Impact of JIA-Related Physiology on Machine Learning-Based Task Prediction Performance from Active Acoustics-Driven Achilles Tendon Sensing: A Proof-of-Concept Study

Luis G. Rosa
School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA, USA
lrosa3@gatech.edu

Quentin Goossens
School of Electrical and Computer Eng.
Georgia Institute of Technology
Atlanta, GA, USA
qgoossens3@gatech.edu

Miguel Locsin
Department of Pediatrics
Emory University School of Medicine
Atlanta, GA, USA
miggsloc234work@gmail.com

Lori A. Ponder
Children's Healthcare of Atlanta
Atlanta, GA, USA
laponde@emory.edu

Sampath Prahalad
Children's Healthcare of Atlanta and
Emory University School of Medicine
Atlanta, GA, USA
sprahal@emory.edu

Omer T. Inan. Fellow. IEEE
School of Electrical and Computer Eng.
Georgia Institute of Technology
Atlanta, GA, USA
omer.inan@ece.gatech.edu

***Abstract*—Juvenile idiopathic arthritis (JIA) and its enthesitis related arthritis (ERA) subtype often present diagnostic challenges due to variable symptomatology and limited accessibility of advanced imaging. Here, we explore a non-invasive approach for diagnostic decision support using active vibrational sensing and machine learning to classify locomotion tasks to characterize symptomatology. By comparing classification performance across JIA subgroups, including ERA and its active/inactive states, we observed that ERA-related physiology reduces ML classifier accuracy, with sample comparison of ERA to the No ERA group yielding p = 0.002 and Cohen's d effect size d = 2.05. Notably, the classification metrics' performance was consistently higher on the No ERA group compared to the three ERA subgroups we considered, indicating that acoustic signatures from inflamed tendons and entheses modulate predictive performance. These findings suggest that task-based classification accuracy might serve as a surrogate biomarker for inflammation severity. Beyond presenting the methodology, ranging from data acquisition with a miniature vibration motor and accelerometer, to PCA-based feature extraction and multi-class classification, our results underscore the potential of integrating vibration sensing into clinical workflows. Ultimately, this study lays the groundwork for potentially enabling more robust, cost-effective diagnostic tools that could support early detection, monitoring, and personalized management of JIA and ERA.**

Keywords—Juvenile idiopathic arthritis (JIA), machine learning classification, non-invasive digital biomarker, pediatric rheumatology, vibration analysis.

I. Introduction

Juvenile idiopathic arthritis (JIA) encompasses several chronic inflammatory conditions that affect children, often leading to pain, swelling, and functional impairment in one or more joints [1], [2]. Within JIA, enthesitis related arthritis (ERA) is a subtype characterized by inflammation at the points where tendons, ligaments, or fascia insert into bone. Accurately identifying ERA is critical because it presents unique diagnostic challenges and may follow a more severe clinical course than other JIA subtypes [3], [4]. However, current methods for diagnosing and assessing disease severity rely heavily on clinical exams, which can be subjective and prone to variability [2], [5]. Furthermore, while imaging tools such as MRI offer more precise insight, they are time-intensive and expensive, underscoring the need for more accessible, quantitative modalities [6], [7], [8].

Recent work on vibration sensing has demonstrated promising utility in detecting inflamed or structurally altered tissues in pediatric rheumatology [9]. Previous studies have predominantly employed machine learning (ML) classification models to distinguish directly between diseased and non-diseased states, focusing primarily on phenotypic markers indicating the presence or absence of inflammation. However, this direct classification approach provides limited insight into disease severity, as classifier confidence in distinguishing between classes has not been shown to reliably correlate with disease severity indicators in JIA [9]. The study presented here introduces a novel approach: rather than classifying solely based on disease presence, we classify locomotion tasks within individual subjects. This method inherently controls for inter-subject variability and aims to serve as a more direct proxy for disease severity, with the hypothesis that greater severity leads to reduced classification accuracy. Specifically, we hypothesize that physiological differences inherent to ERA—particularly during active inflammation—will negatively impact how accurately these tasks are predicted. Thus, our task-based classification performance may provide a quantitative and accessible biomarker, reflecting inflammation severity or progression in ERA more effectively than previous binary disease-state classification methods.

Building on the promising results observed in previous active acoustic (AA) related studies, which demonstrated the feasibility of using vibration signatures to non-invasively characterize both Achilles Tendon (AT) loading and inflammation [9], [10], [11], our study provides an overview of our updated data acquisition and machine learning (ML) pipelines, designed specifically to address JIA diagnosis gaps by

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

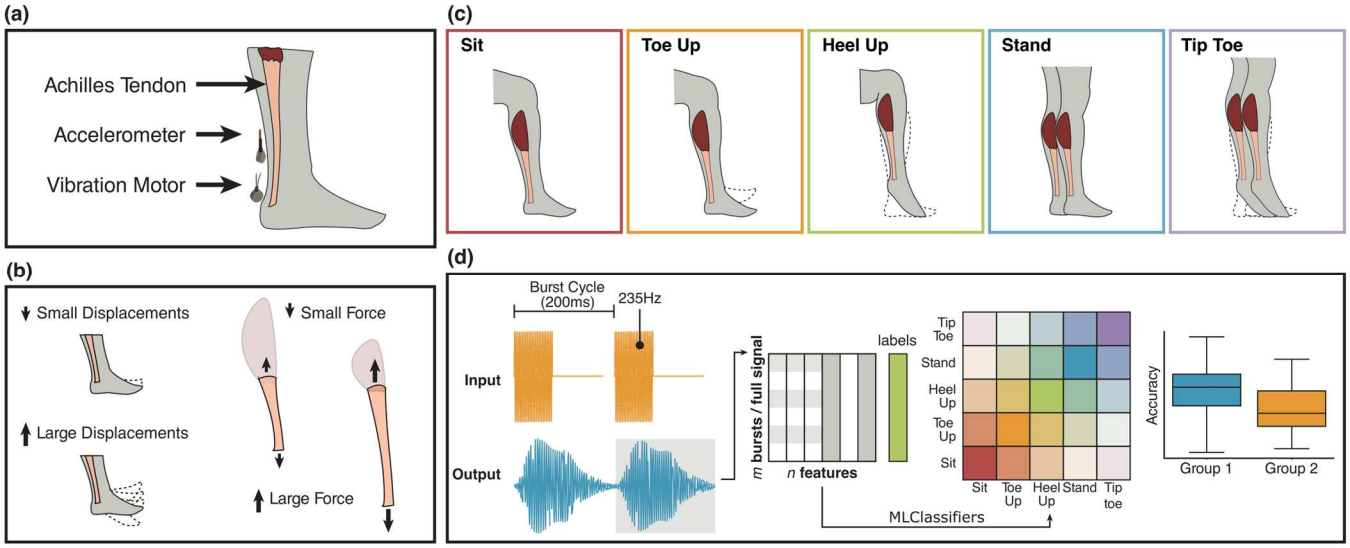

Fig. 1. Experimental Protocol. (a) Representation of our active acoustic (AA) sensing arrangement consisting of an accelerometer and vibration motor placed over the Achilles Tendon (AT). (b) Visual representation of task variability targeting different levels of force and displacement. (c) Our tasks include (red) sitting with no movement, (orange) toe up and down movements while seated, (green) heel up and down movements while seated, (blue) standing with no movement, and (purple) raising and lowering heels to go on tip toes. (d) Sample of our signal processing pipeline starting with input burst vibrations and corresponding output signal readings, followed by features being extracted for either the full signal or individual bursts, and these then being paired with the corresponding labels to train 7 ML classifiers that yielded task performance predictions. Due to the training data separation by group, comparing the task prediction performance gives us an idea of the effects of the underlying physiology for each group.

TABLE I. GROUP INCLUSION CRITERIA

No ERA	All ERA	Inactive ERA	Active ERA
Non-ERA, JIA diagnosed by a physician	ERA diagnosed by a physician	ERA diagnosed by a physician No pointed ankle pain at time of visit	ERA diagnosed by a physician Pointed ankle pain at time of visit

classifying locomotion tasks rather than disease presence alone. Ultimately, this research seeks a more granular understanding of the relationship between inflammatory processes in ERA and biomechanical alterations, paving the way for improved diagnostic and disease-monitoring strategies in pediatric rheumatology.

II. Methods

A. Experimental Protocol

The study was approved by the Georgia Institute of Technology and Emory University Institutional Review Boards (Protocol H15383 – GaTech) and encompasses an expanded dataset to that started by [9]. For this study, we recruited 38 patients diagnosed with JIA according to the International League of Associations for Rheumatology (ILAR) criteria (28 female / 10 male, age: 13.4 ± 4.13 years, height: 154.9 ± 18.01 cm, weight: 53.8 ± 23.87 kg). Subjects were categorized given both a) their responses to a questionnaire capturing the presence of ankle symptoms including ankle swelling, ankle stiffness, and any ankle pain both on the day of recording and in the past 30 days, and b) the evaluation of the attending rheumatologist who we asked about findings of point tenderness and ankle enthesitis from the ankle exam.

Patients were categorized into 4 groups: a) No ERA, b) All ERA, c) Inactive ERA, and d) Active ERA (Table I). The *No ERA* group (n=23), consisted of all patients diagnosed with JIA but not ERA. *All ERA* (n=15) consisted of all patients with ERA at time of visit regardless of specific symptoms. The *inactive ERA* group (n=10) contains patients that were previously diagnosed with ERA based on symptoms and blood markers (HLA B27) but at the time of visit did not exhibit pointed ankle pain. Finally, the *Active ERA* group (n=5) contained the ERA patients who at the time of visit did show pointed joint pain.

Our experimental active vibration sensing system consists of pairing a small vibration motor (LRA Model VG0832012, Vybronics Electronics, 8 mm diameter and 2g weight) with a high-sensitivity and low-noise floor contact accelerometer (3225F7, Dytran Instruments, Inc.) by placing them ~2 cm apart on the skin over the AT with the motor at the level of the medial malleolus (Fig. 1(a)), both attached to the skin using a double-sided adhesive (23 mm Stickie, Rycote) and fabric tape (Kinesio Tex Gold, Kinesio). We used a data acquisition system (USB-4432, National Instruments) to both drive the motor with a 235 Hz sine wave that was multiplied by a 5 Hz square wave (period of 200 ms, 50% duty factor) and to sample the accelerometer output via MATLAB (R2023b, MathWorks). The accelerometer data was sampled at a frequency of 10kHz which captured the tendon response to the sinusoidal stimulation burst train from the motor.

We collected data for 5 tasks to help induce variability in both ankle positions and loading conditions while being easy for pediatric subjects to perform consistently (Fig 1(b)) [12]. Patients performed tasks for about 10 seconds each while mirroring a study team-member who ensured proper execution, all in fixed order as follows: 1) *sit*: sitting with knees at a 90-degree angle, 2) *toe up*: sitting with ankles fully dorsiflexed, 3) *heel up*: sitting with ankles fully plantarflexed, 4) *stand*: standing on two legs with no movement, and 5) *tip toe*: holding a fully extended calf raise.

B. Signal Processing and Machine Learning

Pre-processing for our data pipeline (Fig. 1(d)) starts by taking the output from our active acoustics (AA) system and applying a band pass filter between 220 Hz and 250 Hz to

TABLE II. METRIC PERFORMANCE ACROSS GROUPS FOR DIFFERENT FEATURE QUANTITIES

Number of Features	Balanced			
	No ERA	*All ERA*	*Inactive ERA*	*Active ERA*
2	0.381	0.345	0.338	0.325
3	0.453	0.403	0.375	0.377
5	0.467	0.425	0.428	0.417
10	0.483	0.428	0.427	0.440
15	0.483	0.425	0.408	0.406
Average	**0.453**	**0.405**	**0.395**	**0.393**

[a.] Average accuracy values per group per feature quantities across all classifiers, task combinations, and patient combinations showing consistently lower performance as you progress left to right through the groups for both balanced and unbalanced cases.

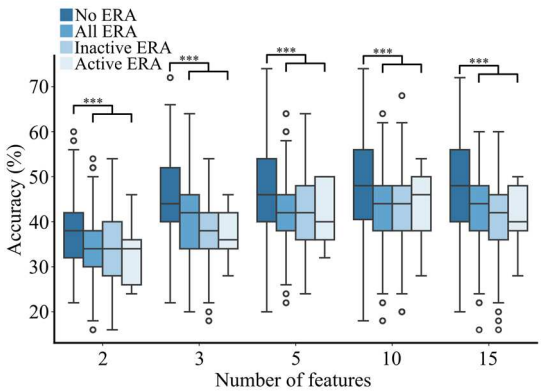

Fig. 2. Accuracy distribution per group across feature quantities (5) and ML classifiers (7) when aggregating all task combinations (26) for balanced scenario (n=5/group). Even if subtle, a consistent pattern emerges across groups in terms of both performance magnitude and distribution for the ERA groups vs No ERA.

eliminate unwanted environmental artifacts and focus on the signals generated around the motor input frequency. We then selected the central 5 second windows within the 10 seconds of each task to ensure we consistently excluded any task transition artifacts and had all the data separated into our 5 main tasks.

After pre-processing our data signals, we proceeded to extract the features to use for our ML classifier training. We extracted 48 features of 4 different types, each to capture a distinct signal component: full-signal, bursts, upper-envelope, and total-envelope. After extraction, we ran a principal component analysis (PCA) for dimensionality reduction while preserving variance.

Once we had the features to use, we proceeded to train 7 different ML classifiers to predict locomotion tasks (Fig. 1(d)) instead of the traditional direct train to predict if disease is present and or its severity. Our ML models were trained with data from one group and predicted for patients within that same group. We used leave-one-subject-out (LOSO) to avoid any data leakage and compiled the results for all patients as the test patient. Because we had groups that had considerably fewer patient data, we ran our analysis in a *balanced* way (limiting all groups to 5 patients [smallest group]). However, to leverage the additional patient data, we repeated the training with 3 randomly selected patient groups taken from the larger pool of available patients per group. Averaging the results from the 3 random patient mixes increases our confidence in the results.

To evaluate our classifier model performance, we calculated model accuracy when iterating over 5 different PCA component/feature quantities (2, 3, 5, 10, 15), 7 ML classifiers (Random Forest, SVM, KNN, Logistic Regression, Decision Tree, Gradient Boosting, and MLP), all 26 unique task combinations (combinations of sit, toe up, heel up, stand, and tip toe with 2 or more tasks at a time), and 3 different random patient combinations. This allowed us to capture overarching performance trends by comparing the distribution of performance values between groups.

Finally, to assess group-wise differences in classification performance, we computed average accuracy performance across classifiers, task combinations, and feature counts. We conducted pairwise comparisons between groups using independent two-sample t-tests with Bonferroni corrections and Cohen's d effect size across accuracy distributions. We also considered a more conservative test case using only two features and the five tasks only once (instead of aggregating all combinations) to capture performance difference between groups all with balanced subject counts (n=5).

III. Results and Discussion

Our work in this study suggests for the first time the possibility of disease-related physiology proportionally affecting ML classification performance, therefore elucidating the possible presence of quantifiable biomarkers able to help quantify JIA/ERA severity.

A. Classification Performance

When aggregating results across different numbers of PCA-derived feature quantities, multiple classifiers, various locomotion task combinations, and subject compositions, significant differences emerged among groups. Classifier performance consistently decreased from *No ERA* to *all three ERA* groups, suggesting a clear link between disease severity and predictive accuracy. Specifically, the *No ERA* group consistently achieved the highest classification performance and *Active ERA* the lowest (Table 2) (Fig. 2).

Statistical analyses corroborated these observations, with significant group-wise differences emerging with corrected $p < 0.001$ and Cohen's $d \geq 0.5$ when comparing *No ERA* vs *all three ERA* subgroups (Fig. 2). Moreover, selecting a conservative sample parameter set for a focused practical approach (Fig. 3) yielded an even greater performance difference $p = 0.002$ and $d = 2.05$ despite the underpowered sample size. These results underscore the sensitivity of active acoustic vibrational responses to physiological alterations linked to inflammation.

Although further exploration of selected features, ML model tuning, and expansion of task are all possible, results did not show significant performance variation between groups when comparing across feature quantities, ML model parameters, or task combinations. Conversely, note that the consistent variability in performance within groups, could be indicative of underlying physiological heterogeneity, symptom burden variability, and more importantly, potential clinical diagnostic inconsistencies. Ultimately, even if slight, the consistent decrease in performance across the wide range of combinations suggests the existence of underlying physiological differences.

TABLE III. ACCURACY PERFORMANCE ACROSS CLASSIFIERS FOR SUBJECT BALANCED CASE WITH 2 FEATURES AND ALL TASKS

Classifier	*No ERA*	*All ERA*
MLP	0.460	0.333
Logistic Regression	0.493	0.380
Random Forest	0.467	0.360
Gradient Boosting	0.487	0.400
KNN	0.433	0.367
SVM	0.433	0.380
Decision Tree	0.400	0.387
Average	***0.453***	***0.372***

[a.] Average accuracy values per group for two features and all five tasks across all classifiers for balanced patient quantity showing consistently lower performance on ERA cases for all classifiers.

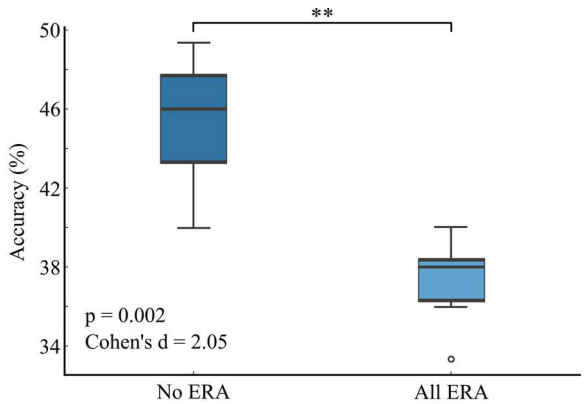

Fig. 3. Accuracy distribution per group across ML classifiers (7) for only 2 features and balanced groups (*n*=5/group) when comparing all 5 tasks with no combinations. This simpler comparison yields a clear differentiation between groups with clearly lower performance for ERA vs No ERA.

B. *Significance and Limitations*

These results suggest that inflammation-related alterations, such as changes in tendon stiffness or damping properties, may modulate vibrational signal characteristics, thus impairing classifier predictions of locomotion tasks. This underscores the potential of ML-based active acoustic sensing as a non-invasive biomarker for disease activity and severity assessment in pediatric rheumatology.

Despite the potential benefits, our current models remain limited by relatively small patient sample size (overall), demographic imbalances (unbalanced disease states), and the risk of overfitting (even if mitigated via cross-validation). Furthermore, the subjective nature of the self-reported data on symptoms has a real effect on classification and is a potential cause of bias as the perceived symptom might defer from the underlying physiological effects. Despite our method classifying tasks instead of symptoms and therefore potentially minimizing the effects of group labeling variability, it is by expanding to larger, more diverse data sets including additional AA and clinical data that we can enable future validation and improved generalizability.

IV. Conclusion and Future Work

Our findings establish that physiological changes associated with JIA and ERA can influence the performance of ML classifiers in predicting locomotion tasks. The consistent decline in classification metric performance with increasing disease severity suggests that active acoustic sensing captures underlying biomechanical alterations linked to inflammation. Despite these promising results, the study has its limitations pertaining to sample size, demographic imbalances, disease-state labeling, and clarity on performance requirements for clinical relevance. Hence, moving forward, larger and more diverse data sets, coupled with objective imaging modalities, will be essential to validate and refine these findings. Optimizing data acquisition protocols, including transducer placement and multi-axis accelerometry, could further bolster classifier robustness. In parallel, investigating specific pathophysiological factors (e.g., tendon stiffness, localized enthesitis) and correlating them with additional clinician, self-evaluation, and imaging data may enhance our understanding of the underlying mechanisms. Ultimately, this work sets the stage for more advanced, non-invasive diagnostic approaches that can improve early detection, monitoring, and personalized management of pediatric rheumatologic conditions.

Acknowledgments and Disclosures

The authors thank IRL Lab members for ideation support. O. T. Inan is co-founder of Arthroba Inc., a start-up focused on commercializing wearable joint health monitoring technologies for osteoarthritis. No other authors have competing interests.

References

[1] A. Ravelli and A. Martini, “Juvenile idiopathic arthritis,” *The Lancet*, vol. 369, no. 9563, pp. 767–778, Mar. 2007, doi: 10.1016/S0140-6736(07)60363-8.

[2] L. N. Zaripova, A. Midgley, S. E. Christmas, M. W. Beresford, E. M. Baildam, and R. A. Oldershaw, “Juvenile idiopathic arthritis: from aetiopathogenesis to therapeutic approaches,” *Pediatr. Rheumatol.*, vol. 19, no. 1, p. 135, Aug. 2021, doi: 10.1186/s12969-021-00629-8.

[3] P. F. Weiss, “Diagnosis and treatment of enthesitis-related arthritis,” *Adolesc. Health Med. Ther.*, vol. 3, pp. 67–74, Jun. 2012, doi: 10.2147/AHMT.S25872.

[4] P. F. Weiss *et al.*, “Enthesitis in an inception cohort of enthesitis-related arthritis,” *Arthritis Care Res.*, vol. 63, no. 9, pp. 1307–1312, 2011, doi: 10.1002/acr.20508.

[5] A. Consolaro *et al.*, “Clinical outcome measures in juvenile idiopathic arthritis,” *Pediatr. Rheumatol.*, vol. 14, no. 1, p. 23, Apr. 2016, doi: 10.1186/s12969-016-0085-5.

[6] M. Tarsia, M. Zajc Avramovič, A. Gazikalović, D. Ključevšek, and T. Avčin, “A clinical perspective on imaging in juvenile idiopathic arthritis,” *Pediatr. Radiol.*, vol. 54, no. 4, pp. 490–504, Apr. 2024, doi: 10.1007/s00247-023-05815-2.

[7] P. F. Weiss, R. Xiao, D. M. Biko, and N. A. Chauvin, “Assessment of Sacroiliitis at Diagnosis of Juvenile Spondyloarthritis by Radiography, Magnetic Resonance Imaging, and Clinical Examination,” *Arthritis Care Res.*, vol. 68, no. 2, pp. 187–194, 2016, doi: 10.1002/acr.22665.

[8] R. Hemke, N. Herregods, J. L. Jaremko, and M. Maas, “Imaging of Juvenile Idiopathic Arthritis,” *Radiol. Clin. North Am.*, vol. 62, no. 5, pp. 889–902, Sep. 2024, doi: 10.1016/j.rcl.2024.02.009.

[9] Q. Goossens *et al.*, “Active Vibrational Achilles Tendon Sensing for Identifying and Characterizing Inflammatory Symptomatology in Enthesitis Related Arthritis,” *IEEE Trans. Biomed. Eng.*, vol. 72, no. 2, pp. 645–654, Feb. 2025, doi: 10.1109/TBME.2024.3466831.

[10] N. Bolus, H.-K. Jeong, B. M. Blaho, M. Safaei, A. Young, and O. Inan, “Fit to Burst: Toward Noninvasive Estimation of Achilles Tendon Load Using Burst Vibrations,” *IEEE Trans. Biomed. Eng.*, pp. 1–1, Jun. 2020, doi: 10.1109/tbme.2020.3005353.

[11] J. A. Martin *et al.*, “Gauging force by tapping tendons,” *Nat. Commun.*, vol. 9, no. 1, pp. 1–9, Dec. 2018, doi: 10.1038/s41467-018-03797-6.

[12] L. G. Rosa, J. S. Zia, O. T. Inan, and G. S. Sawicki, “Machine learning to extract muscle fascicle length changes from dynamic ultrasound images in real-time,” *PLOS ONE*, vol. 16, no. 5, p. e0246611, May 2021, doi: 10.1371/journal.pone.0246611.