Digit Recognition From Wrist Movements and Security Concerns with Smart Wrist Wearable IOT Devices

Lambert T. Leong  
CS Dept., University of Hawaii, USA  
lambert3@hawaii.edu

Sean Wiere  
MBBE Dept., University of Hawaii, USA  
swiere@hawaii.edu

ABSTRACT
In this paper, we investigate a potential security vulnerability associated with wrist wearable devices. Hardware components on common wearable devices include an accelerometer and gyroscope, among other sensors. We demonstrate that an accelerometer and gyroscope can pick up enough unique wrist movement information to identify digits being written by a user. With a data set of 400 writing samples, of either the digit zero or the digit one, we constructed a machine learning model to correctly identify the digit being written based on the movements of the wrist. Our model’s performance on an unseen test set resulted in 100% accuracy when predicting ten writing samples in real-time. The model’s ability to correctly identify all digits via wrist movement and orientation changes raises security concerns. Our results imply that nefarious individuals may be able to gain sensitive digit based information such as social security, credit card, and medical record numbers from wrist wearable devices.

KEYWORDS
Machine Learning, Cyber Security, Wearables, IOT, Sensor, Artificial Intelligence

2 RELATED WORK
Wearable internet of things (IoT) devices provide a constant data stream and result in a considerable amount of data. Machine and deep learning offers many tools and techniques to analyze the vast and copious amounts of wearable IOT data [10]. Various research efforts are aimed at leveraging machine learning models to help make sense of all data and correlate them to particular task and activity related to sports performance, health, and safety [9, 20].

In our work, we focus specifically on wrist wearables and machine learning models built around the corresponding data. Data from just a wrist mounted gyroscope alone has been shown to be adequate for building a machine learning model to detect hand gestures for a novel human computer interaction (HCI) device [11]. Several groups have explored machine learning to build models that perform writing recognition task from wrist wearable device output. In one instance, additional custom sensors were placed on the upper forearm...
and on the finger tip to capture additional information needed to correctly classify hand gestures [24]. They were also able to identify characters written with one’s index finger with an accuracy of 95%. However, the strength of their model is likely attributed to the data coming from the finger sensor more so, than the wrist sensors. Our work aims to perform written digit recognition from sensors placed solely on an individuals wrist. Word level recognition from smart watches was explored by Xia et al. [23]. Their model was able to achieve an accuracy of 48.8% on word level recognition based off wrist movement and they highlighted potential security concerns of their results. Letter level recognition with smart watches was also explored [7]. In that work, writing tasks were performed on whiteboards and audio input from an on-board microphone was used for segmentation, which helped recognition accuracy. These works assured us that sensor data from wrist wearable devices provides sufficient data for building machine learning models to perform written recognition tasks.

Security issues associated with IoT devices is a popular and an ever growing area of research. IoT devices have been shown to be easily compromised [14] and work has been done, using machine learning, to help improve IoT security and detect threats on IoT devices [8, 17]. Our work is not necessarily concerned with the issue of compromised wrist wearable IoT devices rather, it seeks to exploit a nefarious use case for already available device data. An example of security exploits on readily available wearable device data can be seen in [5, 21]. In these articles, restricted areas such as military bases have been mapped out just by having a wrist wearable user, with security clearance, passively walk around secure areas. Wrist wearables were not compromised in those instances but the use of the already available data (e.g. GPS coordinates) posed an alarming security vulnerability. Sensor based attacks involving wrist wearables to capture keystrokes have been shown to be possible [15]. In another instance, Pandelea et al. showed that a machine learning model could be built using data from on-board smart watch hardware to guess the password being entered onto the device [18]. In that paper, they showed that pressing different keys on the smart watch corresponded to a different set of movements and their model was able to map smart watch movements to different key inputs.

3 METHODS

For the purpose of our investigation we focus on handwritten digits. More specifically we focus only on the wrist movements associated with writing the digit zero and the digit one. We formalize our problem, with respect to machine learning, as a binary classification problem. Working with digits mitigates issues that arise with the written differences in English upper and lower case and cursive alphabet characters.

3.1 Hardware Design

Accelerometers and gyroscopes are common amongst wrist wearables and these hardwares are ideal for capturing the wrist movements and orientation. The accelerometer can record the wrist acceleration in three planes, x, y, and z and the gyroscope can capture the wrist angle or tilt during writing, in the x, y, and z plane. We fabricated our own devices equipped with an accelerometer and gyroscope, similar to those seen in popular wrist wearables. Designing our own hardware allowed us to more accurately capture and label data for our experiments.

![Image](b) The LSM9D1 housed the accelerometer and gyroscope

Figure 1: Micro-controller and inertia measurement unit (IMU)

We used Adafruit’s ESP32 Feather [1] micro-controller board and the LSM9D1 inertial measurement unit (IMU) [2]. The accelerometer and gyroscope are housed on the IMU and serial peripheral interface (SPI) protocols are used to communicate data recordings from the IMU to the ESP32 Feather board. The IMU and micro-controller were connected and assembled into a single housing. Acquiring labeled data or labeling data after collection can be expensive and often requires a degree of processing and cleaning. A switch was added to the design of our device for the sole purpose of parsing and labeling data during collection. As a result, the accelerometer and gyroscope data would only be recorded while the switch was engaged and the switch was only engaged during the writing of either of the digits. The switch allowed us to identify when writing began and ended within the accelerometer and gyroscope data streams and also allowed for immediate labeling.

3.2 Data Collection

We recruited participants who are right hand dominant and write with their right hand. The sensor housing was attached to the posterior side of the ulna and radius at the most distal point from the body. In other words, the housing was attached to the top of the participants wrist, common in convention to how a watch and wrist wearables are normally worn. Participants wrote out digits to fill a 10 cm by 10 cm square region. After the writing of each digit, the data was labeled with the appropriate digit and saved to a file. We collected a total of 400 writing samples of digits which breaks down to 200 samples for the digit zero and 200 samples for the digit one.

3.3 Data Processing and Feature Engineering

Seven data fields, as seen in Table 1, were recorded from our device during data collection. These fields include time, acceleration in three planes (x, y, z), and pitch angle in three planes (x, y, z). It is often the case that certain digits can be written in different amounts of time. For instance, writing the digit one usually takes less time to write than the digit zero. To deter our machine learning algorithm from only learning on the time it takes to write a digit, we extracted features which are uncoupled from time. The features we extracted were the minimum, maximum, and mean for acceleration and pitch angle in all three planes (x, y, z).
Using the acceleration features we were able to engineer a velocity and a displacement feature. Integrating over the acceleration yielded the velocity and subsequently, integrating over the velocity yielded the displacement. Velocities and displacements were calculated in all three planes. The minimum, maximum, and mean velocities for all three planes were added to our current feature list. The total displacement in all three \((x,y,z)\) planes was also added to the feature list. Lastly, we calculated the total overall displacement and added that to the new feature list. As a result we transformed our original seven features into 31 new features, shown in Table 2, that are irrespective of time.

### Table 1: List of original features gathered from the device

| Metrics        | Axis      | Value Type | Feature Count |
|----------------|-----------|------------|---------------|
| Acceleration   | x,y,z     | 3          |               |
| Pitch Angle    | x,y,z     | 3          |               |
| Time           | Total     | 1          |               |
| **Total Features** |          |            | **7**          |

#### 3.4 Principal Components Analysis and Class Separability Investigation

Principal components analysis (PCA) was used to evaluate the explained variance of each of the 31 re-engineered features. Data was normalized and Scikit-Learn’s [19] PCA module was used to perform PCA. It was found that the top three principal components explain greater than 99.99% of the variance. Mapping the principal components (PC) back to the original features revealed that the top three components are the maximum pitch angle in the \(z\) plane, the displacement in the \(x\) plane and the mean pitch angle in the \(z\) plane, respectively. Distributions for each of the three features were generated with respect to class and are presented in Figure 2.

Figure 2(a) shows the best separation of classes when compared to the \(x\) displacement and mean \(z\) pitch angle. There is still some, non-negligible overlap in Figure 2(a) and other features may be needed to get clear class separation. Figure 2(b) and 2(c) have a considerable overlap but some separation can be seen and including these features may help define a better decision boundary. Scatter plots were generated to investigate the separability of the two classes and a potential decision boundaries. Plots are shown in Figure 3.

Figure 3(a) shows fairly good separation amongst the classes however there is some overlap in the middle where the two classes meet. Figure 3(b) also presents with fairly good separation but there is also overlap in the middle where the two classes meet. Good separability can be seen with respect to two of the top three principal components. Figure 3(c) plots all three principal components against each other to see if better separability can be seen in a higher dimension. Overlap of the two classes is still present, however, there is more separation seen in three dimensions than seen in two dimensions. This is somewhat expected as greater separability is often observed in higher dimensions and using more features may create more separation between the two classes. Therefore, it is likely that more features are needed to generate a stronger model and thus we chose to build a model utilizing all 31 features rather than just the top three PC.

#### 3.5 Model and Hyper Parameter Tuning

The dataset of 400 writing samples was randomly split into a training, validation, and testing sets via a 60%, 20%, 20% split. The number of samples for each class is shown in Table 3. Our relatively modest sample size led us to first explore simpler models rather than a deep learning approach. We explored ensembling methods which led us to use extreme gradient boosting with the help of Sci-Kit Learn’s xgboost package. To evaluate our model choice we set the following hyper parameters to the package defaults as follows: number of boosting stages at 100, learning rate of 0.1, max tree depth of six, and the auto tree algorithm. The classifier was trained on the training set.
Table 2: List of engineered features. Features used in final model

| Metrics        | Axis | Value Type       | Feature Count |
|----------------|------|------------------|---------------|
| Acceleration   | x,y,z| Minimum, Mean, Maximum | 9             |
| Pitch Angle    | x,y,z| Minimum, Mean, Maximum | 9             |
| Velocity       | x,y,z| Minimum, Mean, Maximum | 9             |
| Displacement   | x,y,z| Total            | 4             |
| **Total Features** |      |                  | **31**        |

Figure 3: Plots of top three principal components versus each other to look for separability

set with all 31 feature and the validation set achieved an AUROC of 88.03%. Preliminary performance was good and we proceeded with gradient boosting as our model choice.

Table 3: Breakdown of sample per class

| Dataset  | Class digit 0 | Class digit 1 |
|----------|---------------|---------------|
| Train    | 122           | 118           |
| Validation | 42           | 38            |
| Test     | 36            | 44            |
| **Total** | **200**       | **200**       |

Hyper parameters, which include the number of estimators, the learning rate, the maximum depth of a tree, and the tree construction algorithm, were optimized using an exhaustive grid search. Models were trained on different combinations of hyper parameters using five fold cross-validation. A list of hyper parameters and their explored ranges are presented in Table 4.

Table 4: Hyper parameters and ranges explored

| Parameter       | Values and ranges |
|-----------------|-------------------|
| n_estimators    | 1000, 2000, 3000, 4000, 5000 |
| tree_algorithm  | hist, exact       |
| max_depth       | 1, 2, 3, 4, 5, 6, 7, 8 |
| learning rate   | 0.1, 0.3, 0.5     |

The best hyper parameters, from the exhaustive grid search, which yielded the best AUROC’s on the validation, set are shown in Table 5. The final models were retrained on the combined, train and validation, dataset using five fold cross-validation. The best hyper parameters, from Table 5 were used in the retraining of the final PCA model and full 31 feature model.

3.6 Performance Evaluation

Model performance was mainly evaluated using the area under the receiver operating characteristic (AUROC) curve. AUROC values are reported as values between zero and one where values closer to one indicate better performance. The test set, which is 20% of the dataset that was never seen by the models, was used to calculate the final AUROC values. Final AUROC values from both the final PCA model and the final full feature model were compared to each other to evaluate model performance.
4 PERFORMANCE RESULTS

Two models were constructed using a different number of features. The first model was constructed using the top three principal components (PC) and the second model used all 31 features. The better of the two models was used to construct our final model which was ported to our device for real-time evaluations.

4.1 PCA Model Results

As mentioned in Section 3.4, the top three principal components explain 99.99% of the variance and we investigated if these features were sufficient to build a good classifier. The test set, which was held out and not seen by the model during construction and hyper parameter tuning, was offered to the model constructed with the three features, shown in Table 6, corresponding to the top three PC. The receiver operating characteristic (ROC) curve and AUROC value is shown in Figure 4.

Table 6: List of PCA reduced features

| Metrics          | Axis | Value Type | Feature Count |
|------------------|------|------------|---------------|
| Pitch Angle      | z    | Maximum    | 1             |
| Displacement     | x    | Total      | 1             |
| Pitch Angle      | z    | Mean       | 1             |
| **Total Features**|      |            | **3**         |

Table 7: Confusion matrix of PCA models performance on held out test set

| Predicted Digit 0 | Digit 1 |
|-------------------|---------|
| Digit 0           | 33      | 3       |
| Digit 1           | 8       | 36      |

Figure 4 shows an improvement in model performance when all 31 features are used to train our model. The second model achieved an AUROC of 1.00 on our test set. This indicates that it was able to predict which digit was written without any errors. Class break downs and predictions by class are shown in the confusion matrix in Table 8. Although the top three PC explain the great majority of the variance, the other components contain the
With the device attached to their wrist, the volunteer wrote their assigned digit and the devices output a prediction. Results from real time predictions could be made. A sample set of ten written digits, which we used, are common hardware on-board wrist wearables. Our findings imply a potential security vulnerability that is associated with wrist wearable devices. Accelerometers and gyroscopes, which we used, are common hardware on-board wrist wearables. While the machine learning model we developed is simple and only for the binary classification of two written digits, it is an important first step and brings awareness to some security vulnerabilities associated with wrist wearables.

4.3 Real Time Performance

The second, full feature, model was retrained on the entire dataset with the same optimum hyper parameters to produce our final model. The final model was loaded onto our hardware device so that real time testing are reported in Table 9.

Table 9: Confusion matrix of final models performance during real time testing

|       | Digit 0 | Digit 1 |
|-------|---------|---------|
| Actual| Digit 0 | 5       | 0       |
| Digit 1| 0       | 5       |

As seen from the confusion matrix, in Table 9, the final model was able to predict all written digits accurately which results in an AUROC of 1. Real time performance results reassures us of our models generalizability and, again, does not lead us to suspect significant overfitting. Performance metrics, including accuracy, precision, recall and F1 score, are reported for the PCA, full feature, and real time full feature models in Table 10.

5 CONCLUSIONS

Our findings imply a potential security vulnerability that is associated with wrist wearable devices. Accelerometers and gyroscopes, which we used, are common hardware on-board wrist wearables. Using machine learning, we were able to identify that the wrist movements involved in writing the digit zero is unique and different to the wrist movements involved in writing the digit one. As a result, a robust machine learning model was constructed which demonstrated perfect real-time prediction performance. Our results imply a plausible reality where sensitive information can be recorded from users during writing while wearing some smart watch or wrist wearable device. In addition, our potential security exploit results from using already available data from smart wrist wearables. Our methods do not involve nor require compromising the wearable devices themselves. While the machine learning model we developed is simple and only for the binary classification of two written digits, it is an important first step and brings awareness to some security vulnerabilities associated with wrist wearables.

6 FUTURE WORK

We hope to explore how our data and model relates to left hand dominant users. It is hypothesized that since your left hand is a mirror image of the right that simply flipping signs or the direction vector will lead to the correct solution. More data, specific to left handed users, is needed to explore how handedness affects generalizability of machine learning writing recognition models.

The size of our data set was modest and only contained wrist movement data for the digits one and zero. Our binary classification problem was not hard to solve and thus a simpler model was sufficient which also raises concerns. The use of a simpler machine learning model implies that capturing users private information may be a trivial task. Our current findings warrant further work to aggregate wrist movement data for the writing of all ten digits, zero through nine. Working with all ten digits presents as a multi-class classification problem and while it may be more difficult there exists more powerful tools which were not explored in our work. In fact, deep neural networks could potentially handle classifying wrist movements for all ten digits fairly easily. More work is needed to explore these concepts to improve and maintain security for the vast variety of wearable IoT devices.

REFERENCES

[1] [n. d.]. Adafruit ESP32 Feather board datasheet. (n. d.). https://www.espressif.com/en/products/hardware/esp32/overview
[2] [n. d.]. Adafruit LSM9DS1 Accelerometer Gyro Magnetometer 9-DOF Breakout. (n. d.). https://learn.adafruit.com/adafruit-lsm9dss-accelerometer-plus-gyro-plus-magnetometer-9-dof-breakout
[3] [n. d.]. Fitbit Development: Accelerometer API. (n. d.). https://dev.fitbit.com/build/reference/device-api/accelerometer/
[4] [n. d.]. Specifications | Samsung Galaxy Watch. (n. d.). https://www.samsung.com/global/galaxy/galaxy-watch/specs/
[5] 2018. Fitness app Strava lights up staff at military bases. (Jan 2018). https://www.bbc.com/news/technology-42853072
[6] Rasha Al-Eidan, Hend Al-Khalifa, and Abdul Malik Al-Salman. 2018. A Review of Wrist-Worn Wearable: Sensors, Models, and Challenges. Journal of Sensors 2018 (12 2018), 1–20. https://doi.org/10.1155/2018/5853917
### Table 10: Performance metrics for the PCA, full feature models, and real time full feature models

| Model                  | Accuracy (%) | Digit 0 Precision (%) | Digit 0 Recall (%) | Digit 0 F1 Score (%) | Digit 1 Precision (%) | Digit 1 Recall (%) | Digit 1 F1 Score (%) |
|------------------------|--------------|-----------------------|--------------------|----------------------|-----------------------|--------------------|----------------------|
| PCA                    | 86.25        | 80.49                 | 91.67              | 85.74                | 92.31                 | 81.82              | 86.75                |
| Full Feature           | 100.00       | 100.00                | 100.00             | 100.00               | 100.00                | 100.00             | 100.00               |
| Real Time, Full Feature| 100.00       | 100.00                | 100.00             | 100.00               | 100.00                | 100.00             | 100.00               |

[7] L. Arðåjser, P. Bissig, P. Brandes, and R. Wattenhofer. 2016. Recognizing text using motion data from a smartwatch. In 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). 1–6. https://doi.org/10.1109/PERCOMW.2016.7457172

[8] Amin Azzmoodeh, Ali Dehghantanha, Mauro Conti, and Kim-Kwang Raymond Choo. 2018. Detecting crypto-ransomware in IoT networks based on energy consumption footprint. Journal of Ambient Intelligence and Humanized Computing 9, 4 (01 Aug 2018), 1141–1152. https://doi.org/10.1007/s12652-017-0558-5

[9] Juan Carlos Davila, Ana-Maria Cretu, and Marek Zaremba. 2017. Wearable Sensor Data Classification for Human Activity Recognition Based on an Iterative Learning Framework. (Jun 2017). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5492798/

[10] R. Fernandez Molanes, K. Amarasinghe, J. Rodriguez-Andina, and M. Manic. 2018. Deep Learning and Reconfigurable Platforms in the Internet of Things: Challenges and Opportunities in Algorithms and Hardware. IEEE Industrial Electronics Magazine 12, 2 (June 2018), 36–49. https://doi.org/10.1109/MIE.2018.2834843

[11] Hobeom Han and Sang Won Yoon. 2019. Gyroscope-Based Continuous Human Hand Gesture Recognition for Multi-Modal Wearable Input Device for Human Machine Interaction. Sensors 19, 11 (2019). https://doi.org/10.3390/s19112562

[12] Apple Inc. [n. d.]. Human Interface Guidelines. [n. d.]. https://developer.apple.com/design/human-interface-guidelines/watchos/user-interaction/accelerometer-and-gyroscope/

[13] Do-Hoon Lee and Hwan-Gue Cho. 1998. The beta-velocity model for simulating handwritten Korean scripts. In Electronic Publishing, Artistic Imaging, and Digital Typography, Roger D. Hersch, Jacques André, and Heather Brown (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 252–264.

[14] M. Lee, K. Lee, J. Shim, S. Cho, and J. Choi. 2016. Security threat on wearable services: Empirical study using a commercial smartband. In 2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). 1–5. https://doi.org/10.1109/ICCE-Asia.2016.7804766

[15] Xiangyu Liu, Zhe Zhou, Wenrui Diao, Zhou Li, and Kehuan Zhang. 2015. When Good Becomes Evil: Keystroke Inference with Smartwatch. In Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security (CCS ’15). ACM, New York, NY, USA, 1273–1285. https://doi.org/10.1145/2810303.2813668

[16] P Mavrogiorgou, R Mergl, P Tigges, J El Husseini, A Schröter, G Juckel, M Zaudig, and U Hegerl. 2001. Kinematic analysis of handwriting movements in patients with obsessive-compulsive disorder. Journal of Neurology, Neurosurgery & Psychiatry, 70, 5 (2001), 605–612. https://doi.org/10.1136/jnnp.70.5.605

[17] H. H. Pajouh, R. Javidan, R. Khayami, A. Dehghantanha, and K. R. Choo. 2019. A Two-Layer Dimension Reduction and Two-Tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks. IEEE Transactions on Emerging Topics in Computing 7, 2 (April 2019), 314–327. https://doi.org/10.1109/TETC.2016.2633228

[18] A. Pandelea and M. Chiriou. 2019. Password Guessing Using Machine Learning wearable. In 2019 22nd International Conference on Control Systems and Computer Science (CSCS). 304–311. https://doi.org/10.1109/CSCS.2019.00055

[19] F Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12 (2011), 2825–2830.

[20] Giulia Regalia, Francesco Onorati, Matteo Lai, Chiara Caborni, and Rosalind W. Picard. 2019. Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands. Epilepsy Research 153 (2019), 79–82. https://doi.org/10.1016/j.eplepsres.2019.02.007

[21] Liz Sly. 2018. U.S. soldiers are revealing sensitive and dangerous information by jogging. (Jan 2018). https://www.washingtonpost.com/world/a-map-showing-the-users-of-fitness-devices-lets-the-world

[22] Daryl J. Wile, Ranjit Ranawaya, and Zelma H.T. Kiss. 2014. Smart watch accelerometer for analysis and diagnosis of tremor. Journal of Neuroscience Methods 230 (2014), 1–4. https://doi.org/10.1016/j.jneumeth.2014.04.021

[23] Q. Xia, F. Hong, Y. Feng, and Z. Guo. 2018. MotionHacker: Motion sensor based eavesdropping on handwriting via smartwatch. In IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 458–463. https://doi.org/10.1109/INFOCOMW.2018.8406879

[24] Chao Xu, Parth H. Pathak, and Prasant Mohapatra. 2015. Finger-writing with a-map-showing-the-users-of-fitness-devices-lets-the-world

[25] Yunze Zeng, Parth H. Pathak, Chao Xu, and Prasant Mohapatra. 2014. Your AP Knows How You Move: Fine-grained Device Motion Recognition Through WiFi. In Proceedings of the 1st ACM Workshop on Hot Topics in Wireless (HotWires-‘14). ACM, New York, NY, USA, 49–54. https://doi.org/10.1145/2643614.2643620