FaceGuard: A Wearable System to Avoid Face Touching

Mohamad A. Eid[1*], Allan M. Michelin[1], Georgios Korres[1], Sara Ba'ara[1], Hadi Assadi[1], Haneen Alsuradi[1], Rony Sayegh[2], Antonis Argyros[3]

[1]Division of Engineering, New York University Abu Dhabi, United Arab Emirates, [2]Cleveland Clinic Abu Dhabi, United Arab Emirates, [3]University of Crete, Greece

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

AMM, SB, HA, and HA developed the prediction component, including the data collection protocol and the CNN model development and evaluation, GK, AA, and ME developed the sensory feedback psychophysical study. RS, AA, and ME contributed to the conceptualization of the solution.

Keywords

Face touching avoidance, IMU-based hand tracking, Sensory feedback, vibrotactile stimulation, wearable technologies for health care

Abstract

Word count: 314

Most people touch their faces unconsciously, for instance to scratch an itch or to rest one's chin in their hands. To reduce the spread of the novel coronavirus (COVID-19), public health officials recommend against touching one's face, as the virus is transmitted through mucous membranes in the mouth, nose and eyes. Students, office workers, medical personnel and people on trains were found to touch their faces between 9 and 23 times per hour. This paper introduces FaceGuard, a system that utilizes deep learning to predict hand movements that result in touching the face, and provides sensory feedback to stop the user from touching the face. The system utilizes an inertial measurement unit (IMU) to obtain features that characterize hand movement involving face touching. Time-series data can be efficiently classified using 1D-Convolutional Neural Network (CNN) with minimal feature engineering; 1D-CNN filters automatically extract temporal features in IMU data. Thus, a 1D-CNN based prediction model is developed and trained with data from 4800 trials recorded from 40 participants. Training data are collected for hand movements involving face touching during various everyday activities such as sitting, standing, or walking. Results showed that while the average time needed to touch the face is 1200 ms, a prediction accuracy of more than 92% is achieved with less than 550 ms of IMU data. As for the sensory response, the paper presents a psychophysical experiment to compare the response time for three sensory feedback modalities, namely visual, auditory, and vibrotactile. Results demonstrate that the response time is significantly smaller for vibrotactile feedback (427.3 ms) compared to visual (561.70 ms) and auditory (520.97 ms). Furthermore, the success rate (to avoid face touching) is also statistically higher for vibrotactile and auditory feedback compared to visual feedback. These results demonstrate the feasibility of predicting a hand movement and providing timely sensory feedback within less than a second in order to avoid face touching.

Contribution to the field

In this paper, we present a system that utilizes machine learning to predict hand movements that result in touching the face, and provides sensory feedback to stop the user from touching the face. The system utilizes an inertial measurement unit (IMU) to obtain features that characterize hand movement involving face touching. A Convolutional Neural Network (CNN) based prediction model is developed and trained with data from 4800 trials recorded from 40 participants. Training data are collected for hand movements involving face touching during various everyday activities such as sitting, standing, or walking. Results showed that while the average time needed to touch the face is 1200 ms, a prediction accuracy of more than 92% is achieved with less than 550 ms of IMU data. As for the sensory response, the paper presents a psychophysical experiment to compare the response time for three sensory feedback modalities, namely visual, auditory, and vibrotactile. Results demonstrate that the response time is significantly smaller for vibrotactile feedback (427.3 ms) compared to visual (561.70 ms) and auditory (520.97 ms). Furthermore, the success rate (to avoid face touching) is also statistically higher for vibrotactile and auditory feedback compared to visual feedback.

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: The studies involving human participants were reviewed and approved by Institutional Review Board for Protection of Human Subjects at New York University Abu Dhabi (Project \# HRPP-2020-108). The patients/participants provided their written informed consent to participate in this study.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

FaceGuard: A Wearable System to Avoid Face Touching

Allan Michael Michelin [1]**, Georgios Korres** [1]**, Sara Ba'ara** [1]**, Hadi Assadi** [1]**, Haneen Alsuradi** [1]**, Rony Sayegh** [2]**, Antonis Argyros** [3]**, and Mohamad Eid** [1,*]

[1] *Applied Interactive Multimedia Lab, Engineering Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates*

[2] *Clinical Associate Professor, Cornea and Refractive Surgery, Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates*

[3] *Professor at the Computer Science Department (CSD), University of Crete (UoC), Crete, Greece*

Correspondence*:
Mohamad Eid
mohamad.eid@nyu.edu

ABSTRACT

Most people touch their faces unconsciously, for instance to scratch an itch or to rest one's chin in their hands. To reduce the spread of the novel coronavirus (COVID-19), public health officials recommend against touching one's face, as the virus is transmitted through mucous membranes in the mouth, nose and eyes. Students, office workers, medical personnel and people on trains were found to touch their faces between 9 and 23 times per hour. This paper introduces FaceGuard, a system that utilizes deep learning to predict hand movements that result in touching the face, and provides sensory feedback to stop the user from touching the face. The system utilizes an inertial measurement unit (IMU) to obtain features that characterize hand movement involving face touching. Time-series data can be efficiently classified using 1D-Convolutional Neural Network (CNN) with minimal feature engineering; 1D-CNN filters automatically extract temporal features in IMU data. Thus, a 1D-CNN based prediction model is developed and trained with data from 4800 trials recorded from 40 participants. Training data are collected for hand movements involving face touching during various everyday activities such as sitting, standing, or walking. Results showed that while the average time needed to touch the face is 1200 ms, a prediction accuracy of more than 92% is achieved with less than 550 ms of IMU data. As for the sensory response, the paper presents a psychophysical experiment to compare the response time for three sensory feedback modalities, namely visual, auditory, and vibrotactile. Results demonstrate that the response time is significantly smaller for vibrotactile feedback (427.3 ms) compared to visual (561.70 ms) and auditory (520.97 ms). Furthermore, the success rate (to avoid face touching) is also statistically higher for vibrotactile and auditory feedback compared to visual feedback. These results demonstrate the feasibility of predicting a hand movement and providing timely sensory feedback within less than a second in order to avoid face touching.

Keywords: Face touching avoidance, IMU-based hand tracking, sensory feedback, vibrotactile stimulation, wearable technologies for health care

1 INTRODUCTION

Coronavirus disease (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread worldwide, with more than 88 million cases and 1.9 million fatalities as of January, 2021 WHO (2020). Maintaining social distancing, washing hands frequently, avoiding touching the face including eyes, nose, and mouth, are the major methods associated with preventing COVID-19 transmission Chu et al. (2020). Contaminated hands have the potential to disseminate COVID-19 especially if associated with touching the face Macias et al. (2009). Face touching is an act that can happen without much thought, and in fact, happens with such a high occurrence that reducing it could mitigate a heavy source of transmission. Beyond simple skin irritations, face touching has been linked to emotional and cognitive processes Barroso et al. (1980); Mueller et al. (2019), increasing with attentiveness while tasks are being performed, as well as with increasing pressure and anxiety Harrigan (1985). For such common underlying motives, it is no surprise to see that on average a person touches their face 23 times in an hour Kwok et al. (2015). Given that the primary source of COVID-19 transmission is through contact with respiratory droplets (via the nose, mouth, or eyes, either directly from another individual or picked up from a surface Pisharady and Saerbeck (2015)), avoiding face touching is of a great value.

Developing a system to avoid face touching outright by stopping hand movement raises two main challenges. First of all, a system must predict rather than detect when a hand movement will result in face touching well before the hand reaches the face. Secondly, once a hand movement is predicted to result in face touching, a sensory feedback must be presented immediately in order to stop the hand movement and thus avoid face touching. Note that the prediction and response components are evaluated separately to better analyze the capabilities/limits of each component.

1.1 Predicting hand movement

Predicting face touching requires precise hand tracking. Two common approaches for tracking hand movement are vision-based approaches Al-Shamayleh et al. (2018) and wearable sensor-based approaches Mummadi et al. (2018); Jiang et al. (2017b). A combination of these have also shown potential for enhanced accuracy Jiang et al. (2017a); Siddiqui and Chan (2020). Vision-based hand tracking utilizes camera networks Pisharady and Saerbeck (2015), and as mentioned, can be supplemented with wearable devices such as motion sensor systems placed along the body, to map either whole body or hand movement Liu et al. (2019). One particular wearable device often used is the inertial measurement unit (IMU), capable of collecting data along 6 degrees of freedom, with three additional angular sensors to enable a total of 9 inputs. Found in many smart watches, the IMU is equipped with an accelerometer and gyroscope, providing an inexpensive option that is not only accurate, taking measures along all three dimensions for each of its components, but also one that does not require complementary infrastructure to operate. This allows the IMU to be versatile yet effective in the context in which it is implemented.

Paired with an appropriate machine learning model, the data from an IMU can be used to notify a user how often they are touching their face, as well as whether they have done so after each movement. IMUs have been used to correctly identify a completed face touch with high accuracy Fu and Yu (2017); Rivera et al. (2017). Even though detecting face touching greatly supports awareness training, it does not prevent face touching from happening. The motivation of the proposed system is to apply machine learning in order to predict face touching and provide vibrotactile feedback to prevent it rather than detecting it.

1.2 Sensory Feedback for Motor Control

Along with the development of hand tracking, the user must also be notified of their impending action before it is committed, with ample time for them to react. The notification must be delivered through a medium that will elicit the fastest response time. The three feedback modalities of relevance are visual,

     2

auditory, and vibrotactile, and it has been shown that vibrotactile feedback produces the fastest response times Ng and Chan (2012). Vibrotactile feedback systems can be used to achieve this, with benefits similar to that of an IMU, being cost-effective, and easily implemented into a wearable device.

A low-cost wearable system that prevents people from touching their face, and in the long run, assist people in becoming more aware of their face-touching, is proposed. The system exploits widespread and off-the-shelf smartwatches to track the human hand and provide timely notification of hand movement in order to stop touching the face. The decision to build the system with just a smartwatch makes it immediately available to people, without the requirement of building or wearing additional hardware. The system assumes a smartwatch with an IMU module and a vibration motor; a reasonable assumption as most commercial smartwatches are equipped with such hardware. Although preventing the spread of COVID-19 is the most evident, the system can be adapted for other applications such as habit reversal therapy (HRT) Bate et al. (2011) and treatment of chronic eye rubbing McMonnies (2008). The main contributions of this paper are summarized as follows:

1. Proposing a conceptual approach that utilizes IMU data to predict if a hand movement would result in face touching and provides real-time sensory feedback to avoid face touching.

2. Developing a model for tracking hand movement and predicting face touching using convolutional neural networks based on IMU data. To train the model, a database of 4,800 hand motion trials recording with 40 users under three conditions, sitting, standing, and walking is built.

3. Presenting a psychophysical study with 30 participants to compare the effectiveness of sensory feedback modalities, namely visual, auditory, and vibrotactile, to stop the hand while already in motion before reaching the face. The response time and success rate were used as the evaluation metrics for the comparison.

2 RELATED WORK

2.1 Understanding Hand Movement

The detection and classification of body activity is a major area of research, with applications and techniques ranging from wearable electrocardiogram recorders to classify body movements in patients with cardiac abnormalities Pawar et al. (2007), recognition and 3D reconstruction of the face using computer vision Chen et al. (2017); Zhao et al. (2018); Yang and Lv (2020); Lv (2020), to activity tracking of remote workers through sensory systems Ward et al. (2006); Manghisi et al. (2020). For instance, a system named HealthSHIELD utilized Microsoft Kinect Azure D-RGB camera to detect high/low risk face touching in order to monitor compliance with behavioral protection practices. Results demonstrated an overall accuracy of 91%. Inertial Measurement Units (IMU) are another particularly common alternative that although can be used in tandem with other systems Corrales et al. (2008), can provide exceptional results on its own Olivares et al. (2011).

In connection to real-time hand movement recognition in virtual reality games, a wearable IMU has been investigated as an alternative to simple button presses on a controller to identify player action intent Fu and Yu (2017). Similar to the IMU implementation of our own study, an accelerometer, gyroscope, and magnetometer are used as the sensor inputs for classification. Once a user moves their hand in a predetermined pattern, a trained long short term memory (LSTM) model identifies the movement, and the relevant in-game controls are carried out.

Detecting the touching of one's face using an IMU has been examined recently Christofferson and Yang (2020). A convolutional neural network is used to identify whether a user had touched their face at the end

111 of a gesture. Once a user made their move, the collected data from the 9 data modes of the IMU are passed
112 through a trained model, with a face touch classification provided simply as true or false. This approach
113 resulted in a 99% accuracy rate.

114   As is seen in previous studies, the deep learning model used alongside the IMU varies. Requiring a time
115 series based solution, recurrent neural networks, particularly LSTM, and convolutional neural network
116 (CNN) models have been implemented with significant success Rivera et al. (2017); Christian et al. (2019).
117 Combinations of CNN layers with LSTM models have also been effective in processing IMU data Silva do
118 Monte Lima et al. (2019). However, in related works where classification time is relevant, a standalone
119 CNN has shown great promise Huang et al. (2017).

2.2   Real-time Sensory Feedback

121   In order to provide a real-time sensory feedback to stop the hand movement and avoid face touching,
122 multiple feedback sensory modalities can be utilized. Sensory feedback is usually presented through
123 visual, auditory, and tactile modalities. Visual modality stimuli such as flashing is common in several
124 warning systems, such as road transport industries Solomon and Hill (2002) and crosswalk warning systems
125 Hakkert et al. (2002). In addition to the use of vision, auditory modality is widely used in transport, heath
126 care, and industrial environments as it has an immediate arousing effect Sanders (1975). For instance, a
127 previous study showed that auditory alarms used in helicopter environments conveyed urgency Arrabito
128 et al. (2004). Comparing the two modalities, it was found that the response time to visual and auditory
129 stimuli is approximately 180–200 and 140–160 milliseconds, respectively Thompson et al. (1992). This is
130 based on a previous finding that an auditory stimulus takes only 8–10 ms to reach the brain whereas visual
131 stimulus takes 20–40 ms Kemp (1973). However, there are several factors that influence the average human
132 response time include age, gender, hand orientation, fatigue, previous experience, etc. Karia et al. (2012).

133   Vibrotactile modality has also been found to improve the reaction time for several applications such
134 as drone tele-operation Calhoun et al. (2003); Macchini et al. (2020), collision avoidance while driving
135 Scott and Gray (2008), and alteration of motor command in progress (such as altering a reach in progress)
136 Godlove et al. (2014). The temporal aspects of visual and vibrotactile modalities, as sources of feedback
137 about movement control, are examined in Godlove et al. (2014). A modified center-out reach task where the
138 subject's hand movement was occasionally interrupted by a stimulus that instructed an immediate change
139 in reach goal is utilized. Results demonstrated that the response for tactile stimuli was significantly faster
140 than for visual stimuli.

141   Utilizing vibrotactile feedback for alarming the user about face touching has recently been studied. A
142 commercial product, named IMMUTOUCH, utilized a smart wristband that vibrates every time the user
143 touches their face Immutouch (2020). A recent research study presented a wearable system that utilizes a
144 smartwatch to provide vibrotactile feedback and a magnetic necklace to detect when the hand comes to a
145 close proximity to the face D'Aurizio et al. (2020). Even though these solutions are a great step forward to
146 reducing the number of face touches and their duration, they do not consider real-time touch avoidance.
147 Furthermore, these studies did not perform any systematic studies to determine the most effective sensory
148 feedback modality to stop the hand movement and eventually avoid face touching. Aside from differences
149 in the type and architecture of the deep learning model used for classification, our study employs a wearable
150 IMU not just to classify a gesture, but to predict a gesture before it happens. The motion input data therefore
151 will not include the final portion of an individual's hand movement, placing a limit on the available data for
152 training. In examining feasibility of success under such constraints, optimal sensory feedback thus plays a
153 significant role.

3 PROPOSED APPROACH

154 A high-level description of the system is visualized in Figure 1. The system utilizes IMU data to measure
155 hand movement, convolutional neural networks to predict, in real time, whether a hand movement will
156 involve touching the face, and vibrotactile feedback to alert the user so they stop their hand movement
157 before touching their face. Note that the system must perform in real time in order to generate response to
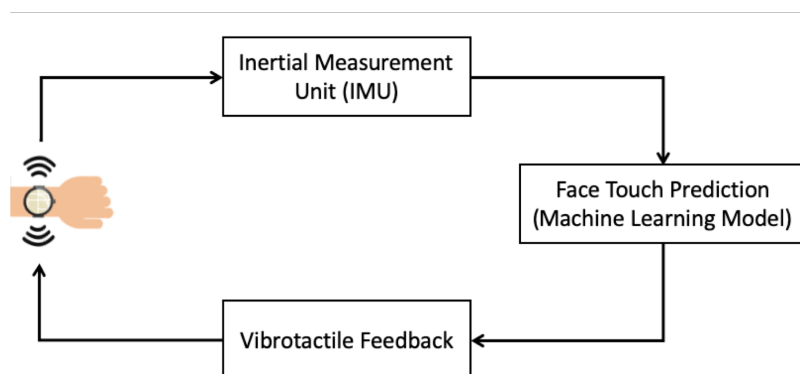158 stop the hand movement before it reaches the face.

Figure 1. The application scenario involves a smartwatch with Inertial Measurement Unit to detect hand movement, a machine learning model to predict when a movement results in touching the face, and a vibration motor to alert the user in order to stop the hand movement.

159 A more detailed description of the system is shown in Figure 2 while a technical description of the system
160 is further analyzed in Section 4. The prediction component involves a sequence of three processes, namely
161 feature selection, data segmentation, and a Convolutional Neural Network (CNN). Three sensory feedback
162 modalities are considered for the response component, namely visual, auditory, and vibrotactile. Section 5
163 presents a psychophysical experiment to compare these modalities and inform the decision about using
164 vibrotactile feedback.

165 A wearable device with an embedded IMU recording 9 different types of hand motion data (x, y, and
166 z components for accelerometer and gyroscope, and rotational pitch, roll, and yaw) makes the input
167 to the prediction component. In the feature selection process, features are extracted and evaluated for
168 relevance to predicting face touching hand movement. These features are used to improve the performance
169 of the prediction model. Feature selection included several data pre-processing procedures such as data
170 augmentation (to increase the size of training data), data filtration to enhance the signal-to-noise ratio, hand
171 orientation calculation, Fast Fourier Transform (FFT) features extraction, and optimization of the combined
172 features.

173 Once the features are identified, the time-series of the selected features are segmented according to a time
174 window. The window size is an extremely important parameter to optimize in this process since it controls
175 the tradeoff between response time and prediction accuracy. Once the time series data are segmented, all
176 the features are fed into a one dimensional convolutional neural network (1D-CNN) model. 1D-CNNs are
177 generally excellent in automatically detecting temporal relationships in multi-channel time-series data with
178 minimal feature engineering. Using the 1D-CNN kernels allows an automatic extraction of the temporal
179 features in IMU data, which is deemed important in recognizing hand movement towards the face through
180 its corresponding IMU data. The model is trained and evaluated with data generated for this purpose that is
181 recorded from 40 participants. Each participant went through a data collection session that consisted of
182 two runs. In each run, the participant had to perform 10 face-touching hand movements during each of the

following everyday activities: (standing, walking, sitting) as well as 10 non-face touching hand movements during the same activities. Thus, each participant contributed 120 trials, yielding a total of 4800 trials. The CNN model provides a binary output, whether the respective hand movement is predicted to result in face touching or not.

As soon as a prediction of face touching event is made, the response component renders a sensory feedback to alert the user, while the hand is in motion, to immediately stop the hand movement in order to avoid face touching. Based on the findings of Section 5, vibrotactile feedback is utilized as the sensory feedback modality as it provided superior performance (measured using the response time and success rate of avoiding face touching), compared to visual or auditory.
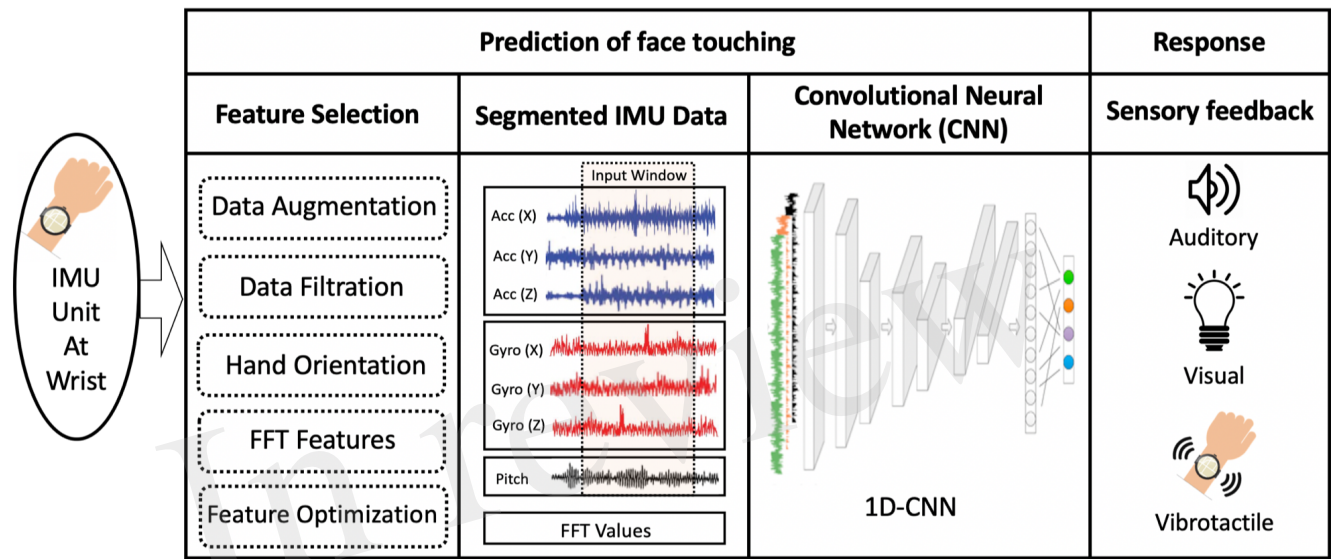

Figure 2. Overview of the FaceGuard system. Abreviations: FFT (Fast Fourier Transform), IMU (Inertial Measurement Unit), CNN (Convolutional Neural Network).

4 PREDICTION OF FACE TOUCHING

4.1 Data Collection

The data collection procedure combines computer and smartwatch interfaces to collect the needed participant data. The hardware used to collect the IMU data is an Esp32-powered, M5Stack development watch known as M5StickC. It has 6 degrees of freedom consisting of a 3-axis accelerometer and a 3-axis gyroscope, with pitch, yaw, and roll being calculated internally.

Using an Arduino IDE, the M5StickC is programmed to read the IMU data and store it into a file through a serial connection to a computer. It relies on input from two buttons: the main button used to start and stop the recording of IMU data, and the side button used for user error correction related to trial invalidation and repetition. The program runs for a predetermined number of trials per session before sending an end signal to the computer that saves all the data to a file. The participation protocol in a session consists of two runs, with 60 trials each (30 face-touching and 30 non-face touching hand movements), that are repeated twice to gather a total of 120 trials. As can be seen in Figure 3, *sitting*, *standing*, and *walking* are considered as the three main activity types due to them being the most common positions taken in our daily lives. Thus, gathering data for *touching* and *not touching the face* for each of those stances would allow the trained model to make accurate predictions regardless of the user's position.

207   The M5StickC is used in conjunction with a computer GUI application developed to ensure a holistic,
208 user-friendly collection protocol. Its main purpose is to guide users through the participation and to store
209 auxiliary user information that may be useful in optimizing the prediction model, including height, arm
210 length, and age group. Users are first asked to fill out the aforementioned optional information fields. Then,
211 the application window displays a list of the different sessions to be completed and their associated number
212 of trials, with instructions on the watch's hardware as well as the next steps. Both interfaces rely on a
213 communication of signals to control the start and end of the data collection process. The M5StickC starts
214 recording the moment the main button on the watch is pressed and stops recording when the same button is
215 pressed again by the participant. A single gesture is recorded in this fashion.

216   The data are then stored into a file. To prevent the loss of data that may occur if the serial connection is
217 interrupted, the user is provided with the option to save their data at any point during the participation upon
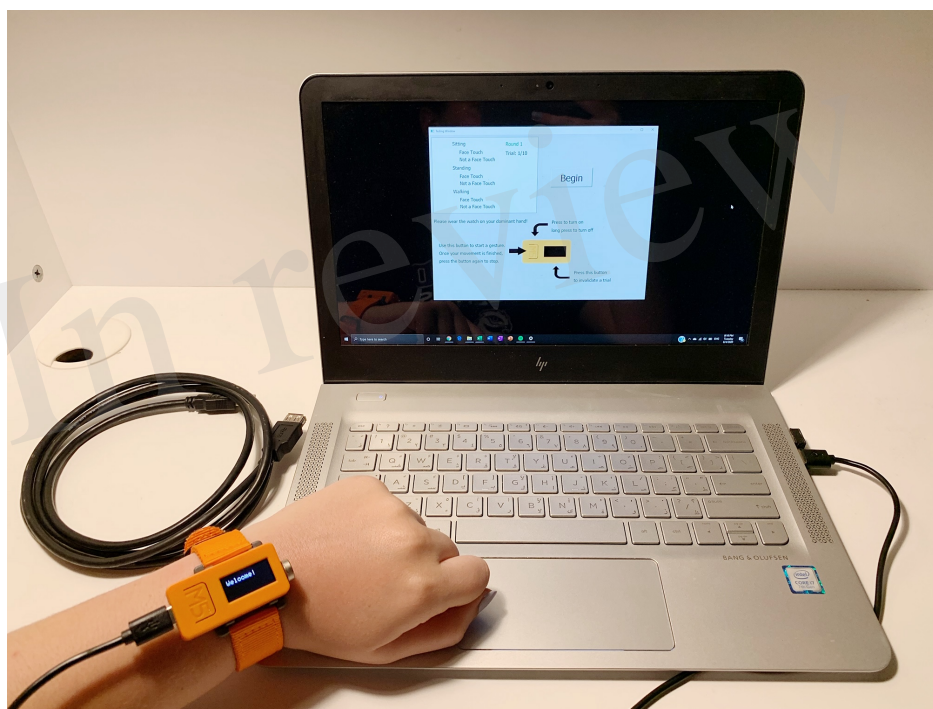218 exiting the computer application.

Figure 3. The general setup of the participation. The user is provided with a Windows laptop, the M5StickC watch, a USB-C cable, as well as a USB extension cable to be used for the walking trials.

219   The general protocol for collecting the data relies on remote participation in compliance with global
220 social-distancing and safety procedures. Users first receive a consent form and statement containing
221 information and instructions pertaining to the participation. If consent is provided, they are given the
222 watch, a compatible laptop, and all needed accessories to complete the required number of sessions in
223 their own homes, as can be seen in Figure 3. The equipment is then sanitized properly before being passed
224 on to the next participant. To ensure overall user anonymity, no identifying information is asked for or
225 stored. Additionally, the protocol is asynchronous, which provides users with the freedom to complete the
226 participation at their own pace as it is not compulsory to complete all trails and sessions in one run, rather
227 users are encouraged to take a break at any point and return later to finish.

228   Overall, 40 sessions were recorded by 40 participants collecting 4800 trials in total as elaborated in
229   Section 3. Of the information disclosed to us, 15 of the 40 participants were female, 15 were male, and 10
230   undisclosed. Additionally, most of the participants were young adults, with the most common age range
231   being 16–20 years followed by 21–25 years.

4.2   Data Preparation and Inputs

233   Once data collection is completed, data are prepared for the training and testing of the CNN-based
234   prediction model. Each gesture lasts varying amounts of time, and therefore, requires a select window
235   size to ensure prediction before a face touch has occurred. However, during data collection, users are able
236   to begin their gestures at any point once the start button has been pressed. As such, each trial recording
237   includes a static component (the duration before the hand movement starts), potentially shifting relevant
238   data outside of the determined window. A script that produces plots displaying averaged sensor values over
239   time identifies the lengths of these gestures. The script is applied to each file individually, providing plots
240   for each feature (IMU sensory data) - split into sub plots for each stance (sitting, standing, and walking).
241   The lengths of the static component of every plot at the beginning of the gesture are recorded and averaged,
242   with the resulting values to be referenced for trimming during data preparation. These plots are also used to
243   observe data trends amongst each feature. It is observed that the roll and yaw did not yield a discriminative
244   pattern for the hand touch condition and thus they are excluded from the analysis. Further confirmation is
245   obtained during the training process of the model; removing these two features improved the accuracy of
246   the model. Furthermore, it is observed that it takes around 1200 ms to complete a hand movement that
247   involves face touching, which marks the upper limit for the total response time of the proposed system
248   (prediction and motor response).

249   From the total number of gestures (4800), the training and testing data sets are formed, randomly split
250   80% to 20% (3840/960 gestures), respectively, and the two 3D input matrices are constructed. Splitting
251   was done by participants; data from a single participant exist either in the training or the test set. This is
252   to ensure the model is resilient to behavioral differences among participants. Filtration is also undergone,
253   where gestures that finish before reaching the time required for the allotted window size are removed. In
254   other words, gestures with very short duration (shorter than the window size of the 1D-CNN) are omitted
255   from the dataset.

256   One challenge for developing a robust prediction model comes from the lack of large-scale data samples
257   (40 participants with 120 trial repetition). To overcome this problem, data augmentation is introduced to
258   prevent overfitting and improve generalization of the model. Augmentation is done by creating copies of the
259   training data set and shifting it in time with 'N' number of steps while maintaining a constant window size.
260   Augmentation is a great tool for populating the training data such that they share the same characteristics
261   of the original set (representing the events of touching or not touching the face).

262   Frequency domain signature of hand movement towards the face can be obtained by taking the Fourier
263   transform of the chosen IMU signals. Frequencies of noise can be learnt and discarded once the frequency
264   domain features are obtained. The fast Fourier transform algorithm which is readily available in NumPy
265   library in python was used towards the calculation of the FFT coefficients for all the gestures. Raw and FFT
266   IMU data are then stacked to form 3D matrices, both for the training and testing data sets. Both sets are also
267   standardized, with the testing data set standardized in reference to the training data set statistics. In other
268   words, the data are transformed to have a mean of zero and a standard deviation of 1 across each feature.
269   This is done in response to differing scales between the components of the IMU, particularly between the
270   accelerometer, gyroscope, and pitch angle. The dimensions of the training and testing matrices are thus
271   $41808 \times W \times 14$, and $844 \times W \times 14$.

272   One dimensional output matrices are constructed to provide the desired output of the model, aligned
273   with each hand movement in the testing and training matrices. The output of the prediction model is set to
274   binary, designating a face touch to (1), or not a face touch to (0).

4.3   CNN-based Prediction Model Architecture

276   The input data used to train the model is arranged into a three-dimensional matrix: the first dimension
277   represents the number of trials in the dataset, the second dimension is the time length of the gesture (each
278   index represents a time step of 11 ms, in accordance to the 90.9 Hz IMU sampling rate), and the third
279   dimension is the number of features. The number of features is defined by 6 degrees of freedom from
280   the IMU (acceleration and gyroscope data), as well as the pitch angle value, and corresponding FFT
281   coefficients to form a total depth of 14 features. These data are used to train and test the model, where
282   first a convolution layer (conv1D) is applied, comprising 64 filters of kernel size 8. This is followed by
283   a rectified linear unit (ReLu) activation function applied to the previous output, a batch normalization
284   layer (BN), and a max-pooling layer with a pool size of 2. A dropout layer of value 0.8 is then applied. A
285   second convolution layer is used, consisting of 128 filters also of kernel size 8, followed by another ReLu
286   activation function. Batch normalization is utilized once more, along with a dropout layer of value 0.9,
287   after which the input at its current state is passed through a flatten layer. Finally, two fully connected layers
288   separated by a third dropout layer of value 0.8 are applied. The first fully connected layer has a dimensional
289   unit of 256, with a softmax activation function, and the second has a dimensional unit of 2, with a ReLu
290   activation function. The last fully-connected layer outputs two probabilities, one for each class (Not a face
291   touch, face touch). The architecture for the CNN-based prediction model is shown in Figure 4.
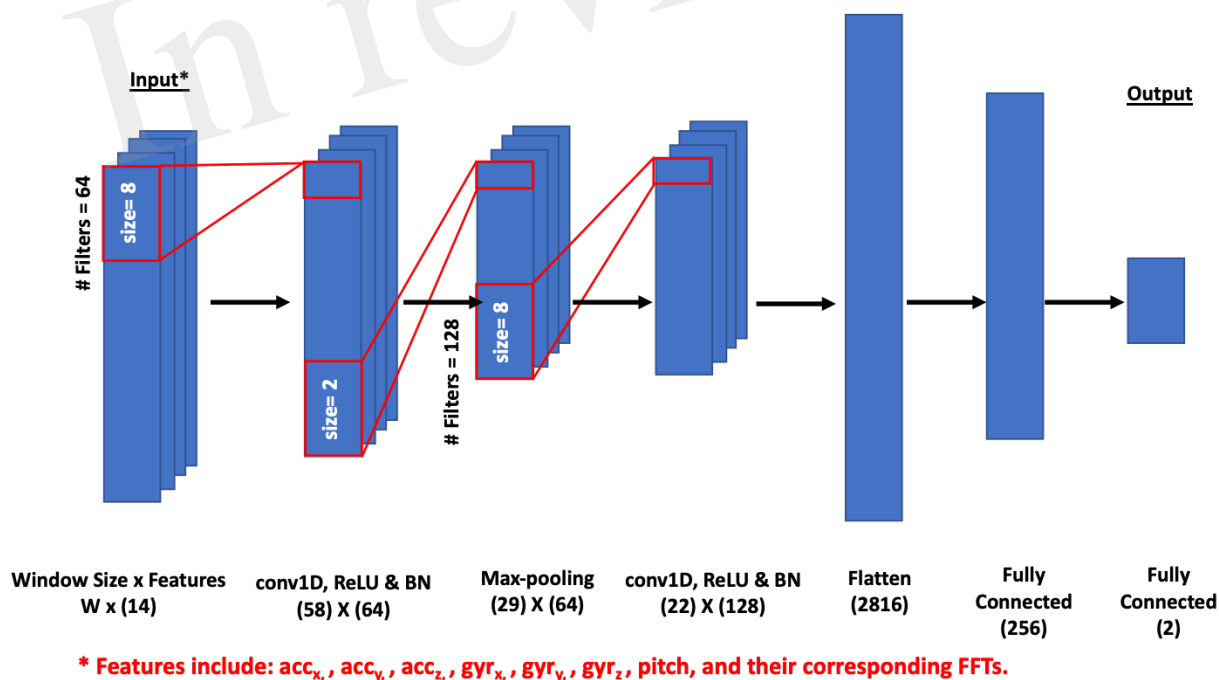

* Features include: acc_x, acc_y, acc_z, gyr_x, gyr_y, gyr_z, pitch, and their corresponding FFTs.

Figure 4. The architecture of the CNN-based prediction model. Note that W represents window size for
the input matrix.

4.4   Training and Performance Measures

293   The model shown in Figure 4 was trained using a categorical cross-entropy cost function with a default
294   learning rate of 0.001, batch size of 512, and 300 epochs. The model was optimized (weights adjustment)

using Adam optimizer Kingma and Ba (2014) during the training process. Batch normalization layers (BN) were used after each of the convolutional layers which basically re-centers and re-scales the input data leading to a faster and more stable training process. To avoid overfitting and prevent co-adaptation of the network weights, a dropout ratio (0.8–0.9) was used in the model. This high dropout ratio proved to work well with our study due to the relatively limited dataset which makes the model more prone to overfitting. The training accuracy reached 96.2% with a loss of 0.1. Table 1 shows the normalized confusion matrix of the results. The trained model has a sensitivity of 0.929 and a specificity of 0.935. This 1D-CNN model was finalized after many optimization rounds for the different hyper-parameters including the number of layers, filters and dropout ratios. An accuracy of 87.89%, 89.7%, 87.31% was obtained for a model with 3, 4, and 5 1D-convolutional layers respectively and thus, a model with 2 layers proved to be more efficient. Reducing the dropout ration to 0.5 reduces the classification accuracy to 90%. Thus, an optimized ratio of 0.8 or 0.9 was used.

Table 1. Normalized confusion matrix of the face touching/not face touching classification.

		Predicted Label	
		Not Face Touching	Face Touching
True Label	Not Face Touching	**0.97**	0.03
	Face Touching	0.11	**0.89**

4.5 Results

With a focus on prediction rather than classification, the period for data collection in real time becomes a significant parameter to select. This window size limits the collection of data from the IMU during a hand movement. Figure 5 displays the resulting prediction accuracy as this window size is varied.

A shown in Figure 5, the prediction accuracy increases as the window size increases, with 95.7% test accuracy reported at around 935 ms. As expected, increasing the window size provides the CNN-based model with further information about the hand movement and thus improves the prediction accuracy. However, increasing the window sacrifices how fast a sensory feedback is presented to the user. When fully implemented, this prediction delay will also be extended by the inference time of the model. At a window size of 700 ms, the average inference time, in which the trained model classifies a single gesture, is 0.313 ms, and at a window size of 990 ms, is 0.446 ms. These values are small enough that they become negligible to the total time delay, effectively reducing time delay before prediction to depend only on window size.

The significance of this delay depends on the application of this device. In a case where it is crucial to keep the user from touching their face, a smaller window size will reduce the time delay before a prediction is made (and before the user can be warned sooner), thereby maximizing time for reaction. This increases the probability that the user will indeed be able to stop their hand movement and avoid touching their face. As is shown in Figure 5, the consequence of this is a reduced prediction accuracy, as reducing the size of the time window reduces the amount of information about the hand movement. With urgency being prioritized, however, false-positives along with ample time to react is still more favorable. In a case where the device is meant to act as a reminder and perhaps a non-essential deterrent, such as may be the case during the COVID-19 pandemic, a larger window size may be excused to achieve higher accuracy. Therefore, finding
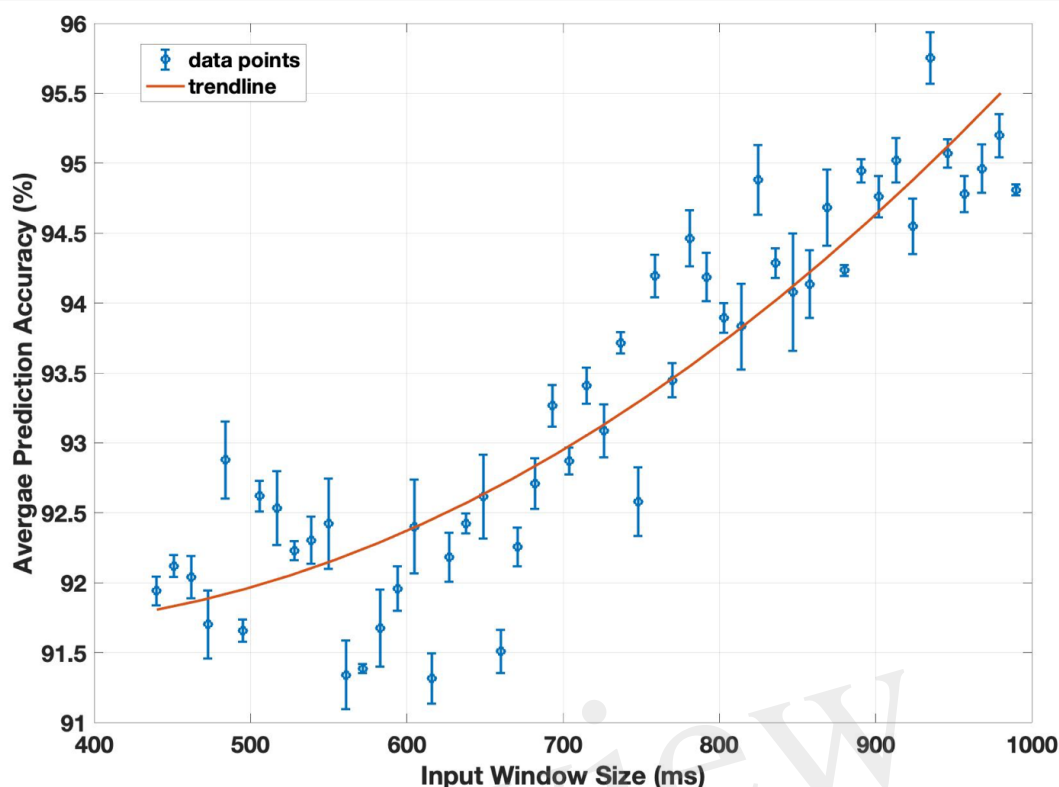
Figure 5. The prediction accuracy of the model against the input window size, averaged for the three conditions (sitting, standing, and walking). The input window size range is 440 ms to 990 ms, in increments of 11 ms, out of the 1200 ms average time needed to touch the face.

329 an optimum trade-off between response time and prediction accuracy through the window size depends
330 largely on the application.

5 SENSORY FEEDBACK FOR MOTOR CONTROL

331 A psychophysical experiment is presented to compare the effectiveness of three different sensory modalities,
332 visual, auditory, and vibrotactile, as sources of feedback to stop the hand movement. The ability of a subject
333 to stop their hand movement when confronted with sensory information is quantified by comparing the
334 response time and success rate (percentage of times the user succeeds in avoiding face touching) for the
335 three sensory modalities ($p < 0.05$). Finally, a questionnaire was introduced to the participants at the end
336 of the experiment to subjectively evaluate their quality of experience.

5.1 Participant

338 Thirty participants (15 female, 15 male, ages 25–50 years) are recruited for the experiment. None of
339 the participants have any known sensorimotor, developmental or cognitive disorders at the time of testing.
340 Written informed consent is obtained from all participants. The study is approved by the Institutional Review
341 Board for Protection of Human Subjects at New York University Abu Dhabi (Project # HRPP-2020-108).

5.2 Experimental Setup

343 A custom wristband is developed to provide the three sensory modalities, shown in Figure 6. A strip of
344 five 3 mm LED's is attached along the top face of the wristband to provide blinking visual feedback. On
345 the bottom face of the wristband a coin type vibration motor is attached to provide vibrotactile feedback
346 (Pico Vibe 310-177, Precision Microdrives vibration motor). At the middle of the top face, a 9 Degrees

of freedom (DoF) IMU is placed in order to sense any movements (displacements and rotations) when the wristband is strapped on a hand. The wristband is connected to a control box which hosts the driving circuit of the vibration motor. A 1 kHz piezoelectric buzzer is utilized to provide auditory feedback. An ATMEGA328 microcontroller unit to control and acquire data from all of the aforementioned hardware components is used. The experimental setup is connected to a laptop through a serial connection over a USB cable.
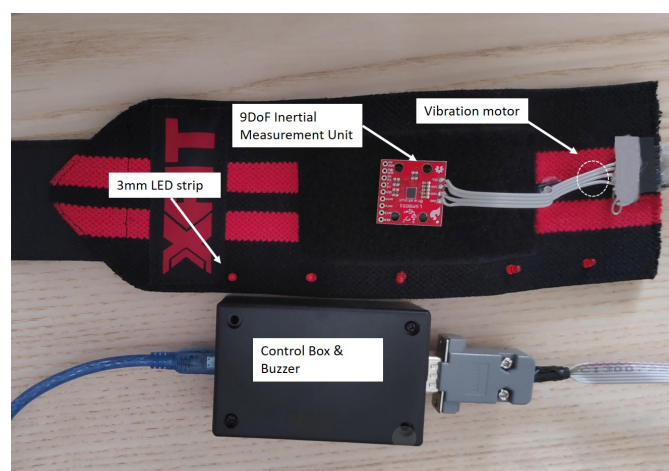

Figure 6. The wristband used in this study with all its components labeled.

Participants sit around two meters in front of the experimenter where they could make unrestricted arm movements. Participants are asked to wear the wristband at their dominant hand and keep their hand in a resting position (on the table). The experimenter instructs the participants through the experiment verbally.

5.3 Experimental Task and Protocol

In this experiment, participants complete a face touching task. Participants are instructed to move their dominant hand to touch their face, during which the hand movement is occasionally interrupted by a stimulus cue that informs the subject to stop the movement in order to avoid touching the face. Each participant completes a total of 100 trials, with 30% of these trials provide sensory feedback while the other 70% of the trials have no sensory feedback and thus result in touching the face. Among the 30% with sensory feedback, 10% are visual, 10% are auditory, and 10% are vibrotactile. To minimize the learning effects that influence the response time, the trials are presented in a counterbalanced fashion.

A trial starts with the experimenter asking the participant to rest their dominant hand on the table with tactile sensing capability to detect the start of the hand movement. The experimenter instructs the participant to move their dominant hand and touch their face. During the hand movement, the sensory cue is applied at the wristband. The hand movement is analyzed based on the recorded IMU data. At the end of the trial, the experimenter prompts the participant to confirm whether they touched their face or not. The sensory stimulus is given at a random time during the movement. The visual stimulus is a blinking red light that shines around the wristband to make it clearly visible, and lasts for 500 ms. The auditory stimulus is a beeping sound at 1000 Hz for 500 ms. The vibrotactile stimulus has a vibration frequency of 200 Hz and lasts for 500 ms. The intensity of vibration is set to be readily detected (defined as $>95\%$ correct in stimulus detection). After completing the experiment, participants fill a questionnaire in order to evaluate their subjective experience.

375   The main quantification is the response time, which indicates how rapidly a subject can respond to a
376   stimulus as a source of feedback and stop the ongoing hand movement. The response time is measured
377   as the time between the onset of the sensory feedback stimulus and the time when the hand reaches a
378   complete stop or reverses the direction of motion. The success rate – the percentage of times the participants
379   succeeds to respond timely to the sensory feedback stimulus and avoid touching their face – is also recorded.
380   The data are analyzed using repeated measures ANOVA (Analysis Of Variance) after confirming normal
381   distribution (D'Agostino-Pearson normality test).

382   It is also worth noting that the experimental protocol followed COVID-19 preventive measures in terms
383   of social distancing, symptom check for all participants, disinfection of study visit area before, and wearing
384   personal protective equipment (surgical mask and gloves).

5.4 Results

386   The average response time for vibrotactile stimulus is 427.3 ms with standard deviation of 110.88 ms.
387   The average response time for visual stimuli is 561.70 ms with standard deviation of 173.15 ms. With
388   regards to auditory stimulus, the average response time is 520.97 ms with standard deviation of 182.67 ms.
389   Response time to vibrotactile stimulus is found to be significantly shorter than that to auditory stimulus
390   (p<0.01) and visual stimulus (p<0.01). Furthermore, the response time to auditory stimulus is found to be
391   significantly shorter than that to visual stimulus (p<0.05). A summary of these findings is shown in Figure
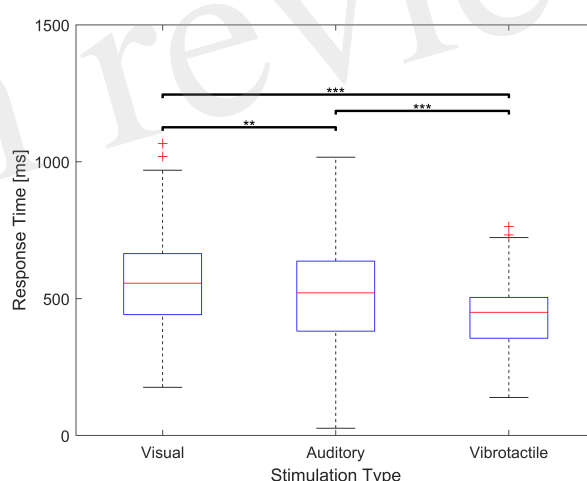392   7.

Figure 7. Response time for visual, auditory, and vibrotactile feedback. The middle red line of the blue box indicate a median value and the bottom and top edges indicate the 25th and 75th percentiles respectively (** means p<0.05, *** means p<0.01).

393   Another important performance parameter to compare is the success rate. The average success rate for
394   vibrotactile stimulus is found to be statistically larger than that of visual stimulus (p<0.05). Furthermore,
395   the average success rate for auditory stimulus is found to be statistically larger than that of visual stimulus
396   (p<0.05). However, there is no significant differences between vibrotactile stimulus and auditory stimulus
397   (p=0.07). Figure 8 shows the differences in success rate among the three groups.

398   The questionnaire is designed to capture the participant's quality of experience. Participants are asked
399   about their favorite modality for feedback, which modality provides the most pleasant experience, whether
400   vibrotactile feedback creates any fatigue or discomfort, and the chance to provide any further feedback.
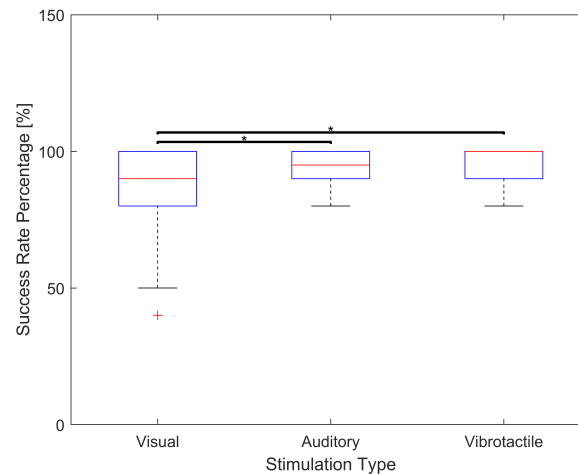
Figure 8. Success rate associated with visual, auditory, and vibrotactile feedback. The middle red line of the blue box indicate a median value and the bottom and top edges indicate the 25th and 75th percentiles respectively (* means *p*<0.05).

As for preference, 25 participants (83.34%) selected vibrotactile as their favorite modality for feedback, 3 (10%) selected auditory feedback, and 2 (6.67%) preferred visual. 29 participants (96.67%) reported that they clearly perceived the vibrotactile stimulation. 28 participants (93.34%) selected vibrotactile feedback as the most pleasant among the three modalities. Finally, none of the participants reported significant fatigue or discomfort during the experiment.

6 DISCUSSION

The CNN-based prediction model requires less than 550 ms of IMU data to predict face touching events with an accuracy greater than 92%. Furthermore, the sensory feedback experiment showed that around 427 ms is needed to stop the hand movement using vibrotactile feedback. Therefore, it will take less than a second from the start of the hand movement until complete stop. Meanwhile, our study suggests that the average time for a hand to reach and touch the face is 1200 ms. Therefore, the proposed system is capable of providing timely response to avoid face touching within less than one second. It is worth noting that there is a tradeoff between the prediction accuracy and the response time. In order to improve the prediction accuracy, the input window size must increase, which implies that it will take more time to stop the hand movement, which causes a decrease in chances to avoid face touching.

Another important factor is the relationship between prediction accuracy and practical usefulness of the system: an increase in the number of false positives would create unnecessary buzzing which may distract/annoy the user while an increase in false negatives would not prevent face touching entirely. Therefore, while the current prediction system is based solely on the IMU data, fusing other sensory modalities into the CNN-based model that are relevant to face touching and hand movement would significantly improve the prediction accuracy. For instance, gender, arm length, hand size, and age group may provide complementary information to improving the prediction accuracy. This involves recruiting a significantly larger number of participants to generate enough data points to train the model. In situations where camera data are available, such as when the user is sitting in front of a PC, computer vision approaches can be applied in order to fine-tune the model for improved performance.

425 A major source for false positives stems from the lack of information about the head posture in reference
426 to the hand movement. Therefore, it would be interesting to augment the current CNN-based prediction
427 model with the head position and/or orientation. With appropriate sensors or camera systems, the head
428 posture can be continuously monitored and used as an auxiliary input to the prediction model to further
429 improve the prediction accuracy. This is an interesting direction for future work. Furthermore, collecting
430 hand movements that are likely to cause false positives (such as eating where the hand movement is very
431 similar to that of face touching) and training the model with such data would significantly reduce the false
432 positives.

433 Although the findings of the present study demonstrate the feasibility of developing a system to avoid face
434 touching, a few limitations should be mentioned. First, the dataset utilized to train the CNN-based prediction
435 model is rather limited. A larger dataset improves the prediction accuracy, including false positives and
436 negatives, which allows for a reduced window size and improved system response. Furthermore, running
437 the CNN model is computationally expensive. Therefore, the inference about the prediction of face touching
438 may have to be performed on a computationally powerful machine such as a smart phone or even the
439 cloud. This adds further delays to the overall system response. Additionally, the current study focused on
440 preventing face touching through the dominant hand. It might be desirable for several applications to avoid
441 touching the face with both hands, and thus evaluating the performance of the system while tracking both
442 hands is necessary for such applications.

443 Another very important limitation is that the collected IMU data were pre-segmented such that each
444 trial is known to have a single hand gesture. As the IMU signals are continuous streaming data, a sliding
445 window must be used to segment the raw data to individual pieces in real time, each of which is the input
446 of the CNN model. The length and moving step of the sliding window are hyper-parameters that need to be
447 carefully tuned to achieve satisfactory performance. This problem is present not only in tasks that require
448 constant gesture recognition Liu et al. (2017), but also in other fields such as continuous speech recognition
449 Palaz et al. (2015). Finally, the participants' behaviour or activities could modulate the hand movement and
450 thus may impact the accuracy of the prediction model. More data must be collected while participants are
451 engaged in various activities/behavior in order to enhance the resilience of the classifications against users'
452 activities/behaviour.

7 CONCLUSION

453 This paper presented a system that utilizes IMU data to predict hand movement that results in face touching
454 and provide sensory feedback to stop the hand movement before touching the face. A 1D-CNN-based
455 prediction model, capable of automatically extracting temporal features of the IMU data through 1D-CNN
456 filters, was developed and trained with IMU data collected from 4800 trials recorded from 40 participants.
457 Results demonstrated a prediction accuracy of more than 92% with less than 550 ms of IMU time series data.
458 Compared to visual and auditory modalities, it was found that vibrotactile feedback results in statistically
459 faster response, better success rate, and improved quality of user experience.

460 As for future work, it is of an importance to evaluate the combined prediction/response system as a
461 whole in a realistic experimental environment (while performing everyday life activities). Furthermore, the
462 authors plan to develop a light-weight CNN-based prediction model that optimizes computational power in
463 order to run the prediction model on a wearable device (with limited computational power). Improving the
464 dataset by collecting more data can immensely improve the model training and performance. In particular,
465 collecting data from tasks that exhibit similar hand movements to face touching but do not involve face
466 touching (such as eating) will improve the system robustness, particularly against false positives.

CONFLICT OF INTEREST STATEMENT

467 The authors declare that the research was conducted in the absence of any commercial or financial
468 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

469 AMM, SB, HA, and HA developed the prediction component, including the data collection protocol and the
470 CNN model development and evaluation, GK, AA, and ME developed the sensory feedback psychophysical
471 study. RS, AA, and ME contributed to the conceptualization of the solution.

FUNDING

472 This project is supported by New York University Abu Dhabi.

DATA AVAILABILITY STATEMENT

473 The dataset used to train the CNN-based prediction model can be found in the Open Science Framework
474 (OSF) platform using this link: Dataset.

REFERENCES

475 Al-Shamayleh, A. S., Ahmad, R., Abushariah, M. A., Alam, K. A., and Jomhari, N. (2018). A systematic
476   literature review on vision based gesture recognition techniques. *Multimedia Tools and Applications* 77,
477   28121–28184
478 Arrabito, G. R., Mondor, T. A., and Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms:
479   A case study. *Ergonomics* 47, 821–840
480 Barroso, F., Freedman, N., and Grand, S. (1980). Self-touching, performance, and attentional processes.
481   *Perceptual and Motor Skills* 50, 1083–1089
482 Bate, K. S., Malouff, J. M., Thorsteinsson, E. T., and Bhullar, N. (2011). The efficacy of habit reversal
483   therapy for tics, habit disorders, and stuttering: a meta-analytic review. *Clinical Psychology Review* 31,
484   865–871
485 Calhoun, G., Draper, M., Ruff, H., Fontejon, J., and Guilfoos, B. (2003). Evaluation of tactile alerts for
486   control station operation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*
487   (SAGE Publications Sage CA: Los Angeles, CA), vol. 47, 2118–2122
488 Chen, Z., Huang, W., and Lv, Z. (2017). Towards a face recognition method based on uncorrelated
489   discriminant sparse preserving projection. *Multimedia Tools and Applications* 76, 17669–17683
490 Christian, M., Uyanik, C., Erdemir, E., Kaplanoglu, E., Bhattacharya, S., Bailey, R., et al. (2019).
491   Application of deep learning to imu sensor motion. In *2019 SoutheastCon* (IEEE), 1–6
492 Christofferson, K. and Yang, R. (2020). *FaceSpace - Apple Watch Application*. Tech. rep., DEVOPS
493 Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., et al. (2020). Physical
494   distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and
495   covid-19: a systematic review and meta-analysis. *The Lancet*
496 Corrales, J. A., Candelas, F., and Torres, F. (2008). Hybrid tracking of human operators using imu/uwb data
497   fusion by a kalman filter. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction*
498   *(HRI)* (IEEE), 193–200
499 D'Aurizio, N., Baldi, T. L., Paolocci, G., and Prattichizzo, D. (2020). Preventing undesired face-touches
500   with wearable devices and haptic feedback. *IEEE Access* 8, 139033–139043
501 Fu, A. and Yu, Y. (2017). Real-time gesture pattern classification with imu data

502 Godlove, J. M., Whaite, E. O., and Batista, A. P. (2014). Comparing temporal aspects of visual, tactile, and
503 microstimulation feedback for motor control. *Journal of neural engineering* 11, 046025

504 Hakkert, A. S., Gitelman, V., and Ben-Shabat, E. (2002). An evaluation of crosswalk warning systems:
505 effects on pedestrian and vehicle behaviour. *Transportation Research Part F: Traffic Psychology and*
506 *Behaviour* 5, 275–292

507 Harrigan, J. A. (1985). Self-touching as an indicator of underlying affect and language processes. *Social*
508 *science & medicine* 20, 1161–1168

509 Huang, J., Huang, Z., and Chen, K. (2017). Combining low-cost inertial measurement unit (imu) and deep
510 learning algorithm for predicting vehicle attitude. In *2017 IEEE Conference on Dependable and Secure*
511 *Computing* (IEEE), 237–239

512 [Dataset] Immutouch (2020). Immutouch wearable device. https://immutouch.com/. Accessed: 2020-09-03

513 Jiang, S., Lv, B., Guo, W., Zhang, C., Wang, H., Sheng, X., et al. (2017a). Feasibility of wrist-worn,
514 real-time hand, and surface gesture recognition via semg and imu sensing. *IEEE Transactions on*
515 *Industrial Informatics* 14, 3376–3385

516 Jiang, X., Merhi, L.-K., and Menon, C. (2017b). Force exertion affects grasp classification using force
517 myography. *IEEE Transactions on Human-Machine Systems* 48, 219–226

518 Karia, R. M., Ghuntla, T. P., Mehta, H. B., Gokhale, P. A., and Shah, C. J. (2012). Effect of gender
519 difference on visual reaction time: A study on medical students of bhavnagar region. *IOSR Journal of*
520 *Pharmacy* 2, 452–454

521 Kemp, B. J. (1973). Reaction time of young and elderly subjects in relation to perceptual deprivation and
522 signal-on versus signal-off conditions. *Developmental Psychology* 8, 268

523 Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint*
524 *arXiv:1412.6980*

525 Kwok, Y. L. A., Gralton, J., and McLaws, M.-L. (2015). Face touching: a frequent habit that has
526 implications for hand hygiene. *American journal of infection control* 43, 112–114

527 Liu, Y., Li, Z., Liu, Z., and Wu, K. (2019). Real-time arm skeleton tracking and gesture inference tolerant
528 to missing wearable sensors. In *Proceedings of the 17th Annual International Conference on Mobile*
529 *Systems, Applications, and Services*. 287–299

530 Liu, Z., Chai, X., Liu, Z., and Chen, X. (2017). Continuous gesture recognition with hand-oriented
531 spatiotemporal feature. In *Proceedings of the IEEE International Conference on Computer Vision*
532 *Workshops*. 3056–3064

533 Lv, Z. (2020). Robust3d: a robust 3d face reconstruction application. *Neural computing and applications*
534 32, 8893–8900

535 Macchini, M., Havy, T., Weber, A., Schiano, F., and Floreano, D. (2020). Hand-worn haptic interface for
536 drone teleoperation. *arXiv preprint arXiv:2004.07111*

537 Macias, A., De la Torre, A., Moreno-Espinosa, S., Leal, P., Bourlon, M., and Ruiz-Palacios, G. (2009).
538 Controlling the novel a (h1n1) influenza virus: don't touch your face! *Journal of Hospital Infection* 73,
539 280–281

540 Manghisi, V. M., Fiorentino, M., Boccaccio, A., Gattullo, M., Cascella, G. L., Toschi, N., et al. (2020). A
541 body tracking-based low-cost solution for monitoring workers' hygiene best practices during pandemics.
542 *Sensors* 20, 6149

543 McMonnies, C. W. (2008). Management of chronic habits of abnormal eye rubbing. *Contact Lens and*
544 *Anterior Eye* 31, 95–102

545 Mueller, S. M., Martin, S., and Grunwald, M. (2019). Self-touch: Contact durations and point of touch
546 of spontaneous facial self-touches differ depending on cognitive and emotional load. *PloS one* 14,

547   e0213677

548   Mummadi, C. K., Leo, F. P. P., Verma, K. D., Kasireddy, S., Scholl, P. M., Kempfle, J., et al. (2018). Real-
549   time and embedded detection of hand gestures with an imu-based glove. In *Informatics* (Multidisciplinary
550   Digital Publishing Institute), vol. 5, 28

551   Ng, A. W. and Chan, A. H. (2012). Finger response times to visual, auditory and tactile modality stimuli. In
552   *Proceedings of the international multiconference of engineers and computer scientists*. vol. 2, 1449–1454

553   Olivares, A., Olivares, G., Mula, F., Górriz, J. M., and Ramírez, J. (2011). Wagyromag: Wireless sensor
554   network for monitoring and processing human body movement in healthcare applications. *Journal of
555   systems architecture* 57, 905–915

556   Palaz, D., Doss, M. M., and Collobert, R. (2015). Convolutional neural networks-based continuous speech
557   recognition using raw speech signal. In *2015 IEEE International Conference on Acoustics, Speech and
558   Signal Processing (ICASSP)* (IEEE), 4295–4299

559   Pawar, T., Chaudhuri, S., and Duttagupta, S. P. (2007). Body movement activity recognition for ambulatory
560   cardiac monitoring. *IEEE transactions on biomedical engineering* 54, 874–882

561   Pisharady, P. K. and Saerbeck, M. (2015). Recent methods and databases in vision-based hand gesture
562   recognition: A review. *Computer Vision and Image Understanding* 141, 152–165

563   Rivera, P., Valarezo, E., and Kim, T.-S. (2017). Recognition of human hand activities based on a single
564   wrist imu using recurrent neural networks. *International Journal of Pharma Medicine and Biological
565   Sciences* 6, 114–118

566   Sanders, A. (1975). The foreperiod effect revisited. *The Quarterly Journal of Experimental Psychology* 27,
567   591–598

568   Scott, J. and Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision
569   prevention in simulated driving. *Human factors* 50, 264–275

570   Siddiqui, N. and Chan, R. H. (2020). Multimodal hand gesture recognition using single imu and acoustic
571   measurements at wrist. *Plos one* 15, e0227039

572   Silva do Monte Lima, J. P., Uchiyama, H., and Taniguchi, R.-i. (2019). End-to-end learning framework for
573   imu-based 6-dof odometry. *Sensors* 19, 3777

574   Solomon, S. S. and Hill, P. F. (2002). *Emergency vehicle accidents: Prevention, reconstruction, and survey
575   of state law* (Lawyers & Judges Publishing)

576   Thompson, P., Colebatch, J., Brown, P., Rothwell, J., Day, B., Obeso, J., et al. (1992). Voluntary stimulus-
577   sensitive jerks and jumps mimicking myoclonus or pathological startle syndromes. *Movement disorders:
578   official journal of the Movement Disorder Society* 7, 257–262

579   Ward, J. A., Lukowicz, P., Troster, G., and Starner, T. E. (2006). Activity recognition of assembly tasks
580   using body-worn microphones and accelerometers. *IEEE transactions on pattern analysis and machine
581   intelligence* 28, 1553–1567

582   [Dataset] WHO (2020). World health organizationn. https://covid19.who.int/. Accessed: 2020-09-06

583   Yang, C. and Lv, Z. (2020). Gender based face aging with cycle-consistent adversarial networks. *Image
584   and Vision Computing* , 103945

585   Zhao, J.-L., Wu, Z.-K., Pan, Z.-K., Duan, F.-Q., Li, J.-H., Lv, Z.-H., et al. (2018). 3d face similarity
586   measure by fréchet distances of geodesics. *Journal of Computer Science and Technology* 33, 207–222
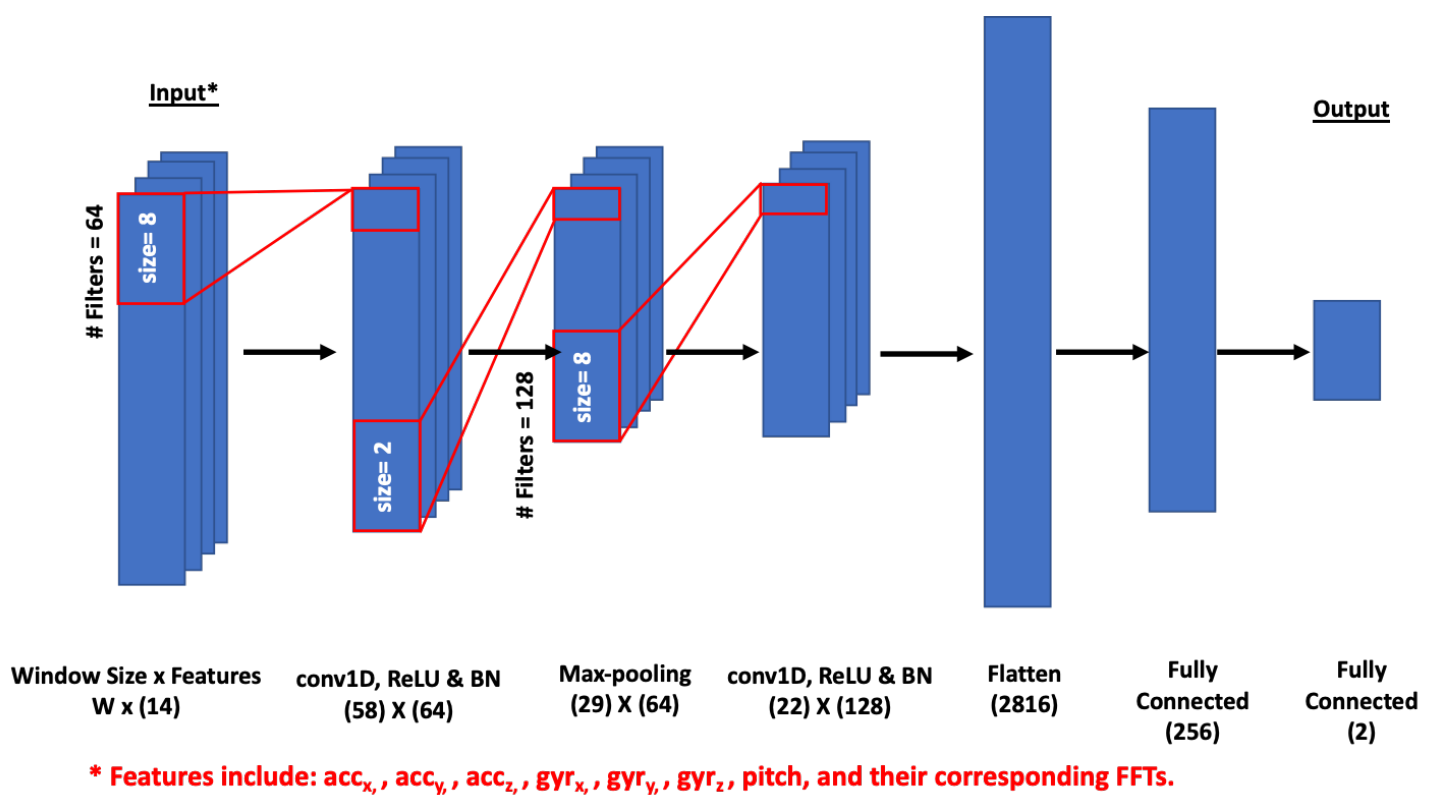
Figure 1.JPEG

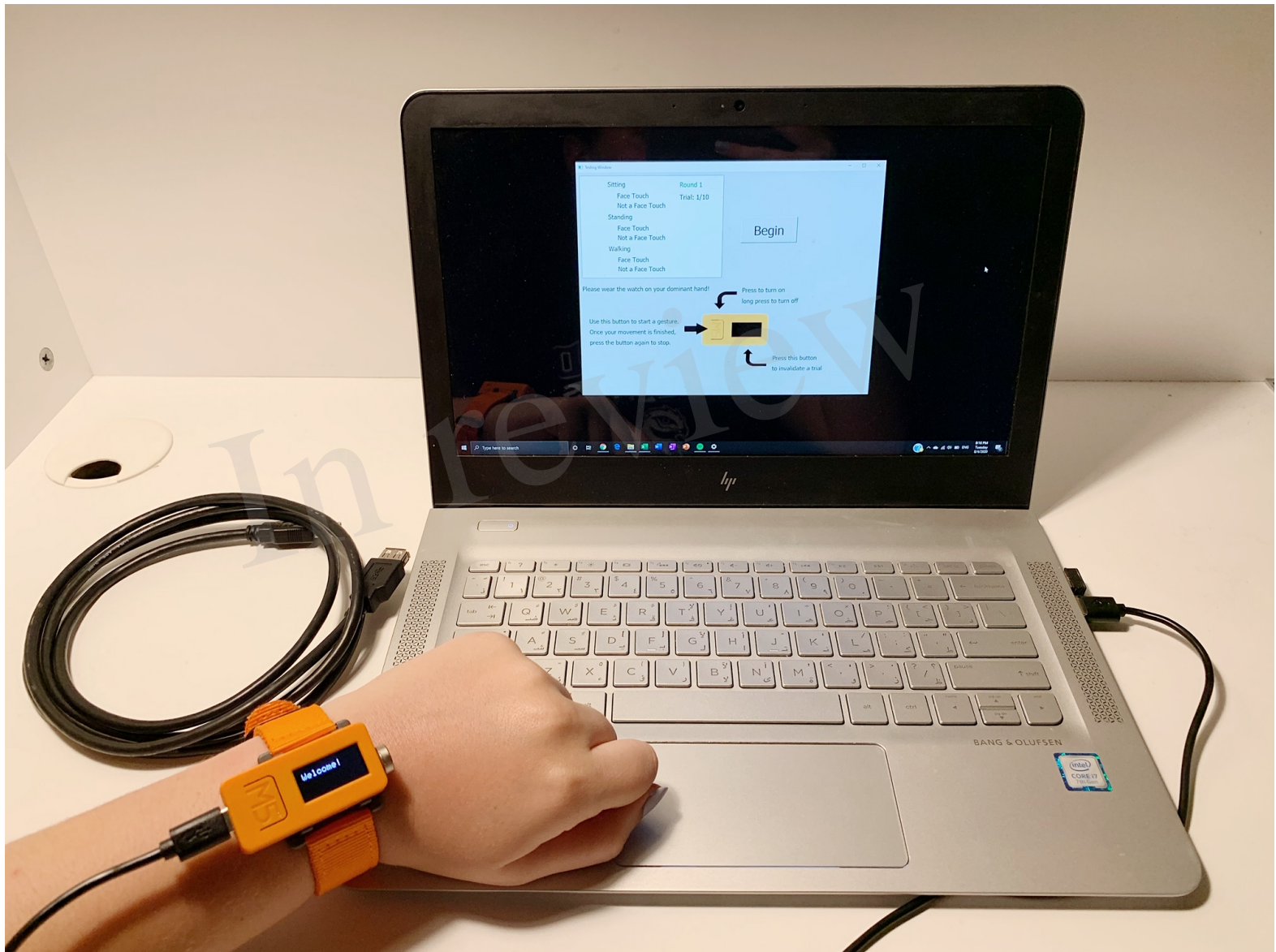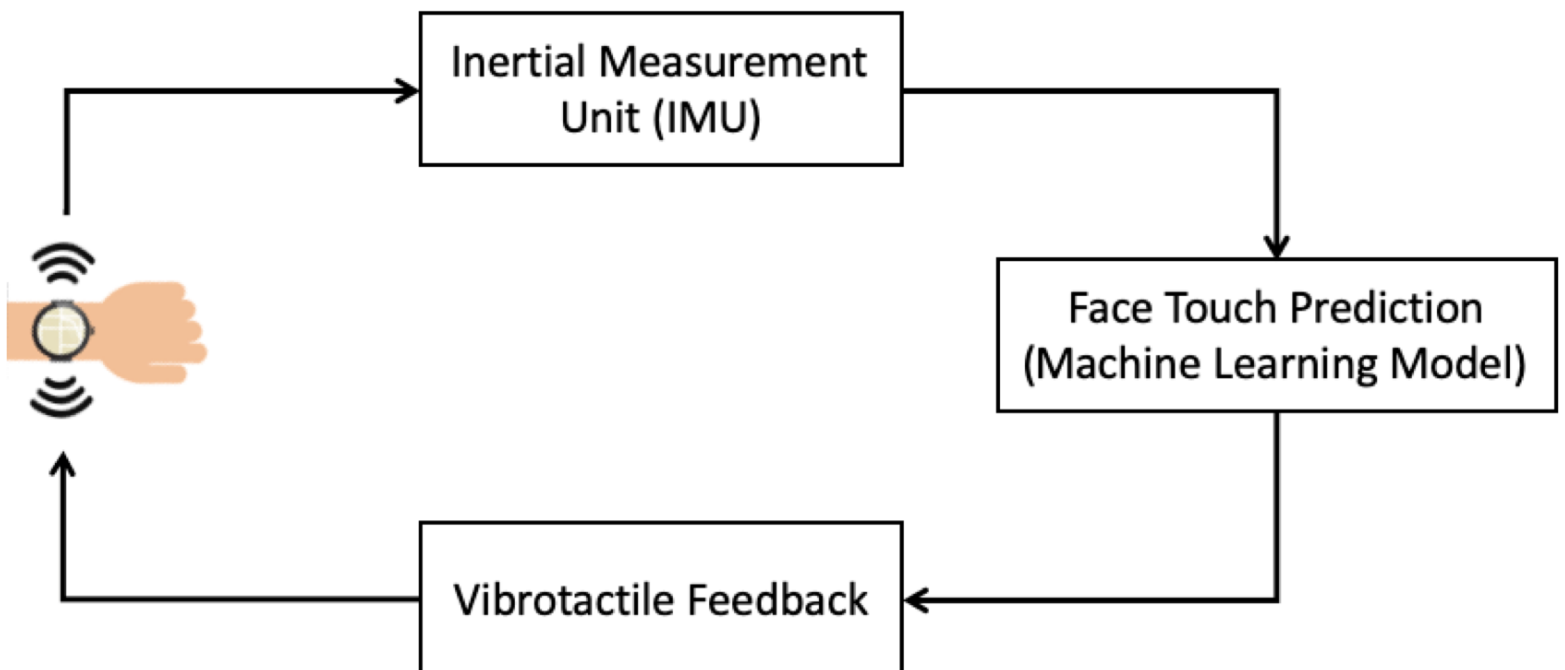Input*

Filters = 64

size= 8

size= 2

Filters = 128

size= 8

Output

Window Size x Features
W x (14)

conv1D, ReLU & BN
(58) X (64)

Max-pooling
(29) X (64)

conv1D, ReLU & BN
(22) X (128)

Flatten
(2816)

Fully
Connected
(256)

Fully
Connected
(2)

* Features include: $acc_{x,}$, $acc_{y,}$, $acc_{z,}$, $gyr_{x,}$, $gyr_{y,}$, gyr_{z}, pitch, and their corresponding FFTs.
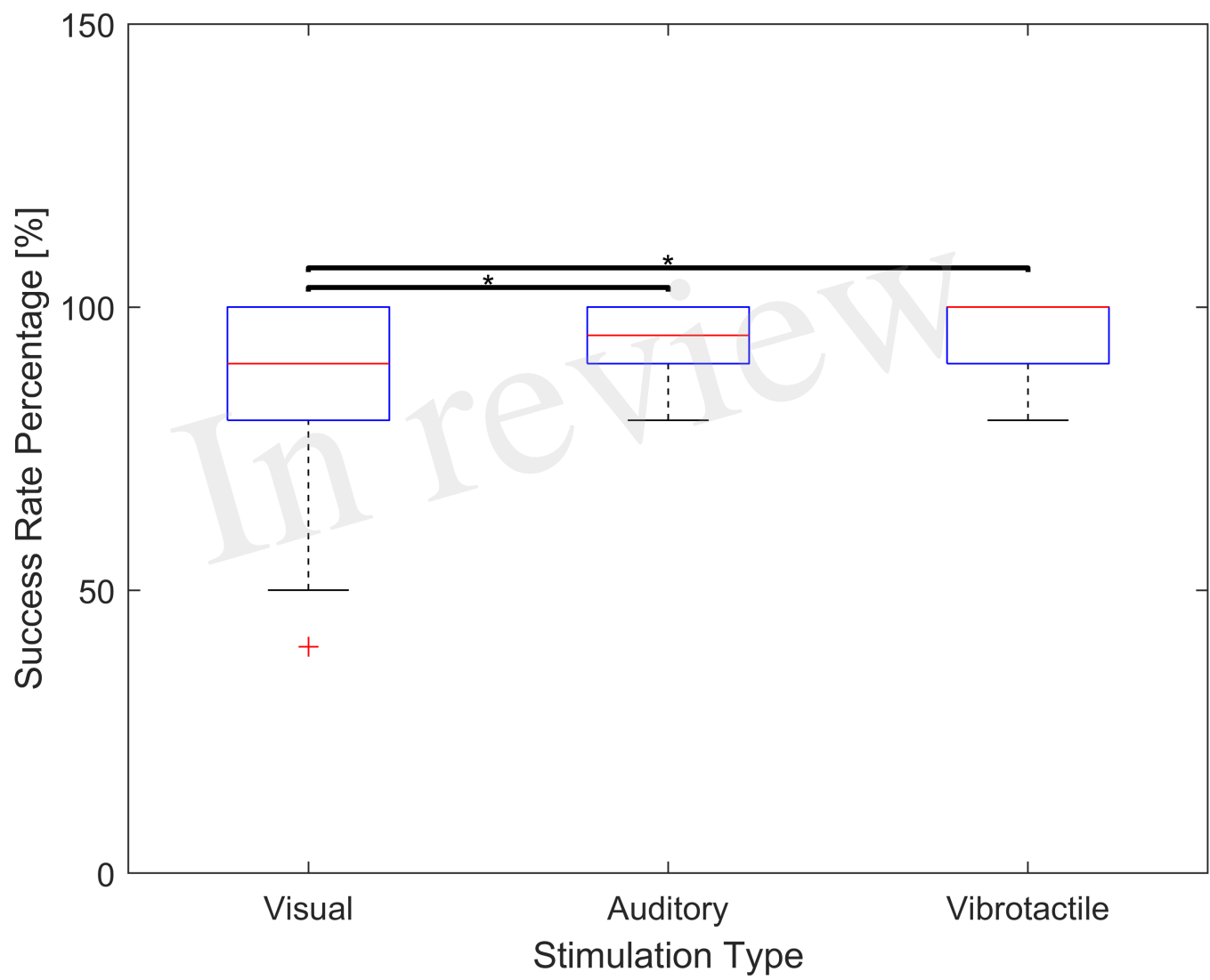
Figure 2.JPEG
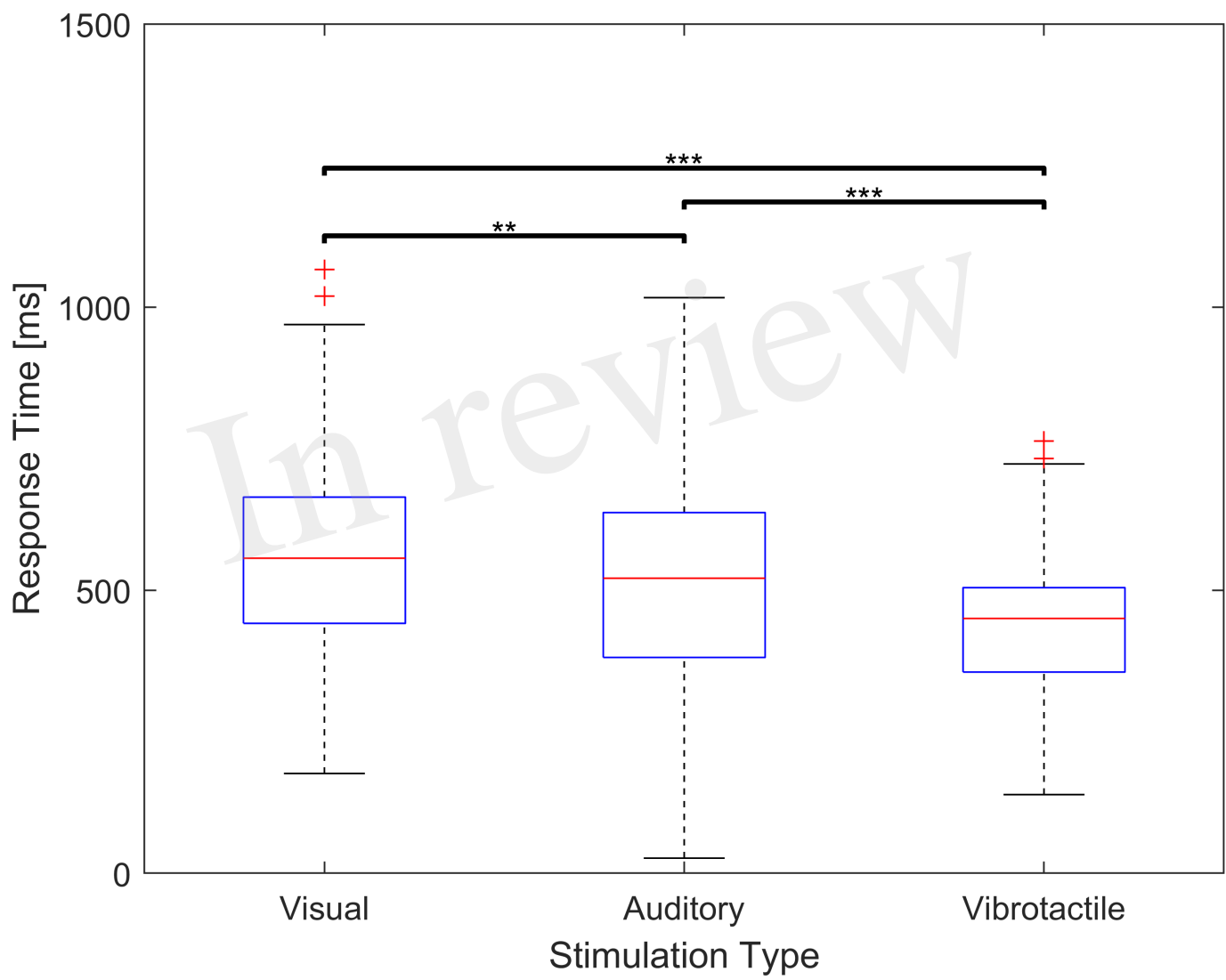
Figure 3.JPEG

Figure 4.JPEG
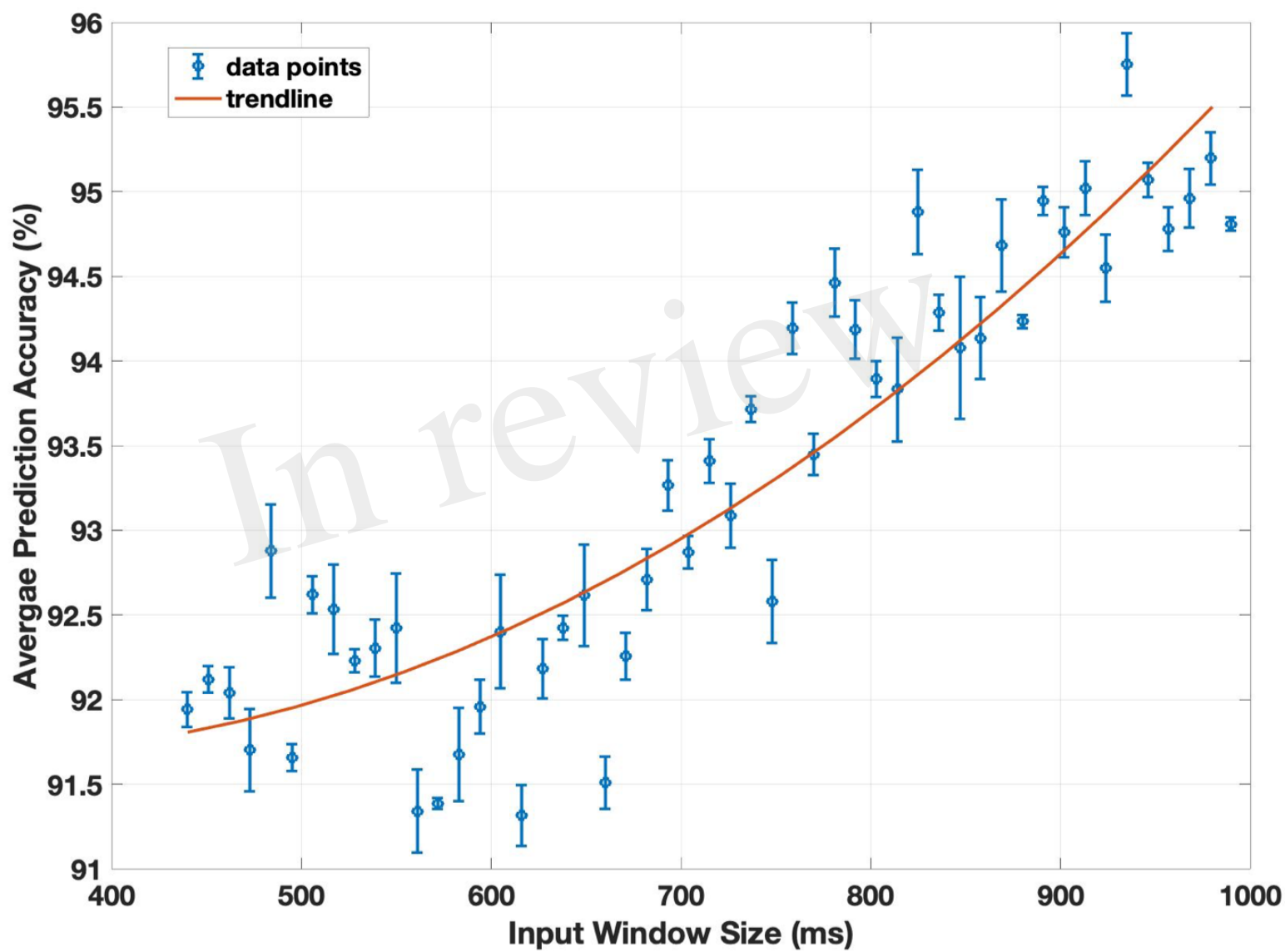
Figure 5.JPEG

Figure 6.JPEG

Figure 7.JPEG

Figure 8.JPEG

9DoF Inertial Measurement Unit

Vibration motor

3mm LED strip

Control Box & Buzzer