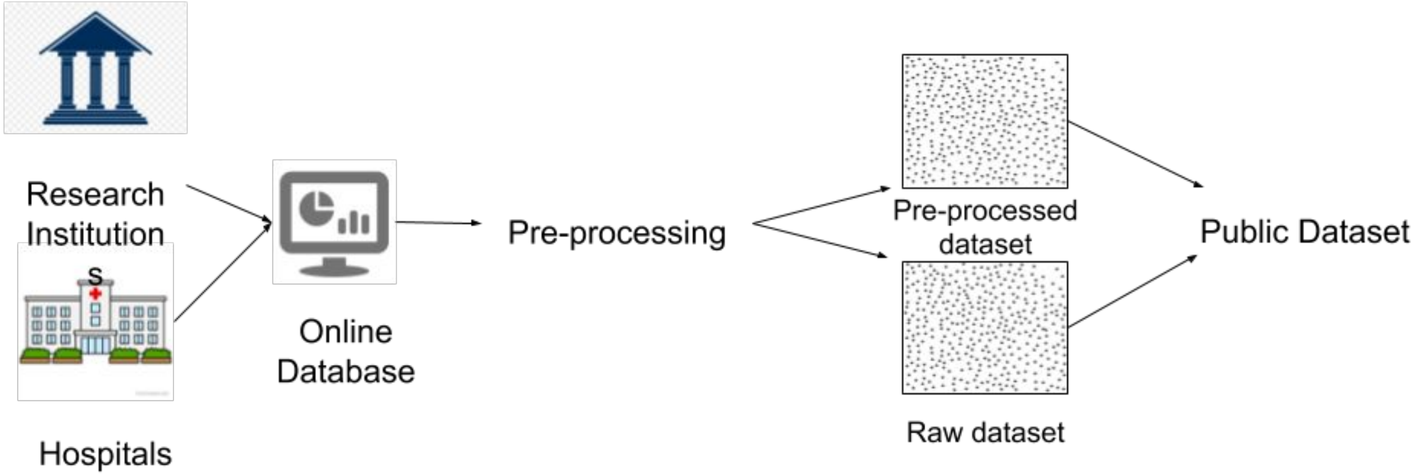
Experimental Design (Aim 1) Shayna

1. Develop an online platform to host image submissions
2. Invite owners of existing datasets to submit unprocessed data
    a. Work directly with public and private hospitals
3. Develop a method to pre-process all images
    a. by adjusting size and pixels to account for brightness, coloring, and clarity
4. Online platform will host two collections: raw and pre-processed

Experimental Design (Aim 1) **Shayna**

1. Plan to be transparent about the source and representation of our dataset(s)
2. Will include statistics regarding demographics of our dataset
3. Will dedicate resources to ensuring the diversity of our dataset
4. Recognize need for compliance with HIPAA and privacy laws
    a. Will allocate funds for a legal advisor to develop a lawful way to gather demographic statistics, however possible

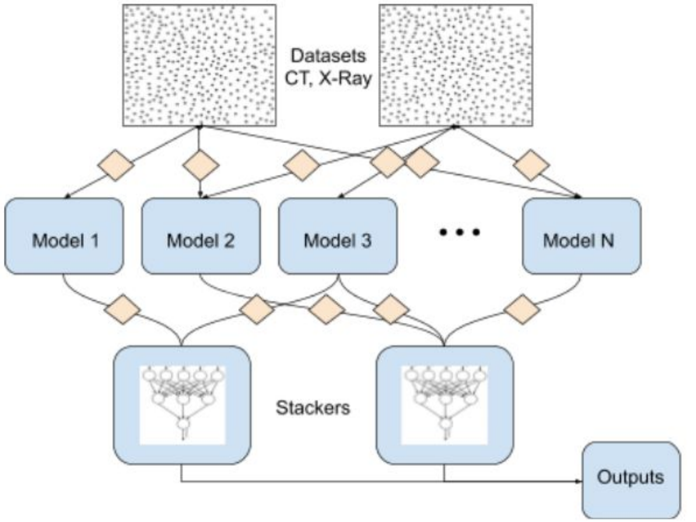
Experimental Design (Aim 1) Shayna



Expected aggregation method: a combination of research institutions and hospitals will submit data to an online database that is then preprocessed and made publicly available online in two forms: raw and pre-processed.

Experimental Design (Aim 2) Abhijit

- Two aggregated datasets, independent of Aim 1
- Preprocessing as per specific model requirements
- Post-processing for standardization
- Feeds into qualitative, quantitative stackers



Potential Pitfalls Megha

Aim One: Data-driven

- Quality of data inputted in order to have an accurately trained system.
    - Randomized testing
- CT and X-ray scans
    - Reporting ratios

Aim Two: Ensemble model

- Splitting data appropriately for testing and training
- Weaker models will get drowned out by more robust models
    - Criteria for each model included in the ensemble

References

1. Lauren M. Kucirka, Stephen A. Lauer, Oliver Laeyendecker, Denali Boon, Justin Lessler, (2020), Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure, *Annals of Internal Medicine*
2. Tahamtan A, Ardebili A, (2020), Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn*, vol. 20(5), 453-454.
3. V. Singh, Y.-J. Chang, K. Ma, M. Wels, G. Soza, and T. Chen, (2014), "Estimating a patient surface model for optimizing the medical scanning workflow," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 472-479.
4. V. K. Singh, K. Ma, B. Tamersoy, Y. Chang, A. Wimmer, T. F. Odonnell*, et al.*, (2017), DARWIN: Deformable patient avatar representation with deep image network, *Medical Image Computing and Computer Assisted Intervention*, 497-504.
5. R. Li, C. Cai, G. Georgakis, S. Karanam, T. Chen, and Z. Wu, (2019), Towards robust RGB-D human mesh recovery
6. G. Georgakis, R. Li, S. Karanam, T. Chen, J. Kosecka, and Z. Wu, (2020), Hierarchical hinematic human mesh recovery
7. *Siemens CT scanner SOMATOM Force, SOMATOM Drive or SOMATOM Edge Plus*. Available: https://www.siemens-healthineers.com/computed-tomography/technologies-and-innovations/fast-integrated-workflow
8. J. Li, U. K. Udayasankar, T. L. Toth, J. Seamans, W. C. Small, and M. K. Kalra, (2007), Automatic patient centering for MDCT: effect on radiation dose, *American Journal of Roentgenology,* vol. 188, 547-552.
9. F. Achilles, A. E. Ichim, H. Coskun, F. Tombari, S. Noachtar, and N. Navab, (2016), Patient MoCap: Human pose estimation under blanket occlusion for hospital monitoring applications, *Medical Image Computing and Computer Assisted Intervention*, 491-499.
10. L. A. Bullock Joseph, Pham Katherine Hoffmann, Lam Cynthia, Luengo-Oroz Miguel A., (2020), Mapping the landscape of artificial intelligence applications against COVID-19
11. Shi,F.,Wang,J.,Shi,J.,Wu,Z.,Wang,Q.,Tang,Z.,He,K.,Shi,Y.,Shen,D, (2020), Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Reviews in Biomedical Engineering
12. B. Wang, Y. Sun, T. Q. Duong, L. D. Nguyen and L. Hanzo, (2020), Risk-Aware Identification of Highly Suspected COVID-19 Cases in Social IoT: A Joint Graph Theory and Reinforcement Learning Approach, IEEE Access, vol. 8, 115655-115661.
13. C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma*, et al.*, (2020), Deep learning-based detection for COVID-19 from chest CT using weak label, *MedRxiv*
14. Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu*, et al.*, (2020), Longitudinal assessment of COVID-19 using a deep learning–based quantitative CT pipeline: Illustration of two cases, *Radiology: Cardiothoracic Imaging,* vol. 2, e200082.
15. A. Narin, C. Kaya, and Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks, (2020), *arXiv:2003.10849*.
16. M. Karim, T. Döhmen, D. Rebholz-Schuhmann, S. Decker, M. Cochez, O. Beyan, et al., (2020), Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images, arXiv preprint arXiv:2004.04582.