Smartwatch-Based Face-Touch Prediction Using Deep Representational Learning

Hamada Rizk[1,2], Tatsuya Amano[1](✉), Hirozumi Yamaguchi[1], and Moustafa Youssef[3]

[1] Osaka University, Suita, Japan
{hamada_rizk,t-amano,h-yamagu}@ist.osaka-u.ac.jp
[2] Tanta University, Tanta, Egypt
[3] Alexandria University, Alexandria, Egypt
moustafa@alexu.edu.eg

Abstract. World Health Organization (WHO) reported that viruses, including COVID-19, can be transmitted by touching the face with contaminated hands and advised people to avoid touching their face, especially the mouth, nose, and eyes. However, according to recent studies, people touch their faces unconsciously in their daily lives, and it is difficult to avoid such activities. Although many activity recognition methods have been proposed over the years, none of them target the prediction of face-touch (rather than detection) with other daily life activities. To address to problem, we propose *TouchAlert*: a system that automatically predict the occurrence of face-touch activity and warn the user before its occurrence. Specifically, *TouchAlert* utilizes commodity wearable devices' sensors to train a deep learning-based model for predicting the variable length face-touching of different users at an early stage of its occurrence. Our experimental results show high accuracy of F1-score of 0.98 and prediction accuracy of 97.9%.

Keywords: COVID-19 · Face touch · Activity recognition · Smartwatch

1 Introduction

The COVID-19 pandemic has impacted all spheres of our life. Infectious exposures of COVID-19 mainly occur in inhalation of virus-containing droplets into their lungs or deposition of the virus to exposed mucous membranes in the nose, mouth, or eyes. Experts said it is like any novel disease, but the critical threat lies in rapidly spreading infecting millions in a single week.

Touching the face with hands soiled is one of the primary ways of the deposition of exposed mucous membranes. Thus, World Health Organization (WHO) urges people to avoid touching their eyes, nose, and mouth to prevent infection not only from COVID-19 but also from other viral diseases such as seasonal

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 493–499, 2022.
https://doi.org/10.1007/978-3-030-94822-1_29

influenza and Ebola virus. However, it was not possible for WHO representatives and specialists to avoid touching their faces.[1] The hand-to-face contact rate for normal people is as high as 10–26 times per hour [6]; this somehow justifies the exponential outbreak of the virus [19]. One study in Germany confirms that people are typically unaware that they are touching their faces, and they unconsciously perform this activity more frequently when they are stressed [4]. In this vein, face-touch is an instinctive activity and a hard habit to break. This motivates us to build a system for helping people **automatically detect and avoid the face-touch activity apriori** and thus reduce the possible infection transmission.

Deep learning has defined the state of-the-art performance in many application domains [1–3,9–17]. Specifically, Deep learning has enabled highly accurate human activity recognition (HAR) using inertial sensor data from widely used wearable devices. HAR is the problem of classifying sequences of multitudes of low-level sensor measurements (e.g., accelerometers and gyroscopes) into predefined movements. Current HAR techniques [8,18] involve walking, jogging, sitting, standing, etc. However, the patterns of these activities are different from the aimed face-touch one; therefore, the direct application of these techniques does set true for identifying face-touch.

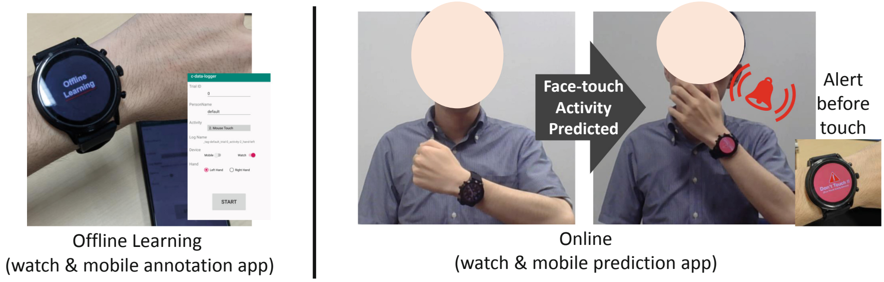

Fig. 1. Smartwatch-based face-touch prediction system

In this paper, we propose *TouchAlert*: a deep-learning-based *face-touch* prediction system to learn the non-linear relation between the sensor measurements captured by wearable devices and the activity in question (Fig. 1). To achieve highly accurate and robust activity detection, we leverage the LSTM autoencoder model to extract subject-invariant features and introduce model regularization techniques to avoid over-fitting. We evaluate the *TouchAlert* system using different smartwatches worn by *ten* different subjects practicing face-touch during pursuing different daily activities. The results show that *TouchAlert* can achieve accuracy and F1-score of 97.9% and 0.98, respectively. This is better than the state-of-the-art techniques in all scenarios.

[1] Even experts cannot avoid face touch activity: https://youtu.be/mA1wqjaeKj0.

2 Related Work

The system in [22] proposed a near-field communication (NFC)-based method for detecting the face-touch activity and warn users when the activity occurs. This method requires the user to wear an NFC reader and NFC tag in his/her wrist and ear, respectively. This design may not be convenient for all users as well as its success rate is not as high as the compared schemes. Few systems [7,21] have been proposed to detect the face-touch activity based on measurements captured by inertial sensors, including accelerometer and gyroscope. These systems either use hand-crafted features (e.g., mean and standard deviation) or neural network-based extracted features. Despite the simplicity of this technique, hand-crafted features cannot compete favorably in scenarios of overlapping activities as the overlapped activities (e.g., walking) will have a dominant sensory effect. Moreover, both techniques consider fixing length input leading to an inconsistent performance when tested with the variable-length activity of different subjects.

In contrast, TouchAlert, leverages a LSTM-autoencoder to automatically extract activity discriminative features from the variable length input stream. Therefore outperforming other schemes and boosts the feasibly and the required safety of this technology.

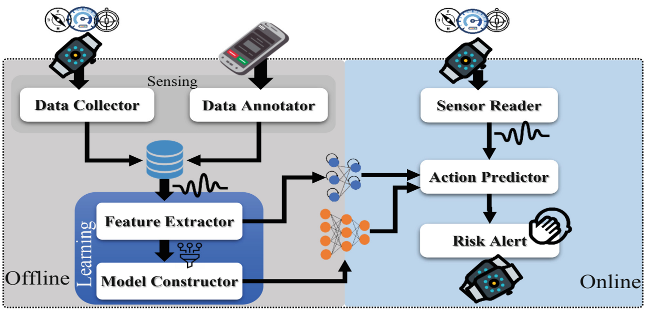

Fig. 2. *TouchAlert* system architecture.

3 System Overview

Figure 2 shows the overall system architecture. *TouchAlert* works in two stages: an offline training stage and an online sensing stage. *TouchAlert* initializes the offline stage by obtaining measurements of the smartwatch's on-board sensors, including both the accelerometer and gyroscope. These measurements represent 3-axial values from each sensor representing the acceleration and rotation along the accelerometer and gyroscope axis, respectively. This is accomplished by the Data Collector service running on the user's smartwatch. To facilitate ground-truth profiling, the Data Annotator app is installed on the connected mobile device. Next, the **Feature Extractor** module automatically learns to extract the main features from the low-level sensor measurements through the use of a deep model. These features are forwarded to the **Activity Detection Constructor** module to optimally build and train a deep model to identify the face touch activity. Finally, the trained models (i.e., *Feature Extraction* and *Activity Detection*) are stored for later use in the online phase.

During the online phase, the sensor measurements are forwarded to the online *Feature Extractor* module. And then, the **Action Predictor** module feeds these temporal features to the deep model constructed by *activity detection constructor* module to estimate the user is going to touch her face (and alert her) or not.

4 The *TouchAlert* System

This section describes the details of the *TouchAlert* system.

4.1 *TouchAlert* Sensing Part

Data is captured with two connected applications; the first is a Wear OS application running on the Android Watch and the other is an Android application installed on the user's smartphone. The watch application continuously collects 3-axis gyroscope and 3-axis accelerometer values from the IMU sensors. The smartphone application is leveraged to label the collected measurements. Both applications are time-synchronized and record the timestamps of the data collection. The data collected by the watch is sent to the smartphone via Bluetooth and processed on it.

4.2 *TouchAlert* Learning Part

The learning part consists of the feature extractor and the model constructor.

The Feature Extractor. The goal of this module is to obtain a fixed-length feature representation from a stream of sensor observations of variable length activities. Therefore, identifying the face-touch activity performed by different users and thus with different behaviors can be facilitated. Towards achieving this, *TouchAlert* has to learn the complex dynamics of input sequences as well as use internal memory to capture the temporal correlation across long and short input sequences. We employ a RNN version of autoencoder called LSTM Autoencoder [20], which is an implementation of an autoencoder for sequence data using stacked LSTM layers for feature extraction. This architecture has shown superior performance in many applications in other domains e.g., [11,12] which suggests leveraging it to automatically extract features of the face-touch activity.

The Classification Model Constructor. This module is responsible for leveraging the extracted features to train a classification model and find its optimal parameters. We construct a deep, fully connected neural network consisting of cascaded hidden layers of nonlinear processing neurons. The output layer consists of a single neuron corresponding to the activity of interest (face-touch). This network is trained to operate as a binary classifier (logistic regressor) by leveraging a Sigmoid activation function in the output layer. The model is trained using the Adaptive Moment Estimation (Adam optimizer [5]) to minimize the average cross-entropy between the estimated output probability distribution and the ground truth. The loss function is selected as binary cross-entropy.

The Online Action Predictor. This phase aims to predict the *face-touch* activity in real-time before the user's hand reaches her face. This can be done by processing the recorded sensor measurements and extracting the corresponding feature vector as described in Sect. 4.2. Thereafter, this vector is then fed to the trained *Activity Detection* model to get early recognition of either the activity is face-touch or not. The face-touch activity is detected if the output probability is higher than 0.5.

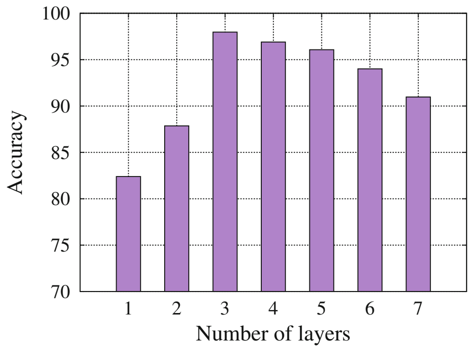

Fig. 3. Effect of number of layers

We evaluate the *TouchAlert* system using different smartwatches FOSSIL Carlyle HR or CASIO Protrek worn by *ten* different users practicing face-touch during pursuing different daily activities. The sampling rate of sensor data 100 Hz. Each face-touching activity lasted about 10–20 s and consisted of three steps: setting the hand away from the face, bringing the hand (watch) extremely close to the face, and returning the hand to its position. Users collected 20 face-touch samples for each of their arms and collected long-term watch sensor data in their daily lives, not including face touches. We used these data to evaluate the accuracy of face touch detection.

Figure 3 shows the effect of changing the number of layers on *TouchAlert* accuracy. The figure shows how increasing the number of layers of the location estimation network increases the location estimation accuracy until it reaches an optimal value at *three* layers. This can be justified as increasing the number of layers increases the model's computing power to avoid underfitting and thus better fit the function. However, the deeper the model, the more likely it is to overfit the training data, reducing its flexibility and accuracy when handling users from different providers. It's worth noting that applying the LSTM-autoencoder simplifies the classification problem in the latent encoded space. As a result, a three-layer network is sufficient for the classification of features.

5 Conclusion

This paper proposed a face touch detection system using acceleration and gyro sensor data from wearable devices. To extract robust features from time-series data of variable-length activity by different subjects, we employed an LSTM autoencoder. Our evaluation results show that the face-touch detection method can obtain 97.9% prediction accuracy with an F1-score of 0.98. These results highlight the feasibility of the proposed system for boosting public safety.

Acknowledgement. The work was supported by "Research and Development of Information and Communication Technologies that Contribute to Countermeasures against Infectious Diseases (222-C03)", the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN.

References

1. Abbas, M., Elhamshary, M., Rizk, H., Torki, M., Youssef, M.: WiDeep: WiFi-based accurate and robust indoor localization system using deep learning. In: Proceedings of the International Conference on Pervasive Computing and Communications (PerCom). IEEE (2019)
2. Alkiek, K., Othman, A., Rizk, H., Youssef, M.: Deep learning-based floor prediction using cell network information. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, pp. 663–664 (2020)
3. Amano, T., Yamaguchi, H., Higashino, T.: Connected AR for combating COVID-19. IEEE Internet Things Mag. **3**(3), 46–51 (2020)
4. Grunwald, M., Weiss, T., Mueller, S., Rall, L.: EEG changes caused by spontaneous facial self-touch may represent emotion regulating processes and working memory maintenance. Brain Res. **1557**, 111–126 (2014)
5. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
6. Kwok, Y.L.A., Gralton, J., McLaws, M.L.: Face touching: a frequent habit that has implications for hand hygiene. Am. J. Infect. Control **43**(2), 112–114 (2015)
7. Michelin, A.M., et al.: FaceGuard: a wearable system to avoid face touching. Front. Robot. AI **8**, 47 (2021)
8. Radu, V., et al.: Multimodal deep learning for activity and context recognition. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 4, pp. 1–27 (2018)

9. Rizk, H.: Device-invariant cellular-based indoor localization system using deep learning. In: The ACM MobiSys 2019 on Rising Stars Forum, pp. 19–23. ACM (2019)
10. Rizk, H.: SoloCell: efficient indoor localization based on limited cell network information and minimal fingerprinting. In: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 604–605 (2019)
11. Rizk, H., Abbas, M., Youssef, M.: OmniCells: cross-device cellular-based indoor location tracking using deep neural networks. In: 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–10. IEEE (2020)
12. Rizk, H., Abbas, M., Youssef, M.: Device-independent cellular-based indoor location tracking using deep learning. Pervasive Mob. Comput. **75**, 101420 (2021)
13. Rizk, H., Shokry, A., Youssef, M.: Effectiveness of data augmentation in cellular-based localization using deep learning. In: Proceedings of the International Conference on Wireless Communications and Networking (WCNC). IEEE (2019)
14. Rizk, H., Torki, M., Youssef, M.: CellinDeep: robust and accurate cellular-based indoor localization via deep learning. IEEE Sens. J. **19**, 2305–2312 (2018)
15. Rizk, H., Yamaguchi, H., Higashino, T., Youssef, M.: A ubiquitous and accurate floor estimation system using deep representational learning. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, pp. 540–549 (2020)
16. Rizk, H., Yamaguchi, H., Youssef, M., Higashino, T.: Gain without pain: enabling fingerprinting-based indoor localization using tracking scanners. In: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, SIGSPATIAL'20, pp. 550–559. Association for Computing Machinery, New York (2020)
17. Rizk, H., Youssef, M.: MonoDCell: a ubiquitous and low-overhead deep learning-based indoor localization with limited cellular information. In: Proceedings of the 27th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM (2019)
18. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst. Appl. **59**, 235–244 (2016)
19. Shen, K., et al.: Diagnosis, treatment, and prevention of 2019 novel coronavirus infection in children: experts' consensus statement. World J. Pediatr. **16**(3), 223–231 (2020)
20. Srivastava, N., Mansimov, E., Salakhudinov, R.: Unsupervised learning of video representations using LSTMs. In: International Conference on Machine Learning, pp. 843–852. PMLR (2015)
21. Sudharsan, B., Sundaram, D., Breslin, J.G., Ali, M.I.: Avoid touching your face: a hand-to-face 3D motion dataset (COVID-away) and trained models for smartwatches. In: 10th International Conference on the Internet of Things Companion, pp. 1–9 (2020)
22. Zhang, J., Kumar, S.: NoFaceContact: stop touching your face with NFC. In: Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services, pp. 468–469 (2020)