How are Patients Moving Outside of the Clinic? Categorizing Activities using Remotely Captured Wearable Data and Machine Learning

Megan V. McCabe
Thayer School of Engineering at Dartmouth College, Hanover, NH
Megan.V.McCabe.20@dartmouth.edu

Douglas W. Van Citters
Thayer School of Engineering at Dartmouth College, Hanover, NH
Douglas.W.Van.Citters@dartmouth.edu

Ryan M. Chapman
Thayer School of Engineering at Dartmouth College, Hanover, NH
Ryan.M.Chapman@dartmouth.edu

ABSTRACT

Introduction: *Over 1M joint replacements occur annually, allowing patients to complete activities of daily living (ADLs). Typical recovery assessments fail to characterize non-clinic performance. Machine learning and wearable inertial measurement units (IMUs) facilitate non-clinic human activity recognition (HAR). Accordingly, we 1) developed/optimized activity classification algorithms from IMU data and 2) investigated training set properties.*

Methods: *10 healthy subjects participated, completing laboratory (training data) and separately out-of-lab capture for one day (8-12 hrs; test data). For all, IMUs were donned on dominant legs (f_s=128Hz) continuously capturing activities of three types (Sedentary: Lying Down, Sitting, Standing; Ambulation: Walking, Jogging, Running; Stairs: Ascent, Descent). Data were labeled, low-pass filtered, and feature extracted. Training features were tabulated (individual) and combined (cohort). Medium decision trees (max splits=20) were trained using individual/cohort sets testing all non-lab data. Three approaches were* *completed: 1) Coarse (sedentary, ambulation, stairs), 2) Fine (all activities), and 3) Optimized (bi-layered). Optimized first separated sedentary from non-sedentary (layer #1), then within each using separate algorithms (layer #2). Confusion matrices were created for each experiment within approach (coarse, fine, optimized), patient quantity (individual, cohort), and feature quantity (full, reduced).*

Results: *Individual training resulted in high coarse approach accuracy (95%), suggesting subject-specific movements. Reducing features improved coarse cohort accuracy (+20%) and fine cohort/individual training (+21%/+7%), indicating full features overfit training data. Optimized approach achieved >80% accuracy during full feature cohort training, but struggled classifying sitting (True positive rate, TPR: 58%).*

Discussion: *High optimized approach accuracy supports decision trees for HAR. Full feature cohort training yielded highest accuracy. Individual training performed best during coarse/fine classification, suggesting coarse before fine classification removes subject-specific movement confusion. Low TPR for sitting implies lab/non-lab performance differs. Future work will capture patient populations and investigate out-of-lab data training out-of-lab classification. This work is an exciting step toward continuous patient monitoring and clinical decision making from 'real-world' data.*

Keywords: Inertial Measurement Units, Machine Learning, Decision Trees, Human Activity Recognition

1. INTRODUCTION:

Osteoarthritis (OA) afflicts about 10% of American adults, causing patients debilitating pain and costing the United States healthcare system an estimated $45 billion annually [1]. Total joint arthroplasty (TJA) is the terminal treatment for individuals suffering from OA. TJA is highly successful at alleviating pain and allowing patients to regain the ability to perform typical activities of daily living (ADLs), like walking and climbing stairs [2]. While most believe postoperative rehabilitation is critical for recovery, existing recovery assessment methods have inherent limitations. Typically, patient reported outcome measures (PROMs) and clinical range of motion (ROM) are used, but PROMs are subjective and prior work measuring ROM before/after surgery does not always show significant change [3–5]. Perhaps most importantly, these methods offer little insight with few discrete data points collected almost exclusively in the clinic. This information largely fails to characterize the entire patient recovery picture. Ideally, clinicians would not only capture clinical data, but also obtain information regarding the quality and frequency of patient movement during ADLs completed in home environments. Small, wearable devices and activity classification of patient motion data via machine learning (ML) are steps towards achieving this goal.

Inertial measurement units (IMUs) are one example of available patient centered, motion-tracking technology. IMUs are highly portable, wearable devices that measure linear acceleration, angular velocity, and magnetic field strength that offer the ability to capture both laboratory and out-of-laboratory motion data. Biomechanics studies in a number of disciplines (e.g. athletic performance, pathologic movement patterns) increasingly use IMUs in place of expensive motion capture systems, which are constrained to well-controlled laboratory or clinic settings [6–9]. Recent work has shown IMU's potential for analyzing longer-term patient biomechanics prior to and following TJA. Chapman et al. developed, validated, and implemented an IMU-based measurement method for capturing longduration ROM (i.e. 8-12 hours/day, >6 weeks continuously) in healthy individuals and patients undergoing total knee arthroplasty (TKA) and total shoulder arthroplasty (TSA), respectively [4,10]. Using IMU data to classify patient activity at similar critical time points could more wholly complete the patient recovery picture.

Human activity recognition (HAR), or classifying motion data by activity type, using ML is a growing field. Motion data is first collected and fed into algorithms for 'training', allowing prediction of classes (or activity types) from newly captured, experimental motion data. IMU technology combined with ML can be leveraged to accomplish this task. Foundational work completing HAR in this way has investigated sensor placement, training data collection/processing, and algorithm development [11–13]. Chinimilli et al. achieved high accuracy (>99%) using IMUs and pressure sensing insoles to classify six ADLs (i.e. sit, stand, walk, jog, ascend/descend stairs). However, the approach utilized six sensors (4 IMUs, 2 pressure insoles) and required significant computation capacity [14]. Achieving high accuracy while reducing computational complexity is likely critical for scalability and real world implementation. Additionally, Attal et al. found supervised learning techniques (those establishing ground truth with labeled data) achieve higher accuracy than unsupervised techniques (those drawing inferences from unlabeled data), but sacrifice computational speed (i.e. slower) [15]. Another ML algorithm type with high accuracy that ***does not*** compromise computational speed is decision trees [16]. During training, decision trees repeatedly partition data creating splits that optimize class separation. Recent work suggests decision trees are promising for HAR due to their implementation ease and sustained performance using reduced feature sets [13]. The studies that exist traditionally implement one decision tree across all activities. To our knowledge, no studies exist developing optimized approaches using separate decision trees for different intensity activities.

Accordingly, we investigated optimizing HAR using decision trees and IMU motion data. Our objectives were 1) to develop and optimize a method for HAR and 2) to consider the effect of two size-related parameters on classification accuracy. We first developed and compared two classification approaches: 1) Coarse (i.e. 3 activity classes: sedentary, ambulation, stairs) and 2) Fine (i.e. 8 activity classes: lying down, sitting, standing, walking, jogging, running, stair ascent, stair descent). The third classification approach (‘optimized’ approach) balanced tradeoffs between specificity and accuracy by training a two-layered algorithm that initially separated sedentary and non-sedentary activity, then classified *within* sedentary and non-sedentary activities, respectively. All approaches used laboratory-captured IMU data to create training sets and develop decision trees to predict activities from non-laboratory, ‘real-world’ data. We also considered two training set parameters including subject quantity within training (individual vs. cohort training sets) and feature set size (full vs. reduced feature sets). We hypothesized 1) coarse classification would yield highest accuracy, 2) ‘optimized’ classification would yield accuracy >80% [15], 3) cohort training would be more accurate than individual training, and 4) reduced feature sets would be as accurate as full feature sets.

2. METHODS:

2.1 Methods – Overview

Activities were chosen spanning a variety of intensities often encountered by young, healthy individuals during typical daily life. Three categories of activities were chosen: 1) Sedentary (lying down, sitting, standing), 2) Ambulatory (walking, jogging, running), and 3) Stairs (ascent, descent). Full study processes are illustrated in Figure 1A. Broadly, this consisted of data collection, data processing, and activity classification.

2.2 Methods – Data Collection

Following IRB approval, 10 young, healthy subjects (28±7 years, 5M) were recruited from the local university student population. Inclusion criteria were age ≥18 years, no musculoskeletal or neuromuscular impairments impacting the lower extremity, no terminal illness resulting in death within one year, clinical full knee extension (<5° flexion) and flexion (>120°), and complete participation in the study [17,18]. For all data captures, IMUs were donned on the dominant lower extremity (as assessed by the Waterloo Footedness Questionnaire) anteromedial shaft and lateral thigh (Figure 1B: APDM, Inc.; Portland, OR; f_s=128Hz) [19]. IMU data were continuously captured during all activities including 3D linear acceleration, angular velocity, and magnetic field strength. Subjects first completed in-lab capture performing prescribed movements for training data creation. Prescribed movements included activities performed for time (standing, sitting, lying down, walking, jogging, running) or completion (stair ascent, stair descent). Specifically, data were captured during 3 repetitions each of treadmill (ProForm 505 CST, Icon Health & Fitness, Logan, Utah) walking at 1.0, 1.5, and 2.0 MPH, respectively, for 45s; standing, sitting, and lying down, respectively, for 10s; and ascending/descending one standard flight of stairs (10 steps). Subjects then returned another day to complete out-of-lab capture for experimental test data creation. They arrived at 8AM, were refitted with IMUs as before, and wore them for the duration of the day (8-12 hrs). Throughout, subjects logged activities in writing (e.g. 0800-0808: Sitting, 0808-0815: Walking, etc.). Activities from out-of-lab capture not collected during in-lab capture (e.g. cycling, squatting) were logged as 'n/a' and excluded from further analyses.

2.3 Methods – Data Processing

Custom MATLAB scripts (MATLAB R2018b, MathWorks, Natick, MA) were used to label, filter, and extract features from IMU data, transforming raw data into feature tables suitable for classification algorithms. Ground truth was established by labeling each data point with an activity classification (e.g. 'walk', 'jog', 'n/a'). For in-lab capture, labels corresponded to each trial's known activity. For out-of-lab capture, data were labeled according to subjects' written activity logs. IMU data were low pass filtered using a Butterworth filter (f_{cutoff}=5Hz) to reduce soft tissue noise. Features were extracted using sliding-window techniques segmenting data by duration from the first data point and sliding the window to capture overlapping data portions. Prior work indicates window lengths should be representative of the activity captured, with shorter window lengths (1-1.5s) sufficient for classifying simple non-gait tasks [12]. More complex activities, such as gait, require longer windows for extracting features [20,21]. Accordingly, sedentary, stair, and ambulation activities were segmented using 1s, 5s, and 15s windows, respectively. Eighteen features from each IMU were extracted for each window (6 per sensing modality: accelerometer, gyroscope, and magnetometer). These features (Figure 1C) were chosen based on prior HAR use [22]. Finally, in-lab and out-of-lab features were tabulated by individual (n=10 tables) and separately combined as a cohort (n=1 table) to create 11 total training and corresponding test tables. Individual training sets were composed of one subject's data used to train algorithms to classify the same subject's out-of-lab data. Cohort training sets combined all individual subject data, training algorithms to classify combined subject out-of-lab data. Full feature sets used all 36 features while reduced feature sets used only the two most important features for training the classifier. Confusion matrices were created for each type of training set (individual vs. cohort) and each classification approach (full vs. reduced features). The series of experiments completed are displayed in Figure 2.

2.4 Methods – Activity Classification

2.4.1 Activity Classification: Overview

Activity classification consisted of training, testing, assessing classification accuracy, and quantifying feature importance for each approach: coarse, fine, and optimized classification. Decision trees were trained using MATLAB Classification Learner with ≤20 splits and 25% of data held out for validation. All algorithms were first trained using full feature sets. Feature importance was then computed dividing feature split by summed square error. Any algorithm with <3 classification features needed no further testing using reduced feature sets. For the remaining algorithms, feature importance was used to reduce feature tables to the 2 most important metrics. Reduced feature tables were used to train/test an additional set of decision trees. Overall and activity specific classification accuracy were computed creating confusion matrices for individual and cohort data using all features and separately reduced features.

2.4.2 Activity Classification: Coarse Classification

Initially, activities were grouped into coarse categories by relative intensity. This included stationary activities (lying down, sitting, standing) grouped in a ‘sedentary’ category, a non-stairs ambulation (walking, jogging, running) category, and a stairs (ascent, descent) category. As described previously, individual and cohort feature tables were generated from in-lab capture and used to assess out-of-lab performance. Furthermore, full feature tables and reduced feature tables were used to assess both individual and cohort coarse classification accuracy.

2.4.3 Activity Classification: Fine Classification

In contrast to coarse classification, fine classification was implemented in an attempt to classify all individual activities (lying down, sitting, standing, walking, jogging, running, stair ascent, stair descent). As before, individual and cohort classification accuracy were quantified for both full/reduced feature tables.

2.4.4 Activity Classification: Optimized Classification

The optimized approach followed the same general classification process as coarse and fine approaches. However, we utilized perceived strengths of each previous approach to develop the optimized approach herein. Specifically, coarse classification yielded high accuracy about few activities, while fine classification struggled distinguishing within sedentary activities (see Results). Accordingly, a bi-layered approach was developed wherein sedentary data was initially separated from non-sedentary data using the original 36-features (coarse: >95% accuracy) followed by subsequent classification within each sub-category. Following first layer decision tree, additional separate decision tree algorithms were developed to train/test within sedentary/non-sedentary activities. Non-sedentary algorithm used the original 36-feature set. However, this was not appropriate for categorizing sedentary activities. Accordingly, sedentary algorithm utilized 6 new features (average acceleration of X, Y, & Z axes of thigh & shank IMUs) over the window extraction length. Again, each classification style (individual vs. cohort) was tested.

3. RESULTS:

3.1 Results – Coarse Classification

Using full feature sets, coarse classification achieved high accuracy for individual training (95%) and moderate accuracy for cohort training (75%) (Figure 3A). Retraining with the two most important features (thigh acceleration Σ^2 <25th and <75th %) yielded significant cohort classification accuracy improvement (+20% to 95%) but did not impact individual classification performance (Figure 3B). With reduced features during cohort training, the algorithm was better able to distinguish ambulation from stairs (stairs recall +66% to 89% and ambulation precision +57% to 88%).

3.2 Results – Fine Classification

In contrast, the fine approach achieved low accuracy (<50%) for all training iterations in both individual (Figure 4) and cohort training (Figure 5). Reduced features performed better than all features, with the individual reduced training set performing best (47%). Distinguishing between sedentary activities (highlighted by dashed box) was the largest error source for all fine algorithms. Because sedentary activities represented the majority of the test data, this likely reduced overall classification accuracy as a result. For cohort training, misclassification of stairs comprised a significant source of error as well (highlighted by dashed box), with stair ascention exhibiting 0% classification accuracy.

3.3 Results – Optimized Classification

The first layer of the optimized approach (separating sedentary/non-sedentary activities) exhibited high overall accuracy (individual: 98%, cohort: 96%). In the second analysis layer, cohort training outperformed individual training within both sedentary (79% vs. 67%) and non-sedentary algorithms (85% vs. 65%) (Figure 6, 7). Cohort training with reduced features performed ~10% worse than training with all features (Figure 8). The true positive rate (TPR; % data correctly identified) within sedentary activities was lowest for sitting for both individual/cohort full feature training (52%/58%) resultant from misclassification as lying down. For individual training with all features, walking had the lowest TPR among non-sedentary activities (TPR: 54%). Within non-sedentary activities with cohort training, the highest error occurred due to misclassification of jogging/running as stairs (TPR: 75%).

4. DISCUSSION:

4.1 Discussion – Overview

Currently, patient function post-TJA is primarily characterized by PROMs and ROM measurements, both of which are discrete and inherently limited measures collected in-clinic. Capturing 'real world', continuous patient motion data likely informs better post-operative care. This is possible using IMUs to capture data and ML to process that data. To this end, we sought to create an HAR process using IMUs and decision trees, ultimately developing a bi-layered optimized classifier balancing tradeoffs between specificity and accuracy from two initial approaches: 1) Coarse approach which showed high accuracy about little information and 2) Fine approach which yielded low accuracy about detailed information. Prior to experimentation of the three approaches and two size related variables (cohort and feature quantity) we hypothesized 1) coarse classification would yield highest accuracy of three approaches, 2) 'optimized' classification would yield accuracy >80% [15], 3) cohort training would be more accurate than individual training, and 4) reduced feature sets would be as accurate as full feature sets.

4.2 Discussion – Coarse and Fine Classification

As hypothesized, coarse classification achieved the highest accuracy for both cohort/individual training (95%) whereas fine classification achieved low accuracy (<50%). Clearly, a tradeoff between accuracy and specificity exists in this implementation. The high accuracy of coarse classification has little meaning because classifying sedentary versus non-sedentary activity provides little information. Moreover, accomplishing this task with ML is likely 'overkill' and could be completed through simpler approaches. However, information gleaned via fine classification is unreliabile due to low accuracy.

Individual training outperformed cohort training in both coarse and fine approaches, countering our hypothesis that cohort training would result in higher accuracy. However, cohort training accuracy approached or equaled that of individual training for both approaches with reduced features, supporting our hypothesis that feature reduction would aid in overall accuracy. Feature reduction ('pruning' decision trees), likely prevented overfitting algorithms to laboratory data allowing improved classification accuracy [23]. Finally, individual training's greater accuracy compared to cohort training as well as improved accuracy of cohort training after pruning suggest movement is subject specific (subject #1 walking ≠ subject #2 walking).

4.3 Discussion – Optimized Classification

Using full feature, cohort training sets, optimized classification achieved accuracy >80%, supporting our hypothesis. In contrast to coarse/fine classification, optimized classification performed best via cohort training, especially for non-sedentary activities (+20% vs. individual). This supports our hypothesis that cohort data would yield higher accuracy and suggests that two-stage classification reduces algorithm confusion due to subject specific movements. The bi-layer approach breaks down the challenge of classifying all activities at once, allowing the algorithm to first create splits that optimize separation of broad activity classes before attempting fine actvitity classification. Optimized approach sedentary classification used only six features, which is comparable to coarse reduced feature training that also achieved high accuracy. The success of the optimized approach separating sedentary/non-sedentary activities indicates few features may be sufficient to distinguish sedentary activities.

The application of wearable technology and ML to HAR research continues to be a challenging and rapidly evolving field. Our approach should be considered in the context of other recent investigations, including those exploring the use of neural networks and/or smartphone wearables [20]. Neural networks are highly accurate classifying moderately detailed activities (>95% for 6 activity classes), but have higher computational demand and are less intuitive than decision trees [25]. Our work shows layered decision trees, classifying broader activity classes with a preliminary algorithm (i.e. sedentary vs. non-sedentary) before classifying finer classes (i.e. lying down, sitting, walking), may represent an opportunity to increase accuracy while keeping algorithms simple. This work is an exciting first step toward continuous patient monitoring and making clinical decisions from non-clinic/non-lab 'real-world' data. The results represent foundational work that can be applied to patient populations in the future and may have implications on establishing non-clinic PT compliance.

4.4 Limitations

Individuals in this study are likely different than patient populations. Future work needs to investigate pathologic cohorts and consider training within demographics (e.g. age). Differences between in-lab and out-of-lab data (i.e. sitting with crossed vs. straight legs) likely caused significant error. Further studies should investigate potential benefits of using training sets created from out-of-lab data compared to in-lab data. Reliance on subjective written patient activity logs for tracking out-of-lab data likely contributes to error as well and may be avoided if patient movement could be tracked by another system, such as a small wearable camera. Finally, the work in this study could be extended to implement a system by which classification occurs in real time to allow for more rapid movement analysis.

5. Conclusions

Improved patient motion tracking outside of the lab/clinic could inform better post-operative care. This is possible through ML and IMUs. This study developed a simple, effective bilayer HAR approach using decision trees, separating coarse activities (i.e. sedentary v. non-sedentary) before differentiating nuanced activities (i.e. Sedentary: lying down, sitting, standing and Non-Sedentary: walking, jogging/running, stairs). The approach achieved >80% accuracy using cohort, full feature in-lab data for training, but may be improved if out-of-lab data were incorporated for training as well.

Figure Captions List

Fig. 1 **A)** Overall study process flow, **B)** IMU placement, and **C)** feature list for decision trees

Fig. 2 Experimental series including coarse, fine, and optimized classification algorithms performed with individual and cohort training as well as full and reduced feature sets during coarse/fine classification

Fig. 3 Results for coarse approach, **A**) full feature set and **B**) reduced feature set with individual (top panels) and cohort (bottom panels) training

Fig. 4 Results for fine classification individual training with **A**) full features and **B**) reduced features. Dashed boxes highlight misclassifications within sedentary activities.

Fig. 5 Results for fine classification cohort training with **A**) full features and **B**) reduced features. Dashed boxes highlight misclassifications within sedentary activities and dot-dash boxes highlight misclassifications within stair ascent/descent.

Fig. 6 Results for optimized approach, individual training. Layer 1 used full 36-feature set to separate sedentary from non-sedentary data. Layer 2 used full 36-feature set to classify non-sedentary activities and modified 6-feature set (average x, y, z acceleration) to classify sedentary activities.

Fig. 7 Results for optimized approach, cohort training. Layer 1 either utilized full 36-feature set (‘all features’) or original reduced feature set (‘reduced features') to separate sedentary from non-sedentary data. Layer 2 utilized either full 36-feature set (top panel) or original reduced feature set (bottom panel) to classify non-sedentary activities and modified 6-feature set (average x, y, z acceleration) used to classify sedentary activities (both top and bottom panels).

Figures

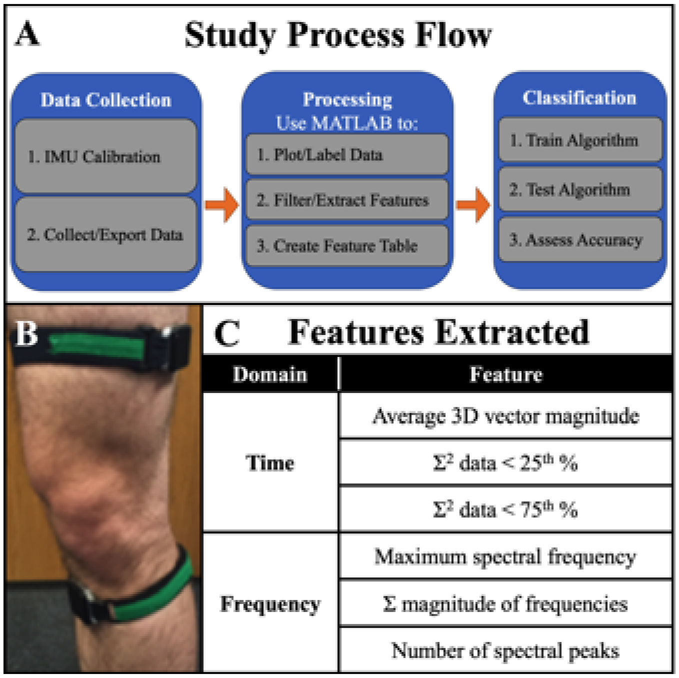


Figure 1. A) Overall study process flow, **B)** IMU placement, and **C)** feature list for decision trees

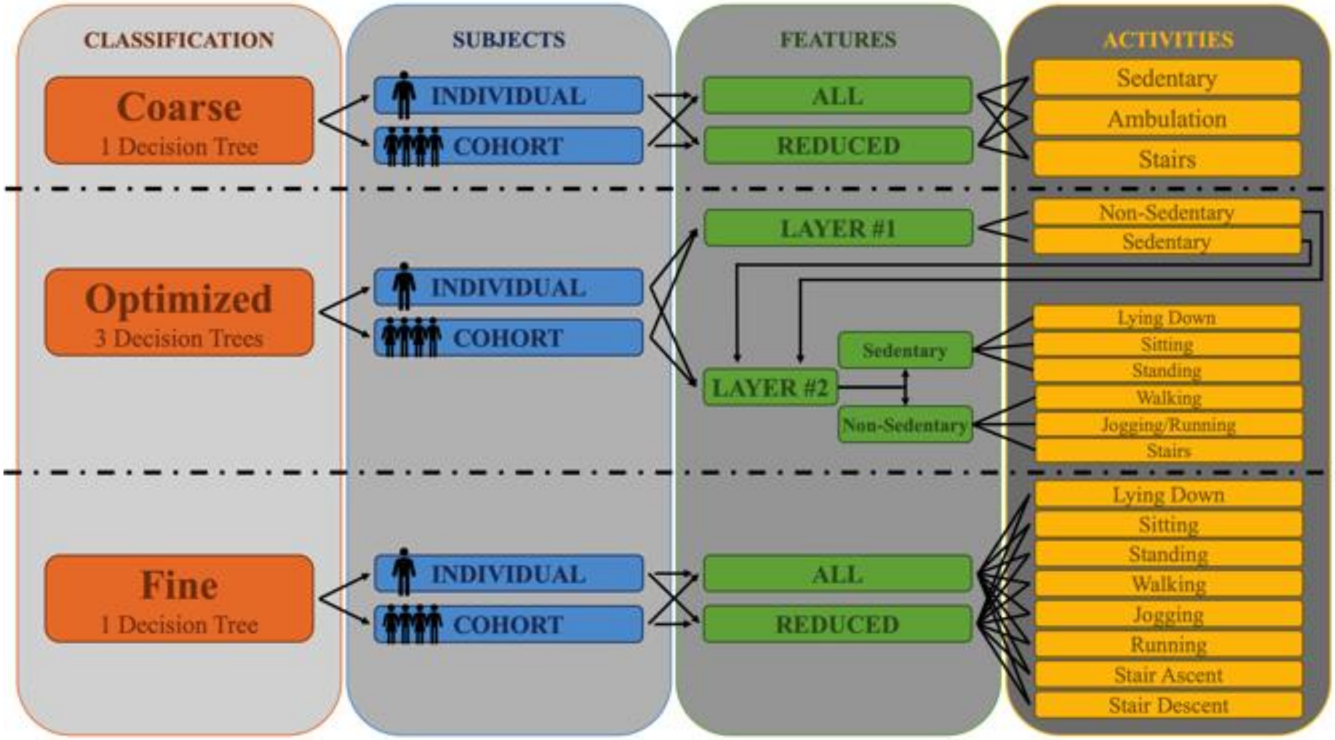
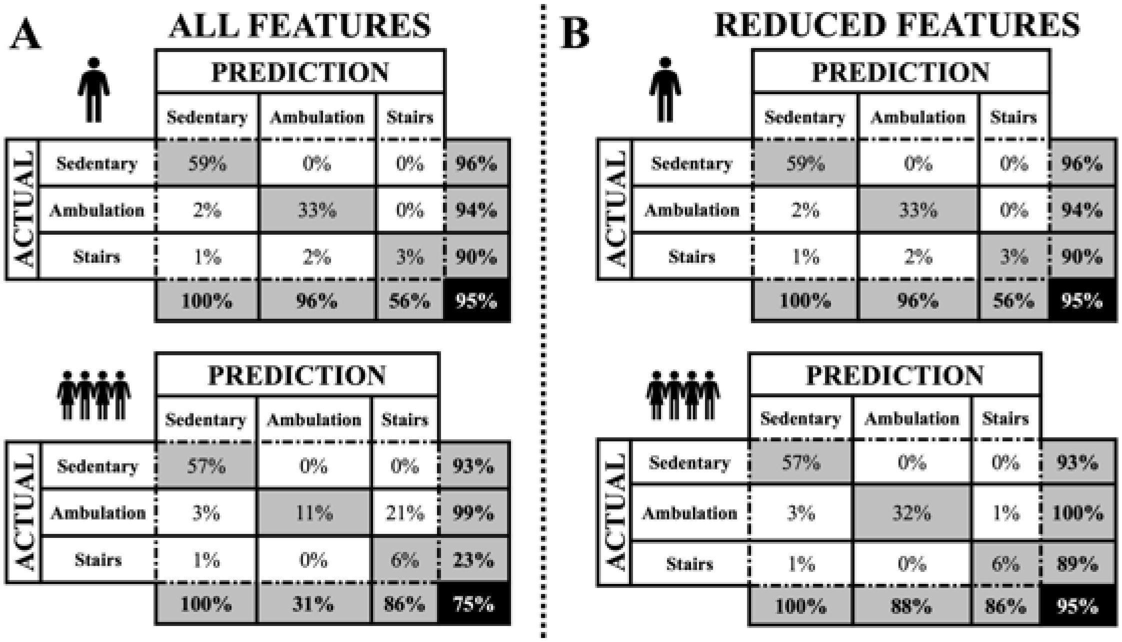


Figure 2. Experimental series including coarse, fine, and optimized classification algorithms performed with individual and cohort training as well as full and reduced feature sets during coarse/fine classification



A — ALL FEATURES (individual)

ACTUAL \ PREDICTION	Sedentary	Ambulation	Stairs	
Sedentary	59%	0%	0%	96%
Ambulation	2%	33%	0%	94%
Stairs	1%	2%	3%	90%
	100%	96%	56%	95%

A — ALL FEATURES (cohort)

ACTUAL \ PREDICTION	Sedentary	Ambulation	Stairs	
Sedentary	57%	0%	0%	93%
Ambulation	3%	11%	21%	99%
Stairs	1%	0%	6%	23%
	100%	31%	86%	75%

B — REDUCED FEATURES (individual)

ACTUAL \ PREDICTION	Sedentary	Ambulation	Stairs	
Sedentary	59%	0%	0%	96%
Ambulation	2%	33%	0%	94%
Stairs	1%	2%	3%	90%
	100%	96%	56%	95%

B — REDUCED FEATURES (cohort)

ACTUAL \ PREDICTION	Sedentary	Ambulation	Stairs	
Sedentary	57%	0%	0%	93%
Ambulation	3%	32%	1%	100%
Stairs	1%	0%	6%	89%
	100%	88%	86%	95%

Figure 3. Results for coarse approach, **A**) full feature set and **B**) reduced feature set with individual (top panels) and cohort (bottom panels) training

A

ALL FEATURES

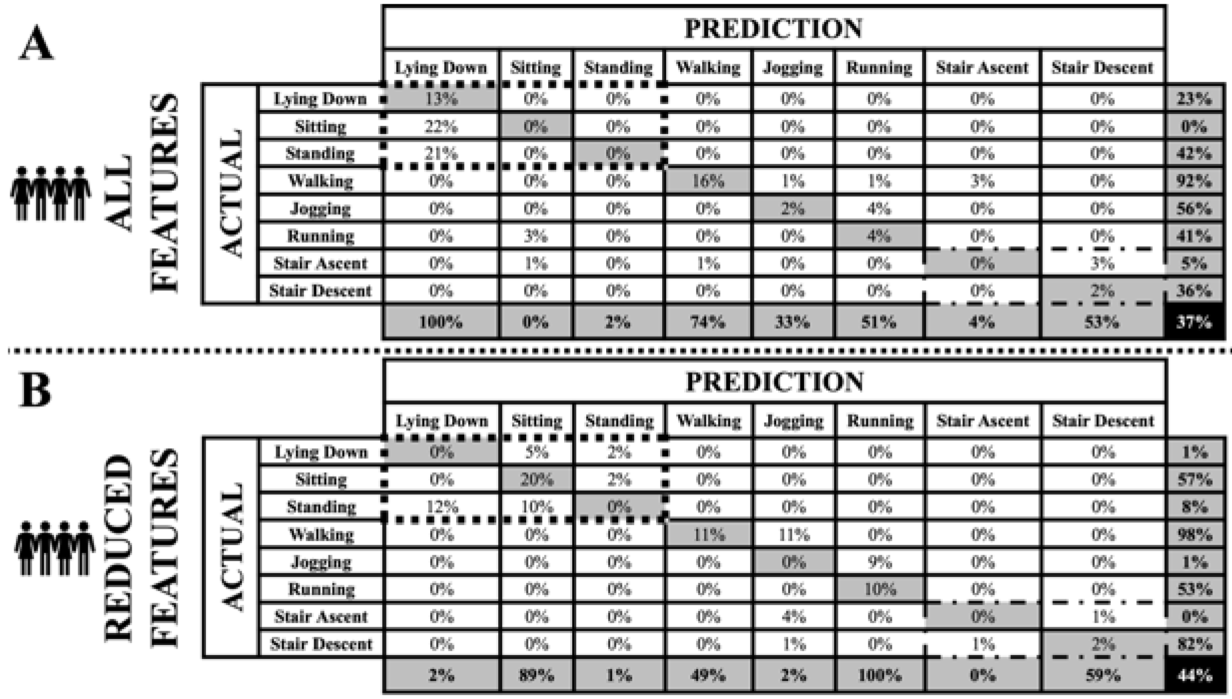
ACTUAL \ PREDICTION	Lying Down	Sitting	Standing	Walking	Jogging	Running	Stair Ascent	Stair Descent	
Lying Down	7%	2%	0%	0%	1%	0%	0%	0%	**11%**
Sitting	14%	3%	3%	0%	1%	0%	2%	0%	**18%**
Standing	15%	5%	1%	0%	1%	0%	2%	0%	**37%**
Walking	2%	0%	0%	9%	4%	0%	4%	4%	**71%**
Jogging	1%	0%	0%	1%	2%	1%	0%	0%	**25%**
Running	2%	0%	0%	1%	1%	1%	0%	0%	**9%**
Stair Ascent	1%	0%	0%	0%	0%	0%	1%	1%	**9%**
Stair Descent	0%	0%	0%	0%	0%	0%	2%	1%	**29%**
	50%	**13%**	**6%**	**41%**	**24%**	**20%**	**33%**	**31%**	**26%**

B

REDUCED FEATURES

ACTUAL \ PREDICTION	Lying Down	Sitting	Standing	Walking	Jogging	Running	Stair Ascent	Stair Descent	
Lying Down	7%	2%	1%	0%	0%	0%	0%	0%	**23%**
Sitting	11%	12%	2%	0%	0%	0%	0%	0%	**40%**
Standing	8%	6%	11%	0%	0%	0%	0%	0%	**59%**
Walking	2%	0%	0%	13%	2%	1%	1%	3%	**77%**
Jogging	1%	0%	0%	2%	1%	1%	0%	0%	**23%**
Running	2%	0%	0%	1%	1%	1%	0%	0%	**33%**
Stair Ascent	1%	0%	0%	1%	0%	0%	2%	0%	**29%**
Stair Descent	1%	0%	0%	0%	0%	0%	1%	1%	**33%**
	52%	**43%**	**39%**	**58%**	**21%**	**44%**	**37%**	**33%**	**47%**

Figure 4. Results for fine classification individual training with **A**) full features and **B**) reduced features. Dashed boxes highlight misclassifications within sedentary activities.



A — ALL FEATURES

ACTUAL \ PREDICTION	Lying Down	Sitting	Standing	Walking	Jogging	Running	Stair Ascent	Stair Descent	
Lying Down	13%	0%	0%	0%	0%	0%	0%	0%	23%
Sitting	22%	0%	0%	0%	0%	0%	0%	0%	0%
Standing	21%	0%	0%	0%	0%	0%	0%	0%	42%
Walking	0%	0%	0%	16%	1%	1%	3%	0%	92%
Jogging	0%	0%	0%	0%	2%	4%	0%	0%	56%
Running	0%	3%	0%	0%	0%	4%	0%	0%	41%
Stair Ascent	0%	1%	0%	1%	0%	0%	0%	3%	5%
Stair Descent	0%	0%	0%	0%	0%	0%	0%	2%	36%
	100%	0%	2%	74%	33%	51%	4%	53%	37%

B — REDUCED FEATURES

ACTUAL \ PREDICTION	Lying Down	Sitting	Standing	Walking	Jogging	Running	Stair Ascent	Stair Descent	
Lying Down	0%	5%	2%	0%	0%	0%	0%	0%	1%
Sitting	0%	20%	2%	0%	0%	0%	0%	0%	57%
Standing	12%	10%	0%	0%	0%	0%	0%	0%	8%
Walking	0%	0%	0%	11%	11%	0%	0%	0%	98%
Jogging	0%	0%	0%	0%	0%	9%	0%	0%	1%
Running	0%	0%	0%	0%	0%	10%	0%	0%	53%
Stair Ascent	0%	0%	0%	0%	4%	0%	0%	1%	0%
Stair Descent	0%	0%	0%	0%	1%	0%	1%	2%	82%
	2%	89%	1%	49%	2%	100%	0%	59%	44%

Figure 5. Results for fine classification cohort training with **A**) full features and **B**) reduced features. Dashed boxes highlight misclassifications within sedentary activities and dot-dash boxes highlight misclassifications within stair ascent/descent.

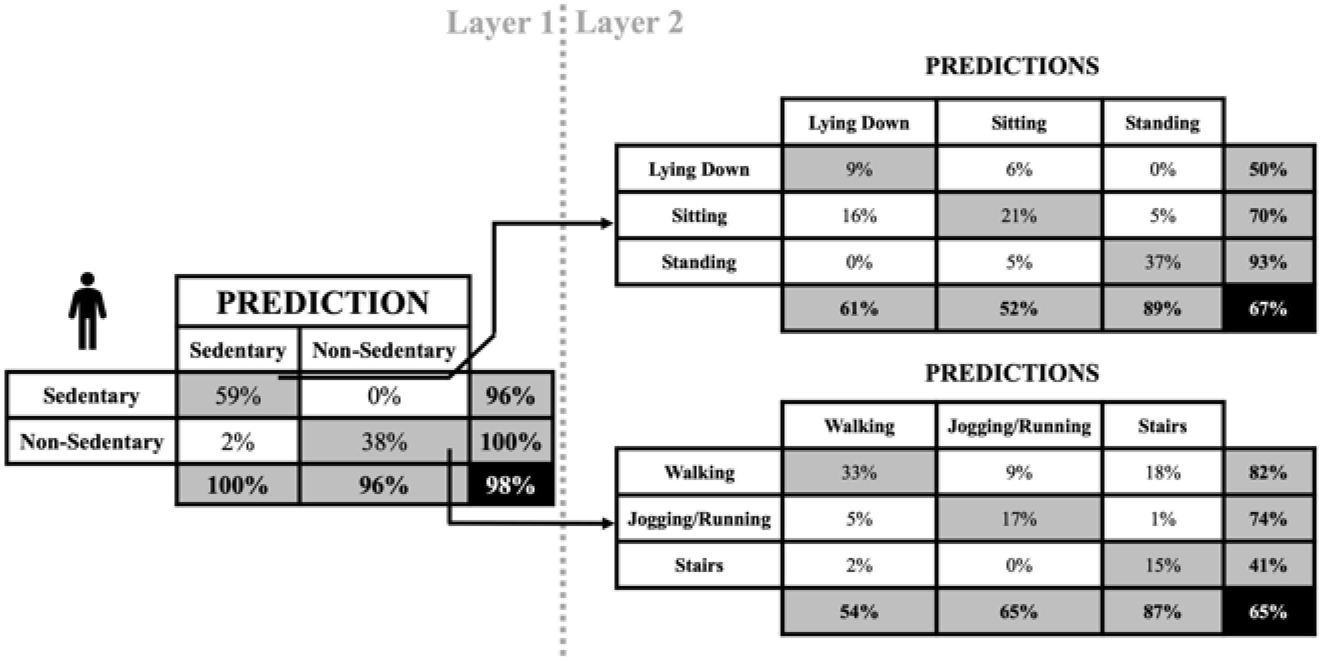


Figure 6. Results for optimized approach, individual training. Layer 1 used full 36-feature set to separate sedentary from non-sedentary data. Layer 2 used full 36-feature set to classify non-sedentary activities and modified 6-feature set (average x, y, z acceleration) to classify sedentary activities.

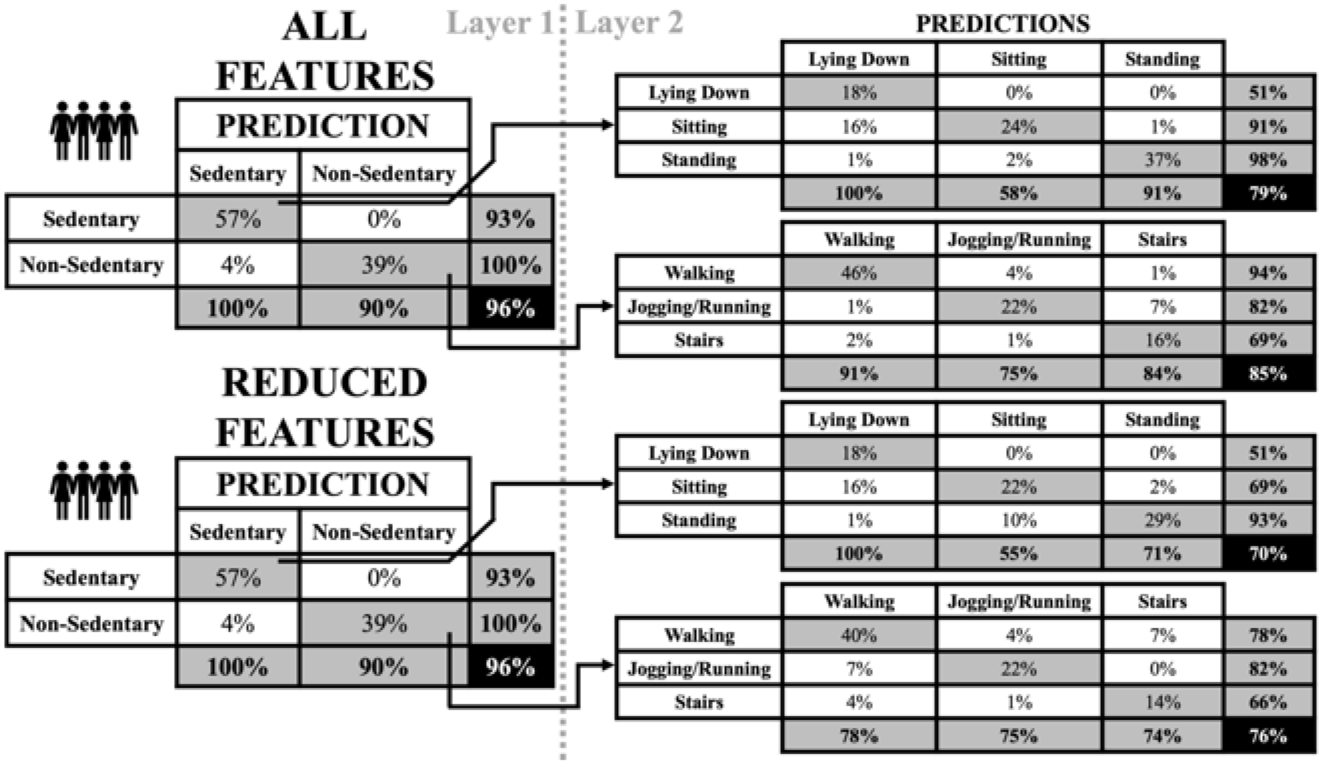


Figure 7. Results for optimized approach, cohort training. Layer 1 either utilized full 36-feature set ('all features') or original reduced feature set ('reduced features') to separate sedentary from non-sedentary data. Layer 2 utilized either full 36-feature set (top panel) or original reduced feature set (bottom panel) to classify non-sedentary activities and modified 6-feature set (average x, y, z acceleration) used to classify sedentary activities (both top and bottom panels).

References

[1] Zhao, X., Shah, D., Gandhi, K., Wei, W., Dwibedi, N., Webster, L., and Sambamoorthi, U., 2018, "Clinical, Humanistic, and Economic Burden of Osteoarthritis (OA) Among Noninstitutionalized Adults in the United States (US)," Value Heal., **21**(xxxx), p. S4. doi: 10.1016/j.jval.2018.04.012.

[2] Steiner C, Andrews R, Barrett M, W. A., 2012, *HCUP Projections. Mobility/Orthopedic Procedures 2003 to 2012*.

[3] Bade, M. J., Kohrt, W. M., and Stevens-Lapsley, J. E., 2010, "Outcomes before and after Total Knee Arthroplasty Compared to Healthy Adults," J. Orthop. Sports Phys. Ther., **40**(9), pp. 559–567. doi: 10.2519/jospt.2010.3317.

[4] Chapman, R. M., Moschetti, W. E., and Van Citters, D. W., 2019, "Stance and Swing Phase Knee Flexion Recover at Different Rates Following Total Knee Arthroplasty: An Inertial Measurement Unit Study," J. Biomech., **84**, pp. 129–137. doi: 10.1016/j.jbiomech.2018.12.027.

[5] Higuchi, H., Hatayama, K., Shimizu, M., Kobayashi, A., Kobayashi, T., and Takagishi, K., 2009, "Relationship between Joint Gap Difference and Range of Motion in Total Knee Arthroplasty: A Prospective Randomised Study between Different Platforms," Int. Orthop., **33**(4), pp. 997–1000. doi: 10.1007/s00264-009-0772-7.

[6] Tan, H., Wilson, A. M., and Lowe, J., 2008, "Measurement of Stride Parameters Using a Wearable GPS and Inertial Measurement Unit," J. Biomech. doi: 10.1016/j.jbiomech.2008.02.021.

[7] Schmidt, M., Rheinländer, C., Nolte, K. F., Wille, S., Wehn, N., and Jaitner, T., 2016, "IMU-Based Determination of Stance Duration during Sprinting," *Procedia Engineering* doi: 10.1016/j.proeng.2016.06.330.

[8] Chen, H., Xue, M., Mei, Z., Oetomo, S. B., and Chen, W., 2016, "A Review of Wearable Sensor Systems for Monitoring Body Movements of Neonates," Sensors (Switzerland), **16**(12), pp. 1–17. doi: 10.3390/s16122134.

[9] Gago, M. F., Fernandes, V., Ferreira, J., Silva, H., Rodrigues, M. L., Rocha, L., Bicho, E., and Sousa, N., 2015, "The Effect of Levodopa on Postural Stability Evaluated by Wearable Inertial Measurement Units for Idiopathic and Vascular Parkinson's Disease," Gait Posture doi: 10.1016/j.gaitpost.2014.11.008.

[10] Chapman, R. M., Torchia, M. T., Bell, J. E., and Van Citters, D. W., 2019, "Continuously Monitoring Shoulder Motion after Total Shoulder Arthroplasty: Maximum Elevation and Time Spent above 90° of Elevation Are Critical Metrics to Monitor," J. Shoulder Elb. Surg., **28**(8), pp. 1505–1514. doi: 10.1016/j.jse.2019.01.003.

[11] Bulling, Andreas, Blanke, Ulf, Schiele, B., 2014, "A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors," Assoc. Comput. Mach., **46**(3), p. 33.

[12] Fida, B., Bernabucci, I., Bibbo, D., Conforto, S., and Schmid, M., 2015, "Varying Behavior of Different Window Sizes on the Classification of Static and Dynamic Physical Activities from a Single Accelerometer," Med. Eng. Phys. doi: 10.1016/j.medengphy.2015.04.005.

[13] De Leonardis, G., Rosati, S., Balestra, G., Agostini, V., Panero, E., Gastaldi, L., and Knaflitz, M., 2018, "Human Activity Recognition by Wearable Sensors : Comparison of Different Classifiers for Real-Time Applications," MeMeA 2018 - 2018 IEEE Int. Symp. Med. Meas. Appl. Proc., pp. 1–6. doi: 10.1109/MeMeA.2018.8438750.

[14] Chinimilli, P. T., Redkar, S., and Zhang, W., 2017, "Human Activity Recognition Using Inertial Measurement Units and Smart Shoes," Proc. Am. Control Conf., pp. 1462–1467. doi: 10.23919/ACC.2017.7963159.

[15] Ferhat Attal, Samer Mohammed,, Mariam Dedabrishvili, Faicel Chamroukhi, L. O. and Y. A., 2015,

"Physical Human Activity Recognition Using Wearable Sensors," Sensors, **15**, pp. 31314–31338.

[16] Jadhav, S. D., and Channe, H. P., 2016, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," Int. J. Sci. Res., **5**(1), pp. 1842–1845. doi: 10.21275/v5i1.nov153131.

[17] Roach, K., and Miles, T., 1991, "Normal Hip and Knee Active Range of Motion: The Relationship to Age," Phys. Ther., **71**, pp. 656–665.

[18] Watkins, M., Riddle, D., Lamb, R., and Personius, W., 1991, "Reliability of Goniometric Measurements and Visual Estimates of Knee Range of Motion Obtained in a Clinical Setting," Phys. Ther., **71**(2), pp. 90–96.

[19] Elias, L. J., Bryden, M. P., and Bulman-Fleming, M. B., 1998, "Footedness Is a Better Predictor than Is Handedness of Emotional Lateralization," Neuropsychologia doi: 10.1016/S0028-3932(97)00107-3.

[20] Liu, Y., Nie, L., Liu, L., and Rosenblum, D. S., 2016, "From Action to Activity: Sensor-Based Activity Recognition," Neurocomputing doi: 10.1016/j.neucom.2015.08.096.

[21] Dernbach, S., Das, B., Krishnan, N. C., Thomas, B. L., and Cook, D. J., 2012, "Simple and Complex Activity Recognition through Smart Phones," Proc. - 8th Int. Conf. Intell. Environ. IE 2012, pp. 214–221. doi: 10.1109/IE.2012.39.

[22] The MathWorks, I., 2014, "Matlab HAR Smartphone Example" doi: 10.1192/bjp.111.479.1009-a.

[23] Quinlan, J. R., 1987, "Simplifying Decision Trees," Int. J. Man. Mach. Stud., **27**(3), pp. 221–234.

[24] Ignatov, A., 2018, "Real-Time Human Activity Recognition from Accelerometer Data Using Convolutional Neural Networks," Appl. Soft Comput. J. doi: 10.1016/j.asoc.2017.09.027.

[25] Ronao, C. A., and Cho, S. B., 2016, "Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks," Expert Syst. Appl. doi: 10.1016/j.eswa.2016.04.032.