2023 IEEE World AI IoT Congress (AIIoT) | 979-8-3503-3761-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/AIIoT58121.2023.10174414

Features Analysis and Classification of Common Gym Exercises

Niko Simon
Department of Electrical and Computer Engineering
University of St. Thomas
St. Paul, MN, USA
niko.simon@stthomas.edu

Cheol-Hong Min
Department of Electrical and Computer Engineering
University of St. Thomas
St. Paul, MN, USA
cmin@stthomas.edu

***Abstract*— This preliminary study examined time domain, frequency domain, and time-frequency domain features for the appropriate classification of four different exercises. Thirty bench presses, squats, crunches, and walking trials were conducted, and data from two other sensor locations were considered (accelerometer data on the right wrist and left ankle). Time, frequency, and time-frequency features were examined across trials with differences in processing methods applied. Multiple classification algorithms were used and validated using 5-fold self-validation and validation against new user data. Results indicated that the features and algorithms investigated could accurately classify exercise routines. Additional exercises and larger data sets are being developed for a more robust study.**

***Keywords*—health, detection, recognition, classification, KNN, SVM, machine learning**

I. Introduction

New wearable gadgets come to market making consumers' lives easier every day. However, the exercise and fitness market has yet to take advantage of the increased integration of accelerometers in these devices and their potential for activity classification [1]. A solution would be a wearable system capable of accurate exercise classification and repetition counting. Many wearable devices on the market (Apple, Fitbit, Samsung, Garmin)are useful for counting steps, activity levels, HR monitoring etc. Some new Garmin devices offer wrist-based exercise classification/counting. Still, no current commercial devices are capable of accurate exercise classification and repetition accounting for both upper and lower body exercises. Similar exercise-tracking devices have proven effective for elite athletes [2].

Different types of exercises have various motion patterns when executed repetitively [3]. However, these different exercises will most likely share similarities in movement, which may be enough to cause a misclassification if not performed as expected by the user. This means having the most accurate method for distinguishing this type of activity data will allow for the best chance of automatically determining an exercise based purely on its activity pattern [3, 4, 5, 6, 7, 8, 9, 10]. Sensors were used on the upper arm, wrist, and ankle (ankle sensor on a different side than the wrist). Thirty trials of data were recorded for the four different repetitively executed exercises, bench pressing, squats, crunches, and walking. Walking picks up movement from all three sensors when viewing these exercises from a movement behavior perspective. In contrast, the wrist and arm sensors expect movement from other activities, excluding walking.

II. Methods

A. Sensors and Sensor Placement

Based on literature findings, sensors placed on the wrist and ankle were found to be the most appropriate for this application [4]. Chest and hip-mounted sensors have been found to provide consistent data. However, the practicality of these sensor locations for users may need to be more convenient for the user. As summarized in Table 1, the sensor was selected to be placed on the wrist based on previous studies and evaluations.

B. Data Collection

Figure 1 shows accelerometer data collected using a MbientLab MMR sensor at a sampling rate of 50 Hz. Each exercise was logged for 30 seconds to ensure data integrity was maintained. 20-second segments per activity trial were usable data, while parts at the beginning and end of each data recording were removed in pre-processing because of inconsistencies in the collection process.

C. Pre-processing

Pre-processing consists of removing unwanted information picked up during the data collection process and removing the DC bias to extract the frequency features. Visual observation indicated a need to filter out the 1 [g] offset due to gravity during collection. This was done by removing the DC bias. This process dramatically increases the distinction between the different activities and also makes clear some of the minor differences between the data collected at various locations. Removing DC offset during the pre-processing stage gives insight into which sensors (based on their placement) may be optimal for classifying the various exercises. However, true insights exist at the activity level.

Table1: Indicates the selection of appropriate sensors being the ankle and the wrist

	Ease of Use	Sensitivity	Discrimination	Overall
Sensor #1 WRIST	++	++	++	+3
Sensor #2 ARM	0	++	-	+1
Sensor #3 ANKLE	++	0	++	+2

Windowing was implemented on each axis, and each stream of data was derived from the various sensors at their respective body positions. This was performed on 100 sample intervals at a sampling rate of 50 Hz (~2 seconds) while ensuring a 50% overlap of the windows. This windowing method was iterated over the entire duration of the data stream to help analyze the data considered.

Additionally, data without windowing was examined to compare its effects on accuracy. Daubechies 5-tap wavelet coefficient features and time domain features were included in a feature matrix to compare with time domain features and frequency features in their own separate feature matrix [11]. Two main differences to note are the processing time of frequency data for practical applications and the differences in the number of data points for windowed and un-windowed data. Processing time domain data is less computationally intensive than frequency domain data.

III. Features Analysis

A. Time Domain

The mean of the acceleration measured was used as a primary means of discriminating between activities. The statistical mean was a distinct feature allowing for sufficiently accurate prediction of the observed exercises, shown in Figure 1. A vital element of this feature is its ability to provide distinct characteristics across all 3-Dimensions (X, Y, Z). During the bench press, minimal acceleration was observed across the z-axis. In contrast, the acceleration measured on the x-axis and y-axis were comparable while the activity was being performed.

The gravitational offset noticed in each axis also played a crucial role in determining the orientation of the sensor. This was also a key feature present in the time domain and

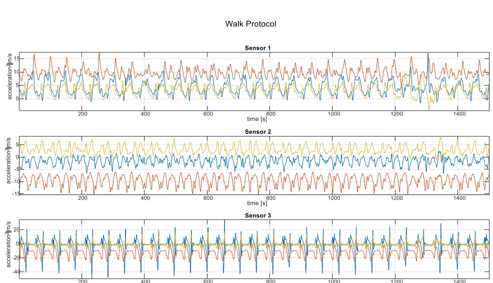

Fig. 1 An example of the X, Y, and Z-axis accelerometer data for various exercise activities.

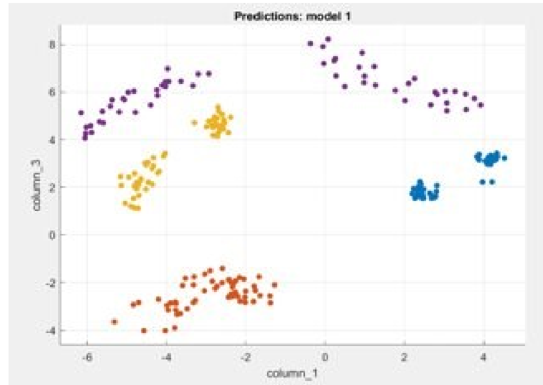

Figure 2. Mean values show distinct clustering and separation, which indicates the feature is valuable for accurate classification.

was retained for this purpose. In addition, though the mean helped distinguish between different activities, its accuracy was still most distinct based on sensor placement. The statistical variance of the acceleration measured was also used as a primary means of discriminating between activities. Observations below show how the exercises vary across means but have relatively similar variances. This was an unexpected observation. Mean and variances were then compared to gain more insight into possible classification methods. Variance showed significant peaks, especially during transitions.

B. Frequency Domain

In addition to the time domain, frequency domain features were also extracted. Fourier transform was used to determine the contribution of frequency components from 0-25Hz. Frequency power bands were created, which represented the power in each of the five frequency bandwidths. Ultimately the bands revealed that the higher frequencies (10-25) did not contribute significantly to classification accuracy and may have reduced the accuracy. Meanwhile, as it can be seen in Figure 2, the low-frequency features exhibit clustering and clear distinction between classes. The lower frequency bands could be split into multiple sub-bands to provide greater resolution of the

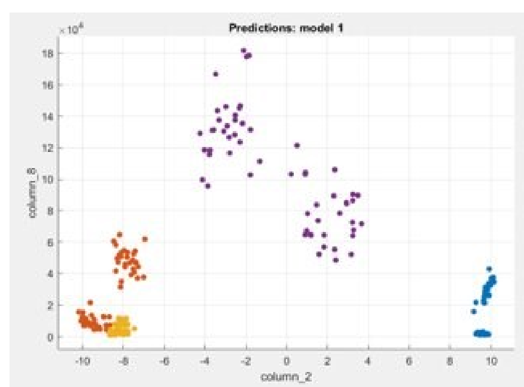

Fig. 3 Clustering is observed in lower frequency bands. Most human activity tends to occur at low frequencies than higher frequency bands. Higher frequency bands may be contributing as noise to the classifiers.

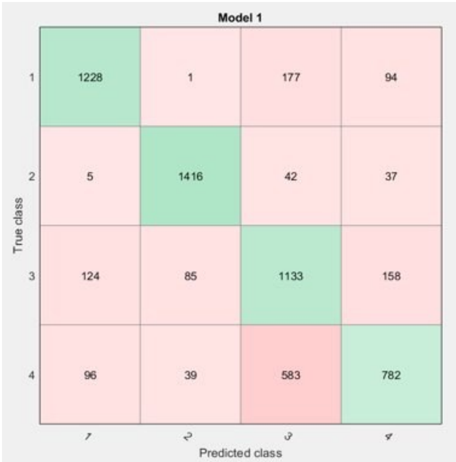

Figure 4. The confusion matrix for wavelet features showing the correct and misclassification of squats and crunches

frequency features.

C. Time-Frequency Domain

Wavelets were the time-frequency domain feature investigated. The Daubechies (5 taps) wavelet is used to compare results to the findings of the literature study [4]. When only using wavelet features, accuracy is reduced, and the major cause of this is a misclassification between crunches and squats. This is observed from the confusion matrix shown in Figure 4. When examining only wavelet features, misclassification was detected between two of the four activities, crunches and squats. As shown in Figure 5, much overlap exists between the activities suggesting lower classification performance. Both motions involve no ankle motion, and misclassification for this feature specifically was consistent across all trials. This suggests that using only the wavelet feature may be inappropriate for classifying this particular set of activities. Because the actions themselves occurred over time (as opposed to static actions), a feature in the time-frequency domain would be

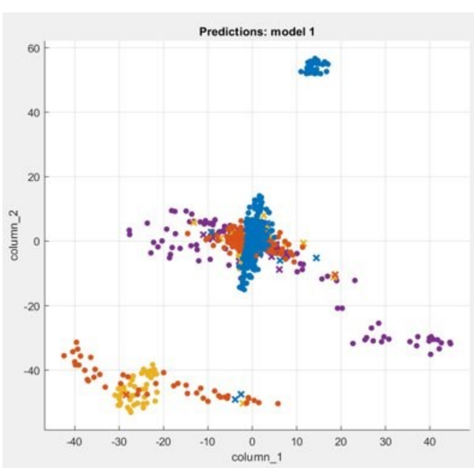

Fig. 5 Clustering for wavelet features only, consistent misclassification is noted across all three trials, which confuse squats with crunches, with accuracy between 70% and 76% across

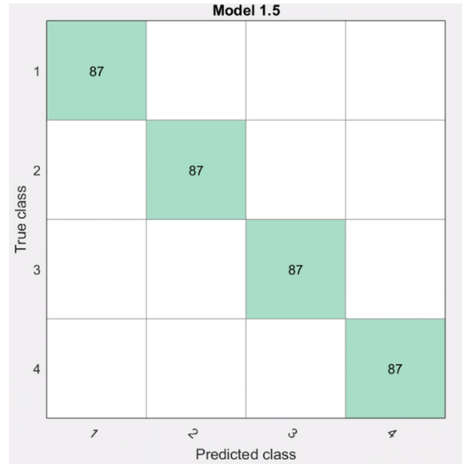

Figure 6. The confusion matrix using combined features showing the correct and misclassification of squats and crunches

expected to contribute to classification.

Further analysis of combined time and frequency features together to produce a much more accurate model with greater than 99% accuracy. For this case, windowing is applied, and the number of data points compared to the time/frequency feature matrix is greater. The accuracy per trial is slightly less than the accuracy of the models, which included frequency features and no windowing, i.e. each multiple sequence of repeated exercise is grouped as one exercise.

IV. Classification

A. Classification Algorithms and Feature Sets

Various classification algorithms were tested in order to compare their relative effectiveness. Classification algorithms tested included Fine, Medium, and Coarse KNN, Linear, Quadratic, Fine, Medium, and Coarse Gaussian SVM. The Coarse KNN and Coarse Gaussian SVM were the least accurate, but most other classification algorithms

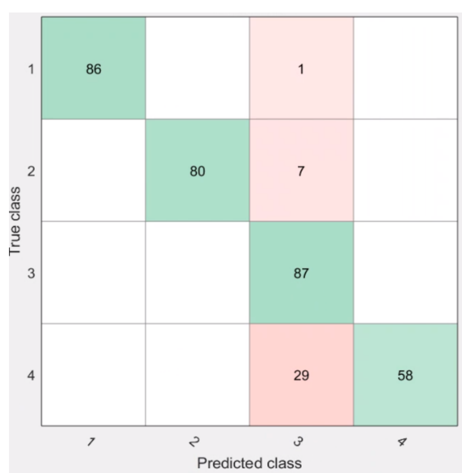

Fig. 7 Coarse KNN confusion matrix showing notable misclassification of walking data.

Table 2. Comparison of classification accuracy for two different approaches. 5-fold cross validation vs Leave One Subject Out (LOSO)

Label	Exercise	5-fold cross validation	LOSO
1	Bench	100%	98.85%
2	Squat	100%	60.92%
3	Crunch	100%	94.25%
4	Walk	100%	95.40%

were very accurate, with an average of 98% accuracy.

Additionally, the classification algorithms were tested with different input feature sets, including time domain, frequency domain, wavelet, and a combination of time and frequency/wavelet domain feature sets. The combination of groups achieved the highest accuracy at 100% in validation.

B. *Validation Methods*

Validation was performed using the function "predict label accuracy", which generates a confusion matrix as an output based on a model and a data matrix for the input. The classification algorithms were validated using two different approaches. First, a five-fold self-validation was performed. In this approach, the entire data set that was collected is randomly divided into five sections, with one section being used as the test set and the remaining four being used to train the model. This approach yielded very high accuracy, at or near 100%. However, because only three subjects were used in this study, a significant portion of the test data is from the same subject as the training data. Therefore, the model is being trained specifically for that user, and the accuracy results are higher than expected in a real-world scenario, where the model cannot be trained for a specific user beforehand.

Therefore, a second validation approach was used. The classification model was trained using data from two of the three subjects and tested on data from the third subject. This validation approach provides a more realistic scenario because the model is not trained on data from the subject that is also being used for testing. Using this approach, accuracy dropped significantly, although some models, such as SVM, still performed well. Table 2 above compares the results for the two validation methods.

Classification accuracy of time/wavelet features showed better accuracy than time/frequency features. Analysis of the combined features was developed using a model trained on two of the three available subjects, with the third being used for test data. The model for the time/wavelet features was developed using the LOSO method, with another being used for the test data, as shown in Table 2 and Table 3. Ideally, many subjects would be used in order to develop a model; however, due to the similarity in the training routine, there may be only a small amount of inconsistencies that would be present in a model using a larger dataset. However, when the time/wavelet feature model did misclassify an activity, the lowest accuracy percentage was about 60%, while the lowest accuracy for the time/frequency was 13% which is unacceptable. When comparing KNN to SVM classification, SVM produces higher accuracy rates.

Table 3. Comparison of classification accuracy for two different classifiers for Leave One Subject Out (LOSO) validation method using time domain and wavelet coefficients as features

Label	Exercise	KNN	SVM
1	Bench	98.85%	100%
2	Squat	60.92%	98.85%
3	Crunch	94.25%	82.75%
4	Walk	95.40%	100%

V. Discussion and Conclusion

Some limitations of the study include the number of subjects and the length of activity data collection. More subjects with an increased time for data collection may provide a more comprehensive base for learning data. In this analysis, it would only be possible to construct a model based on two subjects, and there would only be one subject for test data. It was proven through experimentation that activities performed by different individuals for the trials produced some variations in the data even though the type of activity was the same. Although the variations are small, more data can reduce such concerns.

Considerations for the choice of particular features are governed by what activities and functionalities are being considered for detection or tracking. Adding additional activities and functionalities in order to distinguish individual repetitions may be of interest to the user. In order to optimize this, the choice of features and windowing may need to be reconsidered in order to ensure correct identification. Considering multiple new activities would also require the re-evaluation of appropriate features in order to maximize classification accuracy.

Further focus can be applied to the frequency features. The earlier frequency bands contain more useful information than the later bands. If only the frequency bands with useful information were further segmented, it could provide better classification when considering many different activities, especially if they have similar movement patterns. Additionally, more functionalities, including dynamic time warping, would involve further research and experimentation [12].

The accuracy results for this analysis are somewhat mixed. Time/frequency features, time/wavelet features and features from one domain were examined. In addition, windowed data and non-windowed data were analyzed. Models were developed based on the two major feature combinations of interest (time/frequency and time/wavelet). These models were based on a portion of the trials of activity data that were collected. Because there were only three subjects of data, this means training data for the models could not be comprised of more than two subjects; otherwise, bias would be introduced for the test data. Findings indicated that classification accuracy was comparable between the two different feature combinations (both mostly about 99%). However, the validation of the models with test data was not highly consistent. This is mainly due to the need for more training data available.

References

[1] CRN, “The 10 Coolest Wearable Tech Gadgets for 2022 So Far”, [Online] https://www.crn.com/slide-shows/mobility/the-10-coolest-wearable-tech-gadgets-for-2022/2

[2] CNN Underscored, “The Best Fitness Trackers of 2023”, [Online] https://www.cnn.com/cnn-underscored/reviews/best-fitness-tracker

[3] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu and P. Havinga, "Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey," 23rd International Conference on Architecture of Computing Systems 2010, Hannover, Germany, 2010, pp. 1-10.

[4] L. Palafox, and H. Hashimoto, “Human Action Recognition Using Wavelet Signal Analysis As an input, 4W1H”, in 8th IEEE International Conference on Industrial Informatics, Osaka, 2010.

[5] A. Nagarkoti, R. Teotia, A. K. Mahale and P. K. Das, "Realtime Indoor Workout Analysis Using Machine Learning & Computer Vision," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 1440-1443, doi: 10.1109/EMBC.2019.8856547.

[6] A. Depari, P. Ferrari, A. Flammini, S. Rinaldi and E. Sisinni, "Lightweight Machine Learning-Based Approach for Supervision of Fitness Workout," 2019 IEEE Sensors Applications Symposium (SAS), 2019, pp. 1-6, doi: 10.1109/SAS.2019.8706106.

[7] M. Merenda, M. Astrologo, D. Laurendi, V. Romeo and F. G. Della Corte, "A Novel Fitness Tracker Using Edge Machine Learning," 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), 2020, pp. 212-215, doi: 10.1109/MELECON48756.2020.9140602.

[8] Y. Jain, D. Chowdhury and M. Chattopadhyay, "Machine Learning Based Fitness Tracker Platform Using MEMS Accelerometer," 2017 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2017, pp. 1-5, doi: 10.1109/ICCECE.2017.8526202.

[9] R. Chambers, T. J. Gabbett, M. H. Cole and A. Beard, “The Use of Wearable Microsensors to Quantify Sprot-Specific Movements”, Sports Med. vol 45, 2015, pp. 1065-1081, doi: 10.1007/s40279-015-0332-9

[10] C. O’Brien, and C.-H. Min, “Classification of Various Workout Motions Using Wearable Sensors”, 2022 IEEE World AI IoT Congress (AIIoT), pp. 442-446, 2022.

[11] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean and D. Finaly, “Optimal Placement of Accelerometers for the Detection of Everyday Activities, “*Sensors*”, vol. 13, pp. 9184-9200, 2013.

[12] C. Li, M. Fei, H. Hu and Z. Qi, “Free Weight Exercise Detection Based on Dynamic Time Warping of Acceleration Data,” Intelligent Computing for Sustainable Energy and Environment, pp. 178-185, 2012.