Deep Learning and Wearable Sensing Accurately Classify Tendon Loading in Immobilizing Boot Walking

Stanton Godshall[1], Michelle P. Kwon[1], Eni Halilaj[2], Lorraine AT Boakye[1], Casey Jo Humbyrd[1], Josh R. Baxter[1]
[1]University of Pennsylvania, Philadelphia, PA, [2]Carnegie Melon University, Pittsburgh, PA
stanton.godshall@pennmedicine.upenn.edu

DISCLOSURES: Stanton Godshall (N), Michelle P. Kwon (N), Eni Halilaj (N), Lorraine AT Boakye (1,9), Casey Jo Humbyrd (1,9), Josh R. Baxter (N)

INTRODUCTION: The Achilles tendon is the most frequently ruptured tendon, and incidence rates are increasing [1]. This injury causes permanent functional deficits in two of three patients [2]. To improve outcomes, clinicians have attempted earlier mobilization and loading to limit muscle atrophy and promote tendon regeneration [3]. Despite this innovation, recommendations remain generalized in the absence of evidence-based guidelines. We developed a wearable-sensing paradigm that uses an inertial measurement unit (IMU) secured to the immobilizing boot shortly after surgery to monitor patient activity. Our group uses a similar paradigm to accurately track step count and boot wear time, but those measures fail to capture the variability in tendon loading. Our force-sensing insole data reveal that rupture patients adopt varying strategies that differentially load the Achilles tendon. Some transfer less than 0.5 bodyweights through the tendon, while others exceed 2 bodyweights. These insoles, however, are impractical for clinical care or large studies because of their cost and burdens for patients. To overcome this, we built a custom one-dimensional (1D) convolutional neural network (CNN) that classifies IMU signals into low, medium, and high tendon-loading steps (<1, 1-2, and >2 bodyweights). This approach offers a scalable, low-burden alternative to force-sensing insoles and moves us closer to individualized rehabilitation guidelines after Achilles tendon rupture.

METHODS: We analyzed previously published data [4] from ten healthy adults (3 females, age 25+\-2.4, BMI 23.9+\-6.56) to establish proof-of-concept and plan a follow-up clinical study. This dataset included 6154 steps with paired force-sensing insole data and boot-mounted IMU data that were collected in a controlled laboratory environment. These steps span varying gaits (fast, medium, slow, and pathological) and heel wedges (0, 5, and 30 degrees). We trained 1D CNN models to classify tendon loading using an ordinal classification scheme (low <1 BW, medium 1-2 BW, high >2 BW). IMU signals were segmented and normalized prior to training, and step duration and boot wedge height were included as metadata. Hyperparameters (filter size, kernel size, fully connected nodes, and learning rate) selection and model evaluation were done via nested cross-validation. To reflect our intended deployment and to improve performance, we deliberately incorporated a small portion of each validation/test subject's data into training, enabling subject-specific calibration. This mirrors the clinical use case, where a brief in-clinic snapshot of tendon loading would be used to retrain a generalized model for individual patients.

RESULTS: Across all subjects, classification accuracy was 85.8+/-3.3% and balanced accuracy was 82.1+/-7.9%. Classification accuracy was highest for low- and high- load steps (sensitivities: 96.0%, 93.1%, respectively), with reduced accuracy for medium-load steps (sensitivity: 62.8%) (**Figure 1**). All misclassifications occurred with a neighboring class (**Figure 1**). Errors clustered near the decision boundary, with the average distance from a misclassified step to the nearest boundary being 18.1% bodyweight (**Figure 2**).

DISCUSSION: Our study demonstrates that 1D CNNs are good candidates to estimate Achilles tendon loading in an immobilizing boot with only IMU signals and wedge height. This approach lays the groundwork for future studies to continuously monitor Achilles tendon loading without relying on burdensome sensors that restrict data collection and patient inclusion. The model reliably distinguished low- and high-load gait patterns, suggesting potential to identify patient-specific strategies that modulate tendon loading in the early therapeutic window after rupture repair. Although overall accuracy metrics did not exceed 90% accuracy, most errors occurred near decision thresholds. We suspect that when summarizing over day- and week-long recovery periods, this model will provide sufficient granularity to support clinical decision-making. By connecting motion signals to tendon load, our findings highlight a promising direction for individualized rehabilitation monitoring. This study has several limitations. First, step segmentation relied on force-sensing insoles, limiting complete deployment in IMU-only applications. Additionally, models were trained with healthy subjects, and their generalizability to rupture patients remains unknown. Future work will focus on evaluating tendon load estimation using IMU-only segmentation, external validation in rupture cohorts, and identifying daily loading metrics that may guide long-term healing outcomes.

SIGNIFICANCE/CLINICAL RELEVANCE: This experiment is a necessary step towards developing precision rehabilitative care for patients with Achilles tendon ruptures. We expect that by linking real-world tendon loading to tendon healing and symptoms, we will establish evidence-based tendon-loading profiles that improve patient outcomes by supporting clinical decision making. Additionally, these data provide important practical rehabilitative information that can inform preclinical models and serve as clinically relevant outcome measurements for longitudinal clinical research.

REFERENCES: [1] Whitmore+, *Orthop J Sports Med*, 2025; [2] Brorsson+, *Am J Sports Med,* 2017; [3] Willits+, *J Bone Jt. Surg.,* 2010; [4] Kwon+, *Sci Rep*, 2023

ACKNOWLEDGEMENTS: This research was supported by NIH/NIAMS P50AR080581

IMAGES AND TABLES:

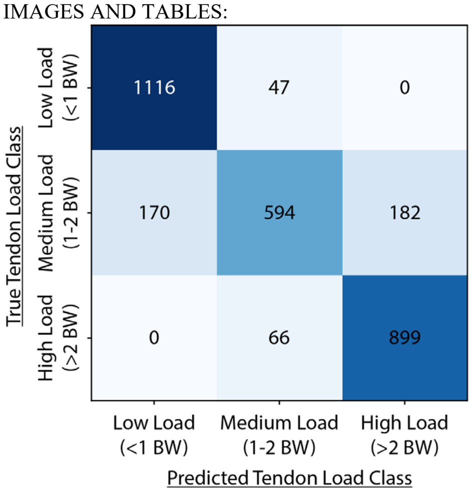

Figure 1. Confusion matrix showing classification performance. Correct predictions lie across the diagonal. Low load is never classified as high load.

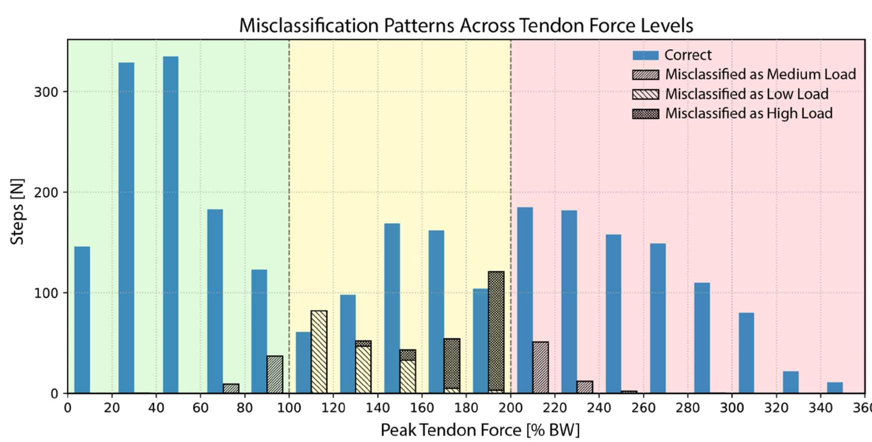

Figure 2. Histogram of classification outcomes by peak tendon force (% BW). Misclassifications are concentrated near threshold boundaries, indicating transitional ambiguity between load regions.