Electronic Preprint for Journal of Information Processing Vol.29

Regular Paper

Person Identification Based on Accelerations Sensed in Smartphones with LSTM

Yoshihaya Takahashi[1,a)] Kosuke Nakamura[1] Takeshi Kamiyama[2]
Masato Oguchi[3] Saneyasu Yamaguchi[1]

Received: February 26, 2021, Accepted: July 8, 2021

Abstract: User identification is an important task for a variety of purposes such as authentication or providing personalized advice for improving user experience. In this paper, we propose a method for identifying a user who is holding a smartphone out of previously given target users from acceleration data obtained from the accelerometer in the smartphone using a deep neural network. This proposed method preliminary creates the model from the acceleration data of each user while walking in its training phase. This method identifies the user from acceleration data for identification based on this model in the classification phase. We evaluated the proposed method with the acceleration obtained from the actual eight and twelve users in two aspects, which were identifications including *no-decision* choice and that without *no-decision* choice. Our evaluation showed that the proposed method achieved accuracies higher than 95% for two- to five-class identification without *no-decision*. The proposed method identified the user with no or little false positive in evaluations with "*no-decision*."

Keywords: accelerometer, machine learning, LSTM, personal identification

1. Introduction

Recent advances in machine learning and deep learning have enabled us to infer many things, such as the activity of the user of the sensor device, from information obtained from sensors [1], [2]. We can expect that the improvement in identification and prediction based on sensed data contributes to many things. For example, activity recognition on a smartphone achieves detection of dangerous usage of the smartphone. Place estimation realizes an automatic explanation to the picture near the user in a museum. User identification achieves background music recommendation based on the favorite of the identified user, user authentication using sensed acceleration from the smartphone, and detection of theft.

The motivation and objectives of this paper are to improve the studies on analyses of sensed data from a smartphone, especially analyses based on artificial intelligence. For this abstract objective in this paper, we focus on user identification from acceleration by a deep neural network, which is one of the most promising artificial intelligent methods.

There are some published papers on identification based on accelerations obtained from an accelerometer. However, most of these works focus on the identification of other things such as activities. A few works focused on user identification. However, the existing methods and works have room for improvement. A work required to use a specialized floor that was full of accelerometers [3], [4]. A work identified users from only two users [5]. A work identified users from several users [6], [7], but the method did not take time sequence feature of acceleration and had room for improvement.

In this paper, we discuss the identification of the user who is holding a smartphone by analyzing 3-axis acceleration information obtained from the smartphone accelerometer. The purpose of this study is to predict users with high accuracy that is close to 100%. We assume the situation wherein the user is holding and watching the smartphone while walking. It is mainly because the noise in acceleration is small and this makes discussion fair by clear definition. We expect that the identification proposed in this paper enables us to improve the user experience in a variety of ways, such as recommending and supporting the use of the system based on user identification, user authentication, and admission management by walking. If the method achieves fairly high accuracy, this can detect theft by non-authorized users.

In our previous work [6], a method for identifying users from the acceleration using DNN (Deep Neural Network) was proposed. However, this did not take the time sequence feature into account. In this paper, we propose a method to identify users from acceleration measured in a smartphone using LSTM (Long Short-Term Memory), which is one of the most popular recurrent neural networks that can extract features from time sequence data. This method mainly supports identifications in a small preliminary given group such as identification in a specified group in a building. This assumes that the acceleration data for training is preliminary available and identifies the user from acceleration data for identification based on the created model. We evaluate the proposed method in two aspects. One is identification with-

[1] Kogakuin University, Shinjuku, Tokyo 163–8677, Japan
[2] Nagasaki University, Nagasaki 852–8521, Japan
[3] Ochanomizu University, Bunkyo, Tokyo 112–8610, Japan
[a)] J017174@g.kogakuin.jp

out the output of "*no-decision*." In this case, the proposed method always outputs the identified user as its identification result. The other is identification including output choice of "*no-decision*." In this case, the method outputs a user or "*no-decision*" as its result. If the neural network does have strong confidence in its output result, the network naturally outputs it. On the contrary, if the network does not have strong confidence, the network aims to avoid outputting an incorrect identification result and abandon to output the result. We call this abandoned "no-decision" in this paper. This is useful for avoiding the false positive and false negative. In this paper, we call these identifications *identification without no-decision* and *identification with no-decision*.

The contribution of this paper to practical uses is as follows. One of the typical applications of the proposed method is a weak user authentication system inside a secured area with a small number of persons. Figure 17 illustrates a sample application. This shows a building of a company and the entrance of the building is secured by a strong security system such as locking via a physical key. Inside the building, some rooms are protected by a weak security system. In a usual system without our proposed method, a user must touch an RFID card reader with the user's ID card, put a pass code, or do another authentication behavior. On the contrary, in a system with our proposed method, a user can be authorized automatically only with holding a smartphone without any operation.

The contribution to the research field is as follows. First, discussion on whether a user can be distinguished by its acceleration and how to identify, if possible, are contributing. Second, identification of users from acceleration may develop new fields. If identification or prediction with a moderate accuracy is achieved, many optional services such as the recommendation of background music based on this user prediction can be provided. If identification with an excellent accuracy is achieved, this identification can be utilized for many purposes such as detection of theft and user identification for unlocking smartphones. Our challenge for identification reveals how accurately the current deep learning method and the current accelerometer in common smartphones can identify the user. Our work showed that high accuracy can be achieved in identification from several persons, such as five persons. Our challenge also pointed out what method can improve the accuracy. We showed that a deep learning network considering time sequence such as LSTM achieved better accuracy than a method without considering this such as a simple DNN. We showed also that the accuracy can improve by increasing the size of input data. In the case of our experiments using the current smartphone accelerometer and a deep learning method, time sequence data with 40 seconds can be enough. In addition, our work shows room for improvement of this challenge. Our work demonstrated that the accuracies did depend on the target persons and those of particular persons were severely low. Thus, specified optimization for particular cases may be important for improving its accuracy.

This work is based on our earlier works [6], [7], [8], [9], [10]. For extension, we added some evaluations and discussions such as identification from three or more users.

2. Related Work

In this section, we refer to the related work.

The followings are on activity recognition or prediction by analyzing data sensed in smartphones using machine learning. Kobayashi et al. proposed a method for recognizing a user's movement by using an accelerometer, microphone, and GPS in a smartphone [1]. In this work, they discussed also a method for reducing power consumption. Song et al. proposed a method for recognizing human activity using a wearable sensor module including an acceleration sensor [2]. They especially discussed the identification of elderly people. Bouten et al. proposed a portable data unit that enables the on-line processing of triaxial accelerometer output to an estimator of physical activity [11]. Eren et al. proposed an approach to understanding driver behavior using smartphone sensors [12]. Using the accelerometer, gyroscope, and the magnetometer, they obtained its position, speed, acceleration, deceleration, and deflection angle sensory information and estimated commuting safety by statistically analyzing driver behavior. The following works are on activity recognition or prediction from data sensed from a sensor such as a smartphone accelerometer. Shoaib et al. surveyed physical activity recognizing methods using embedded sensors, showed their potential and the areas that need further improvements, and then presented various recommendations for conducting future studies on online activity recognition on mobile phones [13]. This survey paper presented excellent opinions, but the authors did not propose a method for identification. Lockhart et al. referred to and classified activity recognition applications. In the works of Refs. [14] and [15], methods based on LSTM were studied. Zeng et al. proposed a method for recognizing human activities based on LSTM, continuous temporal attention, and continuous sensor attention [14]. They showed that the combination of LSTM and Continuous temporal attention provided a fine F1 score [15]. Murahari et al. proposed a method for human activity recognition using the attention model [15]. They inserted an Attention layer to DeepConvLSTM and achieved a high F1 score. In the works of Refs. [16], [17], [18], [19], [20], methods for recognizing user activities based on CNN (Convolutional Neural Network) were proposed. Castro et al. proposed a method for predicting daily activities by analyzing images obtained from a camera device on the user [16]. The method analyzed images via CNN and classified the user's activities, such as chatting and working. Du et al. proposed a method for human activity recognition using cascade neural network [20].

Becker et al. proposed a system called GestEar that recognized activities making sounds, such as clapping or knocking, using CNN based on several sensed data for example acceleration, gyroscope data, and sound [18]. Brock et al. proposed a method for activity recognition in Ski Jumping based on CNN using data sensed from accelerometer and gyroscope [17]. Brunner et al. proposed a method for recognizing swimming styles, such as Crawl and Breaststroke, using a smartwatch [19]. They analyzed data from an accelerometer, gyroscope, magnetometer, barometer, and light sensor by CNN, and achieved a high F1 score. Khan et al. proposed a method for activity recognition using the embedded accelerometer sensor [21]. They argued that only a small number of previous works could perform classification on the device. For this problem, they proposed a real-time/online Smartphone-based activity recognition (SP-AR) system. Lockhart et al. surveyed and categorized a variety of activity recognition-based applications [22]. These works are for recognizing activities and did not provide a method for distinguishing persons who were performing the same activity. However, these pioneering works for recognizing something using data sensed from human activities, methods, and discussion that they showed have been remarkably important for studying another recognition. Especially, the works of Refs. [15] and [18] showed that analyzing the feature of sequence via LSTM is effective for accurate recognition. Our work is also based on these frontier works.

The following works are for user identification by analyzing sensed data with machine learning. Lane et al. comprehensively surveyed sensing technologies on smartphones [23]. They analyzed sensed data and found a large difference between data while walking of young end elderly persons. Shi et al. proposed the user identification framework SenGuard that enables continuous and implicit user identification service using voice, location, multitouch, and locomotion sensor modules [6], [24]. They showed that these four types of information are useful for identification. The work aims identification of a user, but accelerations were used for not user identification but activity recognition in work. Thus, this work is also on activity recognition in an aspect of acceleration analysis. It identified the user using other types sensed data such as voice data. Besides, their user identification was based on not recently advanced highly accurate methods such as deep learning and LSTM but simple statistical values. Their identification was only a two-class identification. Studies on more difficult classification are essential for the next step of researches. Ahmad et al. proposed a method for identification of a user from acceleration obtained from an accelerometer in a smartphone [5]. Their challenge was one of the easiest ones. They tried to identify a user from two target users. They assumed that sensed accelerations could be fingerprints of users and analyzed sensed accelerations using several machine learning methods. They then showed that the ANN (Artificial neural network) presented the highest accuracy. This work focused on the same identification of our work and their findings that a user could be identified with acceleration was remarkably important for the following similar works, e.g., our work. However, this challenge was very simple and identification from two users.

In our previous work, we proposed methods for estimating users' height using accelerometer [10]. These were not for user identification, but these based this work. We proposed a method for identifying the user from a sensed acceleration in Ref. [7]. In these works, we analyzed the accelerations obtained from the accelerometer in a smartphone while walking with deep learning and identified the user who was holding the smartphone. In the work, we challenged to identify the user from the given two users, i.e., two-class identification. The used one was not advanced deep neural networks such as CNN, RNN (Recurrent Neural Network), or LSTM but the original DNN. Namely, this method did not take the feature of time sequence into account. The accuracy of inference for two-class identification was 87% in the highest case. This work is based on and an extended version of these previous researches.

The works of Refs. [25], [26] are for predicting states. Schmidt et al. proposed a method for predicting stress and arousal by analyzing acceleration and skin-temperature by CNN [25]. Rabbi et al. proposed a method for predicting mental health and physical health by analyzing voice and physiological data by a dictions tree [26].

There are several published papers on identifying the person from sensed data from other types of sensors [3], [4], [27]. Corpus et al. proposed a method for accelerometer biometrics with keystroke dynamic features using a neural network and showed that the method could identify the user with high accuracy [27]. Orr et al. [3] and Sousa et al. [4] proposed methods for identifying users using sensing floors. A user for identifying walked on the specified floor that could sense accelerations and their methods identified based on these data. The method in Ref. [3] analyzed the data via ten statistical values such as the mean value and the standard deviation. The method in Ref. [4] utilized small user-bound pedometer devices and identified users in a situation wherein two or more individuals shared the same smart environment. These works provided important findings for analyzing sensed data for user identification. However, these methods required building a sensing floor that was full of accelerometers and could be utilized only on the floor. In addition, Sousa's method [4] could distinguish persons when plural persons were walking but could not detect who the person was.

3. Proposed Method

In this section, we propose a method for identifying the user based on the acceleration obtained from an accelerometer in a smartphone.

3.1 Two Proposed Methods

Here, we explain two methods for identifying the user from the preliminary given target users. These are identifications without and with *no-decision*.

First, we propose a method for identification without *no-decision*. This method is composed of three phases. In the first phase, the accelerations while walking of each user are preliminary measured. In the second phase, this method trains with these data and creates a neural network model by LSTM. In the third phase, this method identifies the user from the acceleration for identifications, which is the testing data, using the model. The user with the highest probability is determined as the identification result. The method always outputs the identification result. Each identification result is accurate or inaccurate.

Second, we propose a method for identification with *no-decision*. This second method is also composed of three phases. The first and second phases are the same as those of the first method. In the third phase, this method outputs *no-decision*, which indicates *no-decision* and *no output*, if the highest probability is less than the threshold. The probability is the output value of the softmax function of the last layer. If the probability is not less, this method outputs the identification result and the

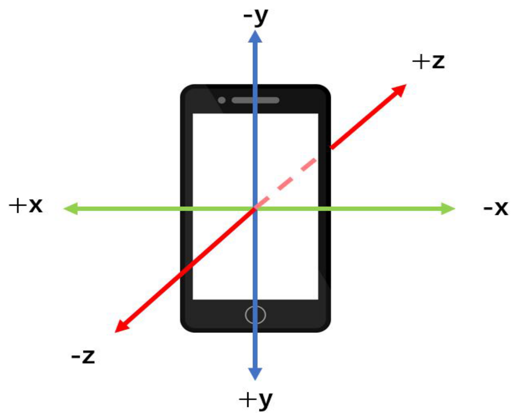

Fig. 1 Acceleration direction of smartphone.

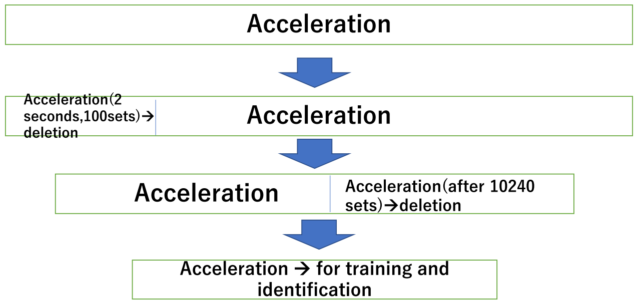

Fig. 2 Input data division method.

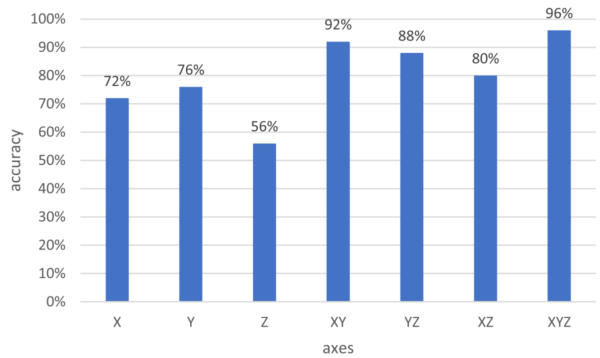

Fig. 3 Accuracies in each axis.

result is accurate or inaccurate. The used neural network will be explained in Section [6].

The first method is for the general purpose. The second method is for avoiding false positives in the cases false positives are a critical issue but *no-decision*, i.e., no prediction, is not.

3.2 Identification Target

In this work, we focus on the identification from given users. The proposed method is provided acceleration data of one user of these n users. We assume that the training acceleration data are available preliminary and the acceleration data are measured while walking.

3.3 Acquisition of Acceleration Data During Walking

The acceleration data while walking of each user were obtained as follows. We asked every user to walk in the target, which was specified, path withholding the smartphone in the user's hand and watching it. We told every user "please walk this path by watching this smartphone as usual." Our staff observed the walking user for the user's safety. The application for recording acceleration was displayed on the screen. This application had only the user interface for starting and stopping recording. We did not forbid touching the screen, but the users did nothing on the screen almost all the time.

3.4 Implementation for Recording Acceleration Data

We implemented an Android application that recorded the acceleration. We then installed this application into the smartphone for recording and asked each user to walk withholding the device. The application obtains acceleration using Google API every 20 ms and stores the data to a file. Three axes, which are x-, y-, and z-axes, of accelerations can be obtained with this API, and the application stores all of them. The directions of these three axes are illustrated in **Fig. 1**. The directions of positive and negative directions depend on the implementation of a device. In the case of our experimental device, the positive x-direction is the left. This may be right in another device. Our discussion in this paper is independent in this direction, i.e., left or right. The x-axis is the vertical axis regardless of the device implementation.

3.5 Input Data

The proposed method provides the constant length of the time sequence of acceleration data to the LSTM for training and classifying. Naturally, LSTM easily trains and classifies variable-length data. We fixed the length of input data for fair comparison and clear discussion. For fixing the length, which corresponds to the number of accelerations, we selected the acceleration data as illustrated in **Fig. 2**. That is, we deleted the first two seconds of data, which are 100 sets of accelerations, for avoiding including the accelerations before walking. We then deleted the last data to make the number of data sets 10,240.

3.6 Input and Output Data of the Neural Network

The proposed method identifies the user using on LSTM network. The number of dimensions of the input layer depends on the number of the axes of acceleration to be input. For example, if only the x-axis is used and input, the number of input data is one. If three dimensions are used and input, the number is three. The number of used dimensions in this paper will be discussed in Section 3.7.

The number of dimensions of the output layer is the number of candidate users. The output value of each dimension depicts the probability of each user.

3.7 The Axes in Acceleration for Input

In this subsection, we propose to use the three axes, i.e., x-, y-, and z-axis, for identification and explain the reason why this method uses these with a basic evaluation.

We performed five class identifications from five users. Namely, the system was provided acceleration data of one of the specified five users. The system then inferred which user's data it was. This system must choose one user from the five candidate users. We call this *five-class identification* in this paper.

Figure 3 shows the accuracies of five class identifications with all the combination of the axes. From these results, we can expect that identification using all three axes achieves the best identification accuracy. Therefore, the proposed method uses all the axes of the acceleration.

In our previous work [6], we preliminary investigated the performance of identification with all the combinations of three axes and chose the set with the highest accuracy. That investigation showed that the difference between the highest and lowest accuracy was around 40% [6]. We then consider that this preliminary investigation is important for achieving fine accuracy.

4. Performance Evaluation

This section evaluates the proposed method. For evaluation, eight users walked on a specified path holding an experimental smartphone and recorded the acceleration. We then picked 10,240 sets of accelerations data as shown in Fig. 2. One set of acceleration is an acceleration vector that is composed of three-dimensional values. For reference, we will present an experimental results using 12 users in Section 4.4.

The recording application used in our experiments obtains 50 acceleration vectors per second. 10,240 sets of accelerations correspond to 204.8 seconds. The eight target users, which are subjective of these experiments, are composed of seven males in his twenties and one male in his forties. We call them User A to User H, respectively.

The LSTM neural network has one hidden layer and the hidden layer has 100 neuron elements. The used device was the Nexus 5X. Its operating system was Android 7.1.1. The directions of the three axes are the same as those in Fig. 1. The used neural network was implementation with PyTorch 1.6. their hyperparameters are described in **Table 1**.

We performed cross-validation for evaluation. For n-clsss, $5 * n$-fold cross-validation was executed. For example, we performed a 10-fold cross-validation for two-class identification. For three, four, and five-class identifications, 15-, 20-, and 25-fold cross-validations were done, respectively.

4.1 Identification without *no-decision*

Here, we evaluate the accuracy of the proposed method without *no-decision*. We performed two- to eight-class identification using the measured accelerations from the eight users. For six- to eight-class identifications, we did cross-validations for every combination of $8Cn$ for n-class identifications. For example, we executed cross-validations with $8C7 = 8$ patterns of user combinations.

Figure 4 shows the average accuracies of two- to five-class identifications. The length of each input acceleration for identification was about 40 seconds, i.e., 2,048 sets of accelerations. These indicate that the accuracies were more than 95% in the cases of two- to five-class identifications. The existing method [3] achieved 87% in the best case, which was the two-class identification.

4.2 Accuracy and Input Data Length

This subsection investigates the relationship between the input data lengths and the accuracies.

Figures 5 to **10** show the accuracies using 10, 20, and 40 seconds of input data of five-class and eight-class identifications. The horizontal axis of each graph indicates the user or the average of all the users. These figures imply that the accuracy decreases

Table 1 Network hyperparameters.

Network hyperparameters	
Activation function	Softmax function
Error function	Cross entropy
Optimization function	Adagrad
Intermediate layer number of dimensions	100
Number of mesospheres	1

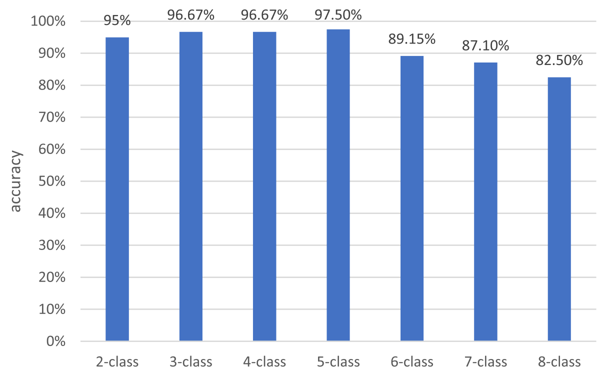

Fig. 4 Accuracies of 2–8 classes identifications.

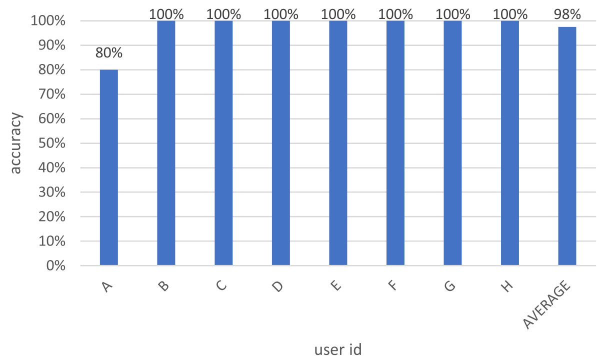

Fig. 5 Accuracies of five-class identification (input data size 40 seconds, without *no-decision*).

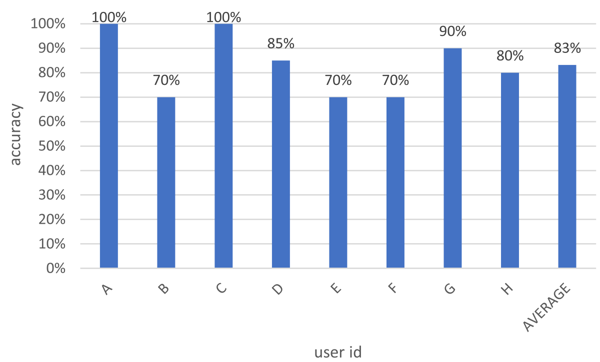

Fig. 6 Accuracies of five-class identification (input data size 20 seconds, without *no-decision*).

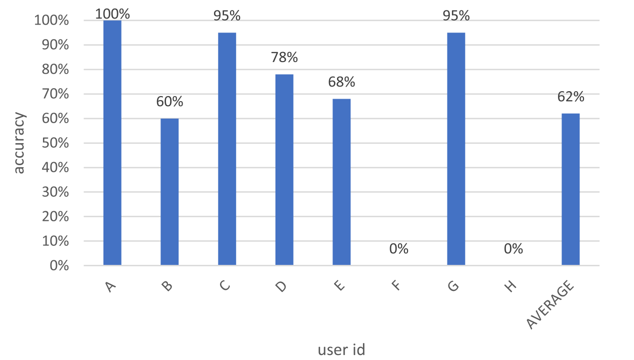

Fig. 7 Accuracies of five-class identification (input data size 10 seconds, without *no-decision*).

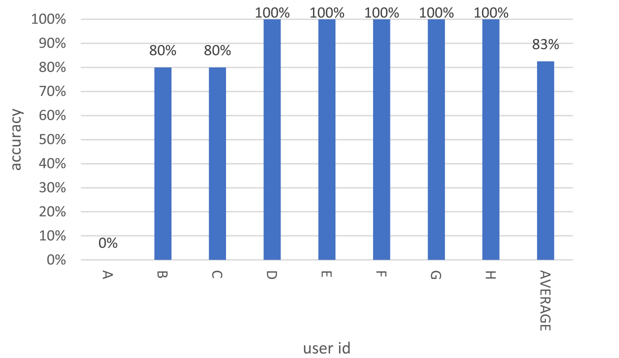

Fig. 8 Accuracies of eight-class identification (input data size 40 seconds, without *no-decision*).

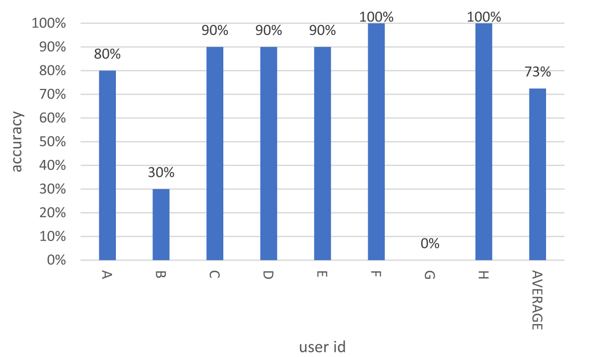

Fig. 9 Accuracies of eight-class identification (input data size 20 seconds, without *no-decision*).

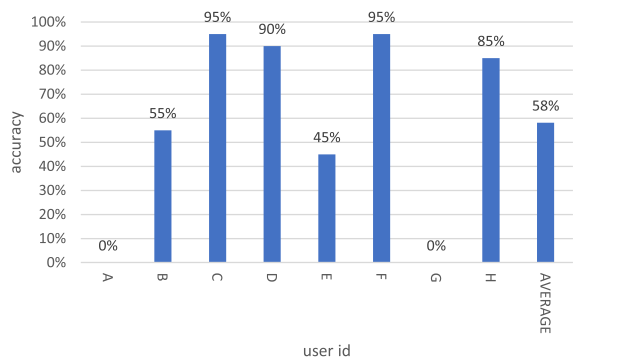

Fig. 10 Accuracies of eight-class identification (input data size 10 seconds, without *no-decision*).

as the length of input data decreases. Also, the accuracies of the particular users declined largely. In Fig. 5 to Fig. 7, the accuracies of the user B, E, F and H declined severely.

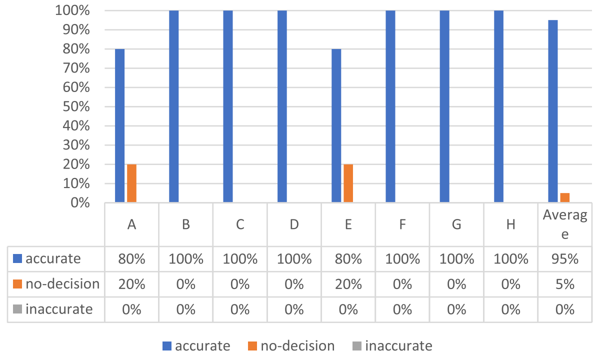

Fig. 11 Accuracies of five-class identification (input data size 40 seconds, with *no-decision*).

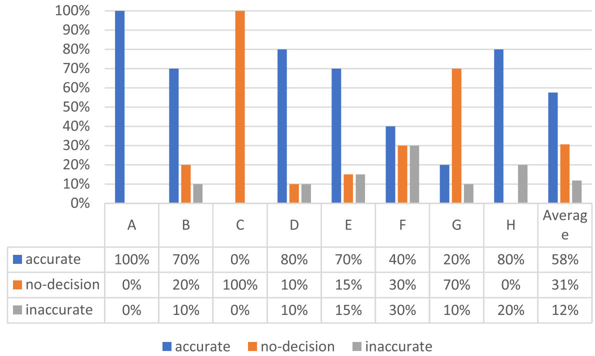

Fig. 12 Accuracies of five-class identification (input data size 20 seconds, with *no-decision*).

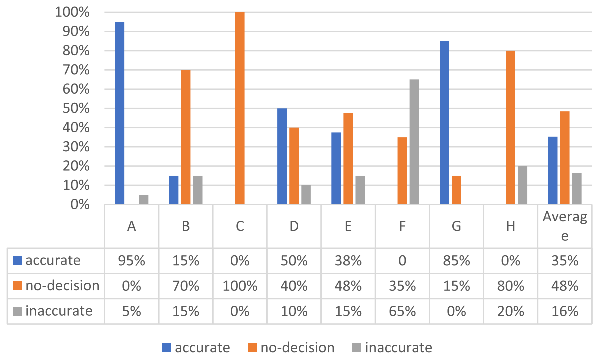

Fig. 13 Accuracies of five-class identification (input data size 10 seconds, with *no-decision*).

4.3 Identification with *no-decision*

This subsection evaluates the proposed method by identifications with *no-decision*. In these identifications, the method may output "*no-decision*" as its result for avoiding false positive and false negative. We expect that true positive and true negative are the best cases for users. On contrary, false positives and false negatives are the worst cases. "*No-decision*" is the middle case.

Figures 11 to **16** show the identification results of five- and eight-class identifications using 10, 20, and 40 seconds of input data. The identification can result in accurate, inaccurate, or *no-decision*. The threshold of *no-decision* or not was 0.9. Namely, the method outputs *no-decision* as the identification result if all the output values of the softmax function in the last layer are less than 0.9. Figure 11 indicates that the proposed method achieved 0% of the inaccurate identification ratio with a side effect of 20% of the *no-decision* ratio. In Fig. 5, the ratio of inaccurate identification was 2% and decreased 2% by including *no-decision*.

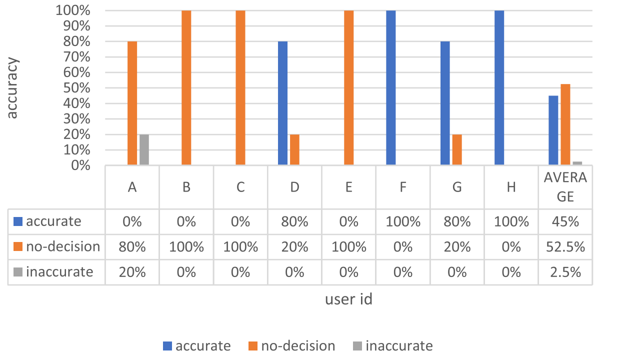

Fig. 14 Accuracies of eight-class identification (input data size 40 seconds, with *no-decision*).

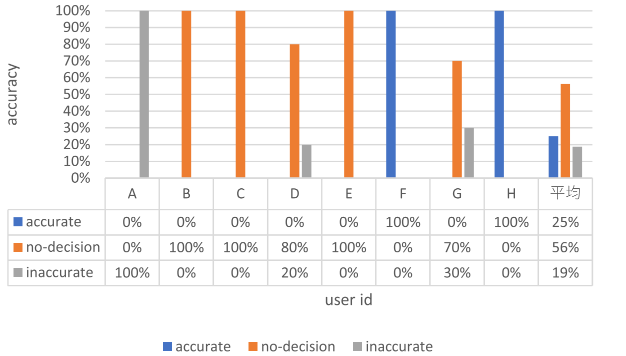

Fig. 15 Accuracies of eight-class identification accuracies (input data size 20 seconds, with *no-decision*).

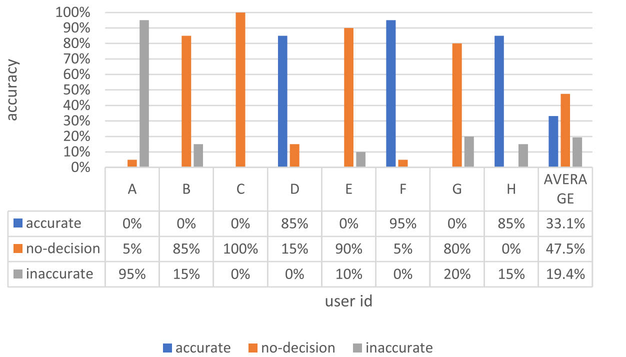

Fig. 16 Accuracies of eight-class identification (input data size 10 seconds, with *no-decision*).

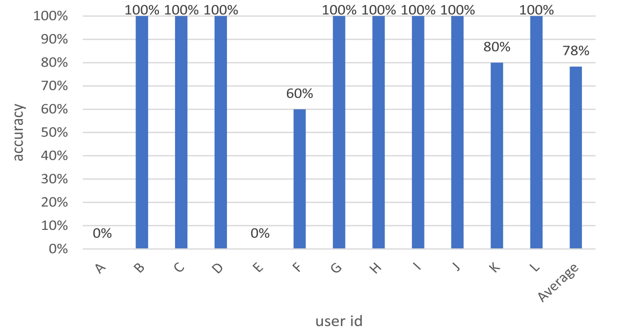

Fig. 17 Accuracies of twelve-class identification (input data size 40 seconds, without *no-decision*).

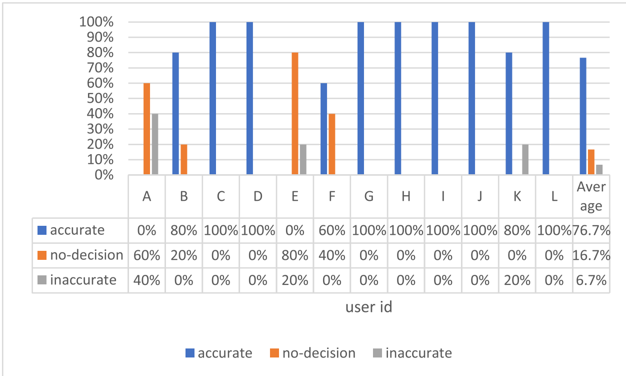

Fig. 18 Accuracies of twelve-class identification (input data size 40 seconds, with *no-decision*).

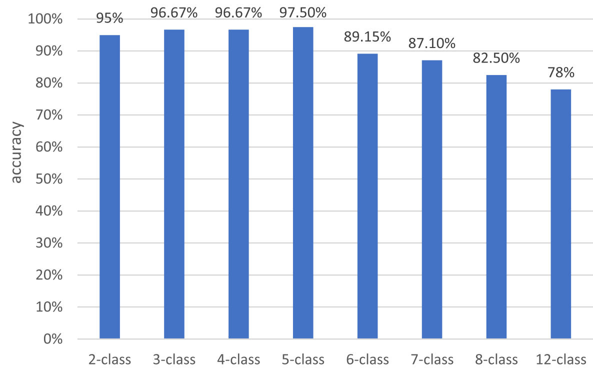

Fig. 19 Accuracies of 2-12 classes identifications.

Comparing the results in Fig. 6 and Fig. 12, the ratio of inaccurate identification decreased by 5%, i.e., from 17% to 12%. These imply that the ratio of inaccurate identification decreased successfully, but the ratio is not 0%. Similarly, we can see the 22% decrease in the ratio of inaccurate identification by comparing the results in Fig. 7 and Fig. 13. However, a 0% ratio was not achieved. Consequently, we can conclude that input data length enough to identify is essential for avoiding inaccurate identification. In the case of our experiments of five users, acceleration data longer than 40 seconds are required.

4.4 12-class Identification

For reference, we presented evaluations for twelve users in this subsection. The twelve users are composed of seven males in their twenties, three female in their twenties, and one male in his thirties. The twelve subjects and the path they walked on were different those of the experiments from Section 4.1 to Section 4.3.

The results with and without *no-decision* are depicted in **Fig. 17** and **Fig. 18**, respectively. **Figure 19** shows the relationship between the number of users for identification and the average accuracy. Please note that the experimental setups for two to eight users and for twelve users were not the same. Thus, direct comparison is not completely suitable and is only for reference.

The results without *no-decision* showed 78% of accuracy. This is not remarkably low for 12-class identification. However, unlike the five-class identification, the accuracy is less than 90%. From these results, we can find the limitation of our proposed method. Namely, the current smartphones and LSTM implementation cannot achieve 90% accuracy for this size of identification. In this experiment, the ratio of inaccurate identification was less than 10% like five-class identification.

5. Discussion

First, we discuss a method for improving the accuracy of identification without *no-decision*. Comparing the results in Fig. 5 to Fig. 7, we can see that the accuracy increases as the number of data increases. Naturally, we can expect that LSTM can more suitably extract the features of each user from larger acceleration data. One of the simplest ways to increase the number of data is to increase the number of obtained accelerations per second. Our application for recording obtained an acceleration vector every 20 ms. Our device support obtaining an acceleration vector every 2 ms. If the increase in the accuracy by increasing the length is mainly caused by an increase of not the measuring time but the number of accelerations, frequent acceleration measurement may increase this accuracy.

The tuning of the optimization function is expected to improve accuracy. In this paper, we used the Adagrad function for the evaluation. We have evaluated the accuracy also using the Adam function and found that the accuracy decreased by more than 15%. This implies that the accuracy is sensitive to the selection of the optimization function. An optimal selection of this function may improve the accuracy.

Second, we discuss the identification with *no-decision*. As we showed in Section 4, including *no-decision* decreased the ratio of 6inaccurate identification. Especially, in the case of 40 seconds of data, the ratio of inaccurate identification was 0%, and that of accurate identification also was high enough. However, in the cases of 10 and 20 seconds, the ratio of inaccurate identification was not 0%. If a user does desire to avoid a false positive, decreasing the threshold is effective. It was 0.9 in our experiment. If it is larger, e.g., 0.95, the method outputs its decision more pessimistically. Naturally, an increase in the threshold increases the ratio of *no-decision*. Thus, tuning of this threshold is important. The results in Fig. 13 showed that the *no-decision* ratios of users B, C, and E were high. **Figures 21** to **Fig. 23** show the output values of the softmax function of the last layer of the identifications that resulted in the inaccurate or *no-decision*. The vertical axes indicate the probability of each user. The horizontal axes indicate the sequence number of identifications that resulted in inaccurate or *no-decision*. In the case of Fig. 18, inaccurate identifications can be avoided with high accuracy by setting the threshold 0.6. On the contrary, the threshold must be high enough, such as 0.99, in order to completely avoid inaccurate identification.

Third, we discuss applications of our proposed identification.

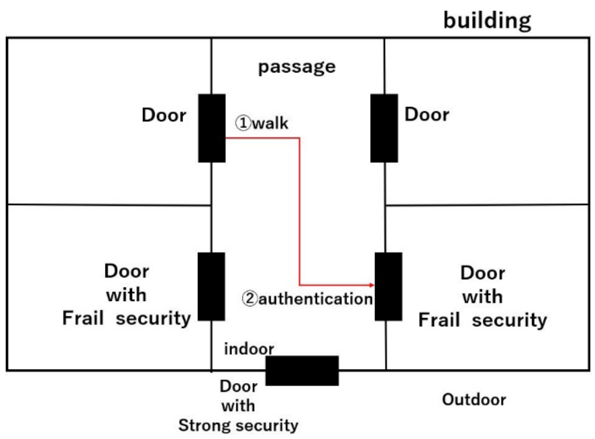

Fig. 20 Authentication system example.

Identification without *no-decision* is useful for improving users' experience. For example, automatically setting or recommending a background music or background colors based on personalized data can be comfortable for a user and an inaccurate identification does not raise a fatal issue. Identifications with *no-decision* can be used for authentication in a small group to which around five persons belong to. This method can be used for distinguishing whether the smartphone is held by the correct user or stolen by a malicious person. For example, notifying the correct user by sending an email on a detection of being stolen is easily performed because inaccurate detection does not result in a fatal problem. It causes only an unnecessary email.

This method enables authentication between these people from acceleration data by walk in a specified path in a specified area, such as in a building of their company. If the method outputs *no-decision*, the user can unlock a door with another method such as using a physical key or an IC card. If the method identifies the user, the user can enter the locked room only by walking. **Figure 20** shows an example of this system.

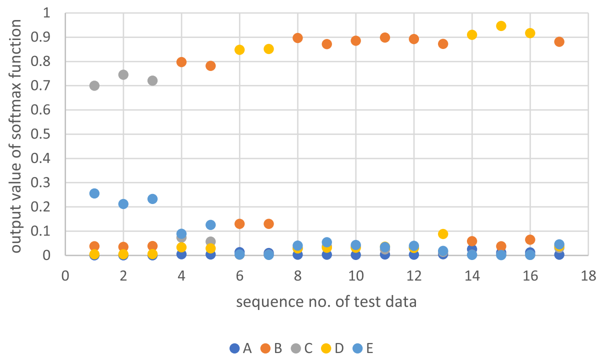

Fig. 21 Output probabilities of users (B is correct).

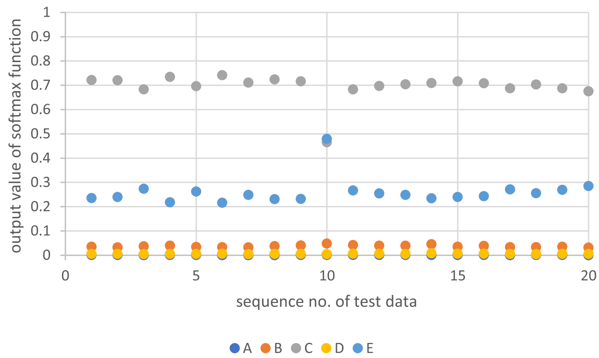

Fig. 22 Output probabilities of users (C is correct).

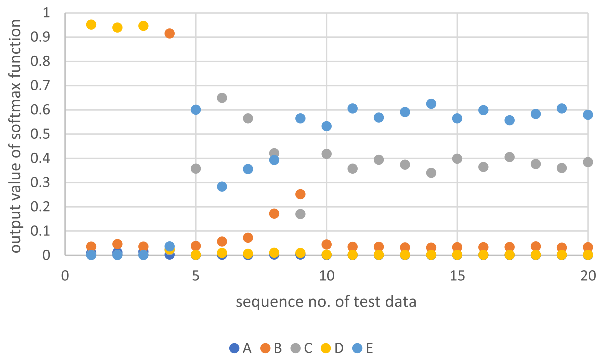

Fig. 23 Output probabilities of users (E is correct).

Third, we discuss the way of holding a smartphone. In this paper, we measured acceleration while walking because we expected that the noise in acceleration is relatively small. Evaluation with acceleration on another way of holding is important future work.

Forth, we discuss the effect of the features of subjects on the identification accuracy. The eight subjects of our experiments are composed of seven males in their twenties and one male in their forties. The twelve subjects are seven males in their twenties, four females for their twenties, and one male in their thirties. Their attributes and characteristics are be very similar. We can expect that this similarity made the identification difficult. If the attributes were completely different, users were expected to be identified more easily as described in Ref. [22]. Our paper showed that the current commodity smartphones and neural networks with a common PC can identify a user from specified eight persons with high accuracy. Also, this work showed also that these couldn't identify accurately from twelve users. Namely, our work showed the current state and provided future works, which is identification from more persons, e.g., more than ten persons. Please note that this discussion is based on comparing two experimental results with the different setups as described in Section 4.4.

Finally, we describe the reason why we focused on the LSTM and discussion on what type of neural network should we investigate. Deep neural networks and deep learning based on these networks have significantly improved since the first stage of the work [28]. In the field of analyzing non-time sequence data, CNN has been successfully processing data and has provided fine performances. In the time sequence data field, RNN (Recurrent neural network) was proposed and achieved good performances in the first stage. This RNN has been improved to LSTM, transformer [29], attention, self-attention, BERT [30]. Thus, we think investigating these advanced neural networks are important. Transformer, attention, self-attention, and BERT are mainly adopted in the natural language processing field. Therefore, we focused on and utilized LSTM in this paper. However, the other state-of-the-art neural network, e.g., attention, may effective for analyzing acceleration. Discussion on the usage of other networks is also important for future work.

6. Conclusion

In this paper, we proposed a method for identifying a user who is holding a smartphone by analyzing acceleration data obtained from the accelerometer using LSTM. Our evaluation showed that the method can identify the user with accuracies higher than 95% for two- to five-class identification. This may be useful for identification inside a small group. The ratio of inaccurate identification decreased low enough by including no-decision. For example, the ratio was 0% if the length of the input data was longer than 40 seconds.

For future work, we plan to evaluate our method with more peoples and larger class identification and discuss a method for improving accuracy by considering the content displayed on the smartphone and the user operations.

Acknowledgments This work was supported by JSPS KAKENHI Grant Numbers 17K00109, 18K11277.

This work was supported by JST CREST Grant Number JPMJCR1503, Japan.

References

[1] Kobayashi, A., Muramatsu, S., Kamisaka, D., Watanabe, T., Minamikawa, A., Iwamoto, T. and Yokoyama, H.: Shaka: User movement estimation considering reliability, power saving, and latency using mobile phone, *IEICE Trans. Information and Systems*, Vol.E94-D, No.6, pp.1153–1163 (2011).

[2] Song, S.K., Jang, J. and Park, S.: A Phone for Human Activity Recognition Using Triaxial Acceleration Sensor, *2008 Digest of Technical Papers – International Conference on Consumer Electronics*, pp.1–2 (online), DOI: 10.1109/ICCE.2008.4587903 (2008).

[3] Orr, R.J. and Abowd, G.D.: The smart floor: A mechanism for natural user identification and tracking, *CHI '00 Extended Abstracts on Human Factors in Computing Systems* (*CHI EA '00*), pp.275–276, Association for Computing Machinery (online), DOI: 10.1145/633292.633453 (2000).

[4] Sousa, M., Techmer, A., Steinhage, A., Lauterbach, C. and Lukowicz, P.: Human tracking and identification using a sensitive floor and wearable accelerometers, *IEEE International Conference on Pervasive Computing and Communications* (*PerCom*), pp.166–171 (online), DOI: 10.1109/PerCom.2013.6526728 (2013).

[5] Ahmad, M., Khan, A.M., Brown, J.A., Protasov, S. and Khattak, A.M.: Gait fingerprinting-based user identification on smartphones, *2016 International Joint Conference on Neural Networks* (*IJCNN*) (online), DOI: 10.1109/IJCNN.2016.7727588 (2016).

[6] Sato, Y., Kamiyama, T., Fukuda, A., Oguchi, M. and Yamaguchi, S.: A consideration of user estimation based on data of acceleration sensors in smartphone, *IPSJ SIG Technical Reports*, 2019-CDS-26, Vol.2, pp.1–6 (2019).

[7] Sato, Y., Kamiyama, T., Fukuda, A., Oguchi, M. and Yamaguchi, S.: Two-Class User Identification Using Accelerometer, *Proc. 81st National Convention of IPSJ*, 7W-08 (2019) (in Japanese).

[8] Sato, Y., Kurihara, S., Fukuda, S., Oguchi, M. and Yamaguchi, S.: Height Estimation based on Sensor Data on Smartphone, *Proc. 15th International Conference on Advances in Mobile Computing & Multimedia* (*MoMM2017*), pp.102–106, Association for Computing Machinery (online), DOI: 10.1145/3151848.3151877 (2017).

[9] Sato, Y., Kurihara, S., Fukuda, S., Oguchi, M. and Yamaguchi, S.: Height Estimation based on Sensor Data on Smartphone, *Proc. 15th International Conference on Advances in Mobile Computing & Multimedia* (*MoMM2017*), pp.102–106, Association for Computing Machinery (online), DOI: 10.1145/3151848.3151877 (2017).

[10] Sato, Y., Kamiyama, T., Fukuda, A., Oguchi, M. and Yamaguchi, S.: User's Height Estimation based on Acceleration from Smartphone Sensors, *2019 IEEE International Conference on Consumer Electronics – Taiwan* (*ICCE-TW*) (online), DOI: 10.1109/ICCE-TW46550.2019.8991753 (2019).

[11] Bouten, C.V.C., Koekkoek, K.T.M., Verduin, M., Kodde, R. and Janssen, J.D.: A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity, *IEEE Trans. Biomedical Engineering*, Vol.44, No.3, pp.136–147 (online), DOI: 10.1109/10.554760 (1997).

[12] Eren, H., Makinist, S., Akin, E. and Yilmaz, A.: Estimating driving behavior by a smartphone, *2012 IEEE Intelligent Vehicles Symposium*, pp.234–239 (online), DOI: 10.1109/IVS.2012.6232298 (2012).

[13] Shoaib, M., Bosch, S., Incel, O.D., Scholten, H. and Havinga, P.J.M.: A Survey of Online Activity Recognition Using Mobile Phones.Sensors 2015, pp.2059–2085 (online), DOI: 10.3390/s150102059 (2015).

[14] Zeng, M., Gao, H., Yu, T., Mengshoel, O.J., Langseth, H., Lane, I. and Liu, X.: Understanding and improving recurrent networks for human activity recognition by continuous attention, *Proc. 2018 ACM International Symposium on Wearable Computers* (*ISWC '18*), pp.56–63, Association for Computing Machinery (online), DOI: 10.1145/3267242.3267286 (2018).

[15] Murahari, V.S. and Plötz, T.: On attention models for human activity recognition, *Proc. 2018 ACM International Symposium on Wearable Computers* (*ISWC '18*), pp.100–103, Association for Computing Machinery (online), DOI: 10.1145/3267242.3267287 (2018).

[16] Castro, D., Hickson, S., Bettadapura, V., Thomaz, E., Abowd, G., Christensen, H. and Essa, I.: Predicting daily activities from egocentric images using deep learning, *Proc. 2015 ACM International Symposium on Wearable Computers* (*ISWC '15*), pp.75–82, Association for Computing Machinery (online), DOI: 10.1145/2802083.2808398 (2015).

[17] Brock, H., Ohgi, Y. and Lee, J.: Learning to judge like a human: Convolutional networks for classification of ski jumping errors, *Proc. 2017 ACM International Symposium on Wearable Computers* (*ISWC '17*),

pp.106–113, Association for Computing Machinery (online), DOI: 10.1145/3123021.3123038 (2017).

[18] Becker, V., Fessler, L. and Sörös, G.: GestEar: Combining audio and motion sensing for gesture recognition on smartwatches, *Proc. 23rd International Symposium on Wearable Computers* (*ISWC '19*), pp.10–19, Association for Computing Machinery (online), DOI: 10.1145/3341163.3347735 (2019).

[19] Brunner, G., Melnyk, D., Sigfússon, B. and Wattenhofer, R.: Swimming style recognition and lap counting using a smartwatch and deep learning, *Proc. 23rd International Symposium on Wearable Computers* (*ISWC '19*), pp.23–31, Association for Computing Machinery (online), DOI: 10.1145/3341163.3347719 (2019).

[20] Du, X., Farrahi, K. and Niranjan, M.: Transfer learning across human activities using a cascade neural network architecture, *Proc. 23rd International Symposium on Wearable Computers* (*ISWC '19*), pp.35–44, Association for Computing Machinery (online), DOI: 10.1145/3341163.3347730 (2019).

[21] Khan, A.M., Siddiqi, M.H. and Lee, S.-W.: Exploratory Data Analysis of Acceleration Signals to Select Light-Weight and Accurate Features for Real-Time Activity Recognition on Smartphones, *Sensors*, Vol.13, pp.13099–13122 (online), DOI: 10.3390/s131013099 (2013).

[22] Lockhart, J.W., Pulickal, T. and Weiss, G.M.: Applications of mobile activity recognition, *Proc. 2012 ACM Conference on Ubiquitous Computing* (*UbiComp '12*), pp.1054–1058, Association for Computing Machinery (online), DOI: 10.1145/2370216.2370441 (2012).

[23] Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T. and Campbell, A.T.: A survey of mobile phone sensing, *IEEE Communications Magazine*, Vol.48, No.9, pp.140–150 (online), DOI: 10.1109/MCOM.2010.5560598 (2010).

[24] Shi, W., Yang, J., Jiang, Y., Yang, F. and Xiong, Y.: SenGuard: Passive user identification on smartphones using multiple sensors, *2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications* (*WiMob*), pp.141–148 (online), DOI: 10.1109/WiMOB.2011.6085412 (2011).

[25] Schmidt, P., Dürichen, R., Reiss, A., Van Laerhoven, K. and Plötz, T.: Multi-target affect detection in the wild: an exploratory study, *Proc. 23rd International Symposium on Wearable Computers* (*ISWC '19*), pp.211–219, Association for Computing Machinery (online), DOI: 10.1145/3341163.3347741 (2019).

[26] Rabbi, M., Ali, S., Choudhury, T. and Berke, E.: Passive and In-Situ assessment of mental and physical well-being using mobile sensors, *Proc. 13th International Conference on Ubiquitous Computing* (*UbiComp '11*), pp.385–394, Association for Computing Machinery (online), DOI: 10.1145/2030112.2030164 (2011).

[27] Corpus, K.R., Gonzales, R.J.D.L., Morada, A.S. and Vea, L.A.: Mobile User Identification through Authentication Using Keystroke Dynamics and Accelerometer Biometrics, *IEEE/ACM International Conference on Mobile Software Engineering and Systems* (*MOBILESoft*), pp.11–12 (online), DOI: 10.1109/MobileSoft.2016.015 (2016).

[28] LeCun, Y., Bengio, Y. and Hinton, G.: Deep learning, *Nature*, Vol.521, pp.436–444 (online), DOI: 10.1038/nature14539 (2015).

[29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I.: Attention is All You Need (2017), available from ⟨https://arxiv.org/pdf/1706.03762.pdf⟩.

[30] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

Yoshihaya Takahashi received his B.E. degrees in Engineering Kogakuin University in 2020. He now stays in Kogakuin University to study electrical engineering and electronics.

Kosuke Nakamura received his B.E. degrees in Engineering Kogakuin University in 2020.

Takeshi Kamiyama joined Nagasaki University in 2020 as an associate professor. His research interests include mobile computing and distributed system based on information and data sciences. Before Nagasaki University, he received Ph.D. in Kyushu University, Japan in 2018. Also, he was co-founder and CEO in e-jis, Inc. from 2003 to 2006 and a researcher in NTT DOCOMO, Inc. from 2006 to 2020. He is a member of IPSJ.

Masato Oguchi received Ph.D. from the University of Tokyo in 1995. He was a researcher in the National Center for science Information System (NACSIS) – currently known as National Institute of Informatics (NII) in 1995. From 1996 to 2000, he was a research fellow at the Institute of Industrial Science, University of Tokyo. He stayed at Aachen University of Technology in Germany in 1998–2000. In 2001, he became an associate professor in Chuo University. He has joined Ochanomizu University in 2003 as an associate professor. Since 2006, he has been a professor at the Department of Information Sciences, Ochanomizu University. His research field is network computing, including high performance computing, mobile networking, big data, and network security. He is a member of IEEE, ACM, IEICE, IPSJ and DBSJ.

Saneyasu Yamaguchi received Engineering Doctor's degree (Ph.D.) at Tokyo University in 2002. During 2002–2006, he stayed in Institute of Industrial Science, the University of Tokyo to study I/O processing. He now with Kogakuin University. Currently his researches focus on operating systems, virtualized systems, and storage system.