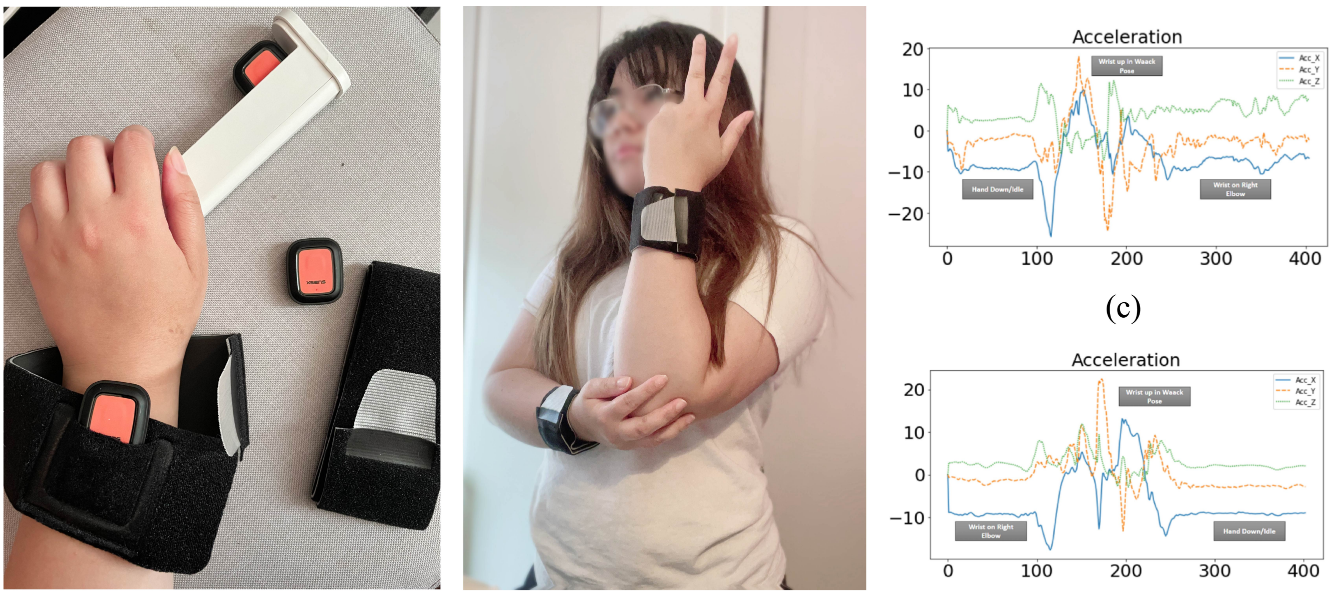
Using a Bidirectional LSTM Approach to Classify Simple Arm Dance Motions from Wearable Inertial Sensors



(a) (b) (d)

Figure 1: *(a)* Apply XSens DOT sensors on both wrists. *(b)* Demonstration of simple waacking arm dance movement; right hand on the left albow and left hand in a waacking motion. *(c)* Acceleration data on the left wrist; the data is for an idle motion, followed by the left wrist in a waacking pose, then placed on the right elbow. *(d)* Acceleration data on the right wrist; the data shows a motion of the right hand placed on the left elbow, followed by the right wrist in a waacking pose, then rests in an idle motion.

1 INTRODUCTION

Using wearable inertial sensors to track everyday human motion has been a popular subject in previous papers in classifying and recognising human activities. Plenty of projects involving human activity recognition have already been explored and published. Sukor et al.[1], Bayat et al.[2] and Gomes et al.[5] investigated a variety of machine learning algorithms for human activity recognition using *accelerometer data* while Pavai [6] applied a bidirectional LSTM approach. Using *motion capture data*, da Silva et al. [3] employed an LSTM approach to classify common human activities. Drumond et al.[4]'s approach is the most relevant to our work as they use an LSTM approach to classify motions from arm movements using only two IMU sensors. However, identifying dance motions from wearable inertial sensors to determine the accuracy of a dance choreography has not been well inspected. Moreover, none of the previous methods have specifically classified *dance motions* using a bidirectional LSTM approach using XSens DOT data.

In this paper, we propose a recurrent neural network approach based on a bidirectional LSTM model to classifiy simple dance arm motions from accelerometer data obtained from wearable inertial sensors.

2 APPROACH

XSens DOT is a wearable inertial sensor device that is used for precise motion tracking. Figure 1 depicts the process of our approach. Two XSens DOT sensors were placed in the small pocket of each straps and they were worn on both the left and right wrists of the performer.

We tested our approach using a hip hop dance style known as *'waacking'* which is renowned for its expressive arm movements and posing. A simple waacking choreography as seen in Figure 1(b) was captured using the sensors and the accelerometer data obtained from the sensors were imported to Jupyterlab for preprocessing.

The original data retrieved from the sensors consisted of acceleration, euler and gyroscope data but for this project, we concentrated on acceleration data so the other columns were dropped. The data was split into training and test datasets and the training data was scaled. A time step of 200 was used to create the sequences for the dataset and a bidirectional long short-term memory approach was used for classifying the activity. Dropout and dense layers were also added to increase the complexity of the model before it was compiled and trained with an epoch size of 100 and batch size of 32.

2.1 Findings

Author	Data	Approach	Accuracy
Bayat et al.	Accelerometer	*Combination of Classifiers:* Multilayer Perceptron, LogitBoost and SVM	91.15%
Gomes et al.	Accelerometer	KNN *Combination of Classifiers*: Random Forest & KNN	79% 78%
Sukor et al.	Accelerometer	Multilayer Perceptron Neural Network	96.11%
da Silva et al.	Motion Capture	LSTM	95%
Drumond et al.	Accelerometer	LSTM	96%
Pavai	Accelerometer	Bidirectional LSTM	90+%
Our Approach	Accelerometer	Bidirectional LSTM	60+%

Figure 2: Results Comparison

After training the accelerometer data using a bidirectional LSTM method, the accuracy results from both wrists yielded an accuracy of over 60%; 64.47% for the left wrist and 60.2% for the right wrist. Although this is lower than da Silva's approach which had a 95% accuracy, Drumond's approach of 96% and Pavai's of over 90%, we may see different results when the model is combined with different classifiers using Bayat et al.'s and Gomes et al.'s decision to use more than one classifier to achieve the best accuracy as an example. Perhaps, an improved accuracy rate can also be anticipated for our approach when more sensors are worn and different variations of the waacking dance choreography are captured.

2.2 Limitations and Future Work

A limitation of this work is that only two sensors were used in the capture of the waacking arm motion. Also, the two sensors were only applied to the left and right wrists.

To develop this work in the future, we aim to enhance our accuracy rate by adding more complexity to the bidirectional LSTM model or by exploring other deep learning approaches. Additionally, using all five of the sensors and applying them to both wrists, both ankles and the waist of the performer will be our next steps for this project as well as using the XSens DOT simultaneously with the motion capture sensors for comparison. Furthermore, we plan to integrate this model into a mixed reality dance training system to determine how accurate the participant is performing the dance technique.

REFERENCES

[1] A. S. Abdull Sukor, A. Zakaria, and N. Abdul Rahim. 2018. Activity recognition using accelerometer sensor and machine learning classifiers. *Proceedings - 2018 IEEE 14th International Colloquium on Signal Processing and its Application, CSPA 2018* April (2018), 233–238. https://doi.org/10.1109/CSPA.2018.8368718

[2] Akram Bayat, Marc Pomplun, and Due A. Tran. 2014. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science* 34 (2014), 450–457. https://doi.org/10.1016/j.procs.2014.07.009

[3] Rogério E. da Silva, Jan Ondřej, and Aljosa Smolic. 2019. Using LSTM for automatic classification of human motion capture data. *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* 1 (2019), 236–243. https://doi.org/10.5220/0007349902360243

[4] Rafael Rego Drumond, Bruno A.Dorta Marques, Cristina Nader Vasconcelos, and Esteban Clua. 2018. PEEK: An LSTM recurrent network for motion classification from sparse data. *VISIGRAPP 2018 - Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* 1, Visigrapp (2018), 215–222. https://doi.org/10.5220/0006585202150222

[5] Eduardo Gomes, Luciano Bertini, Wagner Rangel Campos, Ana Paula Sobral, Izabela Mocaiber, and Alessandro Copetti. 2021. Machine learning algorithms for activity-intensity recognition using accelerometer data. *Sensors (Switzerland)* 21, 4 (2021), 1–12. https://doi.org/10.3390/s21041214

[6] Arumugam Thendramil Pavai. 2018. Sensor-Based Human Activity Recognition Using Bidirectional LSTM for Closely Related Activities. (2018). https://scholarworks.lib.csusb.edu/etd/776