**Breakout Session 4: Track A**

# Development of an AI/ML-Ready Knee Ultrasound Dataset in a Population-Based Cohort

Dr. Katherine Yates
*Rheumatology Fellow, University of North Carolina at Chapel Hill*

# Development of an AI/ML-ready knee ultrasound dataset in a population-based cohort

Presented by: Dr. Katherine Yates, UNC Rheumatology fellow

Johnston County HEALTH STUDY

UNC | SCHOOL OF MEDICINE

# Value to the field

- Systematically performed studies are needed to inform broader use of ultrasound (US) for knee osteoarthritis (OA)

- US is safe, accessible, and efficient, bridging the gap between limited information on x-ray and more expensive MRI for OA

- Demonstrating the utility of US as an assessment tool <u>before disease is detectable on radiographs</u>, is crucial

# Parent R01 summary

- Obtain standardized US and radiography in the diverse population-based Johnston County Health Study (JoCoHS)

- Aim 1: determine the population prevalence of knee US features (i.e., cartilage and meniscal damage, synovitis/effusion, calcium crystal deposition, popliteal cysts and osteophytes)
  - overall and in key subgroups by age, sex, race/ethnicity, and symptom status

- Aim 2: quantify associations between US features and radiographic findings and symptom scores overall and in key subgroups

- Aim 3: apply novel machine learning methodologies to a) develop an overall US score for symptomatic KOA and b) identify the contribution of US variables to phenotypes relevant to KOA based on general health, physical activity, and functional assessments

# US images obtained per knee

| | View | Position | Structure(s) and/or pathology identified |
|---|---|---|---|
| A | **Suprapatellar** | Transverse over the distal femoral condyles, maximal flexion | Articular cartilage |
| B | **Suprapatellar** | Longitudinal over the distal femur and patella, 30° flexion, with and without power Doppler | Effusion/synovitis |
| C | **Suprapatellar** | Transverse over the distal femur, 30° flexion | Effusion/synovitis |
| D | **Medial** | Longitudinal at the medial femorotibial joint, 30° flexion | Osteophytes, meniscus, effusion/synovitis |
| E | **Lateral** | Longitudinal at the lateral femorotibial joint, 30° flexion | Osteophytes, meniscus, effusion/synovitis |
| F | **Posterior** | Transverse medial at the femoral condyle | Popliteal cyst |
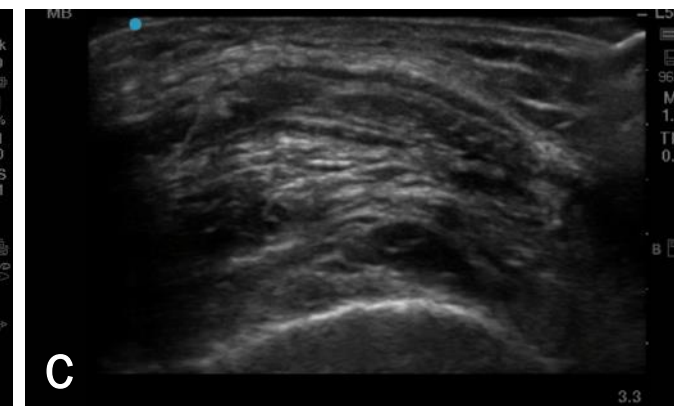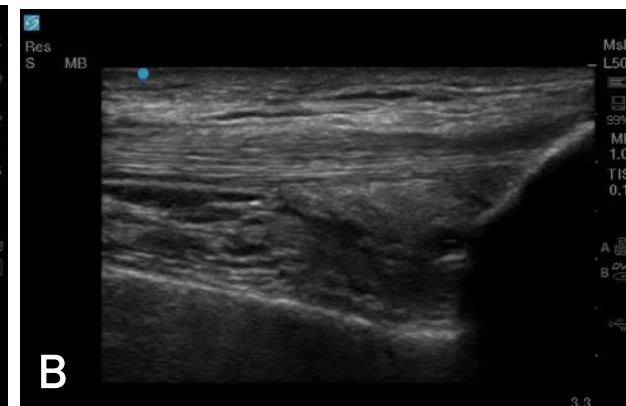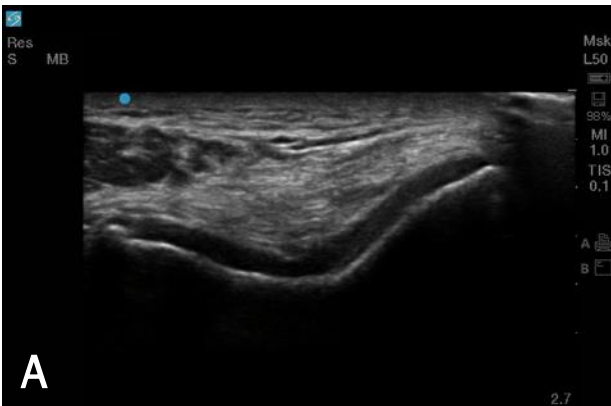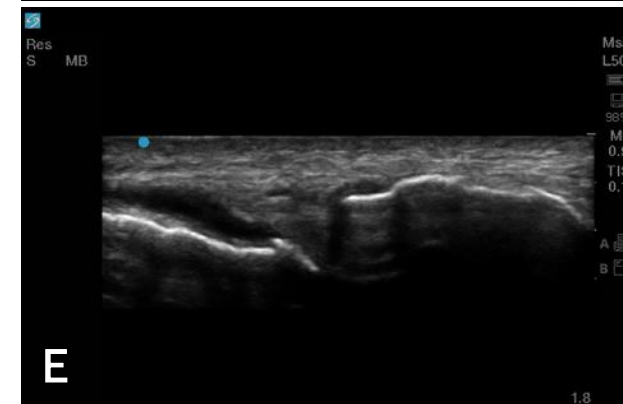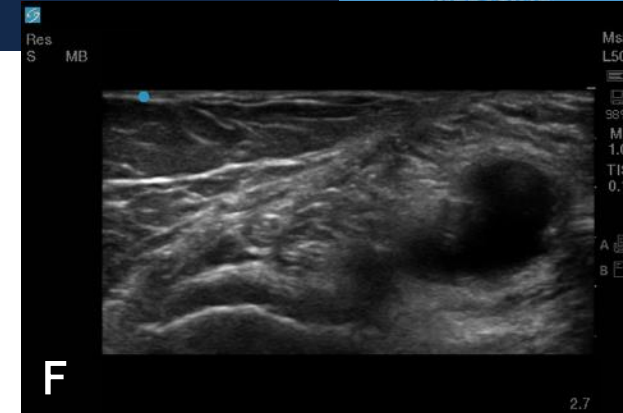
# Supplement Aim 1

- Apply state of the art DL/convolutional neural network (CNN) methods to standardized knee US images obtained in a diverse, representative cohort to determine US features that are cross-sectionally associated with a) knee pain and b) radiographic features of knee OA

# Initial approach

- Combine images into one composite per knee

- Pretrained models: VGGNet, ResNet, etc.
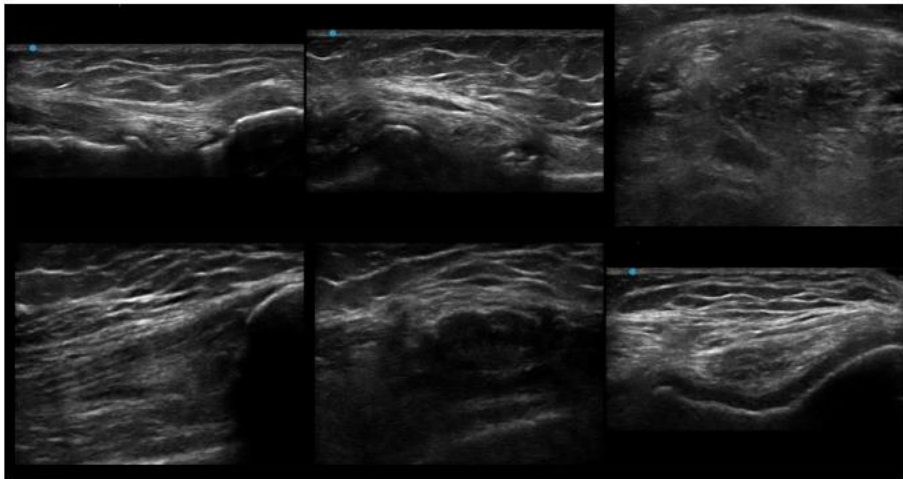


- Result: overfitting
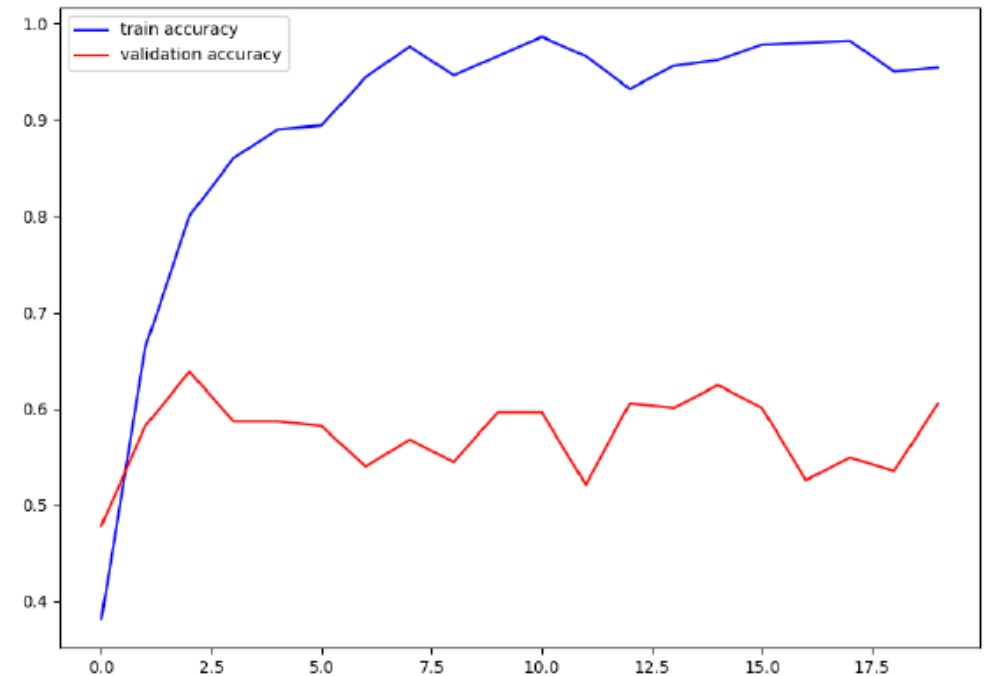
# Data augmentation with pretrained models

- Generative adversarial networks to increase number of (pseudo) images
- Build models for images separately
- Horizontal flips, random sharpness, denoising
- Perspective transformation

- Result: inadequate accuracy



Figure 4: The affine transformation of the images of Fig.3. Points A, B, and C on the checkerboard got mapped to the corresponding arbitrary points marked with red circles. Image by the author.
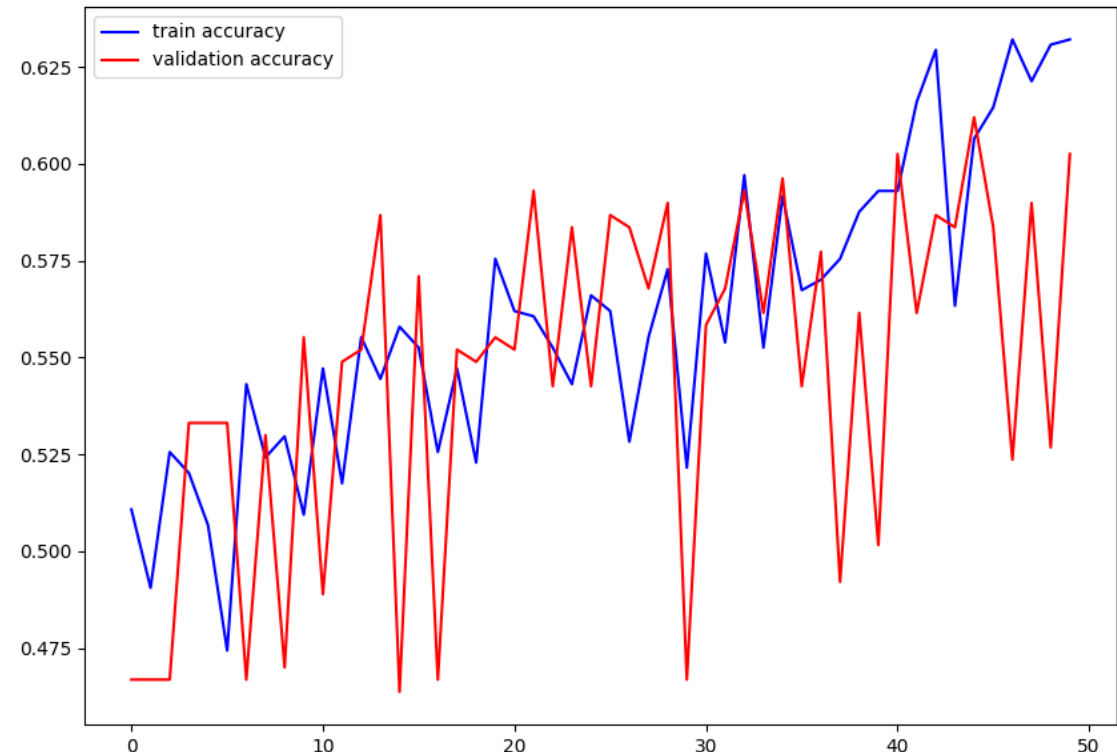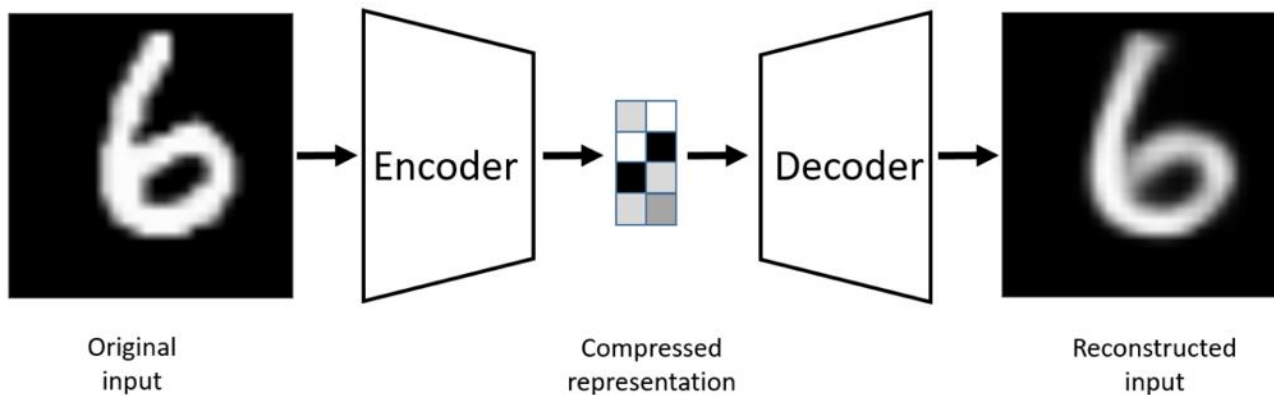
# Autoencoders

- Unsupervised

- Useful for feature extraction and dimension reduction

- Encoder compresses data; Decoder reconstructs data



Original input     Compressed representation     Reconstructed input

Layers

- 3-4 convolutional

- Batch normalization

- Maxpooling

- Rectifier Linear Unit

- Fully connected

# Autoencoders

- Results: training loss and validation loss, unable to capture image features related to pain for classification purposes
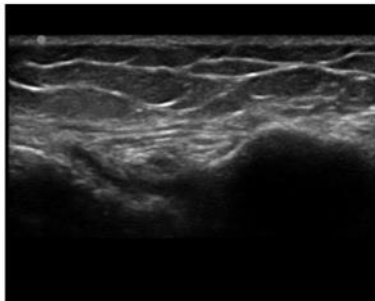- Need for more images (collection ongoing)
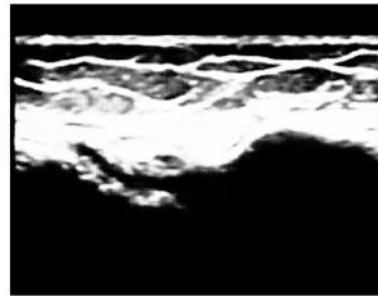
Painful knee
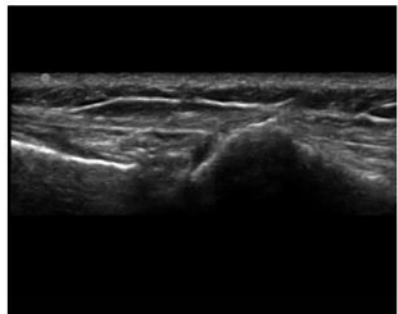


Figure 1: Original
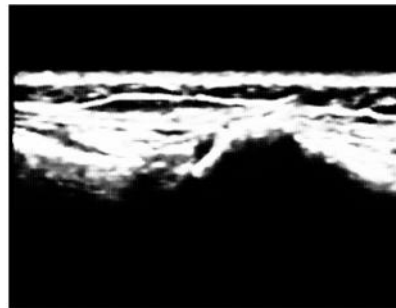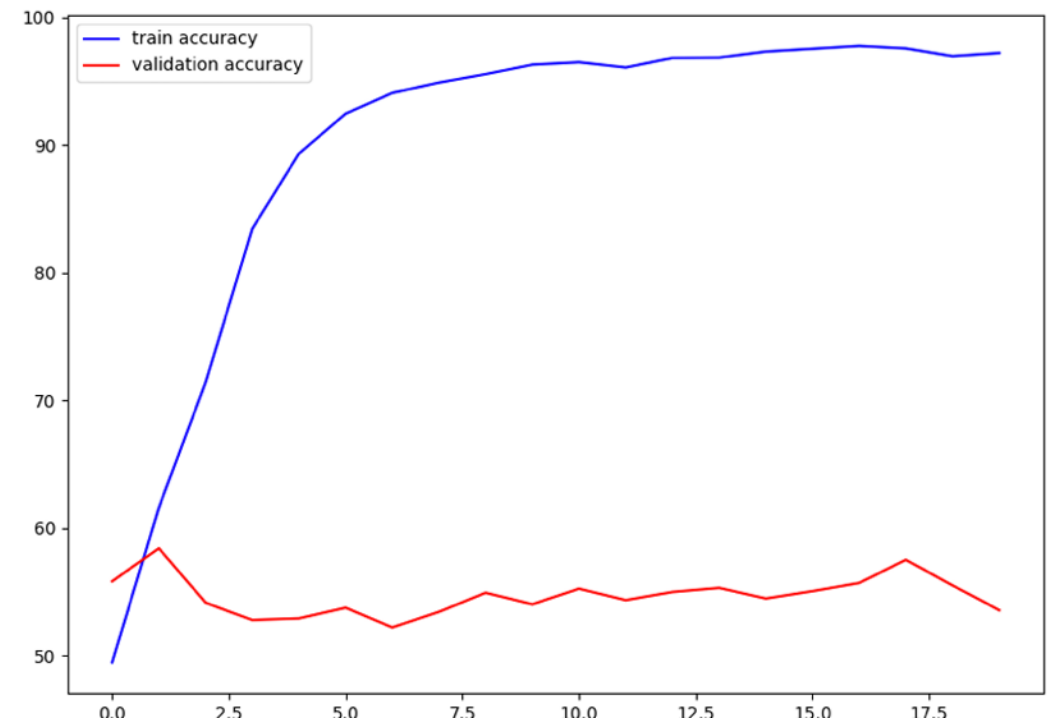


Figure 2: Reconstructed image



Figure 3: Original



Non-painful knee

Figure 4: Reconstructed image

# Supplement Aim 2

- Optimize these US images, in direct response to NOT-OD-22-067, to be ready for AI/ML approaches by our group and others in the OA and US research communities, following the FAIR principles and improving the utility of the data for AI/ML approaches.

- Progress: images are continually uploaded to Dataverse, with ontology keywords and descriptive metadata as well as an instructional Jupyter notebook (next slide)

**The Johnston County Health Study**     More about the study can be found at https://jocohs.unc.edu/

(University of North Carolina at Chapel Hill)

Harvard Dataverse > The Johnston County Health Study >

# Development of an AI/ML-ready knee ultrasound dataset in a population-based cohort

Version 5.1

Nelson, Amanda, 2023, "Development of an AI/ML-ready knee ultrasound dataset in a population-based cohort", https://doi.org/10.7910/DVN/SKP9IB, Harvard Dataverse, V5, UNF:6:e1WuNGzFGakEGlWSMnJalg== [fileUNF]

**Cite Dataset ▾**          Learn about Data Citation Standards.

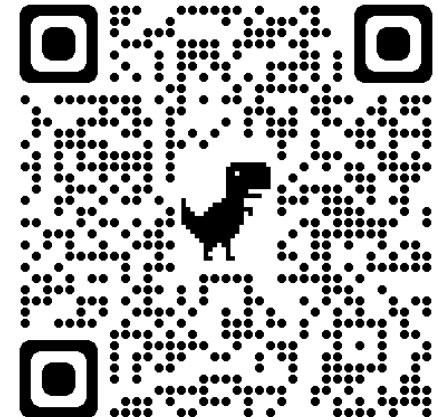**Access Dataset ▾**

Contact Owner     Share

Dataset Metrics ❓

248 Downloads ❓

# Future directions

- We continue to collect US images in the JoCoHS, increasing the size of the dataset
- We have identified some potential external datasets for either increasing image number or eventual validation
- We will continue to explore methods for augmentation and encoding to allow for feature extraction, considering other potential outcomes

# Acknowledgments

US readers: Drs. Samuels, Kohler, Bakewell, and Lin
Database manager: Patrick Gale
Statisticians: Carolina Alvarez, Todd Schwartz
US technologist: Sue Guin

Initial support of the US project (SOAR) from the RRF

JoCoHS Project manager: Tessa Walker
JoCoHS co-PIs: Drs. Nelson and Golightly
Deep learning: Drs. Siyeon Kim and Steve Marron
TARC: Drs. Jonas, Callahan, and Loeser

Initial support of JoCoHS from the CDC

Subsequent support of JoCoHS from UNC and TARC

R01 support of the US project (R01) from NIAMS

Questions? Please email Amanda Nelson, MD MSCR
aenelson@med.unc.edu