Usable User Authentication on a Smartwatch using Vibration

Sunwoo Lee
Korea University
Seoul, Republic of Korea
lswoo92@gmail.com

Wonsuk Choi*
Hansung University
Seoul, Republic of Korea
wonsuk@hansung.ac.kr

Dong Hoon Lee*
Korea University
Seoul, Republic of Korea
donghlee@korea.ac.kr

ABSTRACT

Smartwatches have come into wide use in recent years, and a number of smartwatch applications that improve convenience and user health are being developed and introduced constantly. Moreover, the latest smartwatches are now designed to operate without their paired smartphones, and as such, it is necessary for smartwatches to independently authenticate users. In these current devices, personal identification numbers (PIN) or patterns are entered to authenticate users, but these methods require inconvenient interaction for the user and are not highly secure. Particularly relevant to smartwatch technology, even user authentication based on biometric information needs either special sensors capable of measuring biometric information or user interaction. In this paper, we propose a usable method for user authentication on smartwatches without additional devices. Based on the fact that vibration is absorbed, reflected, and propagated differently according to the physical structure of each human body, our method is designed as a challenge-response scheme, in which the challenge is a random sequence of multiple vibration types that are already built into current smartwatches. The responses to vibrations are measured by the default gyroscope and accelerometer sensors in smartwatches. Moreover, our method is the first working model for commercial smartwatch models with low specifications when vibrating and measuring responses. We evaluated our method using a commercial smartwatch, and the results show that our method is able to authenticate a user with an equal error rate (EER) of 1.37%.

CCS CONCEPTS

• **Security and privacy → Biometrics**.

KEYWORDS

Biometrics; Authentication; Smartwatches; Signal Processing

ACM Reference Format:
Sunwoo Lee, Wonsuk Choi, and Dong Hoon Lee. 2021. Usable User Authentication on a Smartwatch using Vibration. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21), November 15–19, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3460120.3484553

*Co-corresponding authors

CCS '21, November 15–19, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8454-4/21/11…$15.00
https://doi.org/10.1145/3460120.3484553

1 INTRODUCTION

According to recent research, worldwide smartwatch shipments in the first quarter of 2020 increased by 20% to 14 million units annually [27]. The area of smartwatch usage is also expanding continuously, with benefits like payments and health services being the most practical and popular functions facilitated by a smartwatch. Moreover, the latest smartwatches are even able to make phone calls without a paired smartphone, which implies that smartwatches are no longer auxiliary accessories. Because of this standalone mode, there is a growing need for an authentication method that can independently verify users. Presently, such methods available on smartwatches require a user to enter a password or replicate a saved pattern. However, in addition to these traditional methods requiring user interaction and being inconvenient, it is also known that password and pattern-based authentication methods can be vulnerable to shoulder-surfing or brute-force attacks [57, 67].

Researchers have studied biometrics-based methods to provide security and usability for independent user authentication on smartwatches. Among these are methods, many utilize a fingerprint [23], face [55], vein [24], Electrocardiogram (ECG) data [9, 21], iris [11], voice [57], or gesture [36, 37, 45, 53, 56]. However, most of these methods cannot function properly on current smartwatches because they require a special sensor to measure the biometric information. Although the measurement sensors for ECG, voice, or gesture biometrics-based methods are already built into smartwatches by default, the corresponding user authentication procedures are inconvenient because they require direct user interaction when measuring the biometric information. For example, users must place their finger on a sensor to measure ECG signals (i.e., low usability). Furthermore, these biometrics-based authentication methods are *static* and use the same biometric information whenever authenticating users. Accordingly, with the *static* authentication method, it is highly plausible for malicious users to create fake biometric information (e.g., 3D-printed faces or fingerprints) and bypass authentication.

For these reasons, a new type of biometrics-based authentication that enables a challenge-response structure is necessary for smartwatches. In a challenge-response structure, a verifier presents a challenge to a prover and the prover provides a valid response without revealing the secret used to generate the response. In the biometrics-based method, the challenge is in the form of unpredictable stimuli and the response is a bodily reaction to the stimuli. In this paper, we propose a usable user authentication for smartwatches that is designed as a challenge-response structure. Our method measures and analyzes the response to random vibrations from a smartwatch, based on the fact that vibrations are absorbed, reflected, and propagated differently according to the physical structure of each human body [39, 48, 59]. To the best of our knowledge, our method is the first to analyze measurements from default sensors in a smartwatch with a low sampling frequency. Also, vibrations are generated by the default sensor of the smartwatch, which has a low vibration

frequency. Although existing methods have studied user authentication by analyzing the response to vibrations, these methods were not designed for the low specifications of smartwatches [50, 51, 62], and they relied on high-frequency vibrations that cannot be generated on smartwatches. It is noted that high-frequency sensors are needed to measure high-frequency vibrations, but these are not installed by default in current smartwatches, whose sampling frequency is limited to a maximum of 100Hz.

Our detailed contributions are as follows:

- To the best of our knowledge, our method is the first approach to analyze the response to vibrations from a smartwatch, which is sampled by the default sensor of a smartwatch. Unlike existing methods, our method authenticates smartwatch users by generating low-frequency vibrations that are measured at a low-sampling frequency that is available on current smartwatches.
- Our method independently authenticates smartwatch users without the help of the paired smartphone.
- Our method does not require any additional devices (i.e., using built-in sensors on smartwatches), which implies that it can be easily adapted for smartwatches already on the market.
- Our method provides a usable authentication method as the user does not have to interact directly with the authentication procedure. When vibrations are generated by the smartwatch, there is no action that the user must take to complete authentication. Accordingly, our method can be expanded to provide continuous authentication.

2 RELATED WORKS

In this section, we explain biometrics-based authentication methods that can be applied to smartwatches. In addition, we compare existing methods that analyze the response to vibrations for user authentication and our method.

2.1 Biometrics-based Authentication

The information required to authenticate a person's identity is divided into three parts: knowledge (something only the person knows), possession (something only the person has), and inherence (something only the person is). Since biometric information is something only the person is, it falls into inherence parameters. Therefore, unlike traditional user authentication methods using passwords, patterns, or tokens, the biometrics-based methods maximize convenience as there is no risk of forgetting or losing the information required for authentication. Because of the convenience, biometrics-based authentication methods have been studied extensively and applied in various fields [33, 34, 44, 60].

Biometrics-based methods seem to be usable and convenient for smartwatch authentication. Owing to this, smartwatch manufacturers are considering incorporating available biometric information—like a fingerprint, face, vein, or iris—because the measuring sensors can be built into smartwatches [11, 23, 24, 55]. However, this implies that the price of smartwatches would increase due to the additional sensor. Moreover, *static* authentication is vulnerable to fake biometrics, such as a 3D-printed fingerprint or a wax hand showing veins [1, 3, 6, 7, 15, 16, 18, 25, 29]. Therefore, *static* authentication becomes useless once the biometric information is compromised.

In terms of ECG, voice, and user gesture-based methods, current smartwatches are able to measure biometrics using a default sensor [9, 21, 36, 37, 45, 56, 57]. As a result, user authentication can be accomplished without any additional sensors, but these methods still fail to provide high usability during the authentication procedure. For example, the user must speak to complete voice-based authentication on smartwatches [57]. In ECG-based authentication, which analyzes ECG signals for authentication, the user must place a finger on the sensor for the smartwatch to measure the ECG signals [9, 21], and more alarmingly, ECG-based authentication has also been proven vulnerable [40]. Eberz et al. demonstrated this with a mapping function that converted the ECG signal measured by any device to the morphology of the ECG signal measured by the Nymi Band, and this function was used to perform a simulation attack. Similarly, voice-based authentication methods can be bypassed using text-to-speech with machine learning technology [4].

A particular gesture can be also used as biometric information for authentication. Buriro et al. proposed Airsign, an existing smartwatch authentication method, which analyzes a specific gesture acted out by the user [37]. Airsign measures a movement while the user wearing the smartwatch writes their name in the air. Based on the analysis of the measurement, Airsign authenticates a valid user. After Airsign, Buriro et al. subsequently proposed SnapAuth, in which a user is required to snap their fingers [36]. Similar to Airsign, SnapAuth measures a movement while the user wearing the smartwatch snaps their fingers. Nguyen et al. proposed VeriNet to measure movements while the user enters a password on the smartwatch [56]. These methods analyze the distinct movement of the user while a particular behavior is performed, but the disadvantage is that a user must complete the specific gesture to be authenticated (i.e., low usability).

The gait-based method for smartwatch user authentication could provide high usability compared with other methods because the user does not have to concentrate on the authentication procedure [45]. Indeed, the user merely needs to walk normally. However, unfortunately, the gait-based method returns a high error rate [41]. Accordingly, smartwatch users might become frustrated when this method fails to authenticate successfully.

2.2 Analysis of Response to Vibrations

Similar to our method, other approaches have also analyzed a response to vibrations for user authentication [50, 51, 62]. To this point, Liu et al. proposed VibWrite, which is a two-factor authentication method [51] in which a user draws the PIN or a pattern as a secret on a vibrating panel. Absorption or reflection of the vibrations depends on the structure of the finger, such as shape, thickness, or bone density. Accordingly, the response to vibrations is analyzed as the second authentication factor. Being the most widely used algorithm in speech recognition [52, 61], Mel-frequency Cepstral Coefficient (MFCC) is employed for feature extraction from the response to vibrations. Following this, Li et al. proposed VELODY, a challenge-response structure [50] that also uses MFCC for feature extraction. VELODY generates vibrations at a randomly selected frequency, and the response to vibration depends not only on the physical structure of each human body but also on the vibration frequency. VELODY authenticates a user by verifying the response to vibration, which is measured when the user places their palm on the vibrating panel. Accordingly, this suggests that VELODY is secure against replay attacks because any response to vibrations used for authentication becomes useless to an attacker during subsequent

Table 1: Comparison of existing methods and our method

Method	Liu, et al. [51], VibWrite	Li, et al. [50], VELODY	Sim, et al. [62], Sim-2019	Our Method
Vibration Frequency	16kHz - 22kHz	0.5kHz - 10kHz	100Hz - 3kHz	170Hz - 240Hz (±35Hz)
Sampling Frequency	48kHz		6kHz ≤	100Hz
Feature Extractor	Statistical features and MFCC		Deep learning (CNN)	Statistical features, pairs of peak indices and heights, and differences between two adjacent pairs
Purpose	Identification (1:N matching)			Authentication (1:1 matching)

authentication attempts. Sim et al. proposed a method that analyzes raw signals for the response to vibrations with a convolutional neural network (CNN). Because of the deep learning algorithm, it was not necessary to extract any features from the raw signals [62].

Though there are many studies about vibration analysis-based methods for user authentication, it is difficult to apply these methods to current smartwatches because they were designed to generate high-frequency vibrations and measure responses at high sampling frequencies. In other words, existing methods cannot be implemented on modern smartwatches with default sensors, and the features used in their methods cannot be extracted from smartwatch measurements. For this reason, we propose a new method for user authentication based on measurements already available on smartwatches. In [50] and [51], the sampling frequency of the receiver that measures the response to vibration is 48kHz, and in [62], the sampling frequency of the receiver must exceed 6kHz by Nyquist's Theorem [58]. We use the built-in gyroscope and accelerometer sensors to measure the response to vibrations in our method because these sensors are built into smartwatches by default and were originally intended to measure user movement. The sampling frequency of these sensors is limited to 100Hz [8, 49].

Furthermore, the purpose of all existing methods is identification, which ultimately determines membership in a group (i.e., 1:N matching). However, the purpose of smartwatches is to authenticate only one person (i.e., 1:1 matching). In other words, smartwatches have to distinguish one single valid user from all other invalid users. A model for 1:N matching is trained with all the data belonging to the N classes (i.e., members of the group), and when new data is given to the model, it classifies the new data to one of the N classes. On the other hand, a model for 1:1 matching is trained using data belonging to one single valid class (i.e., one smartwatch user) without any invalid class data. Recognizing that 1:1 matching is more difficult than 1:N matching [46], we therefore constructed a novel verification model for user authentication on smartwatches.

Table 1 shows the comparison of existing methods with our method, where it can be seen that existing methods cannot be implemented with default smartwatch sensors. In Section 4, we show that users cannot be accurately classified by the statistical or MFCC features extracted from the responses to vibrations measured in the smartwatch.

3 ATTACK MODEL

We assume that the goal of an attacker is to bypass user authentication on a target smartwatch. If the attacker unlocks the smartwatch, they can steal sensitive and personal user information and make unauthorized payments.

Attackers' Abilities. The following outlines the presumed capabilities of potential attackers:

- An attacker knows how the target smartwatch authenticates its user.
- An attacker knows the target user's height, weight, body fat rate [1], and skeletal muscle rate [2].
- An attacker can steal a target smartwatch but cannot modify or manipulate the smartwatch software or data.
- An attacker can make the smartwatch restart user authentication with a slight physical touch when the smartwatch is locked. In general, the smartwatch detects whether the user is wearing it or not and locks if the user takes it off. When the user puts on the smartwatch again, user authentication restarts. For example, the user is asked to tap the password again to unlock. However, even if the user is not wearing the smartwatch, user authentication restarts with any tactile contact.
- An attacker can obtain the target user's previous response signals to vibrations.
- An attacker cannot physically damage sensors on the smartwatch.
- An attacker cannot attack during vibration enrollment for user authentication, which is performed when the user first purchases the smartwatch.

Attack Types. Depending on attacker ability, there are two potential types of attacks: not-in-wear attacks and impersonation attacks. In not-in-wear attacks, the attacker tries to bypass authentication by leaving the target smartwatch in a stationary position on a desk. When the target user takes off the smartwatch and puts it on the desk, the attacker can initiate the not-in-wear attack just by restarting the authentication process with a physical nudge. Even if the user authentication procedure restarts when an attacker touches it, the smartwatch should not unlock (i.e., user authentication should not be bypassable). Alternatively, in impersonation attacks, the attacker attempts to bypass user authentication on the target smartwatch by pretending to be the target user. There are two ways to perform impersonation attacks: i) by obtaining the target user's previous response signals (i.e., simulation attacks) and ii) by wearing the target smartwatch directly. The height, weight, BFR, and SMR are indicators of physical composition and the response to vibrations depends on the physical structure of the user's body. Accordingly, the attacker who uses the second method for an impersonation attack could have an accomplice with physical indicators similar to the target user wear the smartwatch to increase the probability of success.

4 OUR METHOD

In this section, we present a usable method for user authentication on smartwatches that operates independently and without any additional devices. Our method is based on a challenge-response authentication structure. Since most methods for biometrics-based

[1] Body Fat Rate (BFR) is the total mass of fat divided by total body mass, multiplied by 100 [5]

[2] Skeletal Muscle Rate (SMR) is the total mass of skeletal muscle divided by total body mass, multiplied by 100 [26]

authentication use *static* biometric information, they are vulnerable to presentation attacks using fake biometrics, suggesting the need for a challenge-response authentication method such as ours. In our method, a vibration type is selected at random among those provided on smartwatches by default.

When purchasing a smartwatch for the first time, the user must undergo vibration enrollment. During this step, the user's response signals to each vibration type are measured by the built-in gyroscope and accelerometer sensors, and a verification model to authenticate the user is subsequently generated for each vibration type. After vibration enrollment, user authentication is performed each time a user authentication request occurs. The overall process of user authentication is outlined below, and Figure 1 illustrates an example of how smartwatches authenticate users. We describe our method in three steps: i) vibration generation and measurement, ii) feature extraction, and iii) user verification.

4.1 Vibration Generation and Measurement

If a smartwatch officially provides a total of N vibration types, five are randomly selected from N by allowing reduplication. This random sequence of five vibration types becomes one challenge for user authentication, and this sequence is called a *vibration challenge*. The sequence of selected vibration types is saved. The smartwatch vibrates according to the generated *vibration challenge* and measures a response to *vibration challenge* using the built-in gyroscope and accelerometer sensors. This response is called a *vibration response*.

4.2 Feature Extraction

Each response signal is filtered to remove noise caused by user movements, and features are extracted from the filtered signal.

Filtering. Smartwatches are worn on the wrist, and the built-in gyroscope and accelerometer sensors can measure user movements. When our user authentication procedure is performed on a smartwatch, user movements become noise that must be removed. As the frequency range of human movement is between 0Hz and 20Hz [47], these noises are removed from response signals using a high pass filter of 20Hz.

Features Extraction. We extract features from each response signal after filtering occurs. Table 2 illustrates a list of extracted features, consisting of statistical features along with pairs of peak indices and heights in the frequency domain and in the correlation

Table 2: List of extracted features

IAV	Integral absolute value
MAV	Mean absolute value
Var	Variance
RMS	Root mean square
Std	Standard deviation
MAD	Median absolute deviation
SMA	Signal magnitude area
Skewness	Skewness of frequency domain signal
Kurtosis	Frequency signal kurtosis
IQR	Interquartile range
Energy	Average sum of the squares
Entropy	Signal entropy
Pairs of indices and heights of the five largest peaks in response signals, which are transformed by FFT, DCT, DWT, PSD, and auto correlation	
Differences between two adjacent pairs of peak indices and heights	

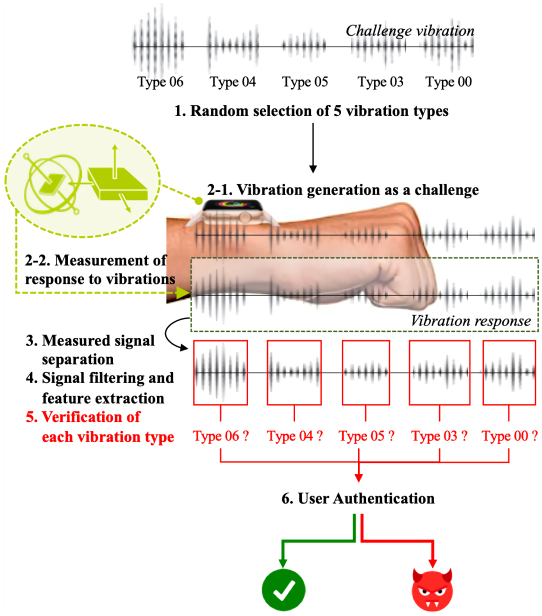

Figure 1: An example visualization of our method

of time domain. The statistical features are integral absolute value (IAV); mean absolute value (MAV); variance (Var); root mean square (RMS); standard deviation (Std); median absolute deviation (MAD); signal magnitude area (SMA); skewness, kurtosis, and interquartile range (IQR); energy; and entropy. After calculating the correlation (Auto Correlation) of a response signal in the time domain, we extract pairs of indices and heights of the highest five peaks in the correlation. Also, we extract pairs of indices and heights of the highest five peaks in a response signal which is transformed using fast fourier transform (FFT), discrete cosine transform (DCT), discrete wavelet transform (DWT), and power spectral density (PSD). We extract the differences of indices and heights between two adjacent pairs in the five extracted peaks. In other words, if the pairs of peak indices and heights are denoted as (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), and (x_5, y_5), the differences between two adjacent pairs of peak indices and heights become $(x_1 - x_2, y_1 - y_2)$, $(x_2 - x_3, y_2 - y_3)$, $(x_3 - x_4, y_3 - y_4)$, and $(x_4 - x_5, y_4 - y_5)$. A total of our features is shown in Table 2. Since each response signal to one vibration consists of six signals (i.e., three-axis signals each for the built-in gyroscope and accelerometer sensors), we have a total of 756 features for each response signal.

We evaluate whether our features are appropriate for the response signal to vibrations, measured by the smartwatch with a low sampling frequency of 100Hz. For evaluation purposes, we use a scatter method, in which the points of features tend to cluster according to participants if the features extracted from response signals for each participant are associated. However, if not, the points of features appear spread out. Since statistical features and MFCC are used in most existing vibration-based methods, we compare the scatter of our features to that of the statistical features and MFCC extracted from response signals sampled at 48kHz and 100Hz, respectively. It is noted that there must be two features to draw the scatter. However, since the dimensions of the features are higher than the two dimensions, we train the random forest

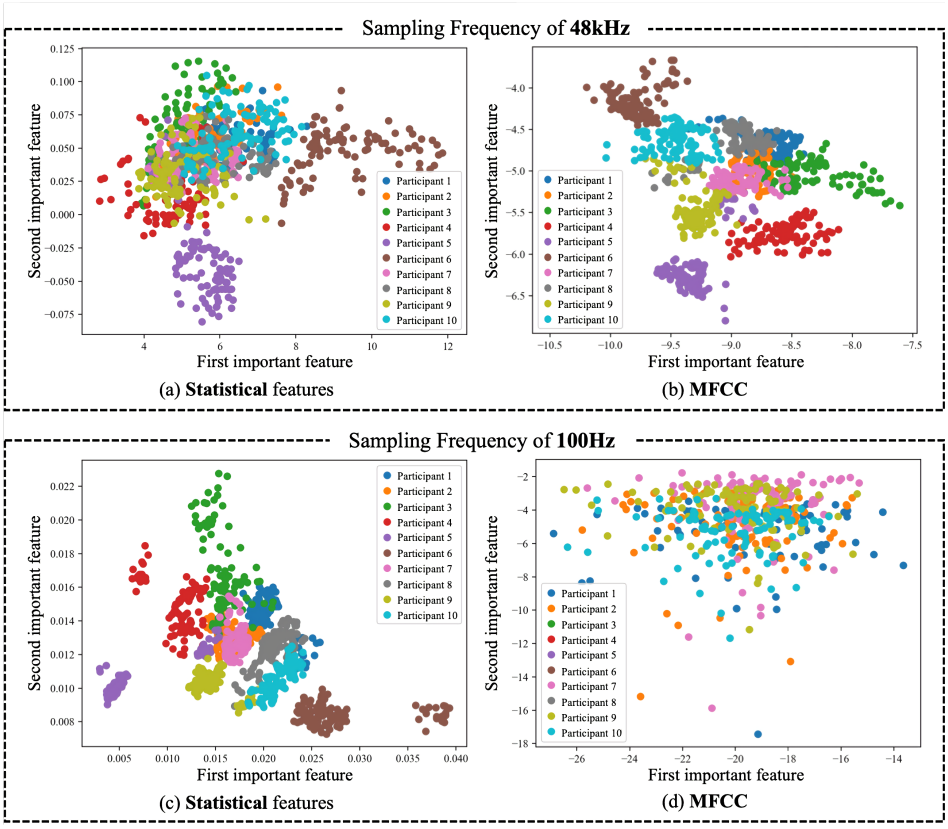

Figure 2: Scatter plots of features ((a) statistical features at 48kHz, (b) MFCC features at 48kHz, ((c) statistical features at 100Hz, (d) MFCC features at 100Hz)

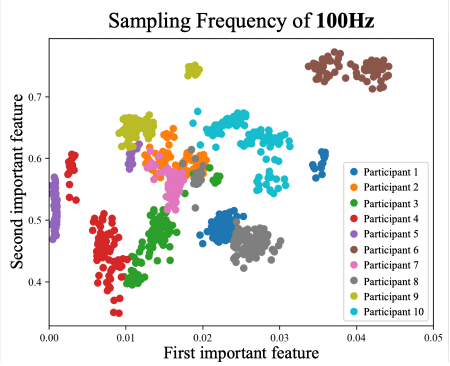

Figure 3: Scatter plots of our features

algorithm to classify ten participants and then select two important features from the total features. Random forest provides feature importance, which is a score assigned to features based on how useful each one is at predicting a participant.

Figure 2 illustrates scatter plots of (a) statistical features and (b) MFCC extracted from response signals sampled at a frequency of 48kHz, and (c) statistical features and (d) MFCC extracted from response signals sampled at a frequency of 100Hz. When response signals are sampled at the high frequency of 48kHz, statistical features and MFCC are associated respectively, as shown in Figure 2 (a) and (b). However, when response signals are sampled at the low frequency of 100Hz, statistical features are associated but MFCC are loosely associated, as shown in Figure 2 (c) and (d). As a result, MFCC cannot be used for low-sampled signals. Figure 3 illustrates a scatter plot of our features, and our features are associated higher than if only statistical features are used. Accordingly, our features are suitable even for response signals measured at the low sampling frequency of 100Hz.

4.3 User Verification

Using a verification model, each response signal constituting the *vibration response* signal is verified to determine whether it is a response signal measured from the valid user. If all five response signals are verified as valid, then user authentication is successful. However, if even one response signal is verified as invalid, user authentication immediately fails. Likewise, if the number of vibration types is N, the number of verification models for user authentication is N in total.

Figure 4 shows one verification model for user authentication. In smartwatches, user authentication is one-class classification, which means that a valid user should be distinguished from invalid users. In this problem, a one-class classifier must be trained using only valid response signals. It can be induced that if invalid response signals are given during user authentication, the classifier may not discover them. To solve this and improve the performance of a

one-class classifier, we devised a verification model for user authentication that combines a multi-class classifier and a binary classifier. As shown in Figure 4, the verification model for user authentication on smartwatches consists of a multi-class model and a binary model. The multi-class model is the part corresponding to the multi-class classifier, and the binary model corresponds to the binary classifier. We set default users to construct the multi-class model. In other words, when the smartwatch is shipped, the smartwatch is already loaded with the default users' response signals by vibration type. The default datasets are needed just to generate our verification model for user authentication. Therefore, when a user purchases a new smartwatch, only response signals to each vibration type are measured from the new user, and the verification model for each vibration type is subsequently generated (i.e., this is vibration enrollment for user authentication, which is performed when the user first purchases a smartwatch). It is noted that information about default users, who are needed to generate verification models at the vibration enrollment step, remains private from the outset to preserve privacy. Details of the multi-class model and binary model as relating to the verification model are as follows.

Multi-class Model. The multi-class model classifies response signals measured from users including default users and new users. It is noted that default users and new users become classes in a multi-class model. If the response signals (i.e., features extracted from filtered response signals) are given to the multi-class model, it computes a probability vector to which class each response signal belongs. The length of the probability vector is the same as the number of classes. Further, the response signal is classified to whichever class is the same as the position of an element in the probability vector that exceeds the threshold. For example, if there are six classes from class 0 to class 5 (i.e., six different users including five default users and one new user) and the threshold is 0.5, and if the multi-class model computes the probability vector of the response signal as (0.1, 0.2, 0.15, **0.8**, 0.1, 0.24) which has six elements, this response signal is classified as class 3. Consequently, the probability vectors are given to the binary model as shown in Figure 4. We use a random forest classifier [35] for the multi-class model as it is less likely to be overfitted on training data because it trains each tree independently using random samples [65]. Importantly, studies have concluded that this is suitable for multi-class problems [63] and biometric information [66]. Finally, to optimize performance of the multi-class model, we use a cross-validation method and tune hyper-parameters.

Binary Model. The binary model ultimately classifies the response signal as valid or invalid. As shown in Figure 4, the probability vectors, which are intermediate outputs of the multi-class model, are labeled according to the default users and new user and are given to the binary model. The probability vectors of the new user's response signals are labeled as a valid class, and the remaining vectors (i.e., probability vectors of default users' response signals) are labeled as an invalid class. We use a support vector machine (SVM) classifier [38] for the binary model. Since only the response signal of the new user is labeled as valid, and the others are labeled as invalid, all response signals, which become training data for the binary model, are imbalanced data. SVM has been evaluated to provide good results when handling imbalanced data [31, 42]. Also, it has been shown that applying different kernels to SVM delivers satisfactory performance even with imbalanced data [32]. Therefore, to optimize SVM, we use a cross-validation method

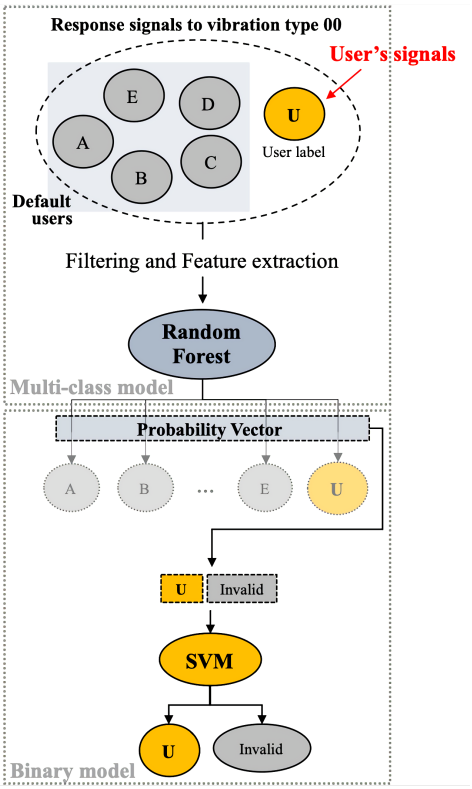

Figure 4: One verification model for each vibration type

and tune various hyper-parameters of SVM, including the kernel. In addition, when training the binary model, accuracy, which is one metric for model evaluation, is not applied to imbalanced data. Instead, precision and recall scores are used.

When response signals to vibrations for another user, who is neither a default user nor a new user, are given to the multi-class model, the probability vectors of response signals from another user differ from the the probability vector being classified into the valid class (i.e., the new user). As a result, the binary model classifies such response signals from another user as invalid.

4.4 Comparison of Our Model and Other Classifiers

We evaluate the authentication performance of one-class classification and binary classification, and then compare these results with our method. The SVM is used to construct one-class or binary classifiers. For a one-class SVM, only data (i.e., response signals to vibrations) from a valid user are used as a training dataset. For a binary SVM, one class is for the valid user and the other is for all other users (i.e., invalid users). Since we could not train the binary SVM with all possible invalid users, we created five participants and used their response signals to train the class for invalid users. We evaluated the authentication performance of models based on the equal error rate (EER), which is the rate at which both acceptance and rejection errors are equal. The lower the EER value, the higher the accuracy of the authentication. Figure 5 shows the EER of the one-class classifier, the binary classifier, and our method. As shown, the one-class SVM and the binary SVM have EERs of 0.523

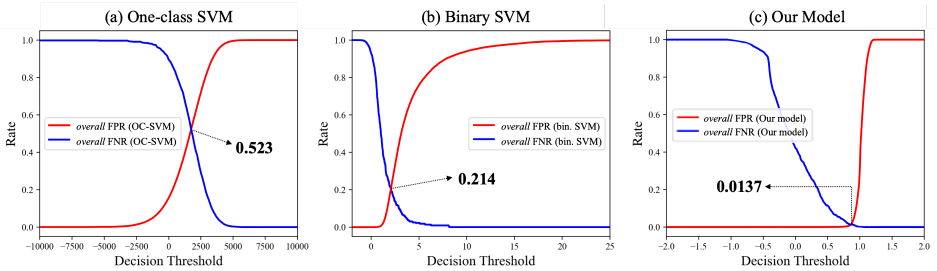

Figure 5: The authentication performance of a one-class classifier, a binary classifier, and our method

and 0.214, respectively. It seems that the reason why the binary SVM has a lower EER than the one-class SVM is because data from valid and invalid users are both trained. However, this rate of 0.214 is still high, and it is not appropriate to use the binary SVM for user authentication on smartwatches. Accordingly, we designed our method by combining multi-class and binary classifications for high-performance user authentication. In our method, data from five participants, called the default users, are used to train the five individual classes in the multi-class model and the class for invalid users in the binary model. Comparing the one-class and the binary SVMs, our method has an EER of 0.0137, which implies that the model combination outperforms single models. In Section 6, we detail the evaluation of our method.

5 EXPERIMENTAL SETUP

The following devices, software, and details were involved in measuring participant response signals and constructing verification models for the vibration types used in our evaluation.

Device. To apply our vibration-based user authentication method, we used the Apple Watch Series 3, which is 38mm in size [2] and comes with standard bands in small and medium. The Apple Watch Series 3 officially has a total of nine default vibration types [17]. The vibration types are named Notification, DirectionUp, DirectionDown, Success, Retry, Failure, Start, Stop, and Click. The last vibration type, Click, has a very low signal strength, and as such, our method is constructed using the other eight vibration types. Although Apple does not release detailed information about vibration frequencies on its public website, it is generally accepted that built-in smartwatch mechanisms vibrate between 170Hz and 240Hz [19, 20], which are low frequencies.

Software. In order to obtain user response signals on the Apple Watch Series 3, we created an application using XCode [30], a comprehensive program for developers to create applications for Apple devices. The application makes the Apple Watch vibrate and simultaneously measure response signals to vibrations using the built-in gyroscope and accelerometer sensors. It is noted that the maximum sampling frequency provided by the default Apple Watch hardware is 100Hz [8]. Finally, we used Python [22] for signal processing and construction of verification models.

Signal Measurements. For the experiment, we recruited 20 students from our graduate school. All participants were in their 20s and 30s, were fully informed of the exact method and reason for the experiment, and gave consent. In addition, each participant was compensated with a coffee shop gift card as incentive for participating. Using the application that we created, we obtained response signals for all 20 participants. For each participant, we first measured 150 response signals to each vibration type. Excluding the last 30 response signals, the remaining 120 response signals were used to train the verification model and tune its hyper-parameters. Seven days after the first measurement day, we measured 30 response signals to each vibration type again on all 20 participants. It is noted that participant recruitment occurred over 10 days, and on the first measurement day, it took an average of two hours including break time per participant to measure the response signals to vibrations. Seven days after the first measurement day, we measured the response signals over an average of one hour per participant. Accordingly, it took approximately six weeks in total to complete the evaluations. During *vibration response* signal measurement, all participants held the back of their hands facing up. Some participants did move their wrists as they spoke or shifted their posture. In general, the wearing location may change slightly when a user puts the smartwatch on. Therefore, we randomly shifted the location of the smartwatches on the participants' wrists by 1cm (i.e., 0.5 cm above and below the initial baseline wearing) and measured the response signals again on the first and second measurement days. Table 3 shows the information of participants, including their band sizes, hole location from the inside, height, weight, body fat rates (BFR), skeletal muscle rates (SMR), age, and sex. Participant heights spanned 150cm to 190cm, weights spanned 49kg to 96kg, BFRs spanned 11% to 44%, and SMRs spanned 32% to 56%.

Table 3: Participant information

ID	Band	Hole	Height	Weight	BFR	SMR	Age	Sex
00	S	2nd	154cm	49.4kg	20.5%	43.4%	27	F
01	S	4th	162cm	54.0kg	22.3%	38.7%	30	F
02	M	4th	180cm	84.0kg	28.2%	49.2%	25	M
03	M	2nd	173cm	85.0kg	33.3%	40.5%	26	M
04	M	3rd	176cm	92.0kg	34.3%	45.3%	27	M
05	S	3rd	174cm	65.0kg	20.6%	54.6%	29	M
06	S	4th	172cm	66.0kg	21.1%	45.5%	27	M
07	S	3rd	186cm	69.0kg	13.9%	55.2%	20	M
08	M	1st	167cm	68.2kg	22.8%	42.6%	39	M
09	M	4th	180cm	91.0kg	27.6%	49.8%	31	M
10	M	4th	172cm	77.1kg	30.3%	47.6%	30	M
11	S	4th	170cm	72.0kg	26.6%	51.5%	30	M
12	M	4th	179cm	94.9kg	29.5%	43.8%	35	M
13	M	1st	176cm	73.3kg	11.7%	44.1%	28	M
14	S	3rd	175cm	76.0kg	26.5%	43.4%	26	M
15	M	Last	168cm	95.7kg	43.1%	40.1%	34	M
16	S	4th	173cm	68.0kg	20.3%	54.1%	30	M
17	S	5th	171cm	72.2kg	28.1%	48.4%	24	M
18	S	4th	156cm	60.5kg	33.6%	32.7%	24	F
19	M	4th	172cm	78.0kg	33.3%	51.3%	24	M

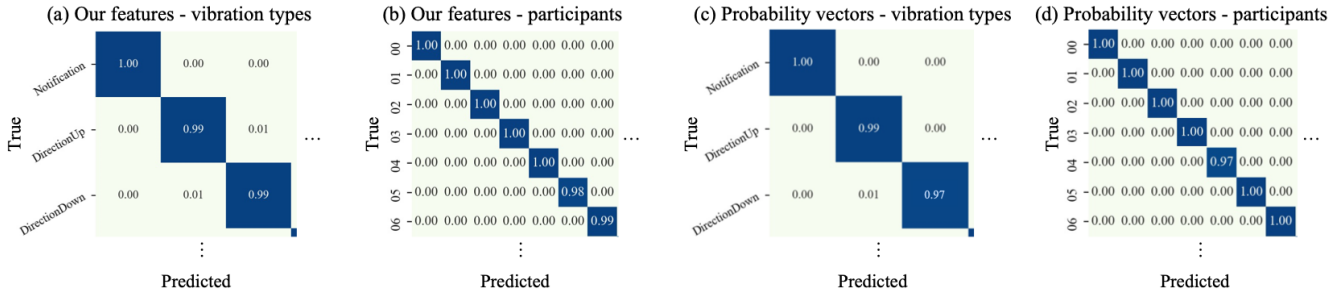

Figure 6: Distinguishability of our features and probability vectors used in our method

Ethical Considerations. Our research is classified as human subject research and has been approved by the Institutional Review Board (IRB) of our institution. In addition, in our method, the default users are required to generate verification models during the vibration enrollment step. Therefore, for privacy reasons, information about default users remains anonymous from the outset. In a similar vein, all response signals measured by the user for model generation are deleted after verification models are generated, and the models are safely stored.

Construction of Verification Models. We selected 5 of the 20 participants to set as default users, and the remaining 15 became new users. The verification models for each vibration type, which include multi-class and binary models, are generated by adding each new user to the default users. It is noted that of the 150 response signals per vibration type measured on the first day, the last 30 response signals are used as test data in the final evaluation, and the remaining 120 response signals are used to train the verification model and tune hyper-parameters. Accordingly, for each participant that becomes a new user, the total number of vibration signals to train in the verification model for each vibration type is 720. As mentioned in Section 4, we extracted a total of 756 features from each response signal. Accordingly, the shape of data to train in the multi-class model for each vibration type is (720, 756).

To optimize multi-class models, we used a five-fold cross-validation method and a `gridsearch` function provided in the Python library to tune hyper-parameters, the number of trees, and the maximum depth of trees. The number of trees and the maximum depth of trees are tuned as one of [50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000] and [1, 2, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, None], respectively. Then, we used a `predic_proba` function to obtain probability vectors of response signals. The probability vectors of the signals become training data for each binary model, and the shape of training data for each binary model is (720, 6). Finally, to optimize the binary model, the hyper-parameters are tuned using the five-fold cross-validation method and the `gridsearch` function, where hyper-parameters are kernels, degree of the polynomial kernel function (which is ignored when using other kernels), kernel coefficients, and regularization parameters. The kernels, degree of the polynomial kernel function, kernel coefficients, and regularization parameters are tuned as one of ['linear', 'poly', 'rbf', 'sigmoid'], [1, 2, 3, 4, 5, 10, 12, 15], ['scale', 'auto'], and [0.0001, 0.001, 0.01, 0.1, 1, 10, 100], respectively. For example, hyper-parameters of the verification model for vibration type 00 are tuned as follows: multi-class model = {'max_depth': 10, 'n_estimators': 600}, and binary model = {'C': 0.1, 'degree': 1, 'gamma': 'scale', 'kernel': 'linear'}.

6 EVALUATION

We evaluated our method with responses to vibrations sampled at low frequencies to show the feasibility of our model in current smartwatches. The results of our evaluation show that our method achieves low error rates when authenticating smartwatch users, and we also show that our method correctly authenticates a user even during attacks defined in Section 3.

6.1 Performance Metrics

In our evaluation, we defined the equal error rate (EER) as a performance metric. The EER occurs when the false negative rate (FNR) and the false positive rate (FPR) are equal. The FNR and FPR are defined as follows.

- **FNR**: The rate of false rejections, which is used to evaluate usability. It is noted that the FNR is equal to 1 − TPR (true positive rate), where the TPR is the rate of true acceptance.
- **FPR**: The rate of false acceptances, which is used to evaluate security.

In our method, user authentication succeeds if the five response signals constituting the *vibration response* signal are all verified as valid by the verification models. Therefore, the success rate of user authentication is the product of all TPRs in the verification models that correspond to the five selected vibration types. We call this rate the *overall* TPR of our method (i.e., *overall* TPR = $\prod_i^5 TPR_i$ (TPR_i is the TPR of the i-th verification model)). On the other hand, if one of the five response signals comprising the *vibration response* signal is verified as invalid, user authentication immediately fails, and our method does not perform verification for the remaining signals. Therefore, the failure rate of user authentication is 1 − *overall* TPR, and we call this rate the *overall* FNR. Likewise, the success rate of an attack is the product of all FPRs in the verification models that correspond to the five selected vibration types, known as the *overall* FPR of our method (i.e., *overall* FPR = $\prod_i^5 FPR_i$ (FPR_i is the FPR of the i-th verification model)). The *overall* EER of our method is calculated using the *overall* FNR and *overall* FPR.

6.2 Distinguishability

We first conducted a random forest classification algorithm to ascertain whether or not our features are able to distinguish between multiple vibration types. Moreover, we checked if our features are able to distinguish between multiple participants. We generated

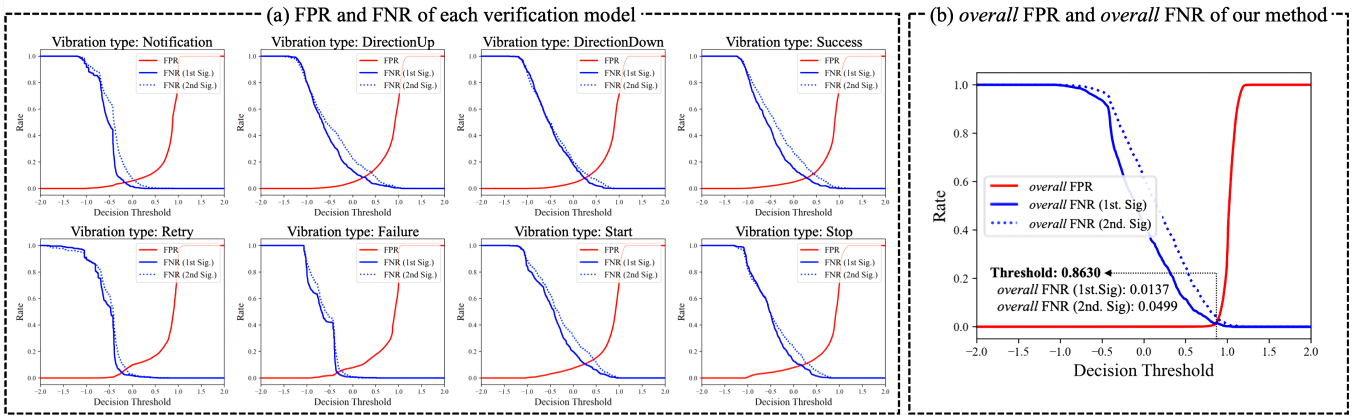

Figure 7: Evaluation results for the verification models by vibration type and our method

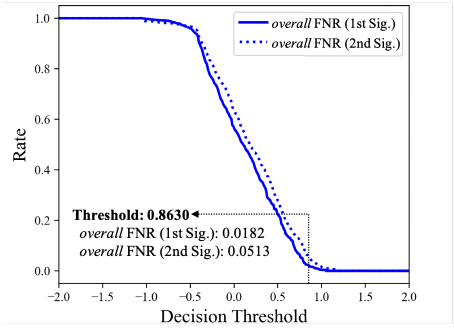

Figure 8: The *overall* FNR of our method when the location of the smartwatch shifts on the wrist

eight types of vibrations and measured their respective response signals on one single participant. Figure 6-(a) illustrates a part of the full confusion matrix, which is the result of classifying our features extracted from response signals by vibration types. In the diagonal components, a minimum rate of 0.96 and a maximum rate of 1.00 are given. This result implies that our features are able to distinguish eight different vibration types with low error rates. We subsequently generated a fixed vibration type on participants and measured the response signals. Figure 6-(b) illustrates a part of the full confusion matrix, which is the result of classifying our features that were extracted from the participants' response signals. In the diagonal components, a minimum rate of 0.95 and a maximum rate of 1.00 are given. Accordingly, our features are able to distinguish different participants when the fixed vibration type is given. The full confusion matrix can be seen in the Appendix.

Probability Vectors. Our method leverages the probability vectors of response signals as intermediate values that are used for the binary model. Accordingly, we also conducted a random forest classification algorithm to check if the probability vectors are able to distinguish between multiple vibration types. Moreover, we checked if the probability vectors are able to distinguish between multiple participants. Figure 6-(c) illustrates a part of the full confusion matrix, which is the result of classifying probability vectors by vibration types. In the diagonal components, a minimum rate of 0.96 and a maximum rate of 1.00 are given. This result implies that probability vectors are able to distinguish eight different vibration types with low error rates. Figure 6-(d) illustrates a part of the full confusion matrix, which is the result of classifying probability vectors by participants. In the diagonal components, a minimum rate of 0.93 and a maximum rate of 1.00 are given. Accordingly, the probability vectors are able to distinguish different participants when the fixed vibration type is given. The full confusion matrix can be seen in the Appendix.

6.3 User Authentication

In this section, we present the evaluation results of our method for user authentication. We first generated vibrations and measured the response signals to those vibrations. Because our method generates a verification model for each vibration type, we calculated the FNR and the FPR on the verification model that is trained with responses to the vibration types from participants. When we set default users as participants 00 to 04, Figure 7-(a) illustrates the FNR and the FPR of the verification model for each vibration type, which are the average rates from 15 participants (i.e., new users who become the rest after selecting five default users). In addition, we measured response signals to vibrations a few days after the first measurement to ascertain whether or not the features used in our method were stable when measured on different days. The blue solid lines indicate the FNR for the response to vibrations measured on the first day, and the blue dotted lines indicate the FNR for the response to vibrations measured seven days later. It can be seen that the FPR increases and the FNR decreases as the threshold increases. For a one-verification model, a high FPR is obtained, but the total FPR of our method is much lower in actuality because FPRs obtained from the model should be multiplied (i.e., *overall* FPR). Figure 7-(b) illustrates the *overall* FNR and the *overall* FPR of our method when five default users are set as participants 00 to 04. In our method, the *overall* FNR and the *overall* FPR are functions of the threshold, which is the point where the *overall* FNR for the first measurement and the *overall* FPR are equal. Accordingly, we determine that the threshold of our method is 0.8630, which means that the average *overall* FNR for the first and second measurements is 0.0137 and 0.0499, respectively.

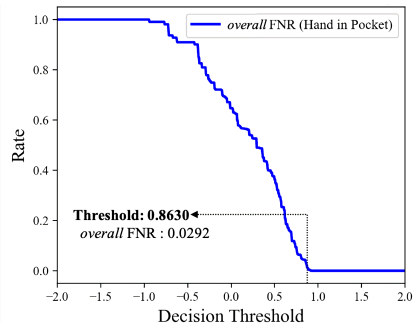

Figure 9: The *overall* FNR of our method when participants' hands are in pockets

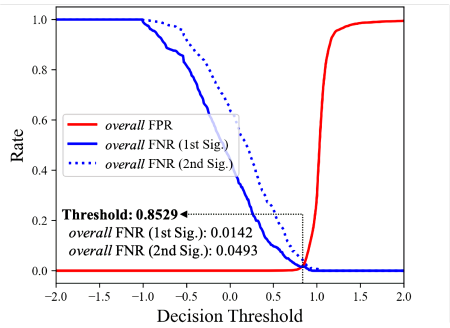

Figure 10: The *overall* FPR and *overall* FNR of our method with randomly selected default users

6.3.1 Shift of Smartwatch Location on the Wrist. We also evaluated user authentication performance when the smartwatch position shifts on the wrist. We measured the response signals by randomly shifting the smartwatches both 0.5cm above and below the baseline of the location where the smartwatches were first worn. After seven days (i.e., the second measurement day), the response signals were measured again in the same way. Figure 8 shows the *overall* FNR for the first-measured and second-measured signals when the location of the smartwatch shifts. Based on the determined threshold of 0.8630, the *overall* FNRs for the first and second-measured signals increase to 0.0182 and 0.0513. However, it can be said that this error rate is still low. Therefore, our method is robust even if the location of the smartwatch varies slightly on the wrist.

6.3.2 Position and Direction of the User's Hand. We evaluated our method for the situation in which the participant's hand takes a different position or changes direction. We measured the response signals to vibrations when participants held five different finger positions: i) all fingers folded to make a fist, ii) one finger making a "pointing sign," iii) two fingers making a "peace sign," iv) thumb extended for a "thumb's up sign," and v) all fingers relaxed. It is noted that the "pointing sign" involves extension of only the index finger with all remaining fingers folded in, the "peace sign" sign indicates that only the index and middle fingers are extended, and the "thumb's up sign" indicates that only the thumb is extended. Next, we measured the response signals for six different wrist directions, depending on i) which of x-axis, y-axis, or z-axis directions changed and ii) the degree of change in direction. It should be noted that when we measured response signals from the built-in gyroscope and accelerometer sensors, we employed the API, which enables measurement of response signals without gravity. Table 4 shows the *overall* FNR as a function of finger positions and wrist directions when we applied the threshold of 0.8630 as outlined in Subsection 6.3. Moreover, we measured the response signals when participants held their hands in their pockets. Figure 9 shows the *overall* FNR of 0.0292 based on the threshold of 0.8630. As a result, we conclude that our method has low *overall* FNRs regardless of hand or finger position and direction, even when the hand wearing the smartwatch is in a pocket.

6.3.3 Arbitrary Default Users. We evaluated user authentication performance when default users were selected at random from the 20 participants. We repeated this random selection five times. Figure 10 illustrates the average *overall* FNR and the average *overall* FPR of our method for the five randomly selected default users. The result shows that the average *overall* EER obtained for first-measured signals is 0.0142. Additionally, based on the threshold of 0.8529, which is the threshold corresponding to the *overall* EER for first-measured signals, the *overall* FNR for second-measured signals is 0.0493. As a result, performance is similar when the five default users are selected as participants 00 to 04 or at random, which indicates that the method for selecting default users does not affect our method.

6.3.4 Number of Default Users. We evaluated user authentication performance by varying the number of default users from four to eight. Figure 11 shows the *overall* FPR and the *overall* FNR according to the number of default users. When four default users are selected for our method, the *overall* EER for the first-measured signals returns as 0.0379. In our method, since the *overall* FNR and *overall* FPR act as functions of a threshold, the threshold is determined as 0.6937, and the *overall* FNR for the second-measured signals is 0.0644. This error rate is higher compared to the existing method [50], which reported an error rate of 0.058. With five, six, seven, and eight default users, the *overall* EERs for the first-measured signals are 0.0137, 0.0126, 0.125, and 0.0126, respectively, revealing a downward trajectory as the number of default users increases. Additionally, after determining the thresholds, the *overall* FNRs for the second-measured signals when there are five, six, seven, and eight default users are 0.0499, 0.0457, 0.0459, and 0.0501, respectively. Unlike the *overall* FNR for the first-measured signals, the *overall* FNRs for the second-measured signals first decrease and then increase again as the number of default users increases. Even with the increase, these error rates are lower than those reported for the existing method [50]. Moreover, since fewer classes would be lightweight in terms of model complexity, we selected five default users even though our method performs marginally better with six default users.

Table 4: The *overall* FNR of our method for different hand positions and directions

Hand Positions	*overall* FNR	Wrist Directions	*overall* FNR
All fingers folded	0.0153	X-axis: 90°	0.0078
Pointing Sign	0.0020	X-axis: 90°, Y-axis: 90°	0.0171
Peace Sign	0.0027	X-axis: 90°, Z-axis: 90°	0.0146
Thumb's Up Sign	0.0145	X-axis: 180°	0.0085
All fingers relaxed	0.0015	X-axis: 180°, Y-axis: 90°	0.0105
-		Y-axis: 90°	0.0208

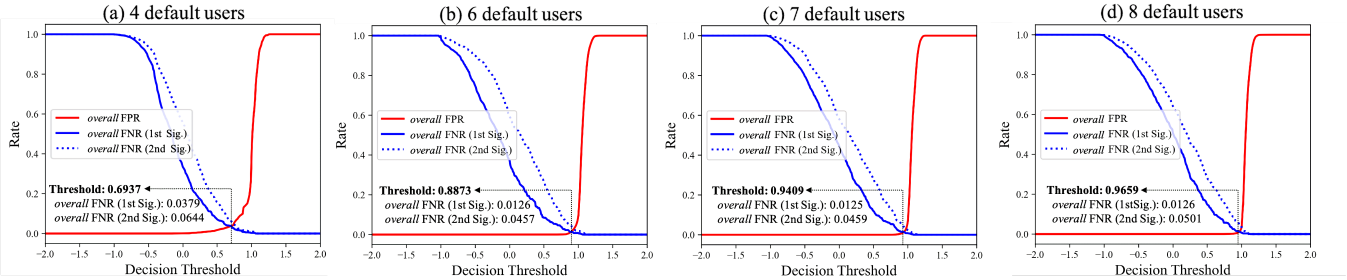

Figure 11: The *overall* FPR and the *overall* FNR of our method with varying the number of default users

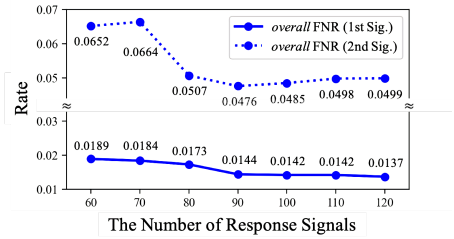

Figure 12: Relation between the *overall* FNR of our method and the number of response signals per vibration type for training

6.3.5 Training Dataset Size. We evaluated the *overall* FNRs of our method as a function of the size of training dataset, which is a trade-off parameter between usability and accuracy. We reduced the number of response signals per vibration type for training the verification model to 60, 70, 80, 90, 100, 110, and 120. Figure 12 shows the *overall* FNRs for the first-measured and the second-measured signals as the training dataset size changes. These results demonstrate that 90 response signals per vibration type are sufficient to generate the verification models while maintaining low *overall* FNRs.

6.3.6 On the Prototype. Because our method is evaluated on a commercial smartwatch, the number of default vibration types is limited. Out of these vibration types, four take 0.5 seconds and the other four take 1 second, and the maximum authentication time is approximate 5 seconds. Since, according to Nah et al. [54], users' tolerable waiting time is up to 2 seconds, we evaluated our method in terms of fastest authentication time. For this reason, we made a prototype with a vibration sensor (ELB060416), a gyroscope sensor, and an accelerometer sensor (MPU6050), as shown in Figure 13-(a). In addition, we leveraged a polycarbonate (PC) plate (3.5cm x 3cm in size and 1.5mm in thickness) with two Arduino UNO boards connected to each sensor. The ELB060416 sensor is a vibration sensor that can control the intensity of vibrations and is set to vibrate a total of eight different types that are all 0.4 seconds in length by controlling the interval and intensity of vibrations. The MPU6050 sensor consists of both gyroscope and accelerometer sensors and is set to measure response signals at a sampling rate of 100Hz. We measured a total of 120 response signals per participant per vibration type. Of the measured 120 response signals, 30 are used as test data, and the remaining 90 are used for verification model training and hyper-parameter tuning.

Like the evaluation in Subsection 6.2, we conducted a classification algorithm to verify whether or not features extracted from the response signals that are measured by our prototype are able to distinguish between eight vibration types. Figure 13-(b) shows a part of the full confusion matrix that implies our features can distinguish between the eight different vibration types with low error rates. In diagonal components, a minimum rate of 0.95 and a maximum rate of 1.00 are given. The full confusion matrix can be seen in the Appendix. Figure 13-(c) shows the *overall* FPRs and the *overall* FNRs, and an *overall* EER of 0.0296 can be seen when the threshold of 0.8239 is employed.

6.4 Not-in-Wear Attack

We mounted a not-in-wear attack in which an attacker does not make an effort to bypass user authentication. Instead, the attacker only focuses on proper placement of the target smartwatch and touches it slightly to restart the authentication process. We measured the response signals to each vibration type when the smartwatch was under one of the following conditions: i) in the air (i.e., non-contact with any object), ii) on a solid wooden desk, and iii) on medium-density fiberboard (MDF). Thirty responses to the vibrations for each condition were measured as signals used for the not-in-wear attack. Figure 14 illustrates the average *overall* FPR for the not-in-wear attack on participants excluding the default users, participants 00 to 04. Based on the threshold of 0.8630, which is the same as that used in the evaluation of user authentication, the highest success rate of not-in-wear attacks is 0.0268, where the attack signals were measured on the MDF. Since the smartwatch detects whether the user is wearing it or not, it is possible to lower the success rate of attacks by requiring authentication only when the smartwatch is worn.

For participant 15, the response signals to vibration types 01, 02, 03, 06, 07, and the not-in-wear attack signals are not clearly distinguishable. As a result, based on the pre-determined threshold of 0.8630, the highest success rate of not-in-wear attacks on participant 15 is 0.527. In other words, unlike other participants, participant 15 is vulnerable to not-in-wear attacks that target certain vibration types. To this point, we analyze the BFR of participant 15 and found that it differs from others as it is the highest. The BFRs of the other participants span 11% to 35%, but participant 15 has a BFR of 43.1%. It can be seen that the higher the BFR, the lower the chance that the physical structure of the wrist is well-reflected in the *vibration response* signals. To solve the case of participant 15, we can construct the user authentication system if we randomly select seven vibration types from vibration types 00, 04, and 05 by allowing reduplication.

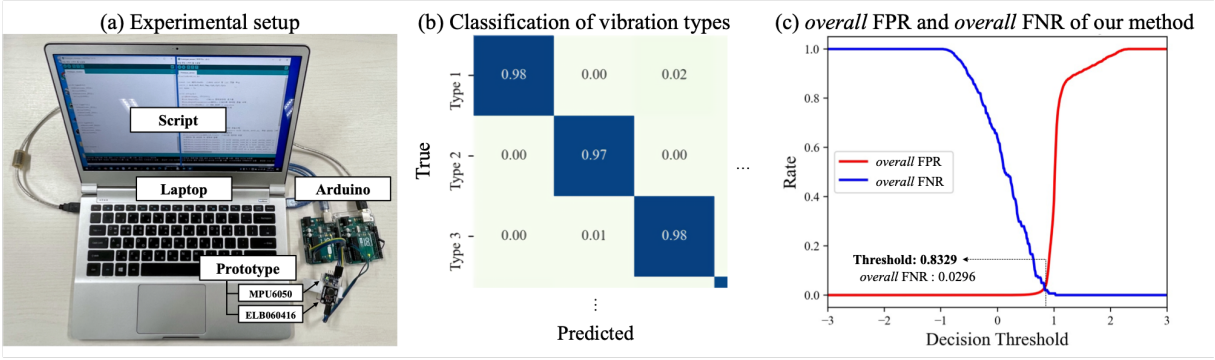

Figure 13: Evaluation with our prototype ((a) experimental setup (b) distinguishability of vibrations (c) results of our method)

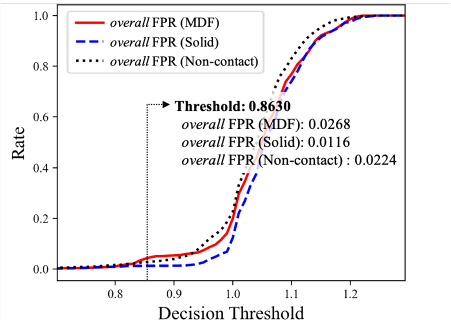

Figure 14: Not-in-wear attacks for various attack signals

6.5 Impersonation Attack

We mounted an impersonation attack, during which the attacker attempts to bypass user authentication by using a participant with similar physical indicators as the targeted user, such as height, weight, BFR, or SMR. In order to evaluate our method for impersonation attacks, we selected a pair of participants with similar physical indicators, excluding the default users, meaning participants 00 to 04 whose response signals to vibrations were already used for verification model generation. We first selected three pairs of participants with similar heights: (Participant 06, Participant 10), (Participant 06, Participant 19), and (Participant 10, Participant 19). Second, we selected three pairs of participants with similar weights: (Participant 08, Participant 16), (Participant 11, Participant 17), and (Participant 12, Participant 15). Third, we selected three pairs of participants with similar BFRs: (Participant 05, Participant 16), (Participant 11, Participant 14), and (Participant 18, Participant 19). Lastly, we selected three pairs of participants with similar SMRs: (Participant 09, Participant 17), (Participant 11, Participant 19), and (Participant 12, Participant 14). In each pair, the first participant is assumed to be the target user. Based on the predetermined threshold of 0.8630, Table 5 shows the *overall* FPRs for impersonation attacks as a function of similar physical indicators, such as height, weight, BFR, and SMR. The average *overall* FPRs for similar height, weight, BFR, and SMR are 3.05e-04, 3.32e-08, 9.38e-04, and 1.64e-03, respectively. Additionally, for the participant 05 and participant 16 pair, even though not only their BFRs are similar to each other but also their SMRs are similar at 54.6% and 54.1%, the *overall* FPR is low as 4.70e-04. Accordingly, it is difficult to bypass our method even if attackers leverage an accomplice with a similar physical composition as a target user.

6.6 Memory Overhead

The size of a verification model for each vibration type is approximately 10.88MB, including the multi-class model and binary model. Accordingly, the total size of our method is approximately 87.04MB. This memory size is similar to or smaller than applications normally installed on smartwatches. For example, Apple Watch application Nike Run Club is approximately 181.5MB, which is larger than our method. Therefore, our method does not have high memory overhead and can be applied to current smartwatches.

7 DISCUSSION

Measurement Time for Model Generation. The measurement of response signals is the main factor affecting how much time it takes to generate verification models in our method. In the evaluation, we showed that approximately nine minutes must be budgeted to efficiently measure response signals per vibration type. The setup process for model generation is necessary if users change their smartwatch bands, just like when they first purchase and set up a new smartwatch. Depending on the type of smartwatch band, the degree of absorption, reflection, and propagation of vibrations may differ. However, our method does not need to measure all response signals at once. For feasibility, we analyzed how long users kept their hands stationary throughout the day. The ExtraSensory Dataset publicly provides the smartwatch accelerometer signals for

Table 5: The *overall* FPR for impersonation attacks on pairs of participants with similar physical indicators

Indicator	Pairs of (Participant ID, Value)	*overall* FPR
Height	(Participant 06, 172cm), (Participant 10, 172cm)	6.15e-04
	(Participant 06, 172cm), (Participant 19, 172cm)	7.47e-10
	(Participant 10, 172cm), (Participant 19, 172cm)	2.99e-04
Weight	(Participant 08, 68.2kg), (Participant 16, 68.0kg)	0.0
	(Participant 11, 72.0kg), (Participant 17, 72.2kg)	9.97e-08
	(Participant 12, 94.9kg), (Participant 15, 95.7kg)	0.0
BFR	(Participant 05, 20.6%), (Participant 16, 20.3%)	4.70e-04
	(Participant 11, 26.6%), (Participant 14, 26.5%)	6.75e-04
	(Participant 18, 33.6%), (Participant 19, 33.3%)	1.67e-03
SMR	(Participant 09, 49.8%), (Participant 17, 48.4%)	5.38e-09
	(Participant 11, 51.5%), (Participant 19, 51.3%)	4.83e-03
	(Participant 12, 43.8%), (Participant 14, 43.4%)	7.74e-05

56 individuals during daily activities [10], and analysis of the open dataset shows that there are at least two minutes of no movement, comprising approximately 12% of the total time. This implies that the setup process can be spread out across a few hours to avoid a nonstop nine-minute vibration measurement session. Therefore, we expect the user to anticipate dedicating only a reasonable amount of time to allowing training data for model generation to be collected.

Simulation Attack. To evaluate simulation attacks against our method, we considered the case of an attacker who can obtain the previous *vibration response* signals of a target user and tries to use these to spoof the gyroscope and accelerometer sensors by injecting acoustic signals, to which the sensors are sensitive. However, previous research shows that it is difficult for the attacker to inject acoustic signals and spoof all three axes of the sensor to their desired values [64]. For example, there are sensors affected by acoustic signals only on the x-axis and others that are vulnerable on all three axes, but all resonant frequencies of the signals affecting each axis are different. In addition, a recent study raised an open-ended question of whether or not it would be possible to use acoustic injections to precisely control the gyroscope and accelerometer measurements on all three axes simultaneously and/or for longer periods of time, indicating the difficulty of spoofing sensors to an attacker's desired values [43].

However, our method is robust even against a strong attacker who is capable of performing sophisticated attacks previously shown to be difficult to carry out in previous studies [43, 64]. In this scenario, it is assumed that attackers would be able to spoof all three axes of the gyroscope and accelerometer sensors to their desired values as they inject acoustic signals into the sensors. Since our method authenticates a user with the *vibration challenge* and *vibration response* pair consisting of five vibration types, and each of the five vibration signals is randomly selected from eight types by allowing reduplication, the probability of this strong attacker succeeding—that is, the probability of predicting exactly all five vibration types—is 0.003%. With results this close to zero, we assert that our method is robust even against strong attackers.

User Authentication on Moving Hands. Users may engage in various activities while wearing a smartwatch, such as walking or shaking hands. However, our method does not have to be performed during each instance of such daily movements. Rather, our method should be performed right when users put the smartwatch on their wrist. If authentication is successful, additional authentication may be required when users engage in a particular activity, such as processing a payment using their smartwatch, which is a type of activity that seems more static than other daily activities. In Section 6, we evaluated our method when the user's hands are stationary, though in varying positions and directions. Accordingly, we believe that our method is not disturbed by a user's movement.

Age Limitation for Participants. We publicly recruited participants in our graduate school for our evaluation, and unfortunately, this limited our participant pool age to individuals in their 20s or 30s. However, it should be considered that smartwatches are mainly used by this demographic. According to the NPD Connected Intelligence [28] in the US, the majority of wearable users are adults in the 25-34 age bracket. Therefore, we believe that our evaluation of these 20 young participants will prove meaningful.

Other Types of Smartwatches. Our method is designed to authenticate users by analyzing their responses to vibrations generated by a commercial smartwatch. We conducted our evaluations on the Apple Watch, but we neither used special functions exclusive to the Apple Watch nor did we physically modify the devices. Generally, other models of smartwatches from different companies commonly have vibration sensors, gyroscope sensors, and accelerometer sensors, and the specifications of these sensors in other models are comparable. For example, Fitbit [12] provides eight vibration types named "alert," "bump," "confirmation," "confirmation-max," "nudge," "nudge-max," "ping," and "ring" [14], and the sampling rate of the accelerometer sensors is 100Hz [13]. Furthermore, we evaluated our method using the our prototype as outlined in Section 6. As such, we expect that our method can be applied to smartwatches regardless of the model.

Combination of Metrics for Physical Composition. We evaluated our method regarding to the impersonation attack using similar physical indicators such as weight, height, body fat rate (BFR), or skeletal muscle rate (SMR). The results show that our method is capable of distinguishing between valid and invalid attempts, regardless of similarities in physical indicators. Accordingly, we expect our method to be able to authenticate users with low *overall* FPRs even if an attacker leverages an accomplice with multiple similar physical indicators as a target user. We plan to conduct further evaluation on this point in future research.

8 CONCLUSION

In this paper, we proposed a vibration-based user authentication method for smartwatches, which is based on a challenge-response structure that does not require user interaction. We evaluated our method on a commercial smartwatch, the Apple Watch Series 3, and default vibration types officially provided in the Apple Watch, which means no additional devices are required to authenticate users. In addition, our method produced a low EER of 0.0137 for first-measured signals and is robust against various attacks, including not-in-wear and impersonation attacks, particularly those involving participants with physical indicators similar to target users. We expect our method to be suitable for a wide variety of smartwatches on the market today.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government Ministry of Science and ICT (MSIT) under Grant NRF-2021R1A2C2014428.

REFERENCES

[1] Apple iphone 5s fingerprint sensor hacked. https://www.cbc.ca/news/technology/apple-iphone-5s-fingerprint-sensor-hacked-1.1864995 [Online; accessed 2020-07-30].
[2] Apple watch series 3. https://www.apple.com/kr/apple-watch-series-3/ [Online; accessed 2020-07-30].
[3] Apple's iphone faceID hacked in less than 120 seconds. https://www.forbes.com/sites/daveywinder/2019/08/10/apples-iphone-faceid-hacked-in-less-than-120-seconds/#3d55683b21bc [Online; accessed 2020-07-30].
[4] Black hat 2018: Voice authentication is broken, researchers say. https://threatpost.com/black-hat-2018-voice-authentication-is-broken-researchers-say/134926/ [Online; accessed 2020-07-30].
[5] Body fat rate. https://en.wikipedia.org/wiki/Body_fat_percentage [Online; accessed 2021-03-09].
[6] CCC members show iris recognition bypass using photo, contact lens. https://techxplore.com/news/2017-05-ccc-members-iris-recognition-bypass.html [Online; accessed 2020-07-30].
[7] A cheap 3D printer can trick smartphone fingerprint locks. https://www.wired.com/story/cheap-3d-printer-trick-smartphone-fingerprint-locks/ [Online; accessed 2020-07-30].
[8] Core motion. https://developer.apple.com/documentation/coremotion [Online; accessed 2020-07-30].

[9] ECG wearables: How they work and the best on the market. https://www.wareable.com/health-and-wellbeing/ecg-heart-rate-monitor-watch-guide-6508 [Online; accessed 2020-07-30].

[10] The extrasensory dataset. http://extrasensory.ucsd.edu/#tutorial [Online; accessed 2021-03-09].

[11] FiDELYS - the world's first iris recognition enabled smartwatch. https://www.youtube.com/watch?v=sw-OCo48F9s [Online; accessed 2020-07-30].

[12] Fitbit. https://www.fitbit.com/global/us/home [Online; accessed 2021-03-19].

[13] Fitbit accelerometer sensor guide. https://dev.fitbit.com/build/guides/sensors/accelerometer/ [Online; accessed 2021-03-19].

[14] Fitbit haptics api. https://dev.fitbit.com/build/reference/device-api/haptics/ [Online; accessed 2021-03-19].

[15] Hackers just broke the iPhone X's Face ID using a 3D-printed mask, note = https://www.wired.co.uk/article/hackers-trick-apple-iphone-x-face-id-3d-mask-security [online; accessed 2020-07-30].

[16] Hackers make a fake hand to beat vein authentication. https://www.vice.com/en_us/article/59v8dk/hackers-fake-hand-vein-authentication-biometrics-chaos-communication-congress [Online; accessed 2020-07-30].

[17] Haptics. https://developer.apple.com/design/human-interface-guidelines/watchos/interaction/haptics/ [Online; accessed 2020-07-30].

[18] A look at how easily 3D-printed heads can hack facial recognition. https://interestingengineering.com/a-look-at-how-easily-3d-printed-heads-can-hack-facial-recognition [Online; accessed 2020-07-30].

[19] LRA coin vibration motor. https://www.digikey.kr/ko/product-highlight/j/jinlong/z-axis-lra-coin-vibration-motor [Online; accessed 2020-07-30].

[20] Murata introduces piezo vibe for wearable device. https://www.murata.com/en-eu/products/info/mechatronics/actuator/2017/0321 [Online; accessed 2020-07-30].

[21] Nymi band. https://nymi.com/ [Online; accessed 2020-07-30].

[22] Python. https://www.python.org/ [Online; accessed 2020-07-30].

[23] Samsung patent illustrates continued work on under-display fingerprint scanning for future smartphones and galaxy watch. https://www.patentlymobile.com/2018/11/samsung-patent-illustrates-continued-work-on-under-display-fingerprint-scanning-for-future-smartphones-and-galaxy-watch.html [Online; accessed 2020-07-30].

[24] Samsung patents smartwatch with vein authentication. https://www.planetbiometrics.com/article-details/i/4121/desc/samsung-patents-smartwatch-with-vein-authentication/ [Online; accessed 2020-07-30].

[25] Samsung's galaxy S10 fingerprint sensor fooled by 3d printed fingerprint. https://www.theverge.com/2019/4/7/18299366/samsung-galaxy-s10-fingerprint-sensor-fooled-3d-printed-fingerprint [Online; accessed 2020-07-30].

[26] Skeletal muscle rate. https://en.wikipedia.org/wiki/Skeletal_muscle [Online; accessed 2021-03-09].

[27] Strategy analytics. https://www.displaydaily.com/article/press-releases/strategy-analytics-global-smartwatch-shipments-grow-20-percent-to-14-million-in-q1-2020 [Online; accessed 2020-07-30].

[28] U.S. wearable ownership by demographics. https://www.npd.com/wps/portal/npd/us/news/press-releases/2015/the-demographic-divide-fitness-trackers-and-smartwatches-attracting-very-different-segments-of-the-market-according-to-the-npd-group/ [Online; accessed 2021-03-19].

[29] We broke into a bunch of android phones with a 3d-printed head. https://www.forbes.com/sites/thomasbrewster/2018/12/13/we-broke-into-a-bunch-of-android-phones-with-a-3d-printed-head/#4bd67e6d1330 [Online; accessed 2020-07-30].

[30] Xcode. https://developer.apple.com/xcode [Online; accessed 2020-07-30].

[31] Rehan Akbani, Stephen Kwek, and Nathalie Japkowicz. Applying support vector machines to imbalanced datasets. In *European conference on machine learning*, pages 39–50. Springer, 2004.

[32] Laura Auria and Rouslan A Moro. Support vector machines (SVM) as a technique for solvency analysis. 2008.

[33] Lucas Ballard, Fabian Monrose, and Daniel P Lopresti. Biometric authentication revisited: Understanding the impact of wolves in sheep's clothing. In *USENIX Security Symposium*, 2006.

[34] Debnath Bhattacharyya, Rahul Ranjan, Farkhod Alisherov, Minkyu Choi, et al. Biometric authentication: A review. *International Journal of u-and e-Service, Science and Technology*, 2(3):13–28, 2009.

[35] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[36] Attaullah Buriro, Bruno Crispo, Mojtaba Eskandri, Sandeep Gupta, Athar Mahboob, and Rutger Van Acker. S nap a uth: a gesture-based unobtrusive smartwatch user authentication scheme. In *International Workshop on Emerging Technologies for Authorization and Authentication*, pages 30–37. Springer, 2018.

[37] Attaullah Buriro, Rutger Van Acker, Bruno Crispo, and Athar Mahboob. Airsign: a gesture-based smartwatch user authentication. In *2018 International Carnahan Conference on Security Technology (ICCST)*, pages 1–5. IEEE, 2018.

[38] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[39] Ren G Dong, Aaron W Schopper, TW McDowell, Daniel E Welcome, JZ Wu, W Paul Smutz, C Warren, and Subhash Rakheja. Vibration energy absorption (vea) in human fingers-hand-arm system. *Medical engineering & physics*, 26(6):483–492, 2004.

[40] Simon Eberz, Nicola Paoletti, Marc Roeschlin, Marta Kwiatkowska, I Martinovic, and A Patané. Broken hearted: How to attack ecg biometrics. 2017.

[41] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. Spoof attacks on gait authentication system. *IEEE Transactions on Information Forensics and Security*, 2(3):491–502, 2007.

[42] Vaishali Ganganwar. An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, 2(4):42–47, 2012.

[43] Ilias Giechaskiel and Kasper Rasmussen. Taxonomy and challenges of out-of-band signal injection attacks and defenses. *IEEE Communications Surveys & Tutorials*, 22(1):645–670, 2019.

[44] Anil K Jain and Karthik Nandakumar. Biometric authentication: System security and user privacy. *IEEE Computer*, 45(11):87–92, 2012.

[45] Andrew H Johnston and Gary M Weiss. Smartwatch-based biometric gait recognition. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE, 2015.

[46] Shehroz S Khan and Michael G Madden. A survey of recent trends in one class classification. In *Irish conference on artificial intelligence and cognitive science*, pages 188–197. Springer, 2009.

[47] Rinat Khusainov, Djamel Azzi, Ifeyinwa E Achumba, and Sebastian D Bersch. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors*, 13(10):12852–12902, 2013.

[48] Luka Knez, Janko Slavič, and Miha Boltežar. A sequential approach to the biodynamic modeling of a human finger. *Shock and Vibration*, 2017, 2017.

[49] Gierad Laput, Robert Xiao, and Chris Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 321–333, 2016.

[50] Jingjie Li, Kassem Fawaz, and Younghyun Kim. Velody: Nonlinear vibration challenge-response for resilient user authentication. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 1201–1213, 2019.

[51] Jian Liu, Chen Wang, Yingying Chen, and Nitesh Saxena. Vibwrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 73–87, 2017.

[52] Beth Logan et al. Mel frequency cepstral coefficients for music modeling. In *Ismir*, volume 270, pages 1–11, 2000.

[53] Chris Xiaoxuan Lu, Bowen Du, Xuan Kan, Hongkai Wen, Andrew Markham, and Niki Trigoni. Verinet: user verification on smartwatches via behavior biometrics. In *Proceedings of the First ACM Workshop on Mobile Crowdsensing Systems and Applications*, pages 68–73, 2017.

[54] Fiona Fui-Hoon Nah. A study on tolerable waiting time: how long are web users willing to wait? *Behaviour & Information Technology*, 23(3):153–163, 2004.

[55] Laurindo de Sousa Britto Neto, Vanessa Regina Margareth Lima Maike, Fernando Luiz Koch, Maria Cecília Calani Baranauskas, Anderson de Rezende Rocha, and Siome Klein Goldenstein. A wearable face recognition system built into a smartwatch and the blind and low vision users. In *International Conference on Enterprise Information Systems*, pages 515–528. Springer, 2015.

[56] Toan Nguyen and Nasir Memon. Tap-based user authentication for smartwatches. *Computers & Security*, 78:174–186, 2018.

[57] Toan Nguyen and Nasir D Memon. Smartwatches locking methods: A comparative study. In *SOUPS*, 2017.

[58] Harry Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

[59] Sébastien S Perrier, Yvan Champoux, and Jean-Marc Drouet. The influence of a human hand-arm system on the vibrational dynamic behaviour of a compliant mechanical structure. *Journal of Vibration and Control*, 23(2):329–342, 2017.

[60] Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3):614–634, 2001.

[61] Md Sahidullah and Goutam Saha. Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition. *Speech communication*, 54(4):543–565, 2012.

[62] Joo Yong Sim, Hyung Wook Noh, Woonhoe Goo, Namkeun Kim, Seung-Hoon Chae, and Chang-Geun Ahn. Identity recognition based on bioacoustics of human body. *IEEE Transactions on Cybernetics*, 2019.

[63] Amit Kumar Singh and Anand Mohan. *Handbook of Multimedia Information Security: Techniques and Applications*. Springer, 2019.

[64] Timothy Trippel, Ofir Weisse, Wenyuan Xu, Peter Honeyman, and Kevin Fu. Walnut: Waging doubt on the integrity of mems accelerometers with acoustic injection attacks. In *2017 IEEE European symposium on security and privacy (EuroS&P)*, pages 3–18. IEEE, 2017.

[65] Rens van de Schoot and Milica Miocević. *Small sample size solutions: A guide for applied researchers and practitioners*. Taylor & Francis, 2020.

[66] Fan Yang, Hua-zhen Wang, Hong Mi, Wei-wen Cai, et al. Using random forest for reliable classification and cost-sensitive learning for medical diagnosis. *BMC bioinformatics*, 10(S1):S22, 2009.

[67] Yue Zhao, Zhongtian Qiu, Yiqing Yang, Weiwei Li, and Mingming Fan. An empirical study of touch-based authentication methods on smartwatches. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 122–125, 2017.

APPENDIX

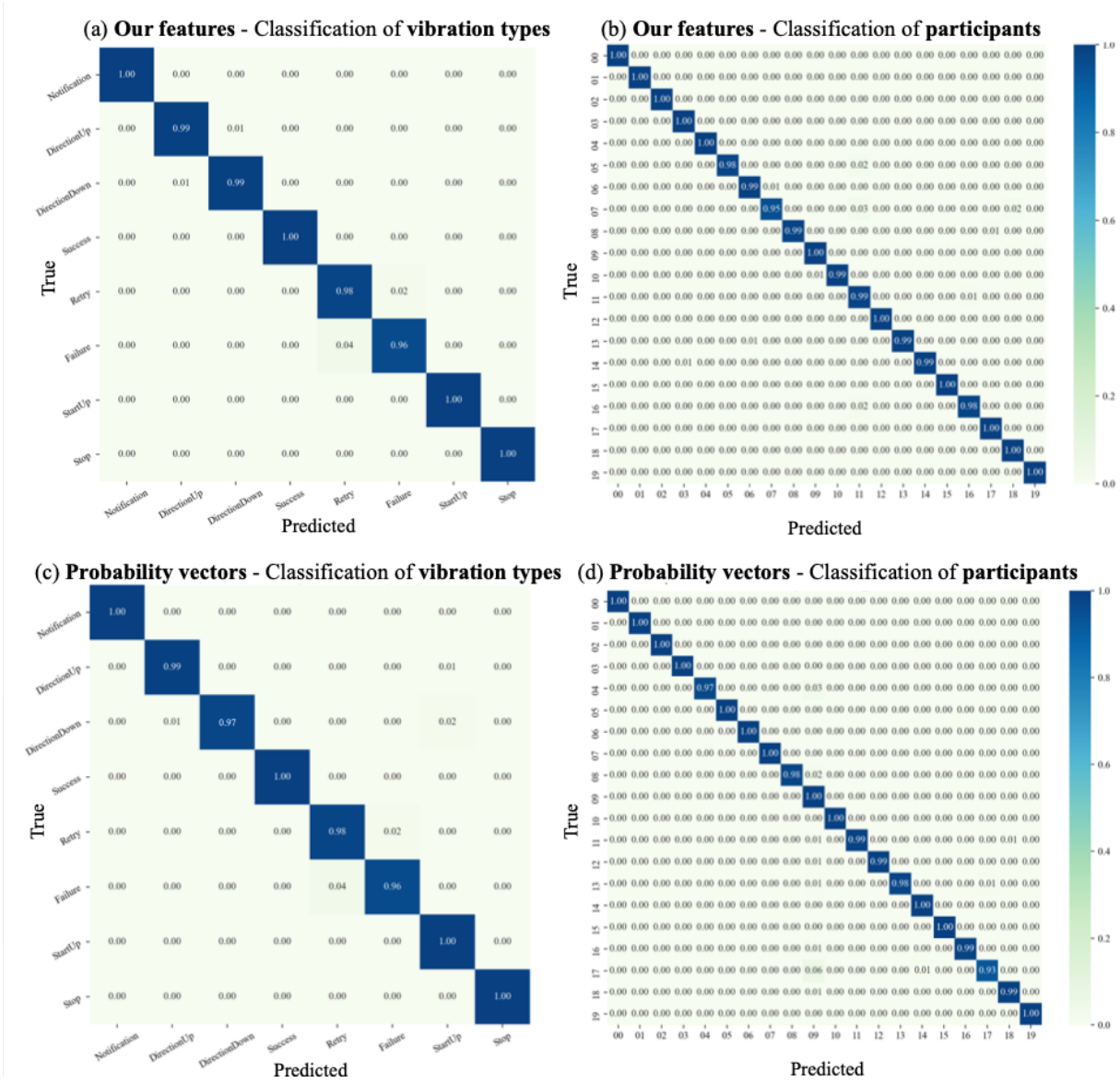

Figure A: Distinguishability of our features and probability vectors used in our method

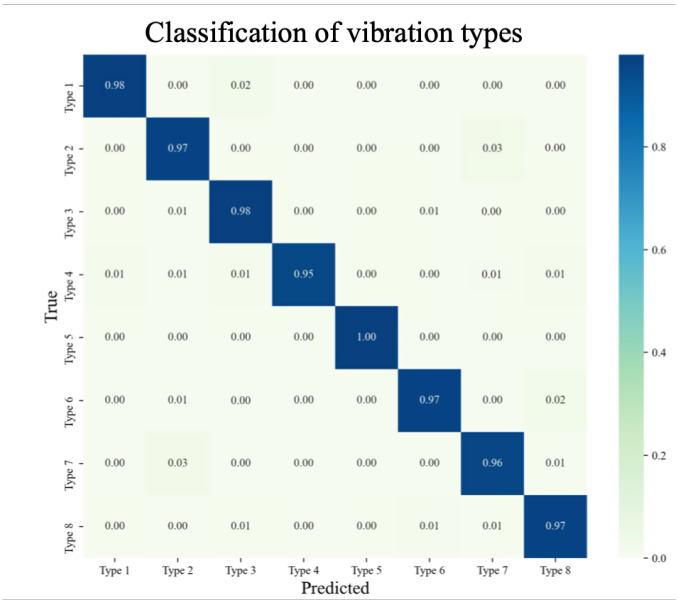

Figure B: Distinguishability of vibration types generated by prototype