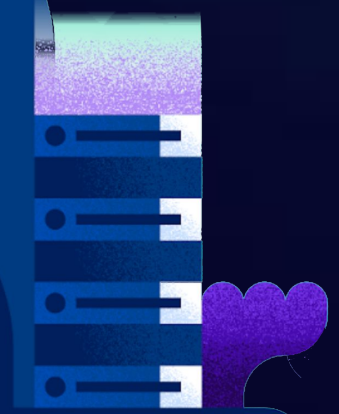
DigitalOcean

# Training Multi-Modal ML Classification Models

## For Real-Time Detection of Debilitating Disease



SCALE

# Nikki-Rae Alkema, PT, DPT

- Doctor of Physical Therapy
- Practicing in Ortho/Pelvic Health
- Movement is Medicine!
- Special interests: Biomechanics, Technology in Healthcare

@nikkidashrae

DigitalOcean

# David vonThenen

- Are you Human or an AI?
- I want 5 Kubernetes
- Virtual Machines are Real
- Cloudy, cloudy, cloudy...
- There is storage for that!

@davidvonthenen

# Agenda

- Medical Case Study for ML
  - Introduce a Disease
  - Discuss Use for AI in Clinical Practice
- Video Classification Model + Demo
- Audio Classification Model + Demo
- Q&A

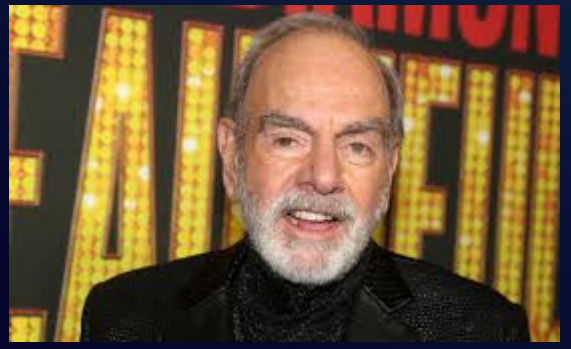
# What's the Common Thread?

Michael J. Fox

Alan Alda

Mohammed Ali

Ozzy Osbourne



Neil Diamond

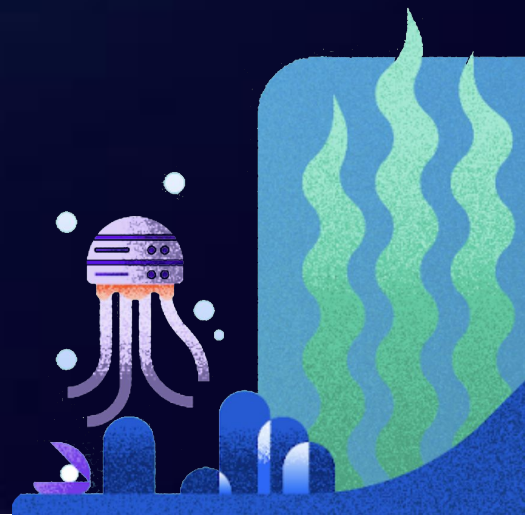
Richard Lewis

Janet Reno

Brian Grant

# Clinical Case Study:

## 74-year old male with R shoulder pain after falling



# 74M: R Shoulder Pain After Falling

- Reason for referral:
  - Pain
  - Difficulty with reaching, lifting, ADLs
- Personal factors:
  - Balance issues
  - Caretaker for his wife
- Clinical observations:
  - Using walker, shuffling steps, soft voice, tremor

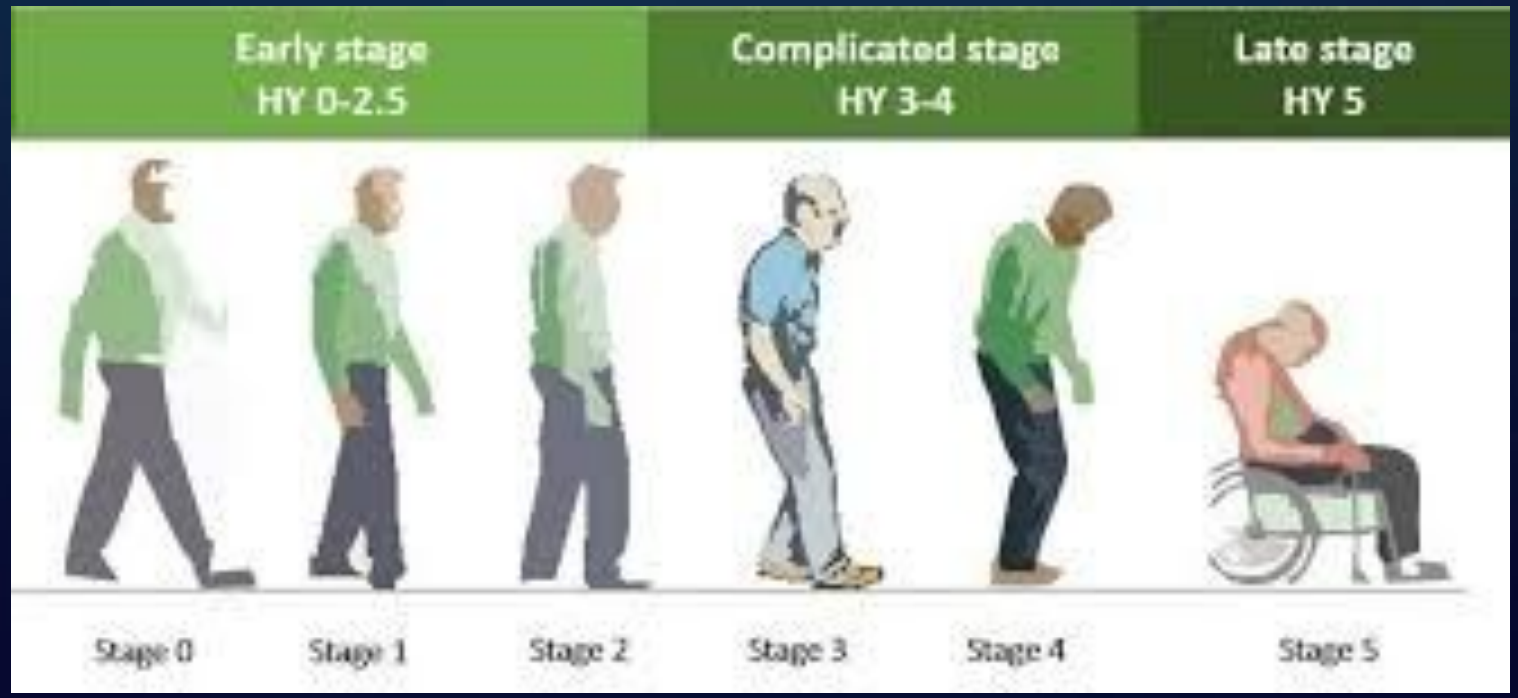
# More Than Meets the Eye?

- <u>He saw ONE problem:</u>
  - “My shoulder hurts.”
- <u>I saw TWO:</u>
  - Mild rotator cuff tear
  - Balance issues*


*Cause of fall → ***<u>undiagnosed</u>*** Parkinson’s Disease

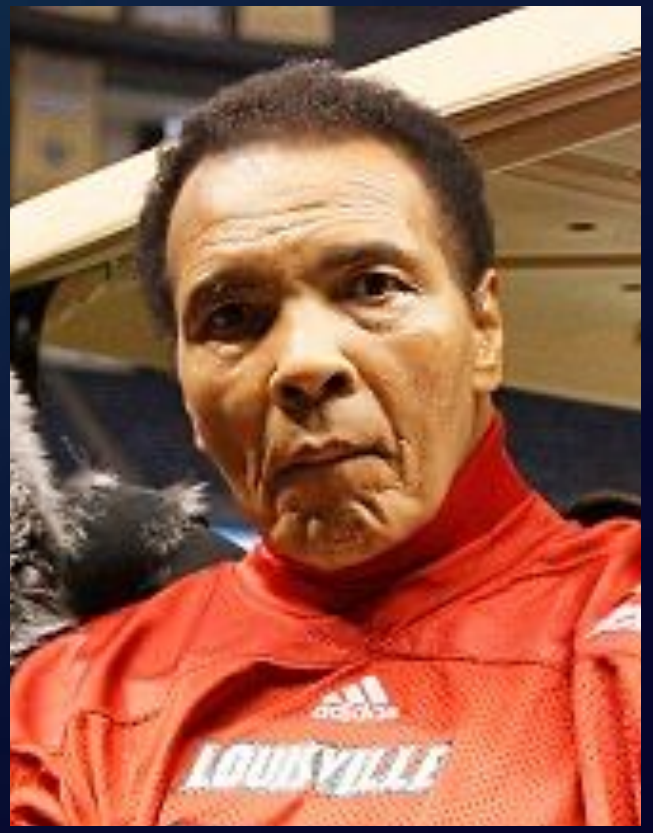
# What is Parkinson's Disease? [1,2]

- Progressive, neurologic movement disorder with no cure



# PD – If You Know, You KNOW [1,2]

- PD affects movement, making it very recognizable
  - Slow, small, rigid movement
  - Shaking or tremors
  - Postural instability and forward flexion
  - “Masked” or flat affect
  - Quiet, slurred speech
- Biomarker testing can support (but not replace) clinical diagnosis




Mohammed Ali

# A PT and an AI Engineer Walk Into a Bar...

- *A (not so) hypothetical discussion began:*

  Given examples of normal vs. abnormal human movement...

  → *Can AI tell the difference?*
  → *If so, how well?*

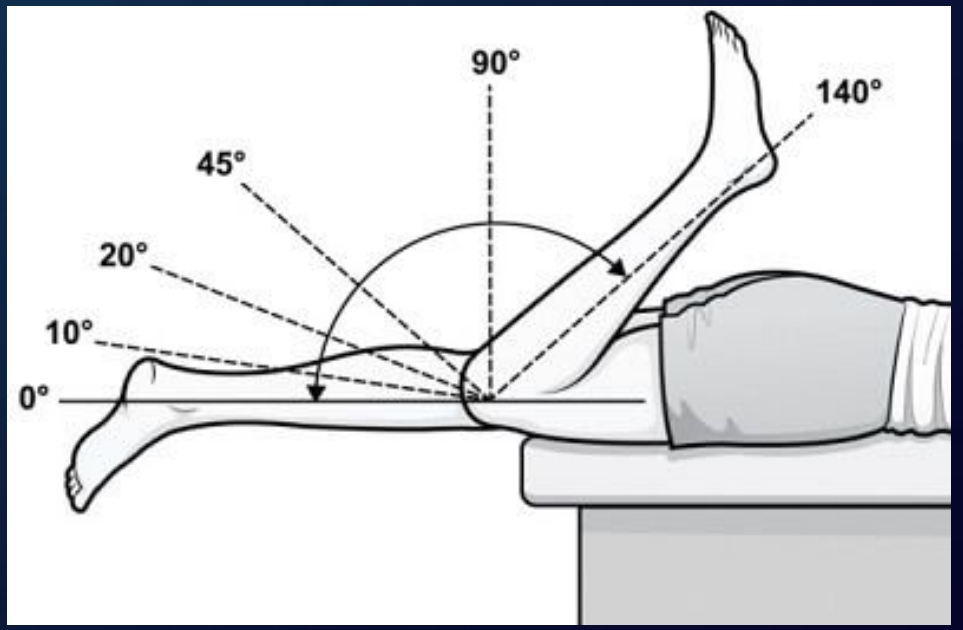
# AI: What's PD Got to Do With It?

✅ AI thrives on pattern recognition

✅ People with Parkinson's demonstrate abnormal yet predictable movement patterns

# Traditional Movement Analysis

- Systematic observation and classifcation of biomechanical characteristics of human movement and posture
  - PTs study normal to know abnormal



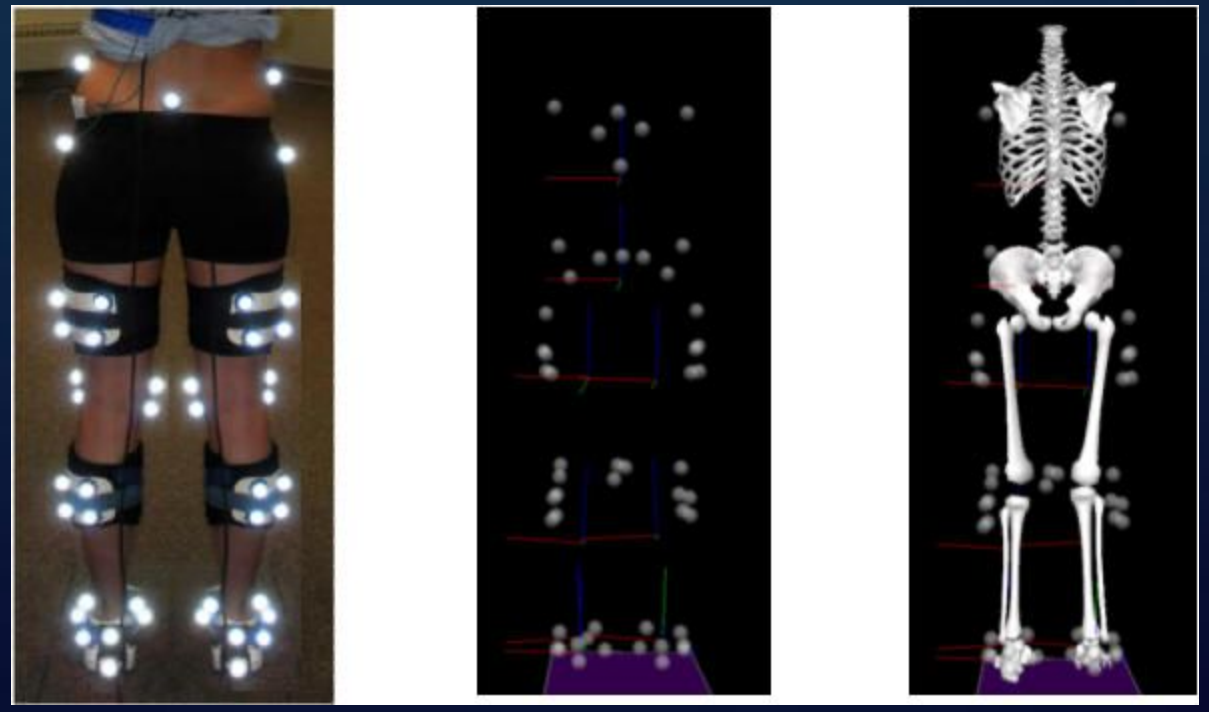

- IT'S PHYSICS: KINETICS + KINEMATICS
  - Joint angles, velocity, fluidity, power, efficiency, etc.

# Movement Analysis Labs

High tech...

OR

...old school?



University of Wisconsin, Lacrosse DPT Program

Both?

SCALE

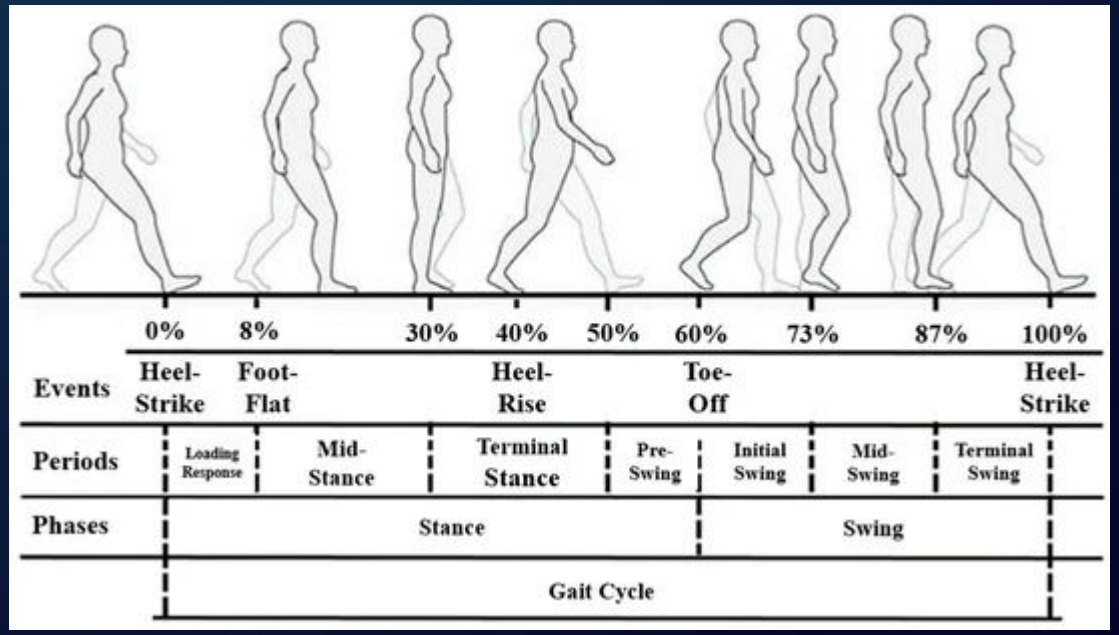
DigitalOcean

# Gait Analysis

## Effects of Parkinson's Disease on Walking

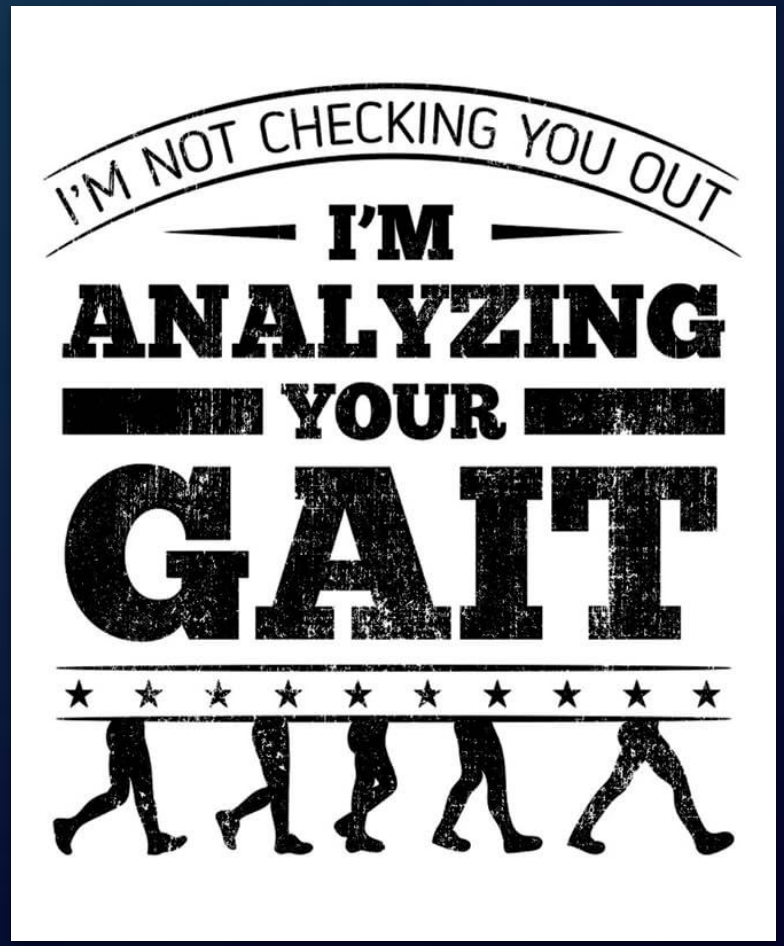
# What is Gait?

- What is gait?
  - An individual's unique pattern of walking
- Why gait?
  - The gait cycle is heavily studied and analyzed
  - Parkinson's gait is highly recognizable



# Relevance of Gait [3,4]

- Gait speed and quality tell me about your…
  - Mobility
  - Independence
  - Fall-risk
- Gait speed: the sixth vital sign
  - Predictive of mortality



# Normal Gait

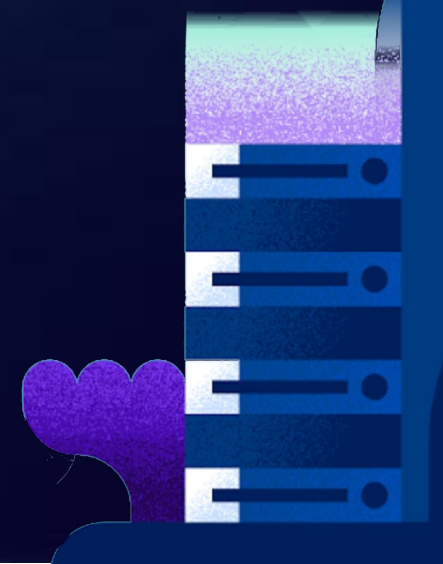
- Relatively symmetric
- Vertical in alignment
- Fluid
- Biomechanics within established norms
  - Speed, cadence, step length, joint angles

# Parkinson's Gait [5]

- Hypokinetic
  - Small step length
  - Reduced arm swing
- Bradykinetic
  - Slow progression
- Unstable
  - Non-fluid cadence
    - Shuffling, freezing
  - Hand tremors
- Rigid
  - Flexed posture

# Gait Classification Model

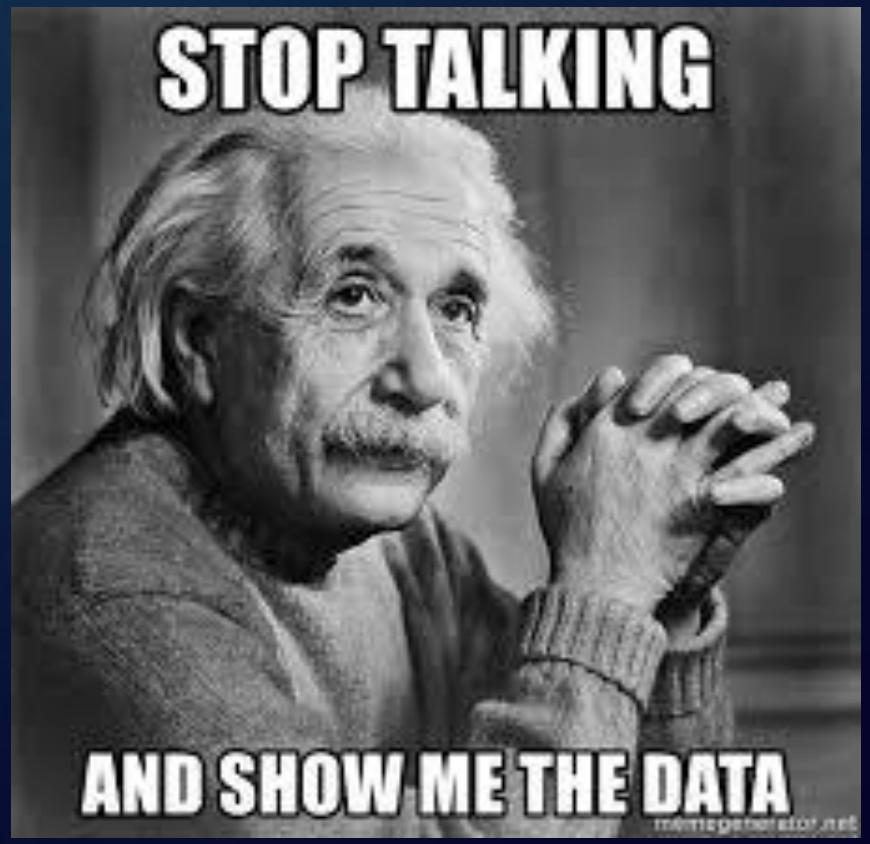
## How to Build a Machine Learning Model for Video



# Show Me the Data!?!

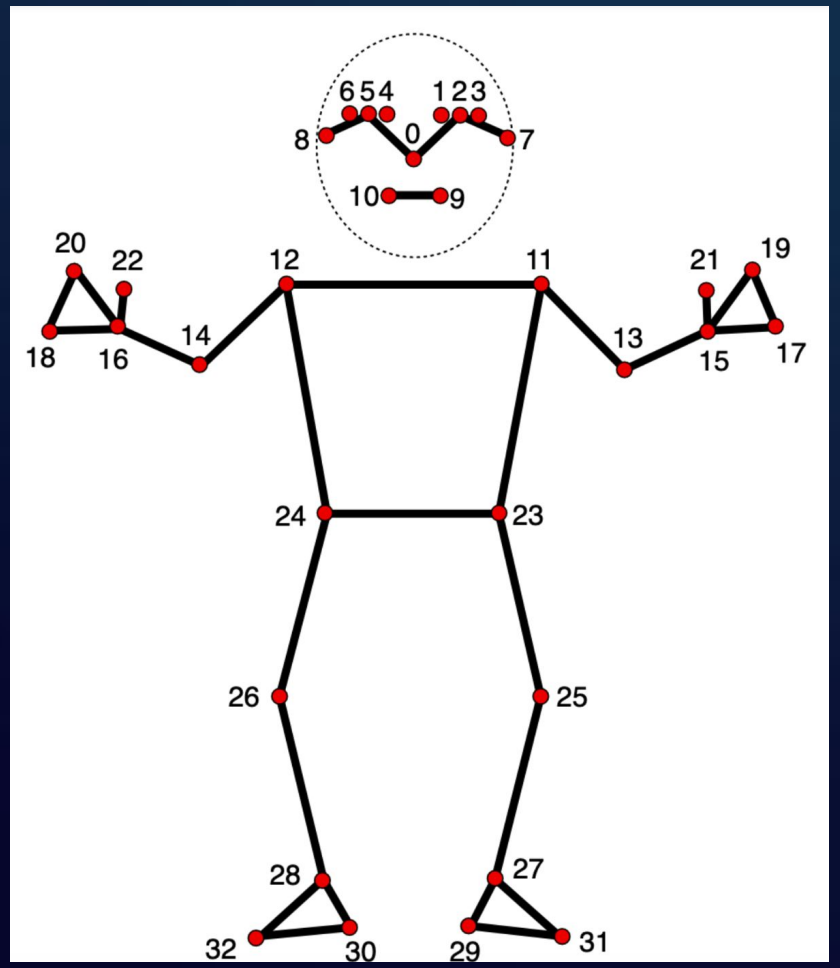
## I Want to Build a Model, Where Do I Get the Data?

- You Have Access to that Data, If...
  - Work in Medical Research
  - Work at a Medical Institution
  - Data Broker – Google, Meta, etc
- That Isn't Me, Now What?
  - Look for Public Datasets
    - Kaggle, AcademicTorrents, etc
  - Get Creative! For This Project...
    - YouTube, Instagram, TikTok, etc



# Convert Video to Data

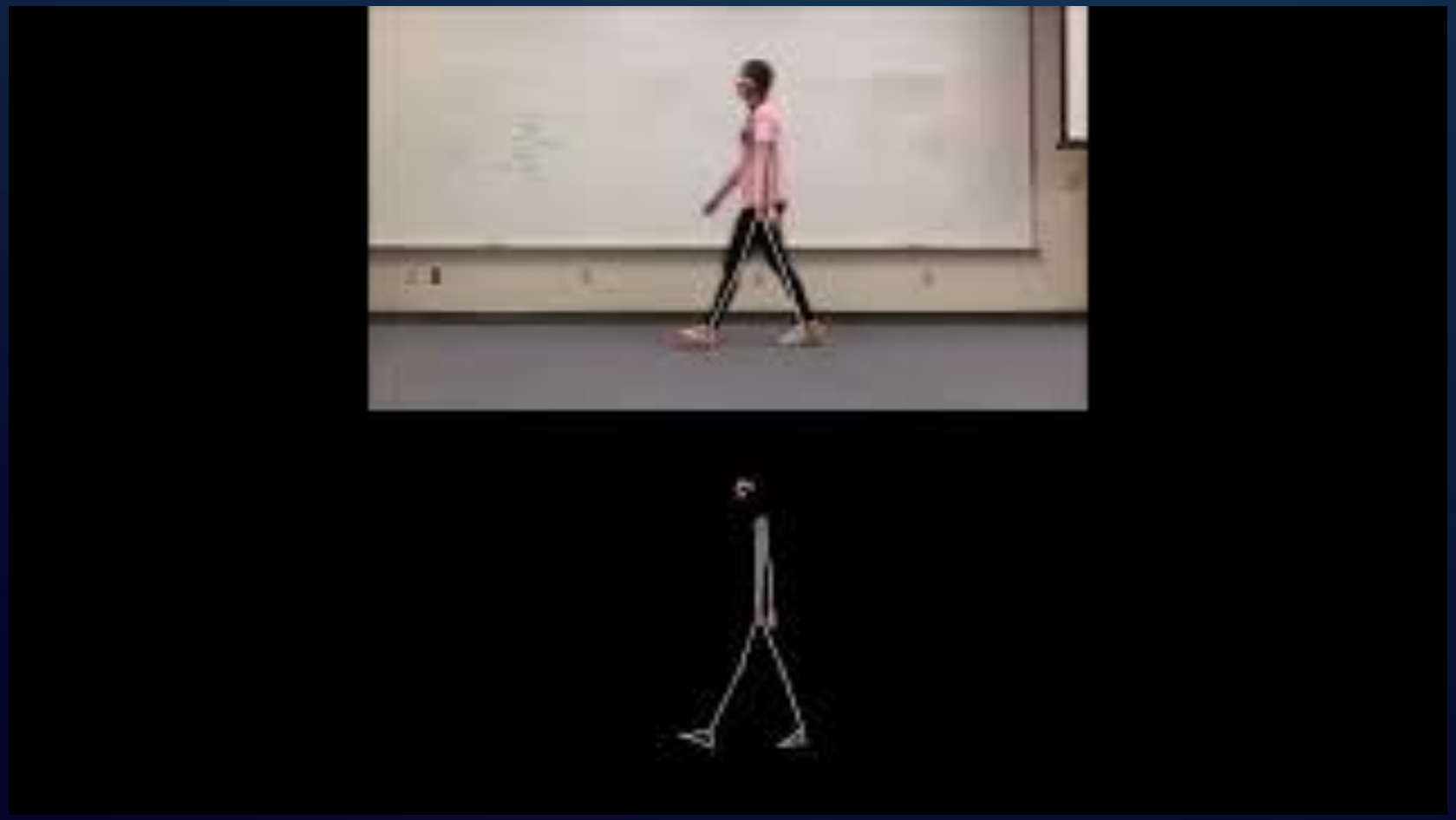
## Google AI Edge: MediaPipe Pose Landmarker




```
0 - nose
1 - left eye (inner)
2 - left eye
3 - left eye (outer)
4 - right eye (inner)
5 - right eye
6 - right eye (outer)
7 - left ear
8 - right ear
9 - mouth (left)
10 - mouth (right)
11 - left shoulder
12 - right shoulder
13 - left elbow
14 - right elbow
15 - left wrist
16 - right wrist
```

```
17 - left pinky
18 - right pinky
19 - left index
20 - right index
21 - left thumb
22 - right thumb
23 - left hip
24 - right hip
25 - left knee
26 - right knee
27 - left ankle
28 - right ankle
29 - left heel
30 - right heel
31 - left foot index
32 - right foot index
```

# Video to Data Demo



```
PoseLandmarkerResult:
  Landmarks:
    Landmark #0:
      x            : 0.638852
      y            : 0.671197
      z            : 0.129959
      visibility   : 0.9999997615814209
      presence     : 0.9999984502792358
    Landmark #1:
      x            : 0.634599
      y            : 0.536441
      z            : -0.06984
      visibility   : 0.999909
      presence     : 0.999958
    ... (33 landmarks per pose)
  WorldLandmarks:
    Landmark #0:
      x            : 0.067485
      y            : 0.031084
      z            : 0.055223
      visibility   : 0.9999997615814209
      presence     : 0.9999984502792358
    Landmark #1:
      x            : 0.063209
      y            : -0.00382
      z            : 0.020920
      visibility   : 0.999976
      presence     : 0.999998
    ... (33 world landmarks per pose)
```

# Codify the Characteristics

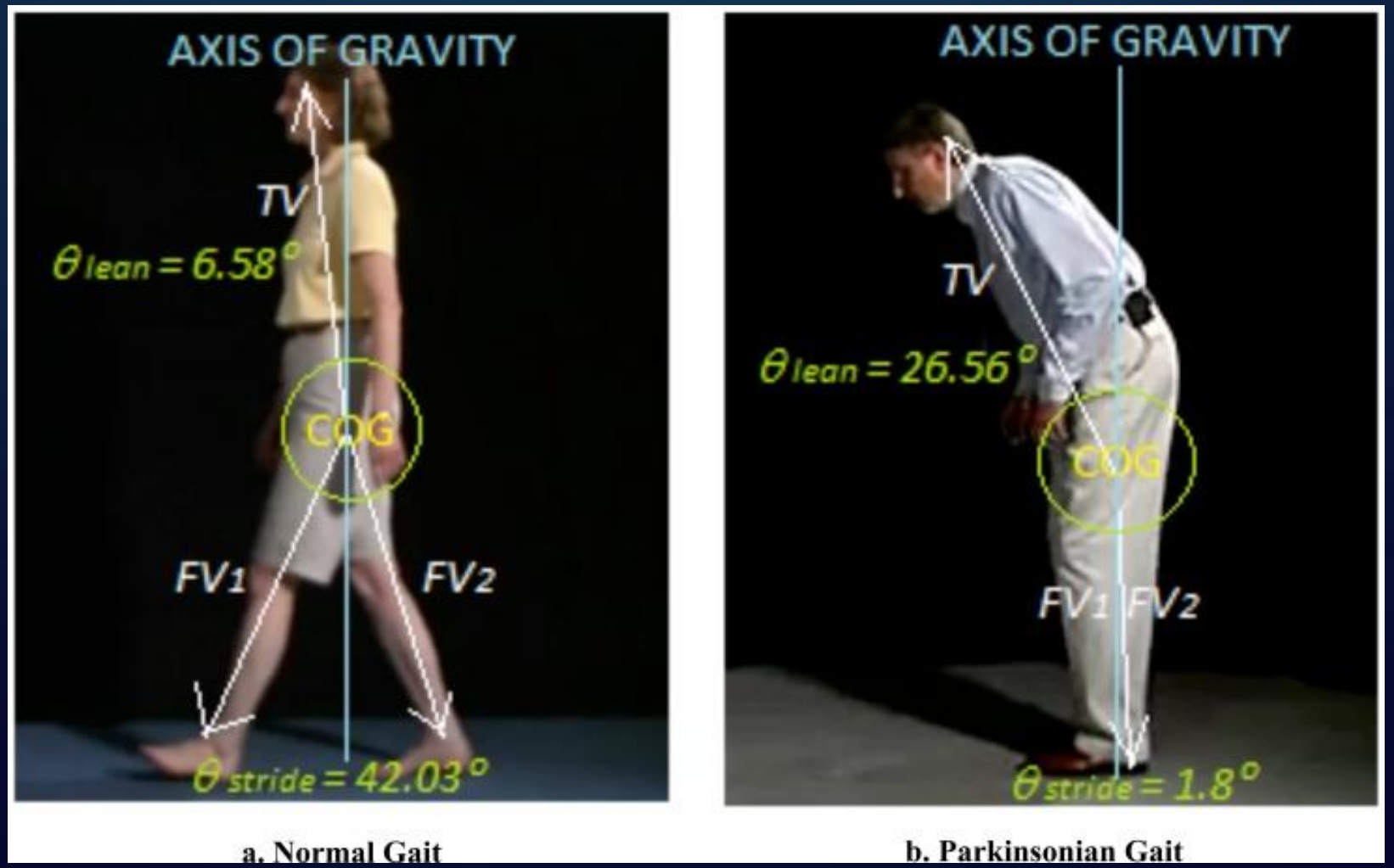
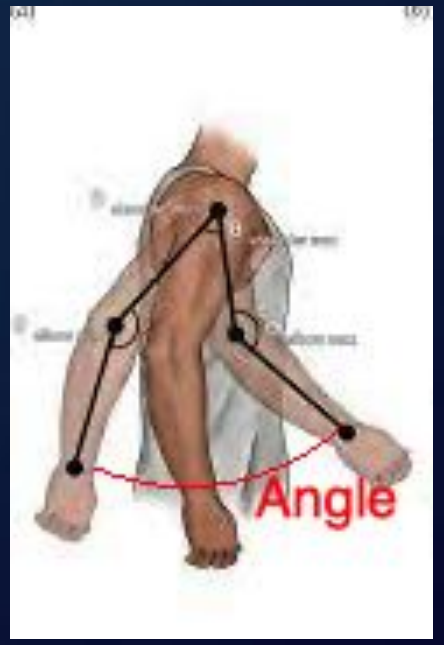
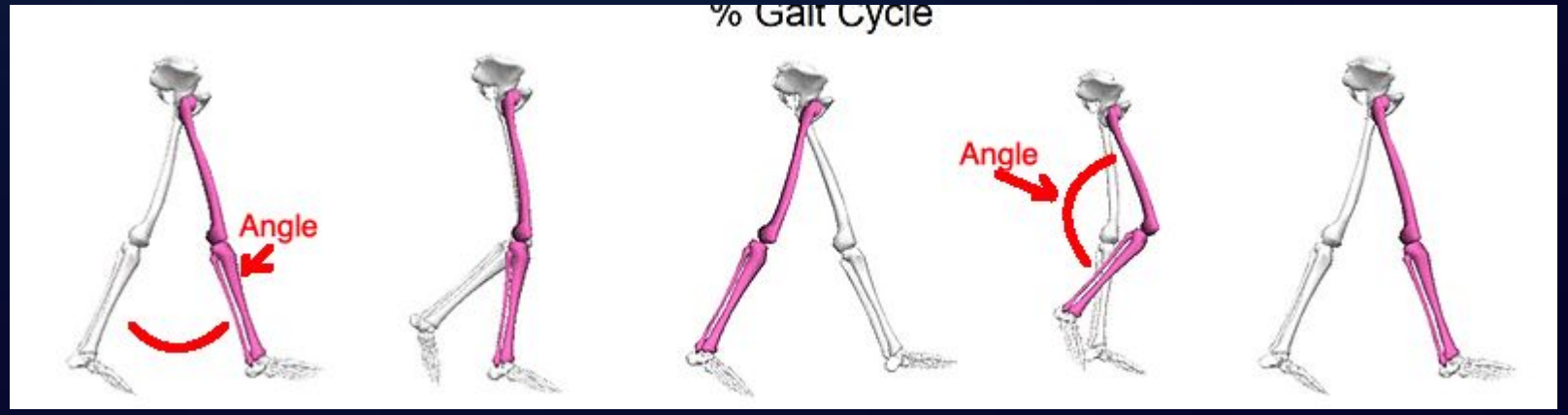



Image Credit:
Motion Cue Analysis for Parkinsonian Gait Recognition
Taha Khan, Jerker Westin, Mark Dougherty

# Feature Engineering

- Acceleration/Velocity of Landmarks
  - Reduce Movement, Rigidity, etc
- Angles:
  - Elbow-Shoulder-Hip
  - Shoulder-Elbow-Wrist
  - Hip-Knee-Ankle
  - Knee-Hip-Knee
  - Etc, Etc
- Step-Length

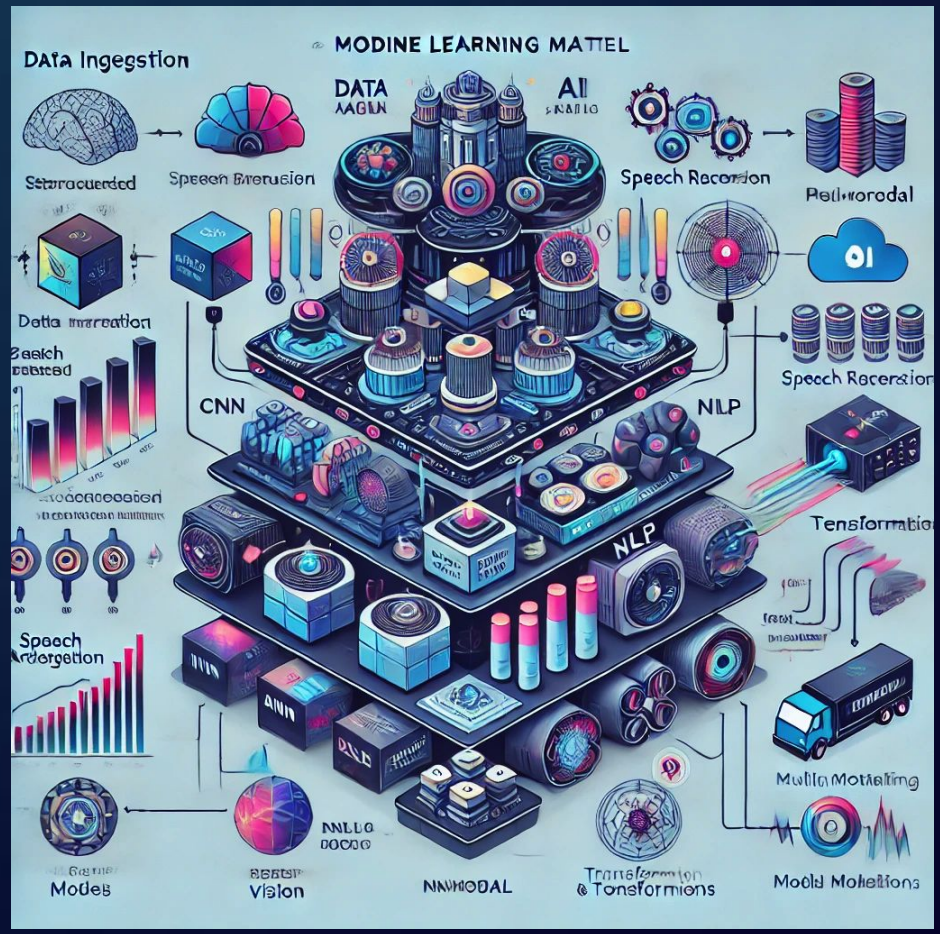



# Data Processing Pipeline

- Video –> Frame Features to CSV
  - Using Media Pipe Pose Landmarker
  - Get 3D (x, y, z) Coordinates
- CSV File <–> Seq. Modeling
  - Movement Data Across Frames
  - Features Captured:
    - Velocity, Acceleration, Angle Measurements, etc
  - LSTM to Capture Temporal Dependencies
- Tuning –> Final Model

# Model Architecture

- Long-Short Term Memory Network
  - Think Time-Series Data
  - KNN Imputation Fills in Gaps
  - Introduce Some Noise
- Training Steps:
  - Hyperparameter Tuning
    - Saved Parameter Grid: hidden size, number of layers, etc
  - Systematically Evaluation
- Test Set Accuracy: 94.35%



DigitalOcean

# Normal – Inference

DigitalOcean

# Parkinson's – Inference

CSV FILE

DigitalOcean

# Demo

https://youtu.be/yz8hNF1Czos

DigitalOcean

# So video worked...

# ...what about audio?

# Speech Analysis

## Effects of Parkinson's Disease on Communication

# What is speech? [6,7]

- Interaction of multiple body systems to produce verbal communication through language
  - Voice
    - Sounds created as air passes through the vocal chords
  - Articulation
    - Motor process of how a sound is formed in the mouth to become words

# Relevance of Speech [8]

- Identity
- Social engagement
- Performance in activities of daily living
- Speech changes can result in withdrawal, isolation, shame, depression

**Non-verbal communication also affected in PD*

# Normal Speech [7,9]

- Clear, fluent, accurate articulation
- Appropriate prosody
  - The “music of language:” stress, rate, rhythm, loudness, intonation

Alan Alda
SAG Awards, 2018
*Dx: Positive*
*Speech: Asymptomatic*

# Parkinson's Speech [8]

- Hypokinetic dysarthria: changes in voice and articulation relating to PD
  - Monotone
  - Monoloud and quiet
  - Hoarse or breathy
  - Rate abnormalities
  - Imprecision, slurring
  + Reduced facial emoting and non-verbal gestures

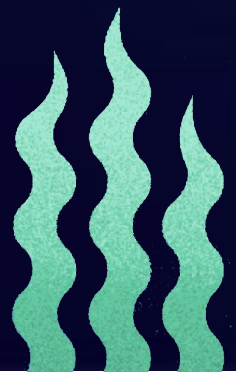
Alan Alda

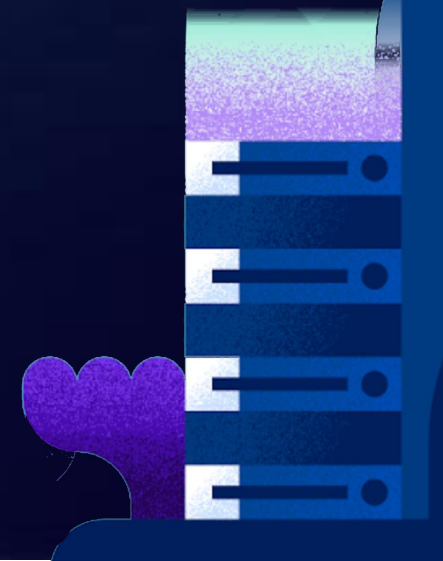
**Everything Happens Podcast with Kate Bowler , 2024**

*Dx: Positive, Speech: Symptomatic*

# Voice Classification Model

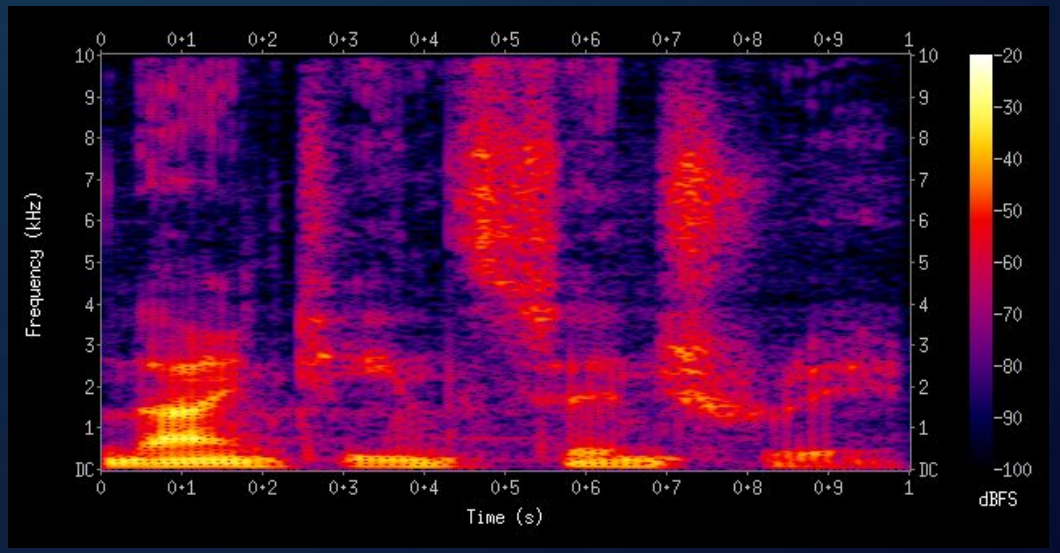
## How to Build a Machine Learning Model for Intonation





# ML Audio Classification?

- MANY Doing This!
- Spectrogram Comparison
  - Visual Representation
  - Similarities Visually
- Example Projects:
  - Cats vs Dogs
  - Environmental Sounds
  - Gunshot Recognition



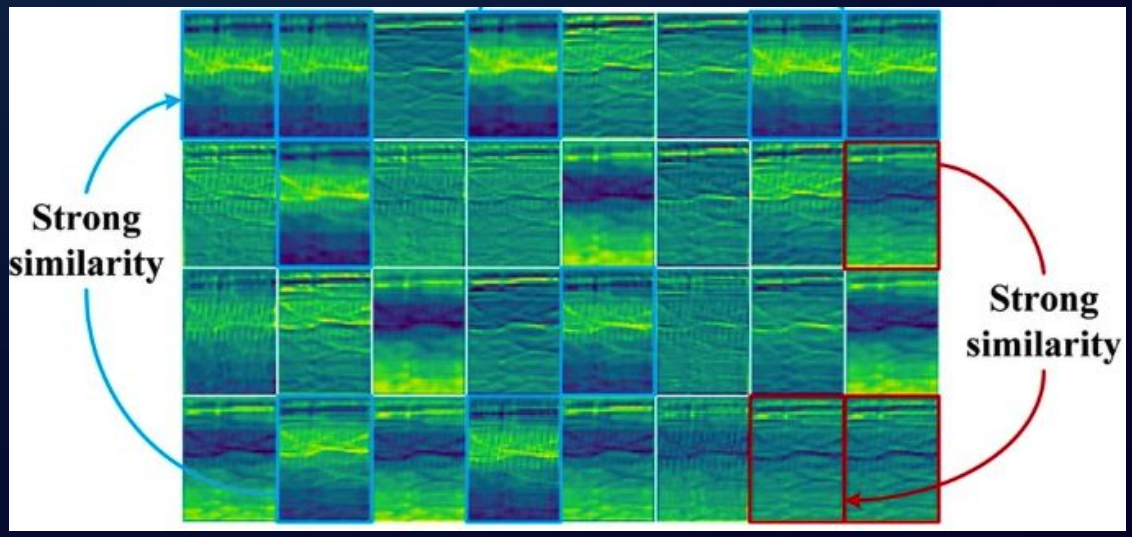




Image Credit:
Fast environmental sound classification based on resource adaptive convolutional neural network DOI:10.1038/s41598-022-10382-x

# Obtaining the Dataset

- Public Datasets:
  - NIH: Mobile Device Voice Recordings at King's College London (MDVR-KCL)
    - Imaging: github.com/CanBul/Parkinson-Disease-Detection
  - SJTU-YONGFU-RESEARCH-GRP
    - Imaging: Enhancing Speech Recognition
- In Addition, Self Curated Dataset from YouTube:
  - Interviews
  - Podcasts

YouTube

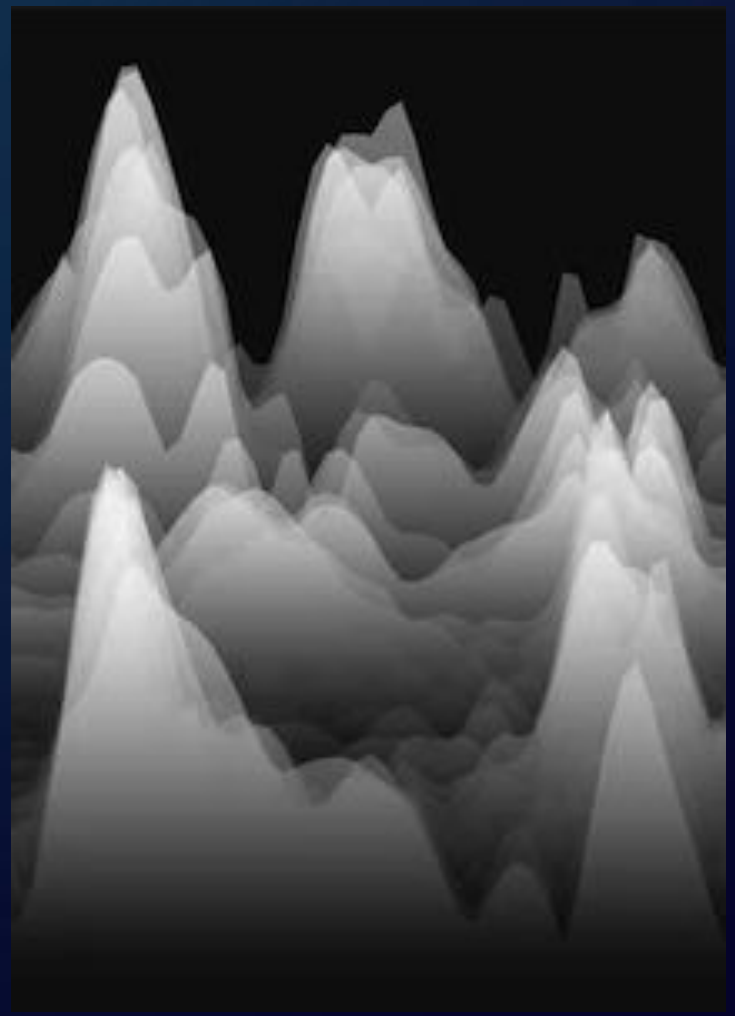
# Unique Dataset

For the Dataset Download From YouTube:

- Obtained Many Hours of Audio Data
  - Pre/Post Diagnosis of the Same Person
- Where Did We Find This Data?
  - Remember That First Slide...
  - Celebrity Interviews, Podcasts, etc
    - Tracking Sheet
- Example: Richard Lewis (2019)
  - Pre Samples: 2000–2015
  - Post Samples: 2023

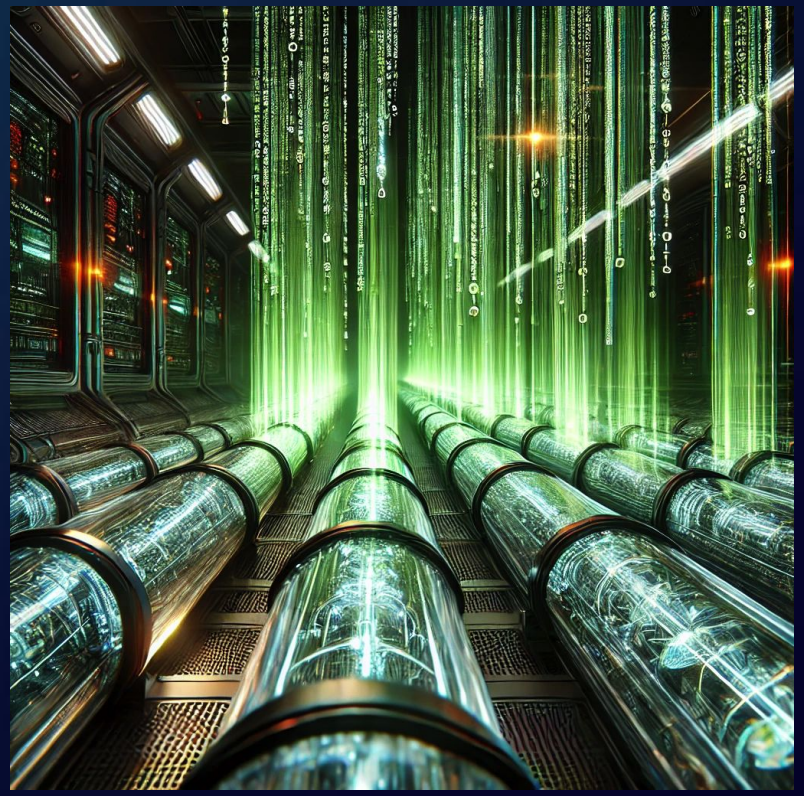
# Feature Engineering

- Extracted Prosodic (Audio) Measurements:
  - Energy (RMS)
    - Root Mean Sq. = Projection, Loudness, etc
  - Formants Freq.
    - Formants = Vocal Resonance Changes
    - Mean, Standard Deviation
  - Harmonics-to-Noise Ratio (HNR), etc
    - HNR = Breathiness/Roughness
  - Jitter
    - Cycle-to-Cycle Variation in Frequency
  - Shimmer
    - Cycle-to-Cycle Amplitude Variations
  - Etc



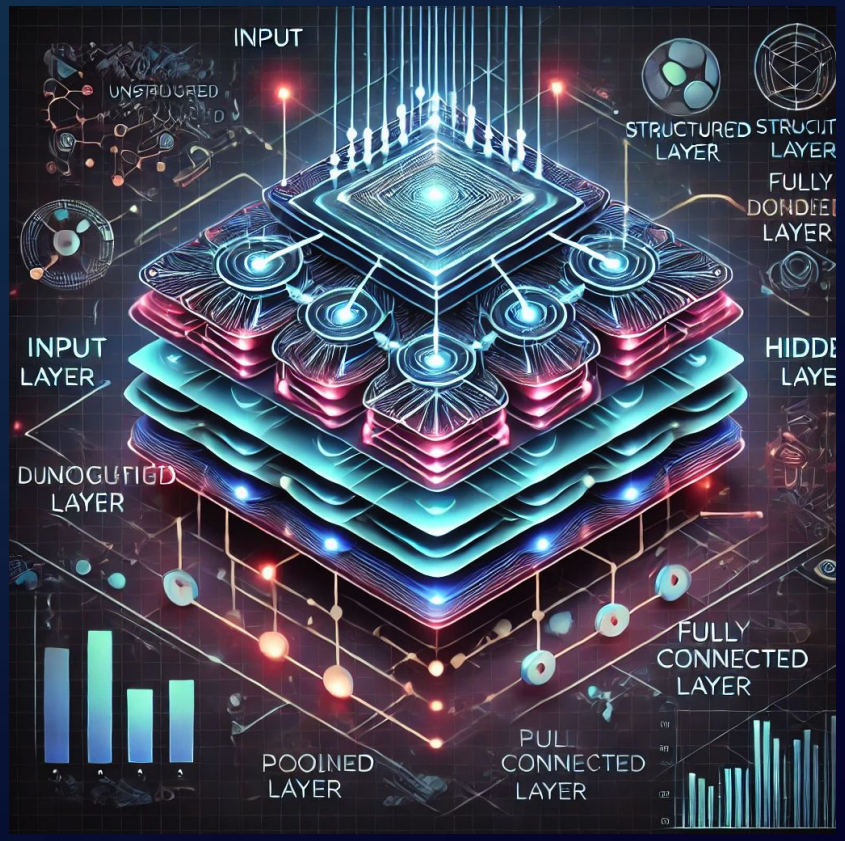
# Data Processing Pipeline

- Recordings -> Frame Features to CSV
  - Clean Noise: SJTU-YONGFU-RESEARCH-GRP
- Extracted Prosodic Measurements:
  - Pitch, Energy (RMS), Formants Freq., Harmonics-to-Noise Ratio (HNR), etc
- Key: Synchronized Audio <-> Words
  - Capture Textual Context of Audio
    - Rhythm, Intonation, Stress, etc
  - Transcription With Time/Word Align
- Tuning -> Final Model



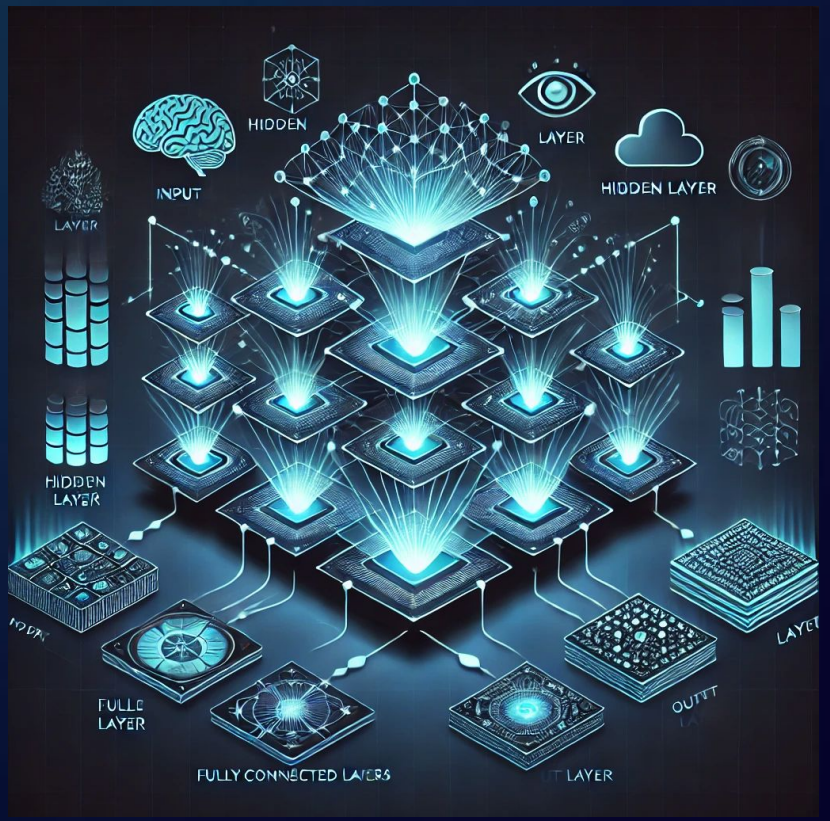
# Model Architecture

- CNN + LSTM Model
  - Convolution Neural Network (Local)
  - Long Short-Term Memory (Time-Series)
- Word/Text Embeddings
  - Word Sequence Appended to CNN
  - Each Utterance is Time Aligned
    - Captures Pacing, Intonation, etc
- Why This is Effective? Merges...
  - Short Terms Acoustics
  - Long Term Linguistic Context
  - Combine Neural Network Techniques
- Final Test Accuracy: 96.30%



# Deep Dive

- Convolutional Neural Network (CNN)
  - 2 Convolution Blocks + Pooling
  - Downsample Time ~4x
- Long Short-Term Memory (LSTM)
  - Downsampled Frames Over Time
  - Captures Continuity
- Hyperparameters:
  - Number of features: 77, CNN filters: 32, Hidden size: 128, Number of layers: 2 Number of classes: 2, Unique vocabulary: 4165
- Results:
  - Highlights Discriminative Features
  - Connects Acoustics to Speech



DigitalOcean

# 4 Years Before – Inference

DigitalOcean

# 19 Years After – Inference


CSV FILE

DigitalOcean

# Demo

https://youtu.be/WGPW_5pClPg

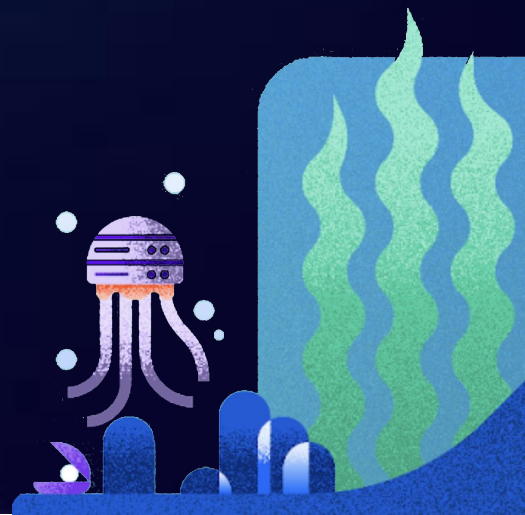
# Creating a Multi-Modal ML Model

## Fusion Approach (By Reusing the Code In This Repo):

1. Video Sub-Network
   - Use your LSTM-based video classifier code up to (but not including) the final classification layer. That is, if your final LSTM outputs a hidden vector of size H, keep that as a video embedding.
2. Audio Sub-Network
   - Likewise, from the audio CNN+LSTM, grab the final LSTM output (before the classifier) as an audio embedding of size A.
3. Fusion Mechanism
   - Concatenate the two embeddings: [video_embedding, audio_embedding] → a single vector of size H+A.
   - Pass this fused vector through a new "fusion head" — for instance, a small MLP or another LSTM that captures temporal modalities.
   - This network ends in a single classification layer outputting the probability of Parkinson's vs. Normal (or 2-class softmax).

# There Is One More Thing...

## Other Future Classification Ideas



# If Quacks Like A Duck...


# The Future is Now [10,11,12]

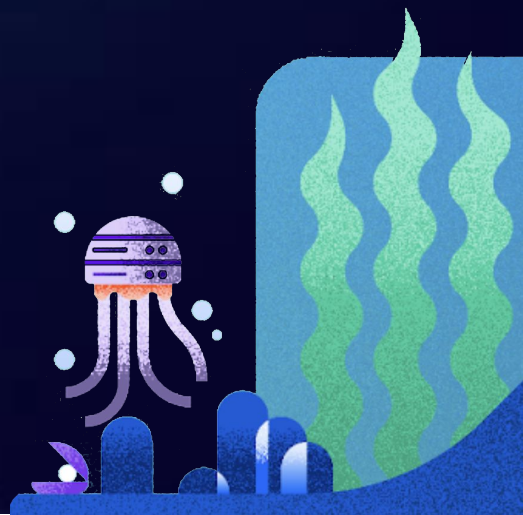
- Attitudes mixed but warming toward AI in medicine
- Chief concerns
  (–) Job security, patient privacy, accurate decision-making
- Great potential
  (+) Reduce administrative burden, improve efficiency, enhance screening

→ *Be Part of the Change...We Need You!*

SCALE

DigitalOcean

# Resources



# Medical References


1. McCormack, R. Understanding the Five Stages of Parkinsons. Parkinson's NSW. Accessed October 7, 2024. https://www.parkinsonsnsw.org.au/understanding-the-five-stages-of-parkinsons.
2. Yun, J. Movement Symptoms. Parkinson's Foundation. Accessed October 7, 2024. https://www.parkinson.org/understanding-parkinsons/movement-symptoms.
3. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. J Aging Phys Act. 2015;23(2):314–322. doi 10.1123/japa.2013-0236.
4. Fritz, Stacy PT, PhD1; Lusardi, Michelle PT,PhD2. White Paper: "Walking Speed: the Sixth Vital Sign." Journal of Geriatric Physical Therapy 32(2):46–9.
5. Moore, K. Trouble Moving or Walking. Parkinson's Foundation. Accessed October 7, 2024. https://www.parkinson.org/understanding-parkinsons/movement-symptoms/trouble-moving
6. What is the Difference Between a Speech Evaluation and an Articulation Evaluation? Child Language and Developmental Speech. Posted October 26, 2022. Accessed March 6, 2025. https://childspeechlanguage.com/what-is-the-difference-between-a-speech-evaluation-and-an-articulation-evaluation.
7. American Speech-Language-Hearing Association. What Is Speech? What Is Language? American Speech-Language-Hearing Association. https://www.asha.org/public/speech/development/speech-and-language/
8. Lansford KL, Liss JM, Caviness JN, Utianski RL. A cognitive-perceptual approach to conceptualizing speech intelligibility deficits and remediation practice in hypokinetic dysarthria. Parkinsons Dis. 2011;2011:150962. doi:10.4061/2011/150962.
9. Applebaum L, Coppola M, Goldin-Meadow S. Prosody in a communication system developed without a language model. Sign Lang Linguist. 2014;17(2):181–212. doi:10.1075/sll.17.2.02app.
10. Moldt, J.-A. et al. (2023) 'Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots', Medical Education Online, 28(1). doi: 10.1080/10872981.2023.2182659.
11. Al-Medfa MK, Al-Ansari AMS, Darwish AH, Qreeballa TA, Jahrami H. Physicians' attitudes and knowledge toward artificial intelligence in medicine: Benefits and drawbacks. Heliyon. 2023;9(4):e14744. Published 2023 Mar 23. doi:10.1016/j.heliyon.2023.e14744.
12. Appel JM. Artificial intelligence in medicine and the negative outcome penalty paradox. J Med Ethics. 2024;51(1):34–36. Published 2024 Dec 23. doi:10.1136/jme-2023-109848.

# AI/ML Resources

[CLICK HERE] for All Material Contained in this Session [CLICK HERE]

DigitalOcean Bare Metal H200 Availability
https://www.digitalocean.com/blog/now-available-bare-metal-nvidia-hgx-h200-gpus

Continue the Conversation – DigitalOcean Discord
https://discord.com/invite/digitalocean

## Code with Instructions for Gait Model:

- Part 1: Processing Videos Features Using MediaPipe
- Part 2: Building a ML Model for Video
- Part 3: Parkinson's Gait Demo

## Code with Instructions for Voice Model:

- Part 1: Processing Audio Features
- Part 2: Building a ML Model for Audio Intonation
- Part 3: Parkinson's Voice Demo

DigitalOcean

# Thank You!


Nikki-Rae Alkema, PT, DPT

@nikkidashrae

https://linktr.ee/nikkidashrae

David vonThenen

@davidvonthenen

https://linktr.ee/davidvonthenen