A Data Mining Approach for Determining Gait Abnormalities in Runners with Patellofemoral Pain Syndrome

Ross Brancati, MS

PhD Candidate

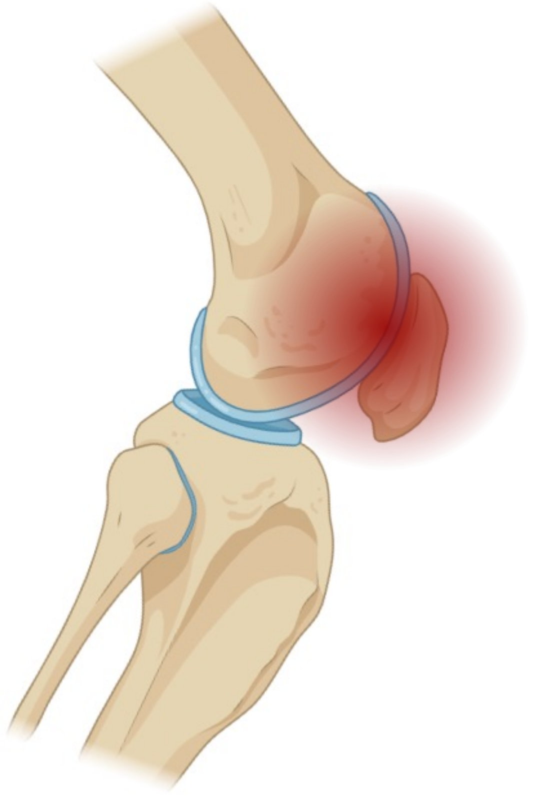
Musculoskeletal & Orthopedic Biomechanics Lab

November 28, 2022

Seminar Presentation

Patellofemoral Pain Syndrome

- Patellofemoral pain (PFP) is a common musculoskeletal pain disorder, often presenting as anterior knee pain
- Experienced by many recreational runners
- High prevalence (25%) and recurrence (up to 90%)
- Prolonged runs typically aggravate the pain
- Pathomechanical model highlights distal and proximal pathways to elevated patellofemoral joint stress



Crossley et al. 2016, Bazett-Jones et al. 2013, Taunton et al. 2002, Powers et al. 2017, Willy et al. 2019

Quick Anatomy Review

Image from Bikesreav

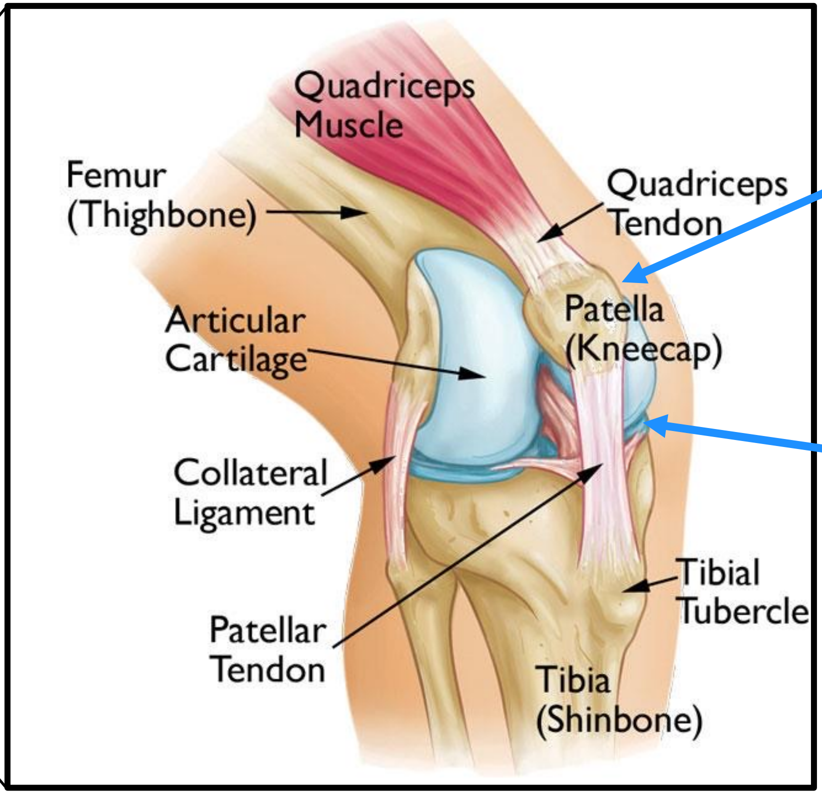


Image from Ortho Info

Patellofemoral Joint

Tibiofemoral Joint

Proposed Pain Mechanisms

Elevated patellofemoral joint stress → Adaptations to patellofemoral cartilage

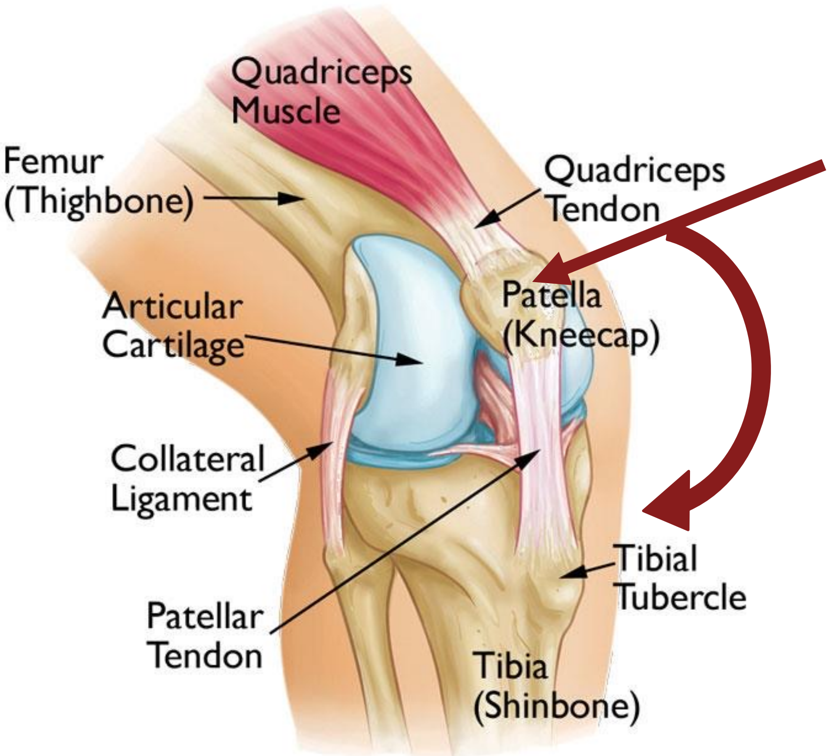


Image from Ortho Info

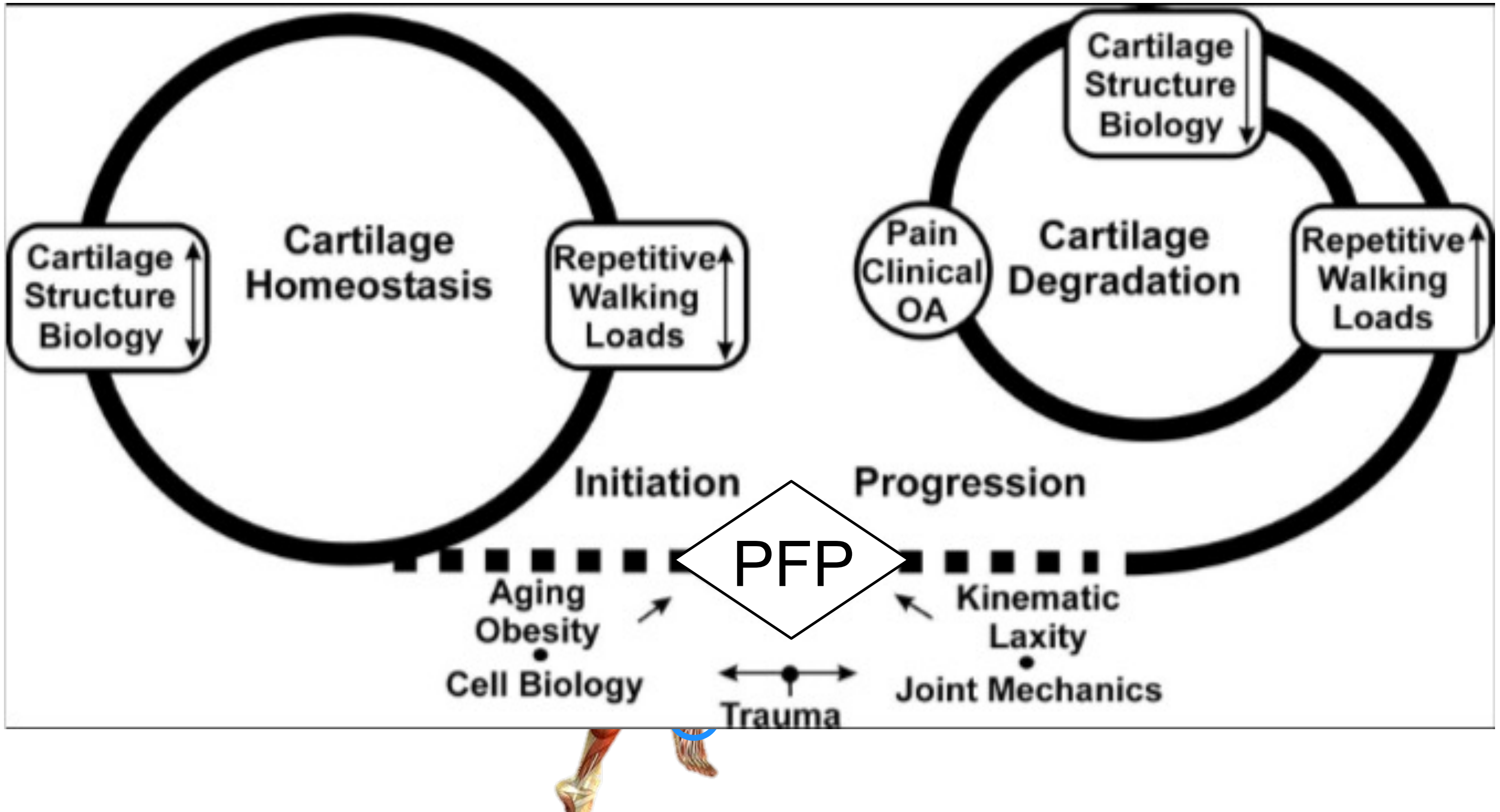
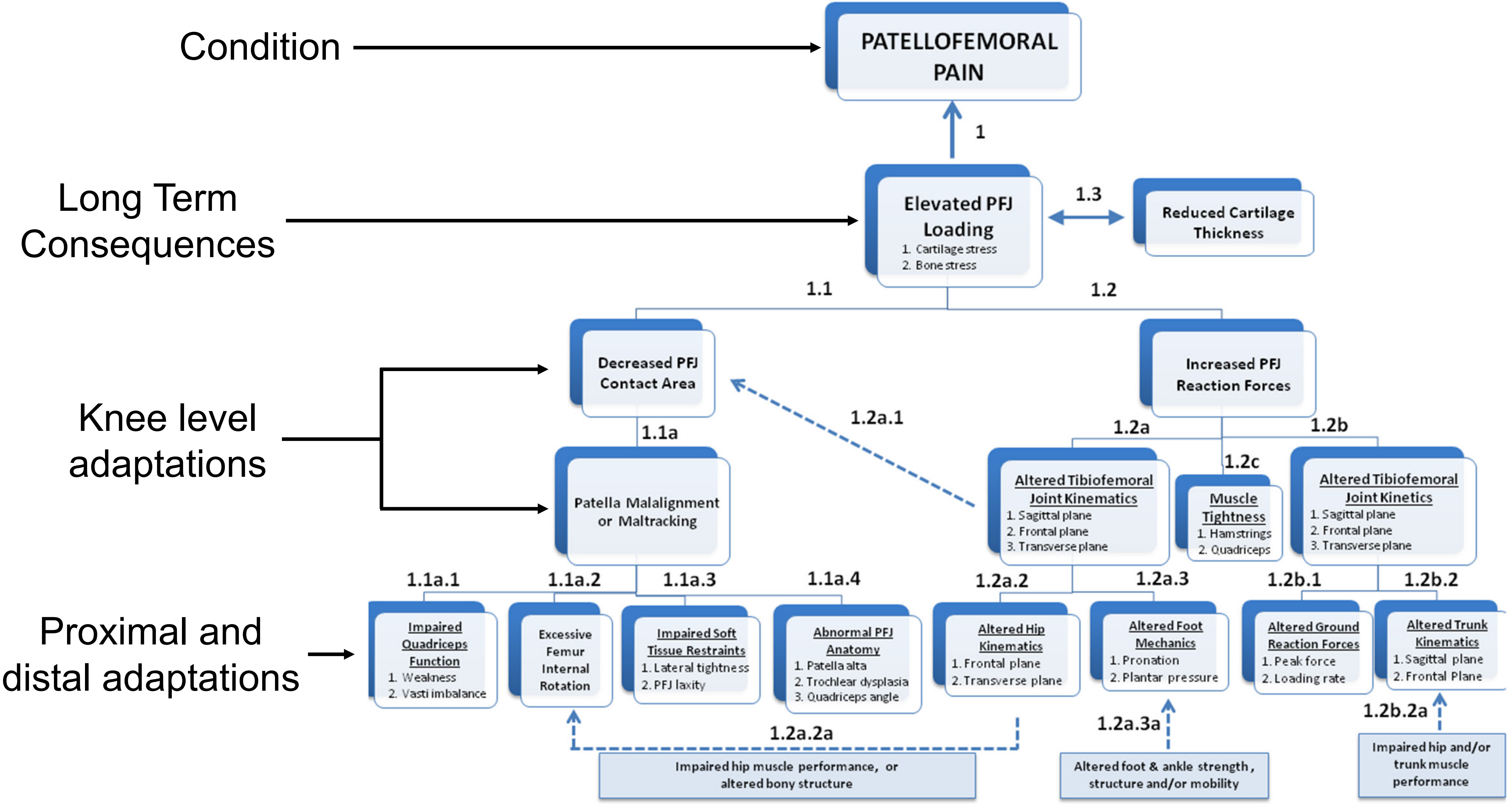


Image from Bikesreav

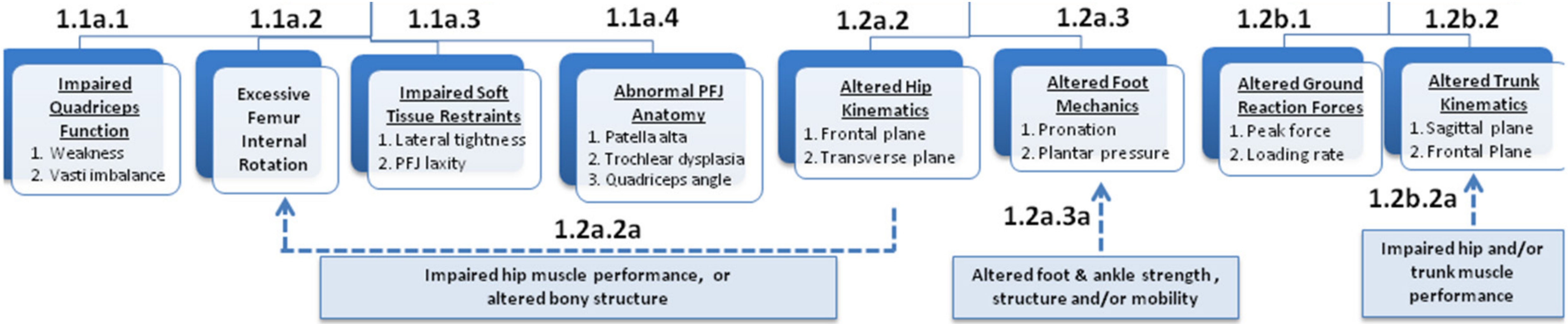
Willy et al. 2019, Powers et al. 2017, Andriacchi et al. 2009

2017 Pathomechanical model adopted from Powers et al.



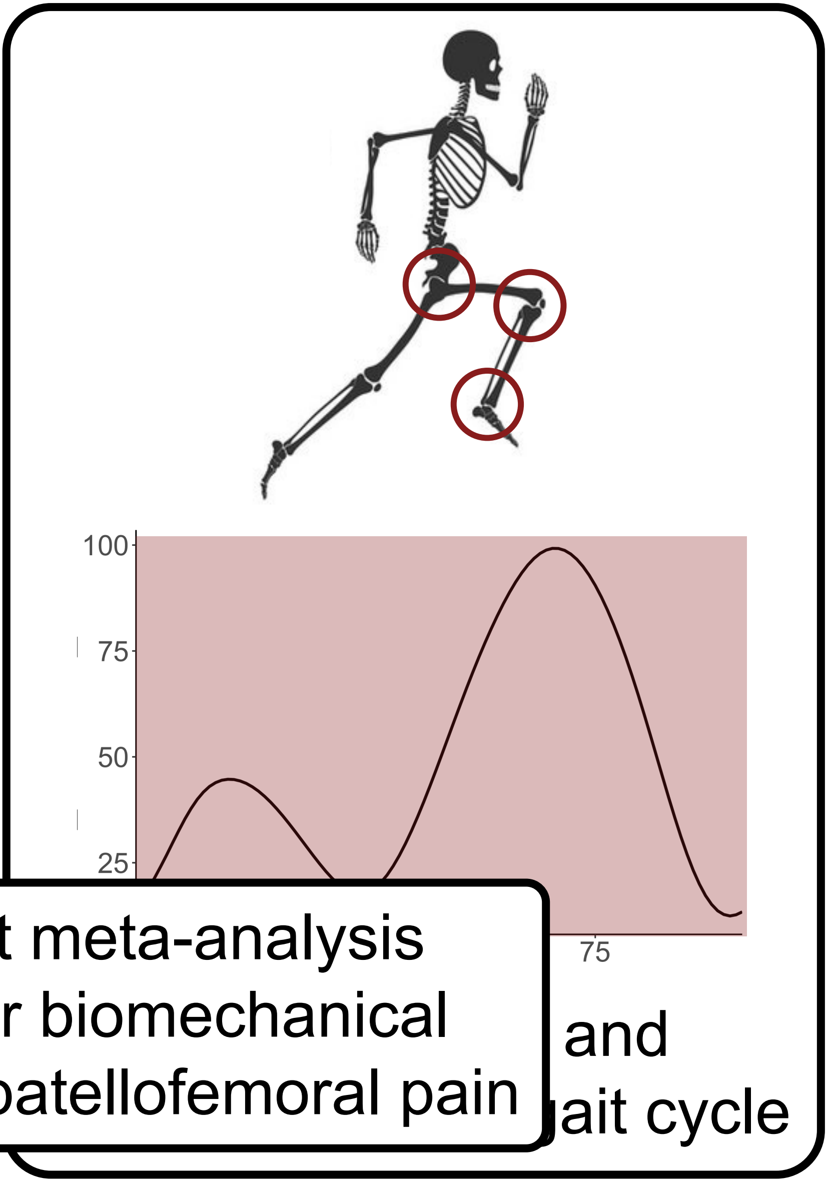
Prior Work

- Many findings are conflicting due to heterogeneity of PFP
- Runners with PFP displayed:
  - Lower peak knee flexion angles
  - Lower peak knee extension moment
  - Prolonged soleus activation



Bazett-Jones et al. 2022, Esculier et al. 2015, Willson et al. 2011, Powers et al. 2017

Prior Work



Conflicting findings: recent meta-analysis suggests limited evidence for biomechanical adaptations in runners with patellofemoral pain

Barton et al. 2012, Dierks et al. 2011, Esculier et al. 2015, Noehren et al. 2012, Bazett-Jones et al. 2022

Review

Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities

Eni Halilaj [a,*], Apoorva Rajagopal [b], Madalina Fiterau [c], Jennifer L. Hicks [d], Trevor J. Hastie [e,f], Scott L. Delp [b,d,g]

[a] Department of Mechanical Engineering, Carnegie Mellon University, United States
[b] Department of Mechanical Engineering, Stanford University, United States
[c] Department of Computer Science, Stanford University, United States
[d] Department of Bioengineering, Stanford University, United States
[e] Department of Statistics, Stanford University, United States
[f] Department of Health Research and Policy, Stanford University, United States
[g] Department of Orthopaedic Surgery, Stanford University, United States

Ina Fiterau Brostean

UMass InFusion Lab

Data Mining: discover new patterns in data

Feature Engineering: creating new features from existing ones to reduce dimensionality

Principal component analysis and machine learning models are becoming more popular in biomechanics research

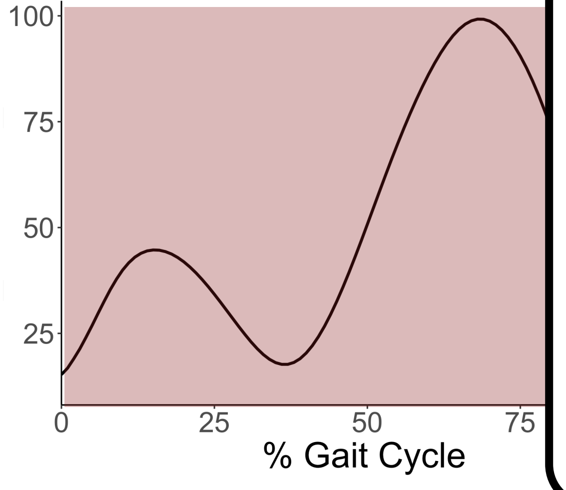
Halilaj et al. 2018

Our Approach

Teasing out the particularly important biomechanical adaptations instead of selecting them apriori could greatly improve our understanding of PFP

Principal Component Analysis

finds important differences
across the entire wavef

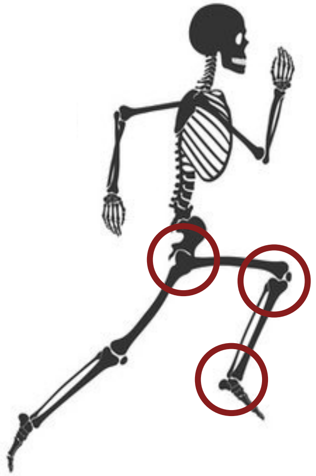


This "data mining" approach finds patterns that differentiate runners with and without PFP

Image from StichData

Machine Learning Modeling

finds important differences
cross all variables



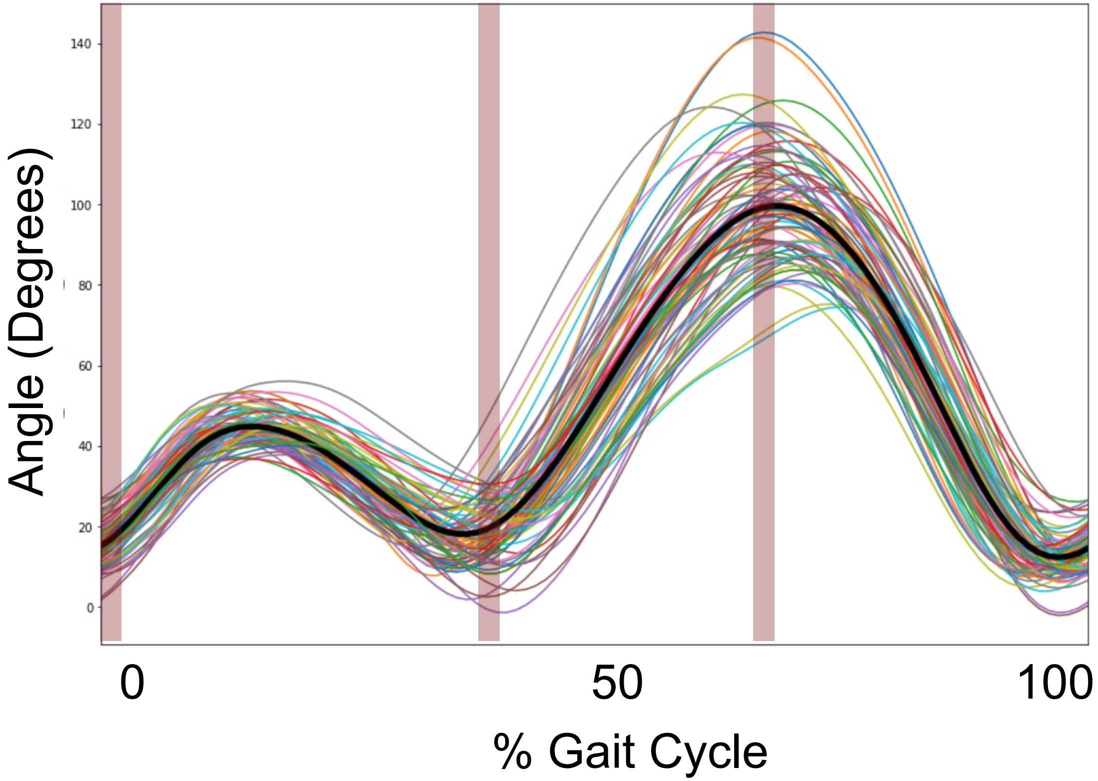
Study Aims

Aim 1: Determine important biomechanical features that differentiate symptomatic for PFP and healthy runners at baseline

Aim 2: Determine important biomechanical features that differentiate symptomatic for PFP and healthy runners following a treadmill run that induces knee pain

Aim 3: Explore if runners recovered from pain still exhibit altered biomechanics by classifying them as either symptomatic for PFP or healthy

Principal Component Analysis (PCA)



100 timepoints of the gait cycle

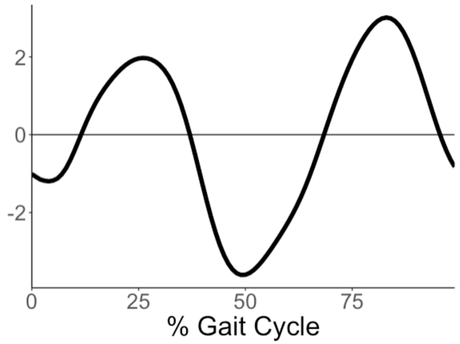
↓

A few principal components explaining most of the variance in the data

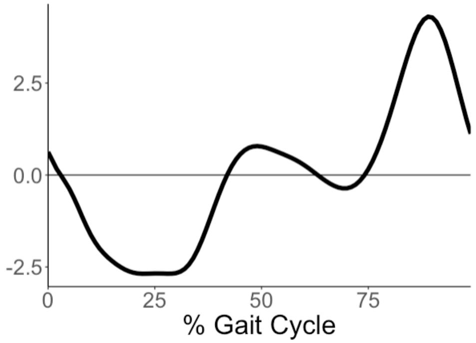
- PCA generates principal components (PCs) and scores for each observation associated with the respective PC

Principal Component Analysis (PCA)

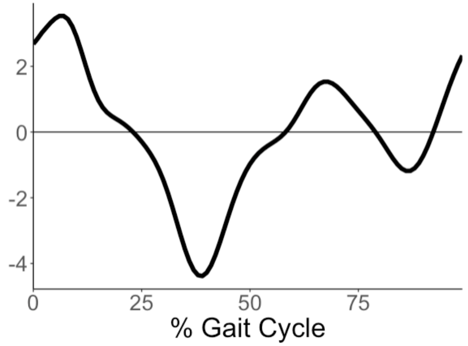
$$Original\ Waveform = \sum_{i=1}^{n}\left[PC_n \times Score_{PC_n}\right]$$



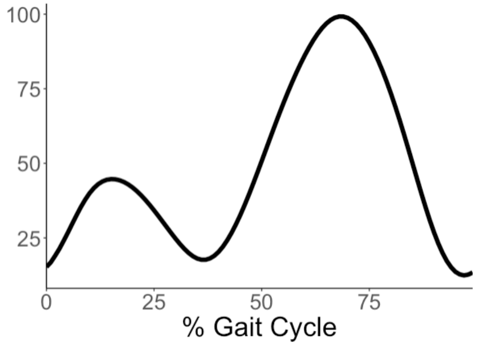
+



+

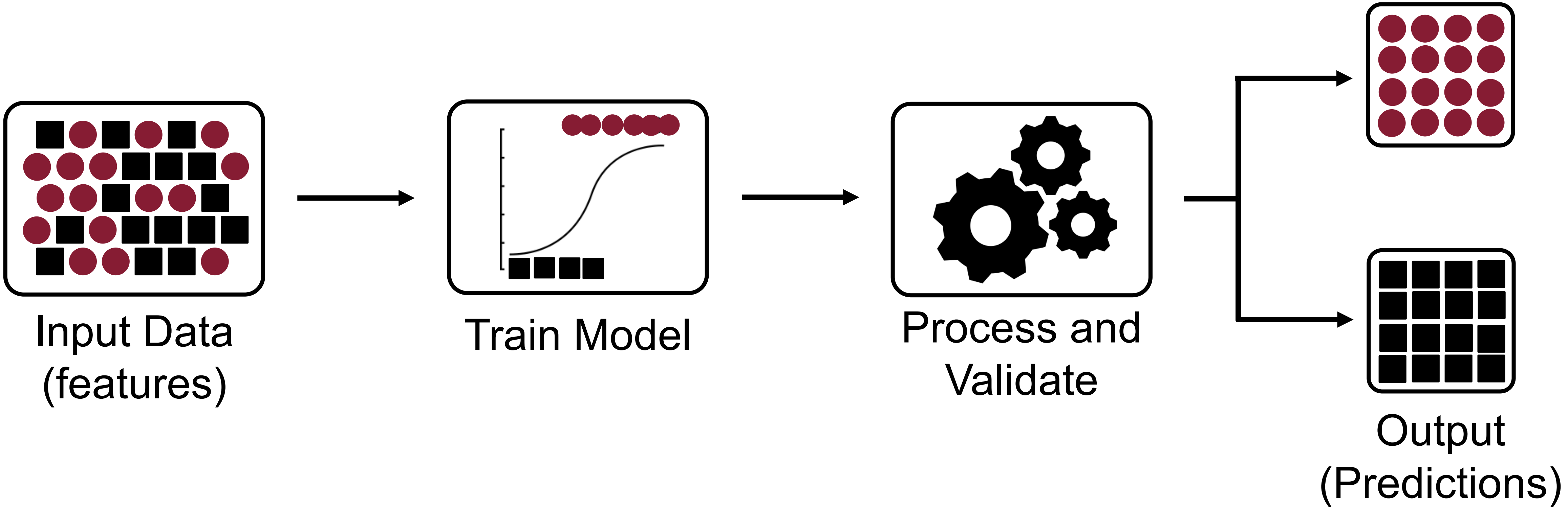


+ . . . +



$\left[PC_1 \times Score_{PC_1}\right]$ + $\left[PC_2 \times Score_{PC_2}\right]$ + $\left[PC_3 \times Score_{PC_3}\right]$ + . . . + $Original\ Waveform$

Supervised Machine Learning Classification



- Healthy Runners
- Symptomatic Runners

<u>Features that Differentiate Groups</u>

- Kinetics
- Kinematics
- Muscle Activation Patterns (EMG)

Interpretable Machine Learning Models

Black Box Approaches

Inputs (features)

↓

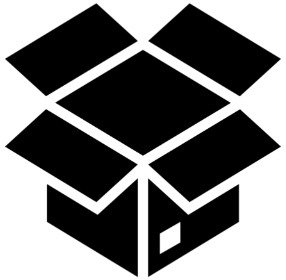
↓

Outputs (predictions)

Interpretable ML Approaches

Inputs

↓



- Informativeness
- Transferability
- Explainability
- Feature Importance

↓

Outputs

Methods – Data Collection

41 recreational runners
- 15 healthy (H)
- 14 symptomatic PFP (S)
- 12 recovered from PFP (R)

Runners completed a 21-minute treadmill run at self-selected speed

- M1 —— 21-minute treadmill run —— M21 -

Kinetic, kinematic, electromyography (EMG), RPE, and subjective pain scores collected at 1st (M1) and 21st (M21) minutes

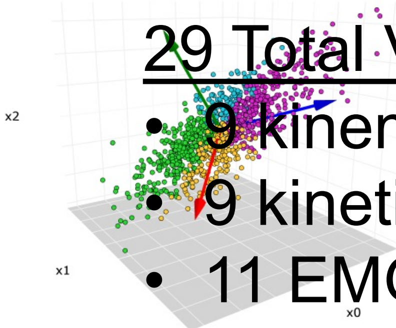
Variable	Healthy	Symptomatic	Recovered
Pain Level Change (M1 > M21)	0.06	3.00	0.33
RPE Change (M1 > M21)	4.80	3.64	2.79

Methods – Data Processing

1. Generate variable matrices

Subject ID	Group	t1	t2	…	t101
Subject 1 M1	H	Datapoint	Datapoint	…	Datapoint
Subject 1 M21	H	Datapoint	Datapoint	…	Datapoint
Subject 2 M1	S	Datapoint	Datapoint	…	Datapoint
…	…	Datapoint	Datapoint	…	Datapoint
Subject 29 M21	S	Datapoint	Datapoint	…	Datapoint

2. Principal Component Analysis



29 Total Variables

- 9 kinematic variables
- 9 kinetic variables
- 11 EMG variables

Generates Principal Component (PC) Scores

3. Logistic regression model: classify healthy and symptomatic runners

Limitation: nearly 100 principal components from all kinetic, kinematic, and EMG variables

Solution: machine learning model to determine the most important principal components

Reduced dataset from 2900 variables to ~100 variables that describe most of the variation in the data

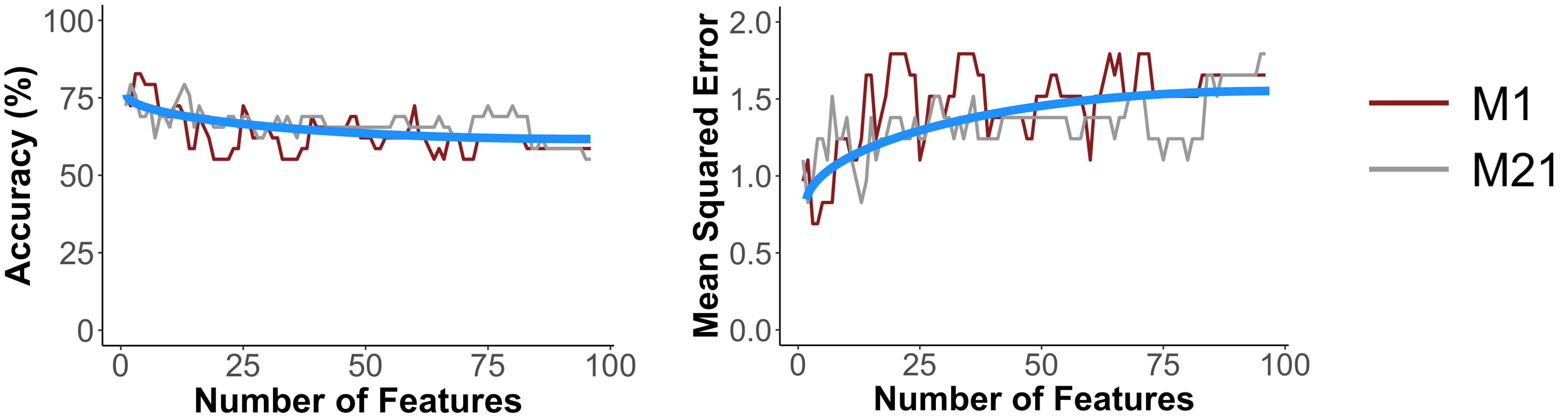
Extract Feature Coefficients

Methods – Machine Learning Modeling

- Generated two models:
  1. Minute 1 Model (M1)
  2. Minute 21 Model (M21)

- Quasi-Forward Stepwise Feature Selection Algorithm:
  1. Select *k* best features
  2. Cross validate model
  3. Calculate model accuracy
  4. Repeat steps 1-3 for $k = k + 1$ features derived from PCA
     - ~100 iterations for ~100 principal components
  5. Select the model with the highest classification accuracy

Results – Machine Learning Modeling



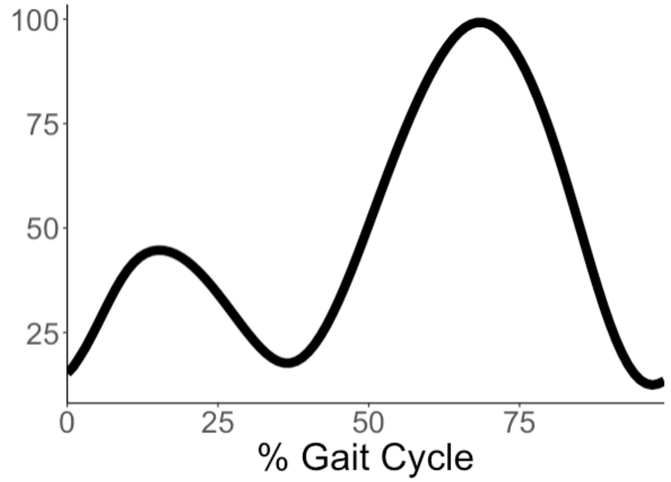
Results

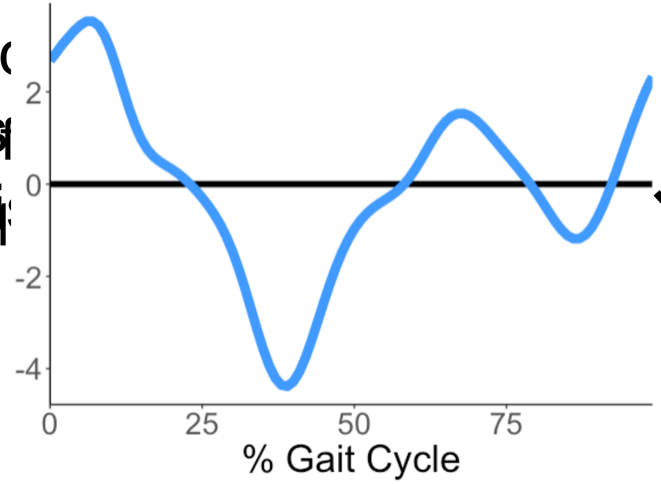
Variable	LR Coefficient	Explained Variance	Healthy Group Score	Symptomatic Group Score	p-value
Minute 1 Model					
Knee Flexion Angle PC3	3.617	8	-0.197 (0.564)	0.346 (0.48)	0.010*
Hip Adduction Moment PC3	1.301	13	-1.119 (2.554)	1.455 (2.988)	0.018*
Gluteus Maximus Activation PC2	0.929	33	1.578 (3.483)	-1.896 (4.415)	0.026*
Hip Adduction Moment PC1	0.801	49	3.731 (4.849)	-2.322 (5.298)	0.003*
Minute 21 Model					
Knee Flexion Moment PC2	2.341	14	-0.211 (0.333)	0.619 (1.831)	0.096
Hip Adduction Moment PC3	1.608	13	-0.66 (2.8)	1.784 (2.7)	0.024*
Knee Flexion Angle PC3	1.515	8	-0.241 (0.448)	0.385 (0.568)	0.003*
Knee Adduction Moment PC3	1.479	13	-0.619 (1.593)	1.32 (3.148)	0.044*
Biceps Femoris Activation PC1	1.452	37	0.934 (4.038)	-1.739 (4.206)	0.092
Ankle Flexion Angle PC3	1.342	13	-0.746 (0.961)	0.023 (1.262)	0.075
Soleus Activation PC2	1.272	23	-1.179 (1.856)	0.777 (3.344)	0.060
Ankle Abduction Moment PC1	0.934	52	-0.936 (2.802)	1.903 (4.285)	0.043*
Ankle Flexion Moment PC2	0.803	7	0.269 (0.604)	-0.186 (0.504)	0.037*
Hip Adduction Moment PC1	0.79	49	2.181 (5.434)	-1.785 (5.728)	0.066
Knee Rotation Angle PC4	0.705	8	0.872 (1.478)	-0.29 (1.299)	0.033*
Hip Rotation Angle PC4	0.604	6	0.977 (2.286)	-1.086 (1.516)	0.008*
Ankle Flexion Angle PC2	0.554	24	0.625 (1.441)	-0.358 (1.473)	0.080

Mean (standard deviation), *p<0.05

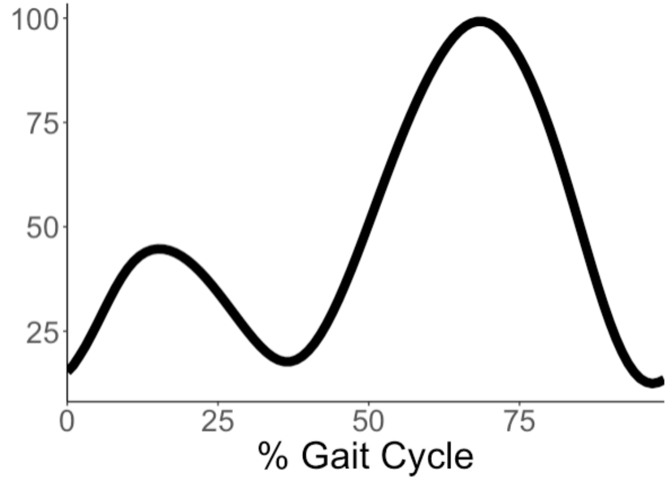
Interpreting Principal Components

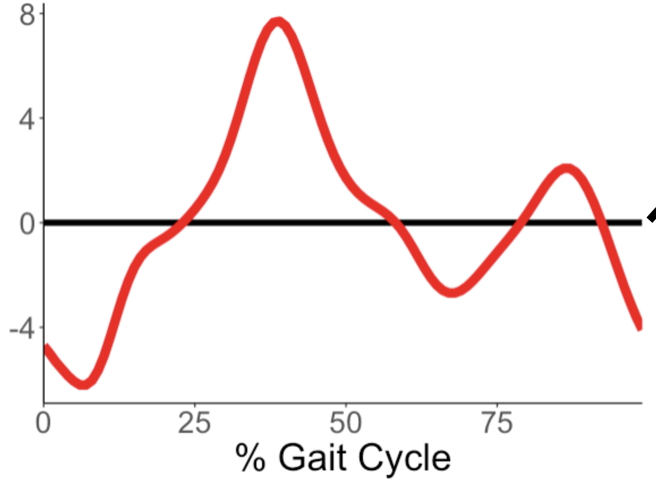
deviation from the mean at
we know what variance that

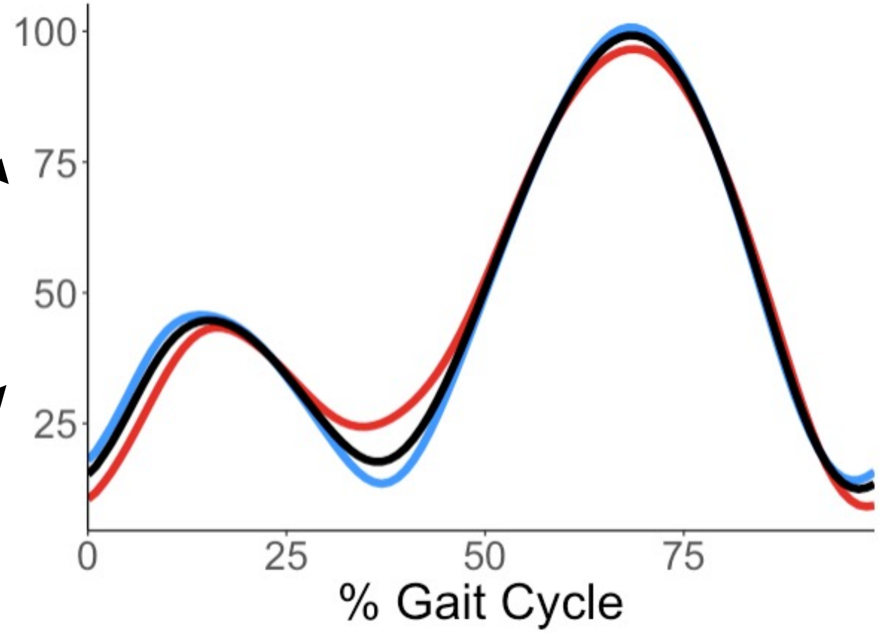




[Ensemble Average Waveform] **+** **[Group PC Score * PC * Scaling Factor]**





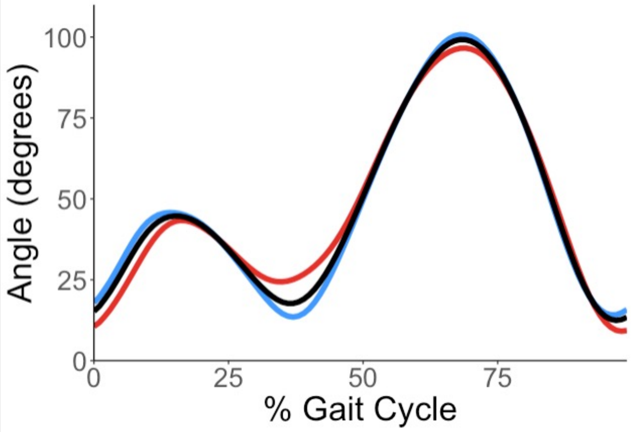
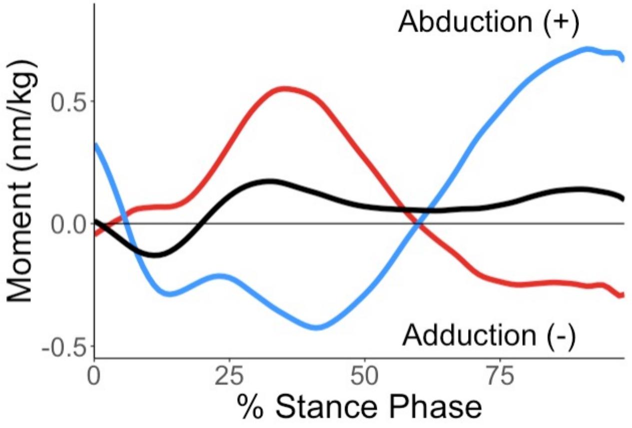
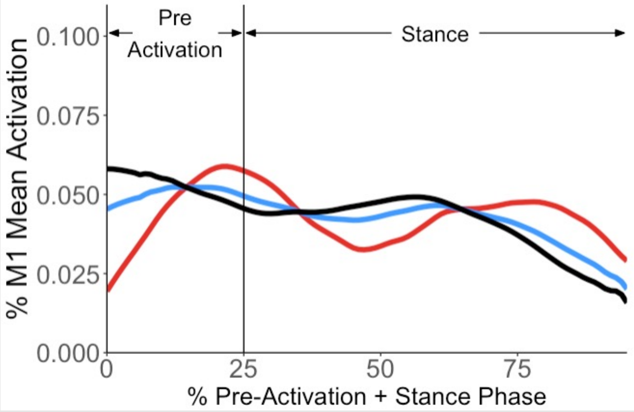


Healthy Group Symptomatic Group Ensemble Average

Results – Minute 1 Model (Aim 1)

Accuracy: 82.76%

MSE: 0.670

Variable	Knee Flexion Angle PC3	Hip Adduction Moment PC1 + PC3	Gluteus Maximus Activation PC2
Principal Component Plot			
Interpretation of symptomatic group	• Larger magnitude of knee flexion angle at toe off.	• Larger magnitude of hip abduction moment in mid-stance. • Larger magnitude of hip adduction moment in late stance.	• Lower pre- and mid-stance phase activation. • Greater activation at heel strike and in late stance.

Discussion – Minute 1 Model (Aim 1)

- Symptomatic runners avoiding full range of knee flexion during stance
  - Smaller patellofemoral contact area?
  - Pain avoidance strategy?

- Altered frontal plane loading patterns at the hip joint
  - Proximal strategy

- Abnormal gluteus maximus activation patterns in pre-activation and stance phases
  - Potential alterations to patellofemoral joint kinematics

Results – Minute 21 Model (Aim 2)

Accuracy: 79.31%

MSE: 0.828

Variable	LR Coefficient	Explained Variance
Knee Extension Moment PC2	2.341	14
Hip Adduction Moment PC3	1.608	13
Knee Flexion Angle PC3	1.515	8
Knee Adduction Moment PC3	1.479	13
Biceps Femoris Activation PC1	1.452	37
Ankle Flexion Angle PC3	1.342	13
Soleus Activation PC2	1.272	23
Ankle Abduction Moment PC1	0.934	52
Ankle Flexion Moment PC2	0.803	7
Hip Adduction Moment PC1	0.79	49
Knee Rotation Angle PC4	0.705	8
Hip Rotation Angle PC4	0.604	6
Ankle Flexion Angle PC2	0.554	24

Same features at M1 model

Results – Minute 21 Model (Aim 2)

Accuracy: 79.31%

MSE: 0.828

Variable	LR Coefficient	Explained Variance
Knee Extension Moment PC2	2.341	14
Hip Adduction Moment PC3	1.608	13
Knee Flexion Angle PC3	1.515	8
Knee Adduction Moment PC3	1.479	13
Biceps Femoris Activation PC1	1.452	37
Ankle Flexion Angle PC3	1.342	13
Soleus Activation PC2	1.272	23
Ankle Abduction Moment PC1	0.934	52
Ankle Flexion Moment PC2	0.803	7
Hip Adduction Moment PC1	0.79	49
Knee Rotation Angle PC4	0.705	8
Hip Rotation Angle PC4	0.604	6
Ankle Flexion Angle PC2	0.554	24

Ankle based strategy

Results – Minute 21 Model (Aim 2)

Accuracy: 79.31%

MSE: 0.828

Variable	LR Coefficient	Explained Variance
Knee Extension Moment PC2	2.341	14
Hip Adduction Moment PC3	1.608	13
Knee Flexion Angle PC3	1.515	8
Knee Adduction Moment PC3	1.479	13
Biceps Femoris Activation PC1	1.452	37
Ankle Flexion Angle PC3	1.342	13
Soleus Activation PC2	1.272	23
Ankle Abduction Moment PC1	0.934	52
Ankle Flexion Moment PC2	0.803	7
Hip Adduction Moment PC1	0.79	49
Knee Rotation Angle PC4	0.705	8
Hip Rotation Angle PC4	0.604	6
Ankle Flexion Angle PC2	0.554	24

Frontal and transverse plane adaptations

Discussion – Minute 21 Model (Aim 2)

- Same knee flexion and hip abduction patterns as minute 1
  - Key strategies regardless of pain status

- Abnormal hip rotation angles, knee extension moments, and knee rotation angles
  - Related to biceps femoris activation patterns

- Ankle-based strategy for maintaining running velocity
  - Potential strategy for mitigating pain

Results – Recovered Runners (Aim 3)

Aim 3: Classify runners previously recovered from PFP as either symptomatic for PFP or healthy

Model Timepoint	# Healthy Classification	# Symptomatic Classification
Minute 1 Model	7	5
Minute 21 Model	5	7

- Recovered runners who experienced any pain classified as:
  - Symptomatic at both timepoints
  - Shifted from healthy at M1 to symptomatic at M21

Summary & Conclusions

- Important adaptations with the onset of pain in runners at both the distal and proximal joints
- Ankle based strategy for maintaining velocity after 21 minutes of treadmill running
- Recovered runners respond heterogeneously
  - Onset of pain important for classification procedure

Acknowledgements

MOBL Group (and others):

- Dr. Katherine Boyer
- Erica Casto
- Skylar Holmes
- Aidan Gross
- Athulya Simon
- Ryan Gladfelter
- Andrew Bourguignon
- Kali Shamaly
- Kiichi Ash
- Carl Jewell
- Bill Johnson

Bailey

DE LUCA FOUNDATION

Funding from the De Luca Foundation

UMass MOBL Musculoskeletal Orthopedic Biomechanics Lab

University of Massachusetts Amherst BE REVOLUTIONARY™

References

Crossley, K. M., Stefanik, J. J., Selfe, J., Collins, N. J., Davis, I. S., Powers, C. M., McConnell, J., Vicenzino, B., Bazett-Jones, D. M., Esculier, J.-F., Morrissey, D., & Callaghan, M. J. (2016). 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. *British Journal of Sports Medicine*, *50*(14), 839–843. https://doi.org/10.1136/bjsports-2016-096384

BAZETT-JONES, D. M., COBB, S. C., HUDDLESTON, W. E., O'CONNOR, K. M., ARMSTRONG, B. S. R., & EARL-BOEHM, J. E. (2013). Effect of Patellofemoral Pain on Strength and Mechanics after an Exhaustive Run. *Medicine & Science in Sports & Exercise*, *45*(7), 1331–1339. https://doi.org/10.1249/MSS.0b013e3182880019

Willy, R. W., Hoglund, L. T., Barton, C. J., Bolgla, L. A., Scalzitti, D. A., Logerstedt, D. S., Lynch, A. D., Snyder-Mackler, L., & McDonough, C. M. (2019). Patellofemoral pain clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy of the American physical therapy association. *Journal of Orthopaedic and Sports Physical Therapy*, *49*(9), CPG1–CPG95. https://doi.org/10.2519/jospt.2019.0302

Taunton, J. E., Ryan, M. B., Clement, D. B., McKenzie, D. C., Lloyd-Smith, D. R., & Zumbo, B. D. (2002). A retrospective case-control analysis of 2002 running injuries. *British Journal of Sports Medicine*, *36*(2), 95–101. https://doi.org/10.1136/bjsm.36.2.95

Powers, C. M., Witvrouw, E., Davis, I. S., & Crossley, K. M. (2017). Evidence-based framework for a pathomechanical model of patellofemoral pain: 2017 patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester, UK: Part 3. *British Journal of Sports Medicine*, *51*(24), 1713–1723. https://doi.org/10.1136/bjsports-2017-098717

Andriacchi, T. P., Koo, S., & Scanlan, S. F. (2009). Gait Mechanics Influence Healthy Cartilage Morphology and Osteoarthritis of the Knee. *Journal of Bone and Joint Surgery*, *91*(Supplement_1), 95–101. https://doi.org/10.2106/JBJS.H.01408

Bazett-Jones, D. M., Neal, B. S., Legg, C., Hart, H. F., Collins, N. J., & Barton, C. J. (2022). Kinematic and Kinetic Gait Characteristics in People with Patellofemoral Pain: A Systematic Review and Meta-analysis. *Sports Medicine*. https://doi.org/10.1007/s40279-022-01781-1

Barton, C. J., Levinger, P., Crossley, K. M., Webster, K. E., & Menz, H. B. (2012). The relationship between rearfoot, tibial and hip kinematics in individuals with patellofemoral pain syndrome. *Clinical Biomechanics*, *27*(7), 702–705. https://doi.org/10.1016/j.clinbiomech.2012.02.007

Dierks, T. A., Manal, K. T., Hamill, J., & Davis, I. (2011). Lower extremity kinematics in runners with patellofemoral pain during a prolonged run. *Medicine and Science in Sports and Exercise*, *43*(4), 693–700. https://doi.org/10.1249/MSS.0b013e3181f744f5

Esculier, J. F., Roy, J. S., & Bouyer, L. J. (2015). Lower limb control and strength in runners with and without patellofemoral pain syndrome. *Gait and Posture*, *41*(3), 813–819. https://doi.org/10.1016/j.gaitpost.2015.02.020

Willson, J. D., Kernozek, T. W., Arndt, R. L., Reznichek, D. A., & Scott Straker, J. (2011). Gluteal muscle activation during running in females with and without patellofemoral pain syndrome. *Clinical Biomechanics*, *26*(7), 735–740. https://doi.org/10.1016/j.clinbiomech.2011.02.012

Noehren, B., Sanchez, Z., Cunningham, T., & McKeon, P. O. (2012). The effect of pain on hip and knee kinematics during running in females with chronic patellofemoral pain. *Gait and Posture*, *36*(3), 596–599. https://doi.org/10.1016/j.gaitpost.2012.05.023

Molnar, C. (2019). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. *Leanpub Publishing*.

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J. L., Hastie, T. J., & Delp, S. L. (2018). Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. In *Journal of Biomechanics* (Vol. 81, pp. 1–11). Elsevier Ltd. https://doi.org/10.1016/j.jbiomech.2018.09.009

Deluzio, K. J., & Astephen, J. L. (2007). Biomechanical features of gait waveform data associated with knee osteoarthritis. An application of principal component analysis. *Gait and Posture*, *25*(1), 86–93. https://doi.org/10.1016/j.gaitpost.2006.01.007

Andriacchi, T. P., Koo, S., & Scanlan, S. F. (2009). Gait Mechanics Influence Healthy Cartilage Morphology and Osteoarthritis of the Knee. *Journal of Bone and Joint Surgery*, *91*(Supplement_1), 95–101. https://doi.org/10.2106/JBJS.H.01408

Cheng, C. (2022) Principal Component Analysis (PCA) Explained Visually with Zero Math. *Towards Data Science.* n.d. Accessed: 8/2/2022

Slide 2 image credit: Adobe Stock. Accessed 8/2/2022 https://stock.adobe.com/ee/search?k=human+skeleton