Classifying Common Knee Rehabilitation Exercise Mistakes Using IMU Data

Fedor Myagkov
Fedor.M.Myagkov.21@Dartmouth.edu

Follow this and additional works at: https://digitalcommons.dartmouth.edu/senior_theses

Part of the Biomechanics and Biotransport Commons, and the Biomedical Devices and Instrumentation Commons

Recommended Citation
https://digitalcommons.dartmouth.edu/senior_theses/218

This Thesis (Undergraduate) is brought to you for free and open access by the Theses, Dissertations, and Graduate Essays at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth College Undergraduate Theses by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
Classifying Common Knee Rehabilitation Exercise Mistakes Using IMU Data

Fedor Myagkov
Advisors: Dr. David Kotz and Dr. Douglas W. Van Citters
Undergraduate thesis in Computer Science
Dartmouth College

June 1, 2021

Abstract

Physical therapy following major surgeries is a branch of medicine that has seen its fair share of technologically inspired advances. One important facet of physical therapy, the “at-home exercises” patients are prescribed to do, is still somewhat of a “black box” to many physical therapists (PTs). PTs have no way of knowing (1) whether the patient is doing the home exercises, or (2) whether the patient is doing the exercises in the correct and healthy manner. This lack of awareness makes it difficult for the PT to guide the patient, which can often lead to prolonged rehabilitation periods or (sometimes) can create life-long health problems for patients. In this thesis, we provide a means for a PT to remotely monitor patient’s performance of at-home exercises. We combined the capabilities of wearable motion sensors with computational algorithms to provide patients feedback on the quality of their performed exercises. We evaluated this approach by asking 20 healthy volunteers to perform popular knee-rehabilitation exercises with various mistakes while wearing motion sensors. After pre-processing and extracting features from the sensor data, we trained machine-learning models on the extracted features. The models showed a high rate of accuracy during testing, which brings us a step closer to giving physical therapists and doctors a tool to automatically and objectively classify certain exercises and mistakes made during those exercises.
1 Introduction

Anterior cruciate ligament (ACL) injuries are one of the most common injuries in the world. In the United States alone, ACL injuries happen about 200,000 times per year and are responsible for about 100,000 surgeries per year [12]. A full or partial ACL rupture usually requires surgery for a patient to get back to their pre-injury capabilities. This specific injury can often end careers for people with physically active jobs (such as athletes, construction workers, firemen) and have life-long impacts for many others (e.g., inability to participate in normal and active daily routines) [2]. The complex surgical intervention, required as a result of the injury, precedes a difficult rehabilitation process. Many patients are never able to recover their full pre-injury physical capabilities.

There are two main factors in whether a person recovers their pre-injury physical capabilities:

1. timing and quality of the surgery, and
2. quality of the post-surgical rehabilitation process.

The contributions of this thesis focus on improving the quality of the post-surgery rehabilitation process. In particular, we examined the processing of data from inertial measurement units (IMU) [16] and developed machine-learning algorithms to evaluate the quality with which a patient is performing a set of rehabilitation exercises.

This contribution is important because during the rehabilitation process patients must perform many exercises at home (away from the clinic) in order to return to their pre-injury physical capabilities. Many physical therapists note that patients have difficulty adhering to the prescribed rehabilitation exercises. Studies show adherence rates for at-home exercises during physical rehabilitation are lower than 50% [4]. Furthermore, some patients perform the exercise poorly or incorrectly – mistakes that can slow their rehabilitation or even cause further injury. Physical therapists are unable to guide the patient or advise about exercise mistakes in their patients’ therapy regime because they are not able to monitor the patients at home. The ability for patients and their therapists to see an objective qualitative representation of the patients’ performed exercises could greatly increase the efficiency and quality of a rehabilitation period – and ultimately give patients the ability to return to their pre-injury physical capacities.

The specific question this thesis addressed is whether it is possible to accurately classify a discrete number of mistakes within a discrete number of knee-rehabilitation exercises, by training machine-learning models on IMU sensor data collected from participants performing such exercises. We invited healthy participants to perform certain exercises – correctly and incorrectly – while having two IMU sensors strapped to one of their legs. We then processed the data collected from the IMU sensors to extract features and train classifiers to recognize correct exercise performance. We evaluated the effectiveness of our classifiers using the overall accuracy score with which they were able to infer exercise variations. We found that our classification models had relatively high accuracy results of around 90 percent.

**Contribution:** we take a step closer to giving physical therapists and doctors an automated system that can objectively assess the quality of at-home rehabilitation exercises through the use of IMU sensors paired with computational algorithms.


2 Related Work

Human activity recognition (HAR) is becoming an important sub-field of human-computer interaction. With the advent of new technologies that are able to track human movements, including mobile phones and cameras, HAR centered systems are becoming more available to the masses. For physical rehabilitation, in particular, both IMU and computer vision have been used. Lee produced a system that classified the depth of squats based on IMU sensors with the help of deep learning [10]. Lim used IMU data to infer specific kinematic problems in people while walking, and had successful results when subjects were walking at lower speeds [11]. Gu and Tang both produced systems based on the ‘kinect’ vision system [7, 13]. They showed that giving subjects feedback during their exercises improved the quality of the performed exercises. Both methods have shown relative success in detecting and classifying particular movements done by subjects mimicking rehabilitation patients. Even though both IMUs and computer vision methods have been used to collect motion data during physical rehabilitation, studies have shown that IMUs produce more accurate and precise motion data. For example, Du produced a system to improve upper-limb exercise technique, a system that integrated both kinect and IMUs [5]. He found that the IMUs were more accurate than the kinect when compared to the “gold standard” motion-capture system also used as a part of the study. This observation is one of the main reasons we chose to use IMUs in the scope of our thesis.

Others have designed systems to help address the adherence problem with physical therapy. Many of these systems involved interactive games that can be played during the rehabilitation period [1]. Alankus developed and tested many different games for patients going through physical rehabilitation following a stroke. He developed games that involved the use of Wii remotes [17]. Alankus showed that incorporating games into the rehabilitation process increase patient motivation to do the exercises, and therefore increases adherence.

3 Methods

This section presents our methods for data collection and initial data classification.

3.1 Participants

We collected data from 20 participants (10 female and 10 male). The subjects were all students at Dartmouth College between the ages of 19-22. All participants were healthy and reported no recent knee injuries. We recruited subjects through an email sent to a subset of seniors at Dartmouth College who were approved for on-campus privileges. The study was approved by the Committee for the Protection of Human Subjects (CPHS), the Institutional Review Board (IRB) at Dartmouth College. All participants were informed of the experimental procedures and gave informed written consent prior to participating.

3.2 Data Collection

We instructed the participants on how to properly do the following three exercises (according to standards set by a physical therapist): squat (Figures 1a–1b), lunge (Figures 2a–2b), and heel slide
For each exercise, we instructed the participants how to perform common movement mistakes that often arise in rehab patients, Table 1 lists the exercises and their corresponding mistakes: squat (Figures 4a–4c), lunge (Figures 5a–5d), and heel slide (Figures 6a–6b).

We placed two IMU sensors (developed by APDM Wearable Technologies, Inc.) on the participants, one below the kneecap on the right shin, the other above the knee cap on the right quadriceps. The sensor placement can be seen in Figure 7. Then the participants would do each exercise 10 times, and every variation of a mistake 10 times, for squats and lunges. The participants would do every variation of heel slides 5 times. The participants were asked to take a 1-minute break between each set of exercises.

Table 1: Exercises and corresponding mistakes classified as part of this thesis.

<table>
<thead>
<tr>
<th>Mistake 1</th>
<th>Squat</th>
<th>Mistake 2</th>
<th>Lunge</th>
<th>Mistake 3</th>
<th>Heel Slide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knees caving inside</td>
<td>Planting foot too far forward</td>
<td>Arched back</td>
<td>Planting foot too far back</td>
<td>Caving knee inside</td>
<td>Rate of slide (too fast/slow)</td>
</tr>
</tbody>
</table>
(a) Beginning and end position of lunge exercise.  
(b) Mid-point position of lunge exercises.

Figure 2: Lunge example

(a) Beginning and end position of heel slide exercise.  
(b) Mid-point position of heel slide exercises.

Figure 3: Heel slide example
(a) Good squat during the mid-point of the exercise.

(b) Squat where the subject caves their knees in.

(c) Squat where the subject arches their back.

Figure 4: Squat mistakes
(a) Good lunge during the mid-point of the exercise.

(b) Lunge where the subject plants their foot too far from the start point.

(c) Lunge where the subject pants their foot too close to the start point.

(d) Lunge where the subject caves their knee in during the mid-point of the exercise.

Figure 5: Lunge mistakes
(a) Good heel slide during the mid-point of the exercise.
(b) Heel slide where the subject caves their knee in during the mid-point of the exercise.

Figure 6: Heel slide mistakes

Figure 7: Sensor placement example. They were placed 13 cm above and below the center of the knee-cap.
After each participant completed all the exercises we uploaded the IMU data to a computer, where it was converted into tabular (comma-separated values, CSV) format for analytical use.

3.3 Data Classification

A physical therapist reviewed the footage of the subjects performing the exercises, classifying each repetition of each exercise to indicate whether that rep was correct or a mistake, and if a mistake, which mistake. The result was that we ended up with a labeled data set of IMU sensor readings from 20 subjects performing the above-mentioned exercises and variations (mistakes) of the exercises.

4 Machine Learning

We chose a machine-learning approach for our study because we anticipated it would be difficult to accurately recognize exercise mistakes using a direct analytical method, due to the many different variations in the motions of the same broader mistake. For example, a knee rehabilitation patient performing a squat may cave their knees inside to more of an extent than another patient caving their knees inside, though both patients are executing the same mistake – caving their knees inside. Another reason to use a machine-learning method is that every knee is unique, in its bio-mechanical movements [8]. Even when performing the same exercise with the same mistake (or no mistake), all knee-rehab patients will have unique movement patterns – which could lead to problems when attempting to classify the movement data using simple deterministic methods.

We were able to use a machine-learning approach for two main reasons. (1) Precision of the data – the IMU sensors are able to detect and collect tiny (±0.05°) differences in exercise motion and movement. (2) Quantity of data – the number of subjects paired with the fast data collection (128 Hz) speed of the IMU sensors gave us thousands of data points for each subject and exercise. The large amount of precise motion data gave us the ability to extract many different features from the data that were representative of the motions performed by the subjects.

In this section we describe the raw data available from our IMU sensors, our methods for pre-processing the data and extracting features, and our methods for training classifiers from that data.

4.1 Raw Data

From the sensors we received raw data from the accelerometer (m/s²), gyroscope (rad/s), and magnetometer (microtesla) in the x, y, and z axes. All of the readings were sampled at 128 Hz. The sensors recorded a timestamp for each data point. Although the sensors also recorded temperature and pressure values, we did not use them in our work.

4.2 Pre-Processing

Due to the oscillating nature of the exercises performed by the subjects, we expected the IMU data to present an oscillating signal – as depicted in Figure 8. As a result, we aimed to split the data into individual windows in accordance with the peaks and valleys as shown in Figure 9. These windows
of data, each representing one repetition (rep) of the exercise motion, gave us the ability to map the data from individual reps to the corresponding classification proposed by the physical therapist as shown in Table 2. We labeled a repetition of an exercise a 0 if the repetition was considered “good” or did not have a mistake and 1, 2, or 3 for subsequent variations of the particular mistake performed during the repetition.

After splitting the data, each time window represented one repetition as a time series of 18-tuples consisting of three axes from each of the three motion sensors (as introduced in Section 4.1) from the two IMUs. Unlike some other applications of ML on temporal data, we divided our data into windows aligned with an exercise repetition, rather than a fixed time period, which means our windows vary in duration. We then performed the feature extraction on each window’s series of 18-tuple data points.
Table 2: Example of initial exercise data classification.

<table>
<thead>
<tr>
<th>Exercise Data</th>
<th>Mistake Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition 1</td>
<td>No Mistake (labeled as 0)</td>
</tr>
<tr>
<td>Repetition 2</td>
<td>No Mistake (labeled as 0)</td>
</tr>
<tr>
<td>Repetition 3</td>
<td>Mistake 2 (labeled as 2)</td>
</tr>
<tr>
<td>Repetition 4</td>
<td>Mistake 1 (labeled as 1)</td>
</tr>
</tbody>
</table>

Table 3: Number of repetitions of each exercise and corresponding mistakes in our data set (n).

<table>
<thead>
<tr>
<th>No Mistake</th>
<th>Squat (200)</th>
<th>Lunge (200)</th>
<th>Heel Slide (100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mistake 1</td>
<td>Knees caving inside (200)</td>
<td>Planting foot too far forward (200)</td>
<td>Caving knee inside (100)</td>
</tr>
<tr>
<td>Mistake 2</td>
<td>Arched back (200)</td>
<td>Planting foot too far back (200)</td>
<td></td>
</tr>
<tr>
<td>Mistake 3</td>
<td></td>
<td>Caving knee inside (200)</td>
<td></td>
</tr>
<tr>
<td>Total n</td>
<td>600</td>
<td>800</td>
<td>200</td>
</tr>
</tbody>
</table>

The resulting data set was large enough to train and evaluate a classifier. The number of repetitions on each variation of an example exercise is shown in Table 3.

4.3 Feature Extraction

We then extracted features from each repetition. We used the tsfresh package, which is a feature extraction and filtering package aimed at time-series data [14]. The tsfresh package was able to extract from each repetition thousands of mathematical features (such as FFT, mean, and sum of squares). We used tsfresh to compute each feature on each axis of each IMU reading (acceleration, rotation, and magnetic field) from both IMU sensors. The tsfresh package works in three phases.

Phase 1 – Feature Extraction

In the first phase, tsfresh extracts mathematical features from the raw time-series data. Examples of the features range from simple statistics such as mean, median, and mean of the data to more complex statistics such as the absolute energy of the time series, and coefficients of the one-dimensional discrete Fourier Transform. We ran the tsfresh package on each of 9 axes from each of 2 IMU devices, in every repetition window, and extracted 18 sets of 82-feature vectors (1,476 total) for each particular repetition window. This step is shown in the change from “raw time series” to “aggregated features” in Figure 10. In our case, “Sample 1 . . . Sample M” corresponds to individual repetitions of an exercise derived from our pre-processing step and “time series type” corresponds to a particular exercise with a particular mistake.

Phase 2 – Feature Significance testing

In the second phase, tsfresh evaluates each feature vector individually and independently in relation to its significance for predicting the corresponding exercise and mistake. The output of these tests are contained in a vector of corresponding p-values. Figure 10 depicts the second phase with the arrows from “aggregated features” to “p-values.” The exact computations to find the vector of p-values can be found in the tsfresh documentation [15].
Figure 10: tsfresh package feature extraction pipeline (from [14])
Phase 3 – Multiple Test Procedure

Finally, in the third phase, the vector of p-values are evaluated on the basis of the Benjamini-Yekutieli procedure [3] to decide which features to keep. The results of this step are depicted in the “selected features” section in Figure 10.

After these three phases and the threshold at which we deemed features significant enough to use in our ML models, we were left with 187 (Squats), 125 (Lunges), and 52 (Heel Slides) total mathematical features of the 1,476 original total features. We used these features to train the classifiers.

4.4 Classifiers

We explored common classifiers from the open-source Python package scikit-learn. The best performing classifiers can be found in Table 4. We found that the best performing classifiers were from the broad class called Decision Trees. We ultimately selected the Extra Trees classifier – one that is a form of a Random Forest classifier and based on running the data through many decision trees and choosing the output which is shared by most of the trees. Our classifiers output 0, 1, 2, or 3 labels which correlated to our ground truth labels.

4.5 Evaluation Methods and Metrics

We evaluated our classifiers, discussed in Table 4, using two different methods. First, we used a random test-train split on the entire data-set, where we used a random third of the data to test our classifier and used the other two-thirds of the data to train the model. Second, we used a Leave-One-Person-Out (LOPO) cross-validation. The LOPO cross-validation method is where one subject’s data is used as the test-data while all the other data is used as the training data. We thus had to train and test the classifiers 20 times, using a different subject as the test data on each iteration as shown in Figure 11.

We believe LOPO is the most useful and objective method for our purposes because the classifier would be tested on data from a subject it had never seen before – much as if a new patient began knee-rehabilitation exercise regime using our system. The problem with a more commonly used test-train split method is that bio-mechanics and gait are unique to all individuals – which means that if we want to make a useful system (one that does not need to be trained for each patient) we do not want to train our classifier on data from the same subjects on which we test the classifier.

While reviewing the results it is important to know the following classic ML evaluation metrics.

\[
\begin{align*}
\text{Accuracy} & = \frac{\text{CorrectPredictions}}{\text{AllPredictions}} \\
\text{Precision} & = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \\
\text{Recall} & = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}}
\end{align*}
\]
Table 4: Accuracy scores across different classifiers and exercises (0.33 test-train split).

<table>
<thead>
<tr>
<th>Classifier Type</th>
<th>Squat Accuracy</th>
<th>Lunge Accuracy Score</th>
<th>Heel Slide Accuracy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra Trees</td>
<td>0.95</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.93</td>
<td>0.89</td>
<td>0.96</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.87</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>K-Nearest Neighbors ($k = 1000$)</td>
<td>0.87</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>Logistic Regression Neighbors</td>
<td>0.75</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>Bagging Neighbors</td>
<td>0.56</td>
<td>0.61</td>
<td>0.69</td>
</tr>
<tr>
<td>Support Vector Neighbors</td>
<td>0.55</td>
<td>0.48</td>
<td>0.52</td>
</tr>
</tbody>
</table>

\[
F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}
\]

Since our data is output in ternary form for squats and quaternary form for lunges, as well as binary form for heel-slides, it is helpful to go over what false positive and negative means in ternary and quaternary forms as they are seen less often than binary outputs.

For ternary outputs, a false positive in the “good squat” row (or an output of 0) in Table 5, is defined by when the model infers an output should be 0 however the ground truth corresponds to a 1 or 2. A false negative in the same row corresponds to a model inference of 1 or 2 when the ground truth is actually 0. The same generalized concept is used for quaternary outputs when calculating precision, recall, and f1 score for a given row in the results seen below.

Our data set did not need to be manipulated to deal with imbalances because every variation of an exercise was performed with an equal amount of repetitions by every subject.

For our purposes, the **Accuracy** metric is the most important since we are not worried about the difference between false positives and false negatives. For many ML applications it is important to dive into the false positives and negatives as both of the mistakes hold vastly different meanings. In our case however, if the classifier misclassified a repetition we are not worried about the type of misclassification because bio-mechanically speaking we are looking at all types of exercise mistakes as equally “bad”. Of course that is not the case in a clinical context, however it is outside the scope of this thesis to determine which type of exercise mistakes will cause more harm during knee rehabilitation. Further, if the model classifies a “good” rep as a “bad” one we will still treat that classification mistake as equal to others in the scope of this thesis, because our main goal is to test how well the ML models are able to classify the repetition features we feed the model. Also when looking at the results we found that the **precision**, **recall**, and **f1-scores** were similar when compared across all exercises. Therefore, we are only worried about whether the classifier was able to, given the exercise, infer the correct type of repetition (labeled as 0, 1, 2, or 3) the subject performed.

## 5 Results

We report results from an evaluation with the Extra Trees classifier because it proved to be the best performing classifier as shown in Table 4.
5.1 Entire Data Set Random Test-Train Split

A random test-train split of 33 percent validation of our method on the full 20-subject dataset produced the results seen in Table 5, Table 6, and Table 7.

Although these results look promising – with accuracy 0.95, 0.93, and 0.96 for the squat, lunge, and heel slide classifications respectively – note that when using a random test-train split the model is based on training data from the same subjects. LOPO is a more appropriate method, as shown next.

5.2 Leave-One-Person-Out Evaluation

The results of LOPO evaluation are depicted in Figure 12. While the accuracy scores for most of the subjects are relatively high (around or above 90 percent accuracy), there are a few subjects whose data showed an under 80 percent accuracy score. We went back through the videos of the exercises, accompanied by a physical therapist, and looked at the three subjects whose accuracy scores were under 0.8 on both squats and lunges. The physical therapist noted that those particular...
Table 6: Lunge classification results (test size=33% of data.)

<table>
<thead>
<tr>
<th>Target</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Lunge</td>
<td>0.86</td>
<td>1.00</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Plant Foot Too Far In</td>
<td>0.97</td>
<td>0.89</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Plant Foot Too Close</td>
<td>1.00</td>
<td>0.93</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>Cave Knee In</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Heel Slide classification results (test size=33% of data.)

<table>
<thead>
<tr>
<th>Target</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Heel Slide</td>
<td>0.92</td>
<td>0.94</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Knee Cave Inside</td>
<td>0.98</td>
<td>0.96</td>
<td>0.92</td>
<td></td>
</tr>
</tbody>
</table>

subjects who made a small mistake in some repetitions – but the degree of the mistake was smaller than for other subjects.

5.3 Biological Sex Assigned at Birth as Output Variable

We also took the opportunity to construct a classifier that could infer whether the data was from a female or male. (The hips and knees of males and females are known to differ biomechanically [6], so it seemed plausible that the IMU data from these exercises might reflect differences due to the biological sex of the participant. For the purposes of this section, we label the data with the ‘biological sex assigned at birth,’ regardless of the participant’s gender identity, because the biomechanics depend on biological sex.) No classifier showed an accuracy of larger than 60 percent. However this type of classification should be explored in future work, as there may be other exercises that have larger variation in movement patterns between biological males and females. It may also be necessary to control for other variables (such as height, weight, or age).

5.4 Conclusion

With the size of the dataset we were able to collect in the scope of this study, the results show that our methods were able to classify specific variations of specific exercises with the use of IMU sensor data we collected with relatively good accuracy. This result leaves the door open for further research to expand on these results and produce a practical system that could be used in physical rehabilitation.

6 Discussion and Future Work

*Why did we choose wearables?* – In a perfect world physical therapists and patients would like to assess the quality of exercises in an objective and consistent manner – which is not always possible to do with the human eye. Further, when patients are doing exercises at home they need feedback
Figure 12: LOPO Results

<table>
<thead>
<tr>
<th>Subject</th>
<th>Accuracy</th>
<th>Subject</th>
<th>Accuracy</th>
<th>Subject</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.92</td>
<td>1</td>
<td>0.86</td>
<td>1</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>0.98</td>
<td>2</td>
<td>0.84</td>
<td>2</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>0.94</td>
<td>3</td>
<td>1.00</td>
<td>3</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>0.88</td>
<td>4</td>
<td>0.83</td>
<td>4</td>
<td>0.89</td>
</tr>
<tr>
<td>5</td>
<td>0.78</td>
<td>5</td>
<td>0.95</td>
<td>5</td>
<td>0.98</td>
</tr>
<tr>
<td>6</td>
<td>0.92</td>
<td>6</td>
<td>0.97</td>
<td>6</td>
<td>0.97</td>
</tr>
<tr>
<td>7</td>
<td>0.92</td>
<td>7</td>
<td>0.78</td>
<td>7</td>
<td>0.92</td>
</tr>
<tr>
<td>8</td>
<td>0.88</td>
<td>8</td>
<td>0.87</td>
<td>8</td>
<td>0.97</td>
</tr>
<tr>
<td>9</td>
<td>0.94</td>
<td>9</td>
<td>0.92</td>
<td>9</td>
<td>0.91</td>
</tr>
<tr>
<td>10</td>
<td>0.98</td>
<td>10</td>
<td>0.93</td>
<td>10</td>
<td>0.91</td>
</tr>
<tr>
<td>11</td>
<td>0.89</td>
<td>11</td>
<td>0.99</td>
<td>11</td>
<td>0.98</td>
</tr>
<tr>
<td>12</td>
<td>0.92</td>
<td>12</td>
<td>0.78</td>
<td>12</td>
<td>0.94</td>
</tr>
<tr>
<td>13</td>
<td>0.94</td>
<td>13</td>
<td>0.86</td>
<td>13</td>
<td>0.88</td>
</tr>
<tr>
<td>14</td>
<td>0.98</td>
<td>14</td>
<td>0.93</td>
<td>14</td>
<td>0.98</td>
</tr>
<tr>
<td>15</td>
<td>1.00</td>
<td>15</td>
<td>0.93</td>
<td>15</td>
<td>0.98</td>
</tr>
<tr>
<td>16</td>
<td>0.82</td>
<td>16</td>
<td>0.89</td>
<td>16</td>
<td>0.95</td>
</tr>
<tr>
<td>17</td>
<td>0.90</td>
<td>17</td>
<td>0.96</td>
<td>17</td>
<td>0.93</td>
</tr>
<tr>
<td>18</td>
<td>0.94</td>
<td>18</td>
<td>0.92</td>
<td>18</td>
<td>0.92</td>
</tr>
<tr>
<td>19</td>
<td>0.94</td>
<td>19</td>
<td>0.97</td>
<td>19</td>
<td>0.98</td>
</tr>
<tr>
<td>20</td>
<td>0.90</td>
<td>20</td>
<td>0.90</td>
<td>20</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Mean 0.92 Mean 0.90 Mean 0.94

(a) Squat classification results.  (b) Lunge classification results.  (c) HSlide classification results.
on the quality of their exercises; even if a physical therapist can observe over a live video connection, it is difficult for them to see what they need to see on a 2D screen. Computer vision is another possibility to capture the motion, however it would be more complex to solve the problem with computer vision as patients have different cameras, lighting, and indoor environments, which all need to be accounted for in computer-vision ML models. For these reasons, we believe wearables are a promising approach for this application.

Number of IMU sensors and other sensors – Future work in this area will certainly focus on expanding the number of exercises and types of mistakes a classifier is able to detect for knee-rehabilitation patients during their post-surgery rehab. Adding IMU sensors (attached to other locations on the same leg, or on the other leg) may increase the accuracy of classifiers due to the addition of data points, but an increase in the number of sensors increases the cost of the system and increases the time patients need to prepare for their exercises. Finding the correct balance between accuracy and ease of use will play an important role in whether the findings of our and related research will make a real-world impact.

Insole sensors have also been used in research surrounding physical rehabilitation and could be used to help classify exercise mistakes [9].

Incorporating a UI for patients and medical professionals – For patients, physical therapists, and doctors to use the feedback from classifiers such as the ones described in this study, they must be integrated into an easy-to-use interface. The research into the usability of these system will pose questions revolving around the most productive ways for patients to receive feedback on their performed exercises. In other words, what user interface will maximize the quality and speed of physical therapy for knee-rehabilitation patients?

Adherence – In addition to patients making mistakes during at-home exercises, non-adherence to prescribed at-home exercises has been proven to be detrimental to the quality of a post knee-surgery rehabilitation period, as stated in the introduction. Further research should be done on increasing adherence. Our sensor-based system may help increase adherence, because many physical therapists told us that patients tend to be more willing to perform exercises when they are receiving constant feedback on their movement.

Injury prevention – Although the research presented here is focused on classifying mistakes during knee-rehabilitation exercises, we can not make the claim that our classifications will help prevent future injuries to the surgically repaired knee. Research on classifying which particular mistakes and movements may lead to injuries further down the line, will have a great impact on the physical therapy and sports science world. The ability to quickly classify specific injury related movements would help medical professionals guide patients down a personalized route on their rehabilitation journey. This type of classification would also give medical professionals more insight into objective movement features that correlate to injuries, and advance our knowledge of bio-mechanics.

7 Summary

Physical rehabilitation after knee surgery is a long and strenuous process – of which prescribed at-home rehabilitation exercises are an important part if the patient would like to get back to their pre-
injury activity level. Current systems leave physical therapists with a difficult time assessing the quality of the at-home exercises performed by their patients, as well as whether they are performing the exercises at all. In this thesis we produced a system that was able to classify common mistakes performed by rehabilitation patients while doing squats, lunges, and heel slides. The system was able to classify the mistakes by extracting features from the data of IMU sensors the subjects were wearing while performing exercises. The results of this thesis showed that IMU sensor data has the ability to decipher between variations of squats, lunges, and heel-slides performed by subjects in a controlled environment. This result gives future researchers the ability to build more robust systems that will be able to classify the mistakes of many different exercises in many different environments.

Acknowledgements

This research results from a research program at Dartmouth College. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

This thesis could not have been done without the help and mentorship from my advisors Professors David Kotz and Douglas Van Citters.

Professor Ryan Chapman helped tremendously by teaching us how to use the sensors and extract the data from them. He also helped with deciding on the set of exercises and mistakes we were classifying as he is an expert bio-mechanist.

Tanguy Nef helped with discussions about the work in this thesis.

The work of this thesis could not have been done without the 20 subjects who were willing to squat and lunge.

References


