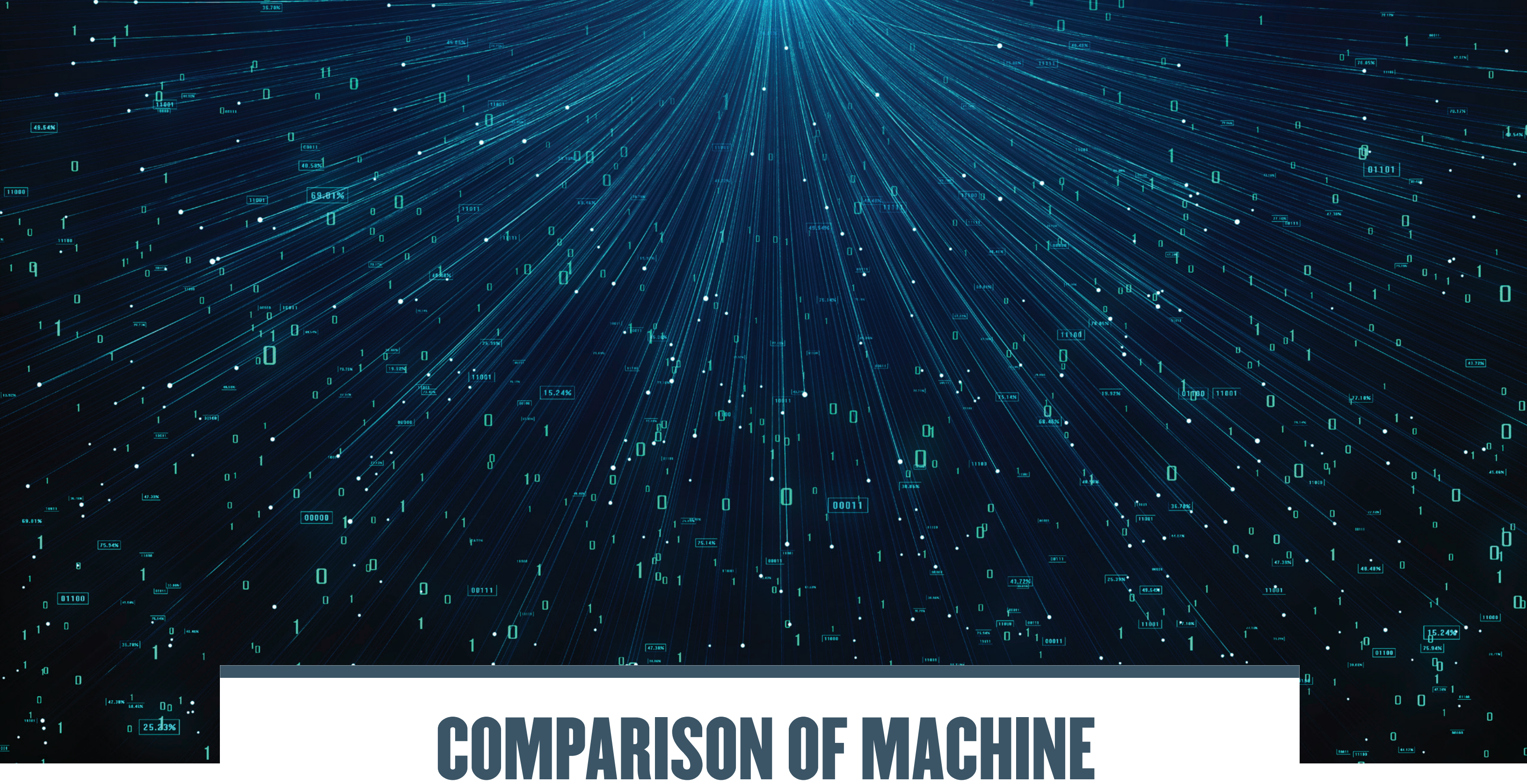

COMPARISON OF MACHINE LEARNING MODELS:

Gesture Recognition Using a Multimodal Wrist Orthosis for Tetraplegics

Student Author

Charlie Martin is pursuing a bachelor's degree in the Weldon School of Biomedical Engineering. Over the summer of 2019, he participated in the Summer Undergraduate Research Fellowship at Purdue University, where he received a Top Research Talk Award for his presentation related to gesture recognition in wrist orthotics. Martin's research interests include neurological treatments and the implementation of assistive devices.

Mentors

Bradley Duerstock is an associate professor of engineering practice in the Weldon School of Biomedical Engineering and the School of Industrial Engineering at Purdue University. He has a courtesy appointment in the Department of Basic Medical Sciences in the College of Veterinary Medicine. Duerstock received his bachelor's degree in interdisciplinary engineering and his PhD in neurotrauma from Purdue University. His research interests focus on overcoming functional limitations through innovative human-technology interfaces and assistive technologies and improving health outcomes through improved physiological sensing and rehabilitation techniques.

Ting Zhang earned her PhD from the School of Industrial Engineering at Purdue University in January 2020. She received her bachelor's degree from the School of Computer Science and Software Engineering at Zhejiang University, China, in 2012 and her master's degree from the School of Industrial Engineering at Purdue University in 2014. Zhang's research interests include assistive technology and human-machine interaction.

http://dx.doi.org/10.7771/2158-4052.1435

Abstract

Many tetraplegics must wear wrist braces to support paralyzed wrists and hands. However, current wrist orthoses have limited functionality to assist a person's ability to perform typical activities of daily living other than a small pocket to hold utensils. To enhance the functionality of wrist orthoses, gesture recognition technology can be applied to control mechatronic tools attached to a novel fabricated wrist brace. Gesture recognition is a growing technology for providing touchless human-computer interaction that can be particularly useful for tetraplegics with limited upper-extremity mobility. In this study, three gesture recognition models were compared—two dynamic time-warping models and a hidden Markov model—in terms of their classification accuracy of gestures from a gesture lexicon known to be accessible to tetraplegics. Gesture data from participants with and without spinal cord injuries was collected using a prototype wrist orthosis. Leave-one-subject-out cross-validation was used to develop a user-independent gesture recognition library. The trained models were then tested using a combination of data from both populations and data separated by population. The classification accuracy and classification time were computed and compared to determine the optimal gesture recognition model.

Martin, C. (2020). Comparison of machine learning models: Gesture recognition using a multimodal wrist orthosis for tetraplegics. *Journal of Purdue Undergraduate Research, 10*, 35–42.

Keywords

spinal cord injury (SCI), gesture recognition, dynamic time warping, hidden Markov model, assistive technology, machine learning

INTRODUCTION

In the United States alone, there are an estimated 280,000 people affected by spinal cord injuries (SCIs), with approximately 17,700 new cases appearing each year (National Spinal Cord Injury Statistical Center, 2019). Out of this population, approximately 59.9% (167,720 people) have complete or incomplete tetraplegia, more colloquially known as quadriplegia (National Spinal Cord Injury Statistical Center, 2019). Tetraplegia is characterized by severe loss of muscular function or paralysis in all four limbs. This causes major difficulties in a person's ability to perform fundamental activities of daily living such as washing, eating, and engaging in leisure activities.

A side effect of tetraplegia is muscle atrophy, which along with nerve damage greatly reduces a person's ability to properly stabilize the wrist; wrist braces, also called wrist orthotics, are commonly used to provide external stabilization to the wrist (Suresh, Manda, Marrero, Jacob, & Duerstock, 2017; Suresh, Madrinan Chiquito, Manda, Jacob, & Duerstock, 2016). Current wrist orthoses are primarily designed for support and have few assistive abilities to help users perform activities of daily living. The few abilities these braces provide are mostly limited to palm cuff pockets into which eating utensils, pens, or toothbrushes can be slotted. The addition of gesture-controlled modules onto a wrist orthosis would greatly increase a person's ability to perform activities of daily living, such as using a flashlight or laser pointer, remotely operating a computer program using Bluetooth, swiping a key card, and signaling for help.

Prior work was conducted to develop a 3D printed wrist brace (Figure 1) equipped with a laser pointer that could be operated using a trained gesture recognition system (Suresh et al., 2017; Suresh et al., 2016). An initial gesture recognition system was tested and was able to detect gestures performed by an individual. The system performed this detection by comparing gathered gesture data with presaved user-specific gesture data sets and was designed for use by a single user.

While this setup provided users with the flexibility to choose their own gesture patterns, an individualized wrist brace training regimen is time-consuming for the end user. By removing the user-dependent training step and replacing it with a user-independent gesture recognition system consisting of predefined gestures, no training by the end user would be required before use of the brace. Multiple machine

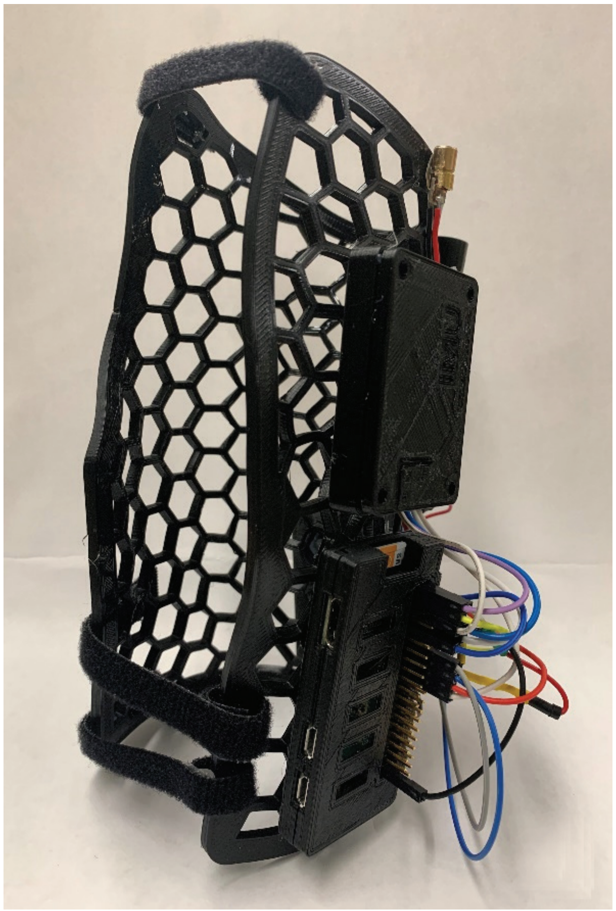

Figure 1. Wrist brace data collection device.

learning algorithms can perform user-independent gesture recognition; however, a high-recognition accuracy is necessary to ensure no inadvertent activation of the wrist brace. The objective of this study is to determine which machine learning model has the highest recognition accuracy while also considering the low computational power of the wrist orthotic systems.

RELATED WORK

The gesture recognition system implemented in the wrist brace incorporates data gathered from a 9 degree of freedom inertial measurement unit (IMU) (Suresh et al., 2017; Suresh et al., 2016). An IMU was selected as the data collection method to allow for long-term daily use in multiple environments (i.e., inside and outside) and to ensure portability and comfort by the user. Other notable methods of gesture recognition, such as computer vision and electromyography, currently fail to best fulfill these requirements (Gupta, Chudgar, Mukherjee, Dutta, & Sharma, 2016; Hussain & Harun-Ur Rashid, 2012; Liu, Zhong, Wickramasuriya, & Vasudevan, 2009; Zhang et al., 2011). Computer vision–based systems are unwieldy and require nontrivial amounts of computational power to return accurate results (Hussain & Harun-Ur Rashid, 2012; Liu et al., 2009). These systems are also highly sensitive to environmental conditions such as inconsistent lighting and variable background colors, which inhibits outdoor use (Chakraborty, Sarma, Bhuyan, & MacDorman, 2017; Zhang et al., 2011). Both wet and dry electrodes used for electromyography recordings have issues with skin irritation after long-term use (Laferriere, Lemaire, & Chan, 2011; Searle & Kirkup, 2000; Yamagami et al., 2018). Wet electrodes have a period of a few hours before skin irritation and signal issues occur due to the drying of the electrode gel, and dry electrodes can have large artifacts in the recorded data due to large movements (Laferriere et al., 2011; Searle & Kirkup, 2000; Yamagami et al., 2018).

Gesture Recognition Algorithm

This research compared three different machine learning models: two dynamic time warping (DTW) models and a hidden Markov model (HMM). DTW models have been used previously in gesture recognition systems with accurate results (Cheng & Zhou, 2019; Hussain & Harun-Ur Rashid, 2012; Liu, Wang, & Ma, 2017; Liu et al., 2009). DTW recognizes gestures by computing the total minimum difference between the two time series data samples, even when the samples are not aligned or of the same length (Gillian, Knapp, & O 'modhrain, 2011). When the minimum distances between the recorded gesture are calculated against a large data set of prerecorded gestures, the lowest minimum difference can be used to determine the gesture performed. This method of gesture recognition, referred to here as Original DTW, was used in prior versions of the wrist brace and is being used in this study as a baseline for comparison (Suresh et al., 2017; Suresh et al., 2016).

Another approach for utilizing DTW for gesture recognition is to compute a template for each gesture that is then used for comparison purposes instead of comparing the test sample with every sample in the data set (Gillian et al., 2011). A template for each gesture is produced by comparing the sample data among itself until the piece of sample data with the smallest total difference is found and identified as the gesture template. The creation of a template drastically reduces the amount of comparisons required to make a gesture determination, as it only requires a single comparison for each gesture that in time will detect a gesture match.

HMM, another widespread algorithm used for gesture recognition, calculates the probability that a future state or event will occur based on a prior state or event (Sutton & McCallum, 2010). For time series data, such as data from an IMU, this ability to determine the probability of a state occurring can be used to determine the overall probability of a time series data sample matching the trained model. This ability has led to HMMs being used in gesture recognition, as they can quantify the probability that a recording from the IMU matches a specific gesture pattern (Liu et al., 2009; Zhang et al., 2011).

METHODS

Gesture Specifications

A predefined set of six gestures is used in this study (shown in Figure 2). The development of this gesture lexicon is based on the lexicon presented by Jiang, Duerstock, and Wachs (2014), which is designed specifically for tetraplegics and considers their physical constraints, including motion fluidity and user fatigue. Specifically, the gesture lexicon was modified by combining the two gestures "upward" and "downward" into a single "up-down" gesture, as shown in Figure 2 (Jiang et al., 2014). The "leftward" and "rightward" gestures were combined into an "in-out" gesture, as shown in Figure 2. These two modified gestures, up-down and in-out, allow orthosis users to start and return their arms to a position of rest before and after the gesture was performed, which alleviates strain and fatigue (Palaniappan & Duerstock, 2019). The gestures also provide for a more fluid arm motion away and toward a person's body.

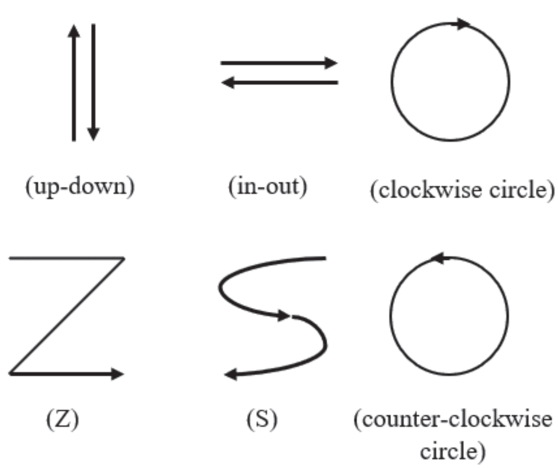

Figure 2. Revised gesture lexicon (Jiang et al., 2014).

Data Collection and Analysis

Raw IMU recordings for each of the six gestures were gathered from 14 participants, of whom 11 were able-bodied and 3 had cervical level SCIs. Participant consent was obtained before commencing data collection using consent forms approved by the Purdue Institutional Review Board (IRB #1802020274). Due to technical issues during data collection, the majority of data from 7 participants was corrupted and unable to be analyzed. Therefore, for this study, only data from the uncorrupted 5 able-bodied and 2 SCI participants was analyzed.

Setup

Participants were asked to wear a 3D printed wrist brace for data collection (see Figure 1). An LSM303DLHC IMU ("LSM303DLHC: e-Compass; 3D accelerometer, 3D digital magnetic sensor, ultra compact, high performance, I2C, SPI interfaces; STMicroelectronics," n.d.) was mounted on the brace to gather acceleration data from the participant's motions. The data was stored on a Raspberry Pi Zero W ("Buy a Raspberry Pi Zero W—Raspberry Pi," n.d.) which was also mounted on the wrist brace. All collected data was transferred to a desktop computer for analysis and training of the machine learning models.

Procedure

Participants were provided with a wrist brace and were instructed to perform 15 iterations of the 6 gestures described above for a total of 90 gestures. These gestures were described, shown through a graphic, and demonstrated before the user commenced testing. The participants were then instructed to perform 15 iterations of the 6 gestures described above for a total of 90 gestures. As performing multiple gestures in a row was known to be fatiguing, the participants were allowed to set their own pace and rest as long as they wished between gestures. To allow for this and to test a real-use activation scenario, an accessible button was used by the participants to initiate the data collection time period for each gesture. Each participant required between 30 and 90 minutes to perform all 90 gestures.

Data Analysis

The data was analyzed in a leave-one-out cross-validation method to determine the overall accuracy of each algorithm. The leave-one-out cross-validation methodology determines overall accuracy by compiling a complete set of data, taking out a single gesture recording, training the machine learning models with the rest of the data, and then testing the accuracy of the machine learning model against the single removed set of data. This process is then repeated until all gesture recordings in the data set have been removed and compared. Confusion matrices, such as the one shown in Figure 3, are common modalities to show the overall accuracy of a machine learning model. The model with the most matches along the central diagonal (indicating correct matches) is the most accurate. Accuracy can also be calculated from these results on a scale of 0 to 1 using the formula

$$ACC = \frac{(TP + TN)}{Total\ Sample\ Size}$$

where ACC is the accuracy, TP is the number of true positive values, TN is the number of true negative values, and Total Sample Size is the total number of samples classified (Saito & Rehmsmeier, 2015).

RESULTS

As can be seen in Figure 3, Original DTW has the highest overall accuracy, with a score of 0.9, compared to HMM (Figure 4), with an overall accuracy of 0.63, and the Template DTW model in Figure 3, with an overall accuracy of 0.26. It is important to note that the overall accuracy of the model does not necessarily correlate to the accuracy for each gesture and instead provides an approximate accuracy for the model as a whole. For example, in the Original DTW confusion matrix (see Figure 3), there is a bias to incorrectly labeling motions as the up-down motion. This bias can also be seen in the HMM, although there is a bit less of a bias. The Template DTW model does not show a trend favoring mislabeling a single motion.

As the motion sensor for the wrist brace is being developed specifically for people with tetraplegia, an analysis was run to differentiate the accuracy between the able-bodied and SCI populations. As with the

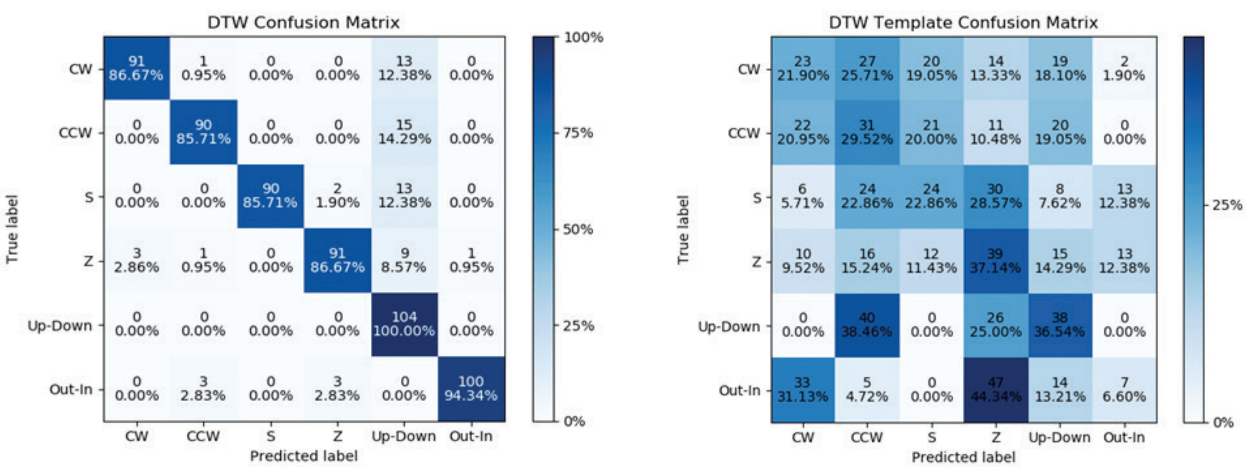

Figure 3. Original DTW and Template DTW confusion matrices.

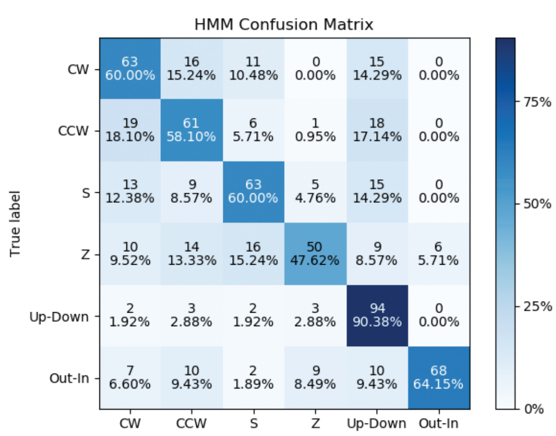

Figure 4. HMM confusion matrix.

general analysis, the models ranked the same for accuracy, with Original DTW achieving the highest, HMM landing in the middle, and Template DTW having the lowest accuracy. For Original DTW, the able-bodied population had an accuracy of 0.84, and the SCI population had an accuracy of 0.62. Figure 5 demonstrates the confusion matrix results for both populations. HMM, which ranked in the middle, had an able-bodied accuracy of 0.64 and an SCI accuracy of 0.52. For Template DTW, the able-bodied accuracy was 0.26, and the SCI accuracy was 0.21. For all of these models, the accuracy of the able-bodied population most represented by the overall accuracy of the model was higher than that of the SCI population. These results are summarized in Table 1.

The time that the gesture recognition system used to distinguish a gesture was calculated. To evaluate a use case for the gesture system, the time from gesture performance to classification was measured, assuming the system had already been pretrained. Table 2 shows the average time in seconds required to classify a gesture of any type: Original DTW performed poorly compared to both Template DTW and HMM. Template DTW and HMM were, however, on par with the amount of time to recognize a gesture.

DISCUSSION

In this study, Original DTW had the highest accuracy, HMM had middling accuracy, and Template DTW had the lowest accuracy. Template DTW possibly had the lowest accuracy due to the variance among gestures performed by different participants. If there is a large amount of variance, this variance would lead to inaccurate templates being created for each gesture. Additionally, during the template creation process, templates could have been created that were matched to multiple gestures instead of a single gesture due to similarities between the templates. This template creation process would throw off the gesture determination by having the templates not be well differentiated between gestures.

For Original DTW and HMM, there is a clear bias toward identifying certain gestures above others. This indicates that there may be many similarities

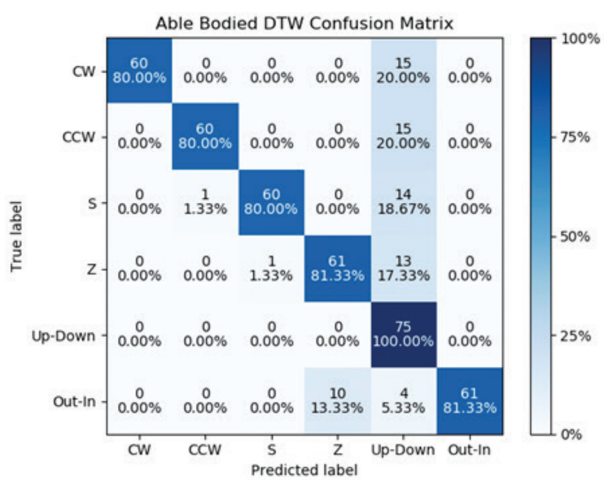

Figure 5. Split population original DTW confusion matrices.

Model	Able-Bodied Accuracy	SCI Accuracy
Original DTW	0.84	0.62
HMM	0.64	0.52
DTW Template	0.26	0.21

Table 1. Accuracy of split population models.

Model	Average Time (sec)	Standard Deviation (sec)
Original DTW	2.25	1.71
Template DTW	0.016	0.00068
HMM	0.040	0.0048

Table 2. Computational time for gesture recognition.

between the gestures performed by the participants that cannot be easily differentiated with these models. For example, in Original DTW (see Figure 3) and HMM (see Figure 4), there is a clear trend of incorrectly identifying gestures as up-down. As many of the motions have vertical components to them, this could indicate that the vertical motion of multiple gestures is detected as being similar, which skews the model toward predicting the motion as up-down.

When comparing the accuracy between the population split models, there is a stark contrast between the able-bodied population and the SCI population. The SCI population sees a lower overall accuracy compared to the able-bodied population. The main reason for this lower accuracy is due to the smaller sample size of the SCI population. There were only two data sets used in analysis for the SCI population compared to the five used in analysis for the able-bodied population. A larger sample size would increase this accuracy by allowing for better training of the model.

CONCLUSION

Three machine learning models for gesture recognition were evaluated for effectiveness in the creation of a user-independent gesture recognition library for tetraplegics. These three models—two DTWs and one HMM—were evaluated for accuracy of gesture recognition and time required to recognize a gesture. While Original DTW and HMM performed well in both a split population and a combined SCI and able-bodied population, Template DTW did not perform well in either scenario. This indicates that Original DTW or HMM would be best used for the creation of a user-independent gesture recognition library. Original DTW, while having the highest computational time by a wide margin, had the greatest accuracy. Therefore, Original DTW is currently suggested as the best model for use in the creation of a user-independent gesture library. All of the models could be improved by gathering a larger data set of gesture data from the SCI population, thereby removing the nuances inherent in each gesture and producing a more robust gesture library. Through the development of a more robust model, incorrect classification of gestures should decrease, which will increase the accuracy of the models explored here. Future work should focus on decreasing the computational load while maintaining accuracy for the Original DTW algorithm. This should be done either by improving the template version of the algorithm or by optimizing the Original DTW algorithm. Focusing on reduction of computational load will decrease the gesture recognition time and increase the usability of the wrist brace by providing a more enjoyable user experience.

ACKNOWLEDGMENTS

Nick Will and Sreeram Nagappa helped with development of the wrist brace 3D structure and some miscellaneous support code. Shruthi Suresh and Shanmugam M. Palaniappan provided technical support and advice on the project. Thank you to the Purdue Office of Engagement for providing funding for this research through the Service Learning Grant Program and to the Purdue Discovery Park for its continued support. This project was made possible through the Summer Undergraduate Research Fellowship and funded through the Purdue College of Engineering, Weldon School of Biomedical Engineering, and B. S. Duerstock.

REFERENCES

Buy a Raspberry Pi Zero W—Raspberry Pi. (n.d.). Retrieved June 25, 2019, from https://www.raspberrypi.org/products/raspberry-pi-zero-w/

Chakraborty, B. K., Sarma, D., Bhuyan, M. K., & MacDorman, K. F. (2017). Review of constraints on vision-based gesture recognition for human-computer interaction. *IET Computer Vision, 12*(1), 3–15. https://doi.org/10.1049/iet-cvi.2017.0052

Cheng, Z., & Zhou, Y. (2019). Finger-worn device based hand gesture recognition using long short-term memory. In *Proceedings—2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018* (pp. 2067–2072). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/SMC.2018.00356

Gillian, N., Knapp, R. B., & O 'modhrain, S. (2011). Recognition of multivariate temporal musical gestures using n-dimensional dynamic time warping. *Nime,* 337–342. Retrieved from http://www.nickgillian.com/papers/Gillian_NDDTW.pdf

Gupta, H. P., Chudgar, H. S., Mukherjee, S., Dutta, T., & Sharma, K. (2016). A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors. *IEEE Sensors Journal, 16*(16), 6425–6432. https://doi.org/10.1109/JSEN.2016.2581023

Hussain, S. M. A., & Harun-Ur Rashid, A. B. M. (2012). User independent hand gesture recognition by accelerated DTW. In *2012 International Conference on Informatics, Electronics and Vision, ICIEV 2012* (pp. 1033–1037). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICIEV.2012.6317364

Jiang, H., Duerstock, B. S., & Wachs, J. P. (2014). A machine vision–based gestural interface for people with upper extremity physical impairments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(5), 630–641. https://doi.org/10.1109/TSMC.2013.2270226

Laferriere, P., Lemaire, E. D., & Chan, A. D. C. (2011). Surface electromyographic signals using dry electrodes. In *IEEE Transactions on Instrumentation and Measurement* (Vol. 60, pp. 3259–3268). Piscataway, NJ:

Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/TIM.2011.2164279

Liu, F. T., Wang, Y. T., & Ma, H. P. (2017). Gesture recognition with wearable 9-axis sensors. In *IEEE International Conference on Communications* (pp. 1–6). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICC.2017.7997414

Liu, J., Zhong, L., Wickramasuriya, J., & Vasudevan, V. (2009). uWave: Accelerometer-based personalized gesture recognition and its applications. In *Pervasive and mobile computing* (Vol. 5, pp. 657–675). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1016/j.pmcj.2009.07.007

LSM303DLHC: e-Compass; 3D accelerometer, 3D digital magnetic sensor, ultra compact, high performance, I2C, SPI interfaces; STMicroelectronics. (n.d.). Retrieved June 26, 2019, from https://www.st.com/en/mems-and-sensors/lsm303dlhc.html

National Spinal Cord Injury Statistical Center. (2019). *Spinal cord injury facts and figures at a glance.* Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26793805%0A http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4064586

Palaniappan, S. M., & Duerstock, B. S. (2019). Developing rehabilitation practices using virtual reality exergaming. In *2018 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2018* (pp. 90–94). Piscataway, NJ: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ISSPIT.2018.8642784

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot is more informative than the ROC Plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Searle, A., & Kirkup, L. (2000). A direct comparison of wet, dry and insulating bioelectric recording electrodes. *Physiological Measurement, 21*(2), 271–283. https://doi.org/10.1088/0967-3334/21/2/307

Suresh, S., Madrinan Chiquito, D. F., Manda, S., Jacob, L., & Duerstock, B. S. (2016). *Motor-activated multi-functional wrist orthotic to assist individuals with cervical spinal cord injuries with activities of daily living.* Retrieved from https://www.resna.org/sites/default/files/conference/2016/pdf_versions/emerging_tech/suresh.pdf

Suresh, S., Manda, S., Marrero, C., Jacob, L., & Duerstock, B. S. (2017). Multifunctional wrist orthotic with universal gesture recognition system. In *Proceedings of the RESNA 2017 Conference, New Orleans, LA, June 28–30, 2017.* https://doi.org/10.1109/ICASSP.2010.549

Sutton, C., & McCallum, A. (2010). An introduction to conditional random fields. Retrieved from http://arxiv.org/abs/1011.4088

Yamagami, M., Peters, K. M., Milovanovic, I., Kuang, I., Yang, Z., Lu, N., & Steele, K. M. (2018). Assessment of dry epidermal electrodes for long-term electromyography measurements. *Sensors (Switzerland), 18*(4). https://doi.org/10.3390/s18041269

Zhang, X., Chen, X., Li, Y., Lantz, V., Wang, K., & Yang, J. (2011). A framework for hand gesture recognition based on accelerometer and EMG sensors. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 41*(6), 1064–1076. https://doi.org/10.1109/TSMCA.2011.2116004