Deep Learning-Based Privacy Preservation and Data Analytics for IoT Enabled Healthcare

Hongliang Bi, *Member, IEEE*, Jiajia Liu, *Senior Member, IEEE*, and Nei Kato, *Fellow, IEEE*

***Abstract*—With the development of the industrial Internet of Things (IIoT), intelligent healthcare aims to build a platform to monitor users' health-related information based on wearable devices remotely. The evolution of blockchain and artificial intelligence technology also promotes the progress of secure intelligent healthcare. However, since the data are stored in the cloud server, it still faces the risk of being attacked and privacy leakage. Note that little attention has been paid to the security issue of privacy information mixed in raw data collected from large number of distributed and heterogeneous wearable healthcare devices. To solve this problem, in this article, we design a deep learning-based privacy preservation and data analytics system for IoT enabled healthcare. At the user end, we collect raw data and separate the users' privacy information in the privacy-isolation zone. At the cloud end, we analyze the health-related data without users' privacy information and construct a delicate security module based on the convolutional neural network. We also deploy and evaluate the prototype system, where extensive experiments prove its effectiveness and robustness.**

***Index Terms*—Data analytics, deep learning, IoT-enabled healthcare, privacy preservation.**

I. Introduction

With the industrial Internet of Things (IIoT) technology, wearable devices can access users' health-related information, upload the information to the cloud for analysis, and give feedback to users, which greatly promotes the development of intelligent healthcare [1]. According to a report, the global Internet of Things (IoT) healthcare market is expected to grow from USD 72.5 billion in 2020 to USD 188.2 billion by 2025 [2]. However, with the collection of health-related data, it is inevitable to collect users' privacy-related behavior information by wearable devices, which also faces the risk of privacy leakage. For example, with wearing smart earphones for healthcare during walking, the uploaded data will be mixed with gait information closely related to the user's identity [3]. Once the attackers obtain the data from the cloud, the attackers can separate the gait information and the users' privacy may be leaked. It is worth thinking about analyzing the health-related data while protecting users' privacy in IoT-enabled healthcare.

With the development of blockchain and artificial intelligence technology, more and more researchers tend to leverage these emerging technologies to build a secure and robust IoT-enabled healthcare platform. In [4], it realizes medical data analysis in a secure way and privacy-assured medical data aggregation on the fog server by improving symmetric homomorphic cryptosystem and fog-based communication architecture. In [5], a scheme using blockchain technology is proposed to solve the security issue of the key management for flying ad-hoc network, which can resist external and internal attacks effectively. Healthbank blockchain [6] can also extract data from wearable devices and store the data securely.

The general privacy protection technologies could ensure that the data are transmitted in nonplaintext mode. However, they are not completely free from loopholes. The attackers can use these loopholes to crack encrypted information. For example, the blockchain technology can be used to protect the access records and logs on the chain from being tampered with. However, network insiders can modify data by adding or deleting sensitive information, which may lead to significant security and trust problems [7]. Therefore, it is necessary to design a new way to strengthen privacy protection. Even if the data stored in the cloud platform or localization server are stolen, the attacker cannot obtain the user's privacy information from the data source level. In order to enhance data security, we need to remove the privacy information in the privacy isolation zone before uploading data to the cloud [8], which can also be used as a supplement to existing privacy protection works.

However, there are also some significant challenges that we need to solve. Because privacy information and health-related information are mixed, it is difficult for people to distinguish them directly. Even if the privacy information can be separated, the health-related data may be distorted and difficult to extract.

Manuscript received February 25, 2021; revised July 13, 2021 and September 5, 2021; accepted September 28, 2021. Date of publication October 8, 2021; date of current version April 13, 2022. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant D5000210598, in part by the National Natural Science Foundation of China under Grant 61771374, Grant 61771373, Grant 61801360, and Grant 62001393, in part by the Natural Science Basic Research Program of Shaanxi under Grant 2020JC-15 and Grant 2020JM-109, in part by the Fundamental Research Funds for the Central Universities under Grant 31020200QD010, and in part by the Special Funds for Central Universities Construction of World-Class Universities (Disciplines), and Special Development Guidance under Grant 0639020GH020114. Paper no. TII-21-0929. *(Corresponding author: Jiajia Liu.)*

Hongliang Bi and Jiajia Liu are with the National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: bihongliang@nwpu.edu.cn; liujiajia@nwpu.edu.cn).

Nei Kato is with the Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan (e-mail: kato@it.is.tohoku.ac.jp).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3117285.

Digital Object Identifier 10.1109/TII.2021.3117285

The effective recognition of separated health-related data is also a problem. In order to solve the abovementioned challenges, we design a deep learning-based privacy preservation and data analytics prototype system for IoT-enabled healthcare. The main steps are as follows: at the user end, we design a privacy-isolation zone to collect health-related data [8]. Specifically, since time series data from wearable devices may be aliased with various action information, we use the data analytics method to separate the aliased privacy information. The health-related data without users' privacy information are uploaded to the cloud for further analysis directly. At the cloud end, we implement the nonprivacy data extraction algorithm to analyze health-related information. Subsequently, the security module is built based on the extracted data using the convolutional neural network (CNN). The analytical result can also be returned to the doctors or the users to monitor the users' health.

In this article, in order to describe our system more specifically, we construct a privacy-preserving scheme for the "office workers" and the "phubbers." As we know, if the head is in an incorrect posture for a long time, people may have a series of health problems due to insufficient blood supply. Therefore, head gesture monitoring is of great significance in IoT health. The emergence of "phubbers" and "office workers" increases the health risks of the people. Given this, we research head gesture recognition during resting and walking states. However, the health-related head gesture data will be mixed with gait information in the walking state, which needs to separate. Therefore, we build a secure prototype system to recognize the head gestures for these subhealthy people. In our prototype system, the raw data are collected from off-the-shelf smart earphones. As private and pervasive wearable devices, more and more smart earphones are integrated with accelerometers, such as AirPods, which are low-cost and universal. It can detect health-related head gestures to prevent neck pain caused by long-term head immobility whether the user is in walking or resting state. Our system can also be extended to other wearable devices, which can realize secure and user-friendly IoT-enabled healthcare.

A. Contribution

With the system designing, the main contributions are described as follows.

1) We propose a prototype system with privacy preservation and data analytics based on deep learning, which can analyze health-related data while protecting the users' privacy.
2) We implement the nonprivacy data extraction algorithm to analyze health-related data after privacy information separation.
3) We use the data augmentation method to avoid overfitting. A customized CNN is used to construct a security module. We implement and evaluate our proposed system with smart earphones on a collected dataset from 20 participants. Different scenarios are also considered to verify the effectiveness and robustness of the system.

B. Organization of the Article

The rest of the article is organized as follows. We first introduce our related work in Section II. In Section III, we present the system overview. Then, we implement the nonprivacy data extraction algorithm and build the security module. Experimental results are given in Section IV. Finally, Section V concludes this article.

II. Related Work

Given the popularity of wearable devices in pervasive and personalized healthcare, there are some IoT-enabled healthcare works. In [1], a wearable feedback system, which can help therapists monitor swimmers' physical recovery and injury prevention. In [9], a smart indoor anticollision system based on radio frequency identification (RFID) is proposed to help visually impaired people guide from obstacles. In [10], a real-time biomonitoring method has been proposed to monitor facial surface electromyography, reflecting the pain intensity of patients. The wearable data with the biosensor mask can be uploaded to the cloud server for analysis to realize the medical care of patients. In [11], a home care system based on a wearable accelerometer is proposed to measure the respiration rate and daily volume variability, where the measured data can be stored in the cloud and timely feedback to patients can be provided. The data involved in health-related information is susceptible, and security solutions are also proposed in some works. In [12], the Bayesian network algorithm is applied to human sensor networks to detect and eliminate fault sensor data and prevent medical diagnosis errors. In [13], Bodyedge, an edge-based architecture consisting of a tiny mobile client module and performing edge gateway, is proposed to support healthcare applications, which can ensure flexibility, robustness, and adaptive service level on the private cloud and public cloud platforms. In [14], a cloud-based user authentication scheme is proposed, which can realize secure communication with a secret session key. In [15], a privacy protection method based on RFID is proposed to protect the consistency and synchronization of authentication information in the medical environment. In [16], a lightweight data integrity verification technology based on an edge server is proposed to verify the integrity of patients' health data stored in the cloud and prevent disease misdiagnosis. In [17], a reliable scheme to distinguish patients from noisy ECG signals is proposed to provide differential privacy protection. In [18], a large-scale privacy protection scheme based on blockchain technology is proposed to protect the privacy of data stored in the hospital database or cloud.

However, the data mixed with privacy information are stored in the cloud server, which is still the possibility of being attacked. Therefore, it is necessary to propose the privacy-preserving IoT-enabled heathcare system to ensure data security further.

III. Proposed Prototype System

In this section, we first described the overview of the system architecture. Then, a privacy-isolation zone was designed to

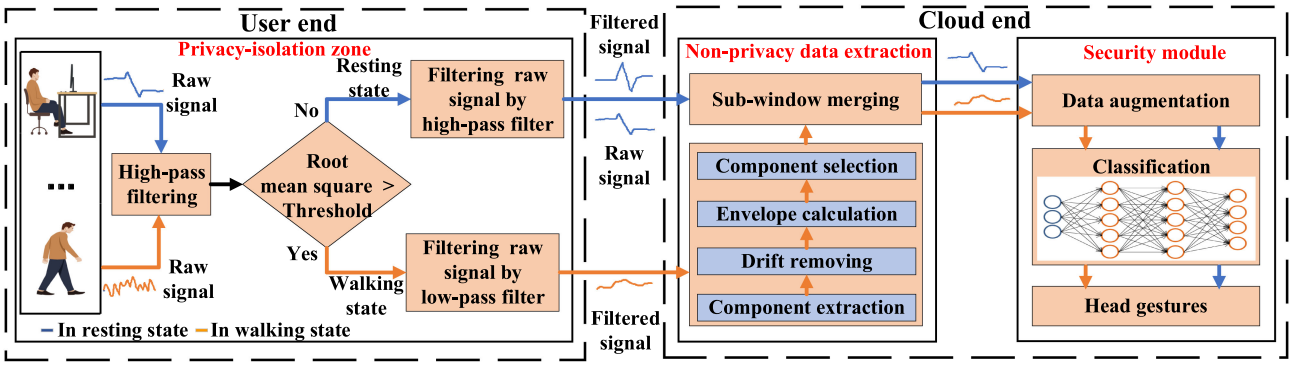

Fig. 1. Flowchart of the system architecture. (The user end determines whether the user is in walking or resting state and removes the gait information during walking. The nonprivacy data extraction and security module can be realized at the cloud end.).

detect and separate gait signal from the acceleration stream at the user end. Finally, the nonprivacy data extraction and security module were implemented at the cloud end, which analyzed users' head gestures both in the resting state and in the walking state.

A. Overview of System Architecture

We design the system for IoT-enabled healthcare with built-in accelerometers in the smart earphones. The sampling frequency of acceleration is 100 Hz. The X-axis is parallel to the side of the face and perpendicular to the ear handle. The Y-axis is the opposite of the direction of the smart earphone handle. The Z-axis is perpendicular to the plane where the X- and Y-axes lie. At the user end, we open a privacy-isolation zone to receive data from smart earphones and separate privacy information, such as gait information. At the cloud end, we deploy algorithms to realize the nonprivacy data extraction and security module construction, which can reduce the burden at the user end [8]. Fig. 1 illustrates the flow of our system architecture, which mainly consists of the following processes.

1) We distinguish between walking and resting and separate gait information from collected data during walking in the privacy-isolation zone. First, we leverage a high-pass filter to eliminate the influence of gravity on raw data. Then, the root mean square (rms) value of the signal is calculated to determine the user's state. If the user is in the walking state, we remove the gait information through low-pass filtering of the raw data. The filtered data without user's privacy information are uploaded to the cloud end. If the user is in the resting state, the raw data are sent to the cloud together with the high-pass filtered data.
2) We also propose a nonprivacy data extraction algorithm for health-related head gesture detection. If the user is in the resting state, we utilize a subwindow merging algorithm to detect head gestures directly. If the user is in the walking state, we determine the head gesture boundary through a series of processes, including component extraction by principal component analysis (PCA), drift removing, envelope calculation, component selection, subwindow merging, etc.
3) We expand the samples through the data augmentation to avoid overfitting. The CNN is used to build the security module, which will be used for head gesture recognition.

Our goal is to design a system for privacy preservation and data analytics in IoT-enabled healthcare. The system can achieve the following design goals.

1) *Privacy Protection:* By separating the users' privacy information in the privacy-isolated zone at the user end, the security of data transmission and storage at the cloud end can be further guaranteed.
2) *Passive Sensing:* We can develop a monitoring system for the users' health without impairing the users' adherence.
3) *Reliable Analysis:* Our system can ensure that the health-related data without privacy information can be obtained and analyzed.

B. Design of Privacy-Isolation Zone

As abovementioned, while collecting the user's health-related head gesture data by wearable devices during walking, the data are inevitably mixed with the gait information associated with the user's identity. Once the data in the cloud are stolen, the attacker may separate the gait information from the aliased data, which will cause the user's privacy disclosure. Therefore, we need to segment the data for analyzing the user's state in the privacy-isolation zone before uploading the data to the cloud.

The window function, a smooth function that goes to zero at the border, makes the signal outside the boundary approximate to zero and retains the signal within the boundary by multiplying with the collected signal. We can extract the signal within the boundary based on a fixed threshold. However, the collected data include not only the gait information but also the gravity information. The gravity is a downward force with a fixed value of 9.8 m/s^2. As shown in Fig. 2(a), the projection of gravity on each axis will change with the head movement, which is hard to find a fixed threshold. In addition, the interference of gait on the head gestures also causes a low signal-to-noise ratio (SNR), which makes it difficult to distinguish between noise and head gestures. Moreover, they are aliased in the time domain. Therefore, we cannot separate the aliased signals directly with the window function.

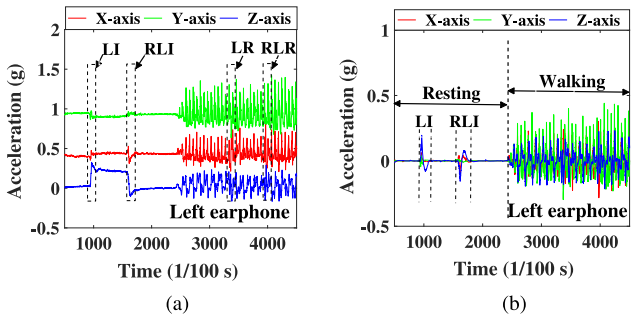

Fig. 2. Raw acceleration signal and linear acceleration signal by high-pass filtering of left smart earphone.

Fortunately, different behavior information shows different frequency characteristics. In order to separate the data, we analyze the signals in the frequency domain through Fourier transform. The gravity is a direct-current component in the frequency domain. The frequency of the gait ranges from 1.4 to 2.1 Hz, which are all at the high-frequency band relative to the gravity [19]. Therefore, we use a high-pass filter to filter out the gravity information and a low-pass filter to filter out the gait information. The filters that can realize high-pass or low-pass filtering include Butterworth, Chebyshev, Elliptic, and Wavelet filters [20].

The Wavelet filter has a complicated wavelet decomposition process, which is unsuitable for deployment on the user terminal with limited resources. Butterworth filter has the flattest passband frequency response curve and relatively slow stopband attenuation, which has a more stable amplitude frequency characteristic than the Chebyshev filter and Elliptic filter. We also use the SNR to analyze filtered neck extension (NE) signals. The average SNRs of signals using Butterworth, Chebyshev, and Elliptic are 11.99, 11.50, and 11.88, respectively. The higher the SNR is, the more effective the head gesture can be extracted. Therefore, we separate mixed signals using the Butterworth filter.

As shown in Fig. 2(b), after eliminating the gravity information, we can use a window function to analyze the user's state by calculating the rms of noise signal within the window width of 1 s.

$$\text{rms}_I = \sum_{i \in I} \frac{\sqrt{LX_i^2 + LY_i^2 + LZ_i^2 + RX_i^2 + RY_i^2 + RZ_i^2}}{6}$$

where LX, LY, LZ are the three axes acceleration in left earphone, and the RX, RY, RZ are the three axes acceleration in right earphone. We also calculate the average rms from the collected data. The average rms of noise in the walking state is greater than 4, whereas the average rms of noise in the resting state is less than 0.4. Therefore, we can set a threshold to distinguish between walking and resting.

Since the gait frequency is at high-frequency band relative to the gravity frequency, we use a high-pass filter to remove the gravity information and a low-pass filter to separate the gait information. If the user is in the walking state, the filtered signal is directly uploaded to the cloud for analysis. If the user is in the resting state, the raw data are sent to the cloud together with the filtered data. All the Butterworth filters use a cut-off frequency of 0.45 Hz.

C. Nonprivacy Data Extraction

After eliminating the privacy information in the privacy-isolated zone, we can extract the health-related head gestures at the cloud end. There are 12 kinds of head gestures, including NEand recovery (RNE), neck flexion (NF) and recovery (RNF), neck left inclination (LI) and recovery (RLI), neck right inclination (RI) and recovery (RRI), neck left rotation (LR) and recovery (RLR), and neck right rotation (RR) and recovery (RRR). Because the gravity direction is vertically downward, the angle between gravity and acceleration will change with head movement. However, for the head gestures LR, RLR, RR, and RRR, gravity has little effect on the acceleration signal, and the gait plays a major role during walking. The change of angle is also very small. The separation of gait will also lead to signal distortion. It is difficult to detect the LR, RLR, RR, and RRR during walking. To solve this problem, we design a nonprivacy data extraction algorithm during walking. The flow of algorithm is shown in Fig. 3.

Compenent Extraction: The different head gestures have different degrees of influence on each axis of acceleration. Especially for the head gestures LR, RLR, RR, and RRR, they are difficult to extract because of low SNR. In order to select the component, most affected by the head gesture, we map the filtered data into mutually independent components in the walking state by PCA [21], [22]. Before PCA, we need to normalize the data, which can make the data comparable among axes and avoid the information loss due to amplitude difference

$$D = [LX', LY', LZ', RX', RY', RZ']$$

$$D^* = \frac{D - \min(D)}{\max(D) - \min(D)}$$

where $LX', LY', LZ', RX', RY',$ and RZ' represent the six axes filtered data; D is a matrix of these six column vectors; D^* is a normalized matrix of D with a dimension of Len $\times$ 6, where Len represents the number of sample points. We calculate the correlation coefficient matrix C of D^* according to the formula $C = D^* \times (D^*)^T$. We solve the eigenvectors of the C and put them into a matrix R by row. Each PCA component can be obtained by the formula $Y = P \times D$. In Y, each row represents the projection of the raw data on each component.

Drift Removing: The signal drift caused by the internal noise of sensors and the occurrence of some outliers will increase the signal variance. When the signal is projected to the orthogonal direction with the largest variance through PCA, the variance can be amplified and more significant drift appears [23]. We can remove drift of each component through the linear regression fitting method [24], [25]. First, the trend term of each component can be calculated by minimizing the square sum of errors between the fitted component and original component. Then, we subtract the trend term from the component and focus the

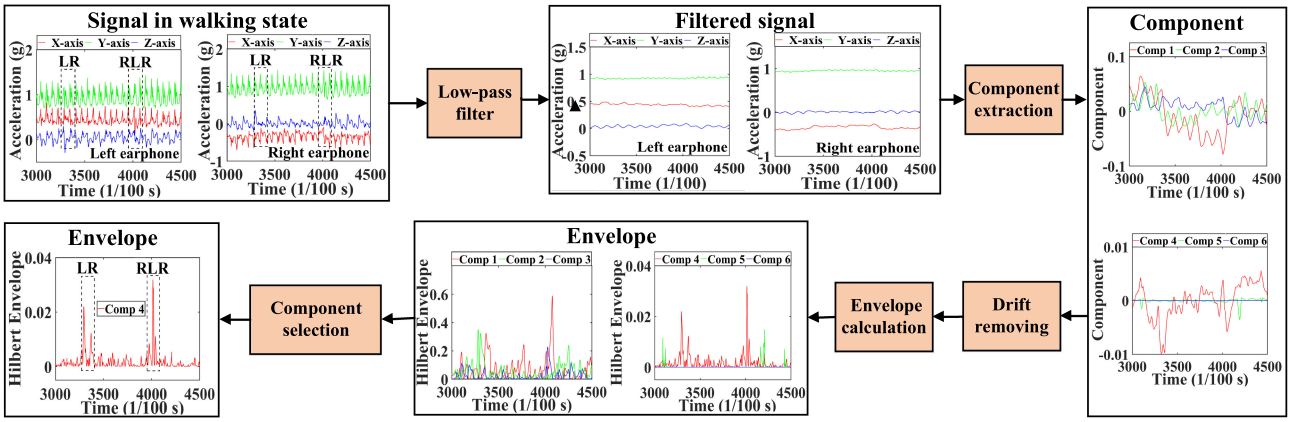

Fig. 3. Flowchart of the nonprivacy data extraction algorithm during walking. (We take the LR and RLR as examples. The Comp1, Comp2, Comp3, Comp4, Comp5, and Comp6 are six components by PCA, respectively.).

analysis on the fluctuation of the data itself.

$$Y'_k = Y_k - f_k,\ k \in \{1, 2, \ldots, N\} \tag{1}$$

where Y'_k is the component after drift removing, f_k is trend term of the component Y_k, and N is the total number of components.

Envelope Calculation: Because the plane of the head rotation is hard to be strictly perpendicular to the direction of gravity, there will be also a signal fluctuate. Gravity still plays a small, albeit small, role in the signal. Compared with random noise, the head gesture signal has a larger fluctuate. We can distinguish them from random noise by amplifying the fluctuate difference by slope calculation

$$y_k = |\Delta Y'_k(t) = Y'_k(t) - Y'_k(t-1)| \tag{2}$$

where y_k is the absolute value of slope of the solved component signal Y'_k in (1).

Even so, the random noise will lead to a larger slope and easily be mistakenly detected as the head gesture. To detect the head gesture signal more efficiently, we extract the envelope A_k of the slope signal y_k by Hilbert transform, which can suppress the influence of larger slope caused by partial noise and further smoothen the head gesture signal [26]

$$A_k(t) = \sqrt{y_k^2(t) + H_k^2(t)},\ k \in \{1, 2, \ldots, N\} \tag{3}$$

where H_k is the value by Hilbert transform of the slope signal y_k in (2), and N is the total number of components.

Component Selection: The component with a higher SNR can help us to extract the head gesture signal more effectively. Therefore, we select the component with the maximum SNR for analysis. However, the SNRs of different components with the same head gesture are different from each other. The SNRs of different head gestures on the same component are also different. We cannot directly determine an optimal component to extract head gesture signals. Therefore, we implement an adaptive component selection method using the (4) to find the optimal component for head gesture extraction.

$$\mathrm{SNR}_k = \frac{A_k^2}{A_{\text{noise}}^2},\ k \in \{1, 2, \ldots, N\}$$
$$\hat{k} = \arg \max_{k \in \{1,2,\ldots,N\}} \mathrm{SNR}_k \tag{4}$$

where SNR_k is the SNR of the envelope A_k in (3), $\hat{k}$ is the component number of the envelope with the maximum SNR, and N is the total number of components. Finally, we leverage the subwindow merging algorithm to segment the selected envelope signal.

Subwindow Merging: The traditional fixed sliding window algorithm usually faces a fixed window width problem. If the window's width sets unreasonable, it is easy to cut off the signal or contain multiple redundant signals. In order to cover the whole head gesture signal, we utilize small-sized subwindows for continuous detection and merge them into a parent window [27]. With this subwindow merging algorithm, we can adaptively extract signals with different widths in time domain, which is beneficial to retain the complete head gesture information. If the rms is greater than that of random noise, the subwindow will be retained. Otherwise, we proceed to the next head gesture analysis. Considering that the 50% overlap rate of subwindow, only the first $\frac{N}{2}$ sampling points of each subwindow I are retained

$$E = E \bigcup \left\{ P_1^I, P_2^I, \ldots, P_{\frac{N}{2}}^I \right\}$$

where E is the detected signal in the merged window. After determining the boundary of event signal during walking, we input the low-pass filtered six axes acceleration data within the boundary to the security module for classification. In particular, we can directly use the subwindow merging algorithm to determine the boundary based on the high-pass filtered data in the resting state.

D. CNN-Based Security Module

Data Augmentation: Limited by the time and cost, it is impossible to collect all sample data. We can use the data augmentation

TABLE I
NETWORK STRUCTURE OF SECURITY MODULE

Number	Layer Name	Input	Output
1	Convolution 1D	(None,600,6)	(None,600,800)
2	Pooling 1D	(None,600,800)	(None,150,800)
3	Convolution 1D	(None,150,800)	(None,150,800)
4	Pooling 1D	(None,150,800)	(None,38,800)
5	Convolution 1D	(None,38,800)	(None,38,800)
6	Pooling 1D	(None,38,800)	(None,10,800)
7	Convolution 1D	(None,10,800)	(None,10,800)
8	Pooling 1D	(None,10,800)	(None,5,800)
9	Flatten	(None,5,800)	(None,4000)
10	Dense	(None,4000)	(None,12)

method to generate more training samples for improving the generalization of the module as follows [28].

1) Time warping processing, which transforms the time-domain position of the signal, represents data collection with different speeds.
2) Amplitude distortion processing, which changes the amplitude of the data randomly, represents data collection with different forces.
3) Time scaling, which changes the signal width, represents data collection with different amplitudes.
4) Permutation, which changes the time position of the signal in the window, represents data collection with different segments.
5) Rotation processing represents data collection by wearing smart earphones at different angles.
6) Adding random noise to the data represents data collection with different noise environments.

Network Structure: The existing common neural networks include long short-term memory (LSTM), deep neural network (DNN), CNN, etc. The LSTM is suitable for the samples with precedence and dependency relationships. The order of input will affect the output of LSTM. The DNN has a large training cost due to too many parameters. The CNN has a parameter sharing mechanism, which can reduce the number of parameters for training. Each adjacent point in the extracted signal has a significant correlation, which is also suitable for convolution processing of CNN. Moreover, the predicted samples have no sequence relationships between the past and future, which are not suitable for the LSTM. Therefore, we use CNN to build our security module [29].

Because the lengths of acceleration data extracted by the variable window function are different and most people complete the head gestures within 6 s, we regularize the input data into 1×600 1-D format by the ways of truncating and filling zero at the end (with the frequency of 100 Hz). The label of each class is encoded in one hot mode. And we use the z-score standardization method to preprocess the 1-D data, which can eliminate the influence of data unit.

As shown in Table I, the network structure includes four 1-D convolutional layers, four 1-D pooling layers, one fully connected layer, and a softmax layer [30], [31]. Each convolution layer has 800 convolution kernels, which are used for feature extraction. For the first convolution layer, we input the six axes time series data with a length of 1×600. The input and output of each layer are shown in Table I. The size of each convolution kernel is 1×4. The moving step size of each convolution kernel is 1×1. The output of the convolution kernel is processed by the activation function Relu to improve the expressiveness of the model. Then, the feature dimensionality reduction is performed by extracting the maximum value in 1×4 moving window. The moving step size is set to 1×4 except the last layer that is set to 1×2. After flattening the output data of the last convolution layer, the predicted probability of each class can be solved through the fully connected layer and the softmax layer. Based on the predicted category and the true category, we calculate the rms error (RMSE) and propagate it back to the network. The Adam gradient descent method is used to update the network parameters with the learning rate of 0.0001.

IV. EXPERIMENTAL RESULTS

We implemented the nonprivacy data extraction algorithms and then built a security module using CNN. In this section, we conducted different experiments to evaluate and verify the effectiveness and robustness of our system.

A. Experimental Settings

In the experiment, we recruit 20 participants (five females) from 20 to 30 years old. All the participants are healthy students. Each participant performs the 12 kinds of head gestures in resting and walking states respectively. Each head gesture is repeated 20 times. Participants comfortably perform head gestures with an interval of 10 s. To collect data in the resting state, we randomly select ten participants to sit in front of the computer to browse the website, and the rest participants need to keep standing posture. The data collection during walking is carried out on a treadmill with a speed of 4 km/h and a slope of 0°. Besides, each participant collects walking data of 6 s, 20 times as a new category *W*. The video recordings are also provided by the camera. The ELAN software is used to label each head gesture. Finally, we have collected a total of 10 000 samples ($20 \times 20 \times 12 + 20 \times 20 \times 13$). Among them, 80% of the samples are used for training, and the others are for testing. For the training set, we use tenfold cross validation as the verification method.

B. Evaluation Metrics

To evaluate the recognition performance, we use accuracy, precision, recall, F-score, and confusion probability matrix as the evaluation metrics, where the true negative (TN), true positive (TP), false negative (FN), and false positive (FP) are basic metrics [32], [33].

1) Accuracy is the ratio of the true positive samples to the total number of samples, defined as $\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FT}+\text{FN}}$.
2) Precision is the ratio of true positive samples to predicted positive samples, defined as $\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$.
3) Recall is the ratio of true positive samples to all positive samples, defined as $\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$.
4) F-score is a harmonic mean of precision and recall, defined as $\frac{2}{\text{F}-\text{score}} = \frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}$.

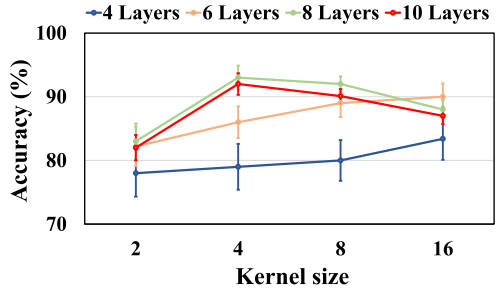

Fig. 4. Accuracy comparison under different kernel size and layer number.

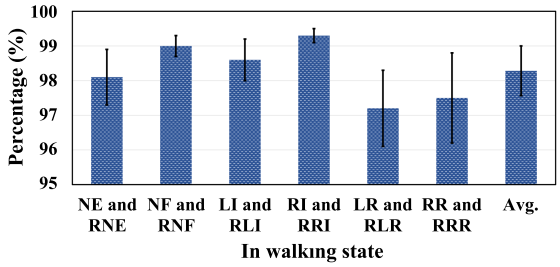

Fig. 5. Performance of component selection.

5) The confusion probability matrix is used to analyze the relationship between the predicted results and actual results. Each row of the confusion matrix corresponds to the true label; each column of the confusion matrix corresponds to the predicted label.

C. Parameter Selection

The more layers the CNN has, the stronger the learning ability of the model has. However, too many layers will easily lead to overfitting of the model. If the size of the convolution kernel is too small, it will not be able to obtain complete signal characteristics; otherwise, it will introduce too much noise. To select suitable model parameters, we analyze the impact of different kernel sizes (2, 4, 8, and 16) and layer numbers (4, 6, 8, and 10) on the recognition performance by grid search method to find the best parameter combination [34]. As shown in Fig. 4, the model has the accuracy of 93% with the kernel size of 4 and the layer number of 8, which is the best performance.

D. Effectiveness Evaluation

Performance of Component Selection: The effective component selection is the premise of signal segmentation. In order to prove that we can determine the signal boundary based on the components of PCA, we analyze the ratio of head gestures with the effective determining components and all head gestures (the higher the ratio, the better the performance). The head gestures and corresponding recovery gestures are collected together. Their signal detection can be carried out on the same component and analyzed together. As shown in Fig. 5, because head gestures LR and RR have more serious distortion after filtering the gait signal, the performance of head gestures LR and RR are slightly lower than those of other kinds of head gestures. Even so, the ratios of all kinds of head gestures that can select effective components exceed 97%. Therefore, we can segment the signal based on the components of PCA.

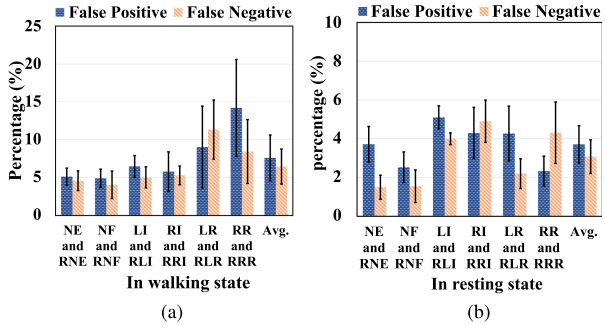

Fig. 6. False positive rates and false negative rates of the nonprivacy data extraction algorithm.

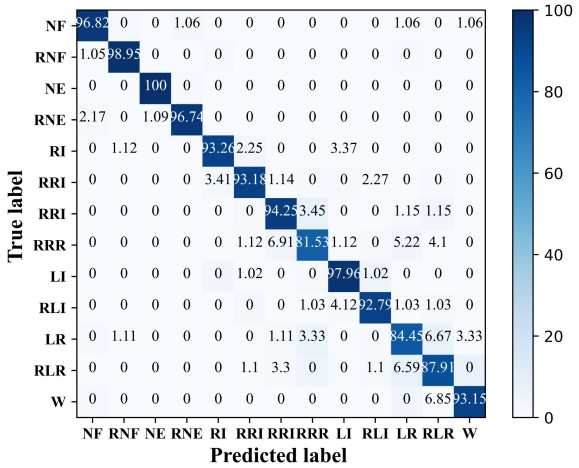

Fig. 7. Confusion probability matrix of 12 kinds of head gestures.

Performance of Nonprivacy Data Extraction: The nonprivacy data extraction will directly affect the performance of the model. We evaluate the performance of nonprivacy data extraction with the false positive rate and the false negative rate. When there is no head gesture happening, but the signal is detected, we define it as a false positive. When the head gesture happens, but no signal is detected, we define it as a false negative. As shown in Fig. 6, the experimental results show that the average false positive and false negative rates of nonprivacy data extraction in the resting state are 3.71% and 3.08%, respectively. The average false positive and false negative rates of nonprivacy data extraction in the walking state are 7.59% and 6.46%, respectively. We can see that the false positive and false negative rates in the walking state are higher than those in the resting state. This is because the walking has a more significant interference on head gesture detection. The average false positive and false negative rates of head gestures LR, RLR, RR, and RRR are higher than those of the other eight kinds of head gestures. To further reduce false positives, we take walking as a category.

Confusion Probability Matrix: We analyze the accuracy of misclassification using confusion probability matrix. As shown in Fig. 7, the error rates between head gestures RR and RRR

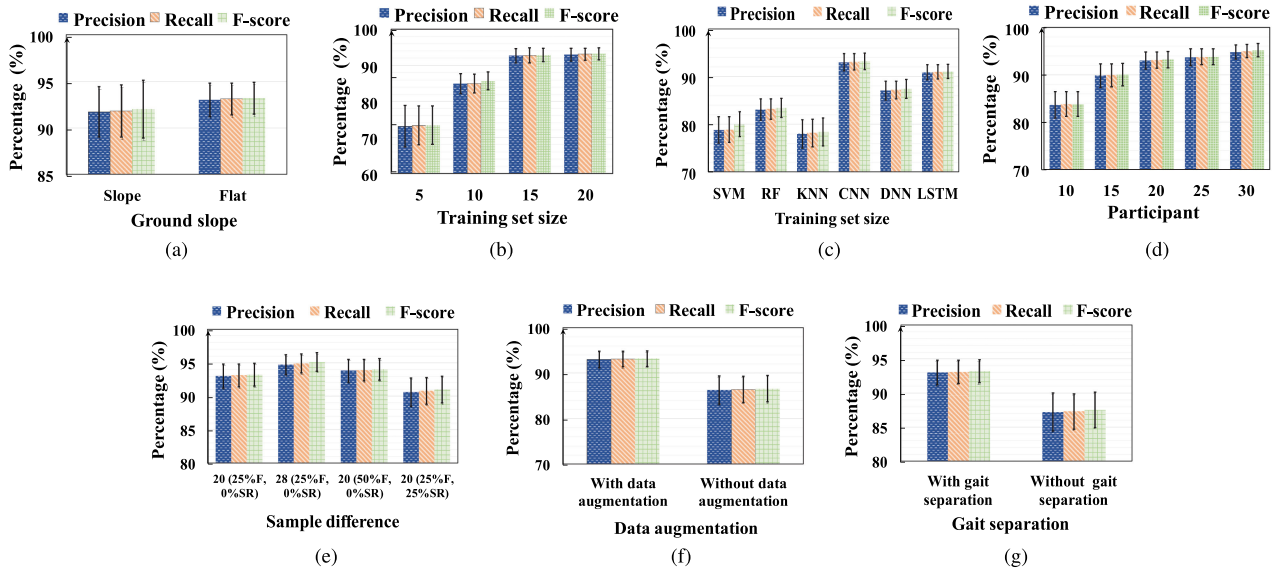

Fig. 8. Robustness evaluation (%F and %SR mean the proportion of female and senior man in sample, respectively).

are 6.9% and 3.5%, respectively. The error rates between head gestures LR and RLR are 6.6% and 6.7%, respectively. These error rates are relatively high because they are similar with each other. Nevertheless, the average recognition accuracy of model exceeds 93%. At the same time, the recognition accuracy of walking (W) is 93.2%, which can help us further reduce the impact of false positive rate.

E. Robustness Evaluation

In this section, in order to prove the robustness of the model, we evaluate the effects of different factors on performance of security module, including ground slope, classifier, data augmentation, gait separation, data sample,training set size, etc.

Impact of Ground Slope: In order to evaluate the performance in different ground slopes during walking, we set the treadmill to three kinds of slopes, including uphill slope of 4°, downhill slope of −4°, and flat ground of 0°. For both uphill and downhill scenarios, all participants do each kind of head gesture 10×. With the 80–20 split rule, 80% samples of both uphill and downhill are used to build the model, and the rest samples are used to test the model. As shown in Fig. 8(a), the precision, recall, and F-score of the model with slope are 91.7%, 91.8%, and 92%, respectively, which is slightly lower than that of the model during walking on the flat ground. When the user walks on a larger slope, her/his body shakes more violently. It makes the event more difficult to detect. Nevertheless, the slope has little effect on the performance, showing that the model has good robustness to the ground slope.

Impact of Training Set Size: In order to ensure effectiveness in the new environment, the model needs to be updated in time. The more training samples we collect, the better the generalization the model has. However, too frequent data collection will bring participants a bad experience. To find a suitable data collection scheme, we train the model with different training set sizes. As shown in Fig. 8(b), by comparing the performance of the different proportion of training set size, we can find that when we train the model with five samples per class, the precision, recall, and F-score are 73.1%, 73.3%, and 73.4%, respectively. The model has the best performance when we have a training set size of 20 for training. When the training set size exceeds 15, the performance of the model tends to be stable. Therefore, the data we collect are enough to build the model.

Impact of Classifier: We also use other machine learning methods to construct the model for performance comparison. The training set and testing set are the same as in Section IV-A. As shown in Fig. 8(c), the accuracies of the models constructed by traditional machine learning methods cannot exceed 85%. The accuracies of the models constructed by DNN and LSTM are also lower than that of our model. Therefore, the model built with CNN can realize the best recognition performance.

Impact of Data Sample: In order to prove the impact of sample difference on the number of participants, age, ratio of sex, we rerecruit ten participants, including five senior men (SR), with ages from 50 to 65, and five young females (F), with ages from 20 to 30. Each volunteer performs each kind of head gesture 20 times. We retrain and evaluate the models based on the different data samples. As shown in Fig. 8(e), the precision, recall, and F1-score are above 90%. Therefore, age and gender factors have little effect on performance. Furthermore, when the recruitment numbers increase to 28 (including seven females), the accuracy is almost unchanged. In addition, we also build different models based on 10, 15, 20, 25, and 30 participants, respectively. As shown in Fig. 8(d), the result shows that as the number of participants exceeds 20, the performance of the model tends to be stable. Therefore, it is enough to build our model based on the 20 participants.

Impact of Data Augmentation: In order to study the impact of data augmentation on performance, we train the models with data augmentation and without data augmentation, respectively.

The result is shown in Fig. 8(f). The precision, recall, and F-score of the model without the data augmentation are 86.3%, 86.4%, and 86.6%, respectively. The result shows that the performance of the model with data augmentation is 7% higher than that of model without data augmentation. Therefore, even if we cannot obtain a large number of training samples, we can use the augmentation method to cover uncollected data and avoid overfitting.

Impact of Gait Separation: In order to evaluate the impact of gait separation on the performance, we build the model based on the original data without filtering. As shown in Fig. 8(g), the precision, recall, and F1-score are lower than those of our model. This is because that gait has a significant influence on the data distribution, which reduces the model performance.

F. Delay and Energy Consumption

Large time delay and high power consumption will reduce the user experience. We evaluate the running time of algorithms, including head gesture detection and recognition. We deploy the algorithm on an 8-core Intel (R) Core (TM) i7-9700 CPU and 16-GB RAM desktop computer. The total delay time of the algorithm is 4.7 ± 0.1 ms, which has a short delay time and is acceptable by users. We also use a voltmeter to measure the average power consumption of data transmission from earphones for 10 min. The result shows that the power consumption is 7.13 ± 0.21 mW/min. with the sampling rate of 100 Hz. Therefore, the proposed algorithm has low energy consumption and short delay [35].

V. Conclusion

In this article, we presented our prototype system for privacy preservation and data analytics in IoT-enabled healthcare based on deep learning, which can separate the privacy information mixed in the raw data and analyze health-related data. It works by isolating privacy-sensitive content, extracting, and recognizing nonprivacy data. We also evaluated the performances in different scenarios and validated the effectiveness and robustness of the system. The system can be used as a supplement to future intelligent healthcare. With system architecture design, we can also expand to other existing wearable devices for IoT-enabled healthcare.

However, there are also some deficiencies in our work, which will be further improved and studied in the future. First, because of the time and cost constraints, the number of volunteers is small. In future, we can recruit more volunteers to improve the generalization performance of the system. We will also introduce more policies to strengthen the protection of user identity [36]. Besides, the system needs to collect training samples in advance to be trained, it is inconvenient in face of changeable scenarios. In the future, we can use meta-learning technology to build the system directly in different scenarios, which can reduce the adaptive burden. Finally, we aim at the disease prevention of "office workers" and "pubbers" with earphones. The nontime series health-related data, such as text data, cannot be protected directly using our scheme. There will be some privacy-related behavior characteristics for people who have already suffered from cervical spondylosis. The health-related data may be mixed with other behavior information in other scenarios. They are not included in the scope of our research. However, our scheme can still provide a reference for other scenarios. We will improve our technological solutions to build a more secure e-Health system in the future.

References

[1] A. Kos and A. Umek, "Wearable sensor devices for prevention and rehabilitation in healthcare: Swimming exercise with real-time therapist feedback," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1331–1341, Apr. 2019.

[2] MarketsAndMarkets, "IoT in healthcare market by component, application, end user, and region-global forecast to 2025," 2021. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/iot-healthcare-market-160082804.html

[3] Y. Sun, J. Liu, J. Wang, Y. Cao, and N. Kato, "When machine learning meets privacy in 6G: A survey," *IEEE Commun. Surv. Tut.*, vol. 22, no. 4, pp. 2694–2724, Oct.–Dec. 2020.

[4] Y. Yang, X. Liu, and R. H. Deng, "Lightweight break-glass access control system for healthcare Internet-of-Things," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3610–3617, Aug. 2018.

[5] Y. Tan, J. Liu, and N. Kato, "Blockchain-based key management for heterogeneous flying AdHoc network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7629–7638, Nov. 2021, doi: 10.1109/TII.2020.3048398.

[6] Healthbank. Sep. 2021. [Online]. Available: https://www.healthbank.coop/

[7] T. McGhin, K.-K. R. Choo, C. Z. Liu, and D. He, "Blockchain in healthcare applications: Research challenges and opportunities," *J. Netw. Comput. Appl.*, vol. 135, pp. 62–75, 2019.

[8] H. Zhang, C. Song, A. Wang, C. Xu, D. Li, and W. Xu, "Pdvocal: Towards privacy-preserving parkinson's disease detection using non-speech body sounds," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1–16.

[9] F. Xiao, Q. Miao, X. Xie, L. Sun, and R. Wang, "Indoor anti-collision alarm system based on wearable Internet of Things for smart healthcare," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 53–59, Apr. 2018.

[10] G. Yang *et al.*, "IoT-based remote pain monitoring system: From device to cloud platform," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 6, pp. 1711–1719, Nov. 2018.

[11] A. R. Fekr, K. Radecka, and Z. Zilic, "Design and evaluation of an intelligent remote tidal volume variability monitoring system in e-health applications," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 5, pp. 1532–1548, Sep. 2015.

[12] H. Zhang, J. Liu, and N. Kato, "Threshold tuning-based wearable sensor fault detection for reliable medical monitoring using Bayesian network model," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1886–1896, Jun. 2018.

[13] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 481–489, Jan. 2019.

[14] J. Srinivas, A. K. Das, N. Kumar, and J. J. Rodrigues, "Cloud centric authentication for wearable healthcare monitoring system," *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 5, pp. 942–956, Sep./Oct. 2020.

[15] K. Fan, W. Jiang, H. Li, and Y. Yang, "Lightweight RFID protocol for medical privacy protection in IoT," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1656–1665, Apr. 2018.

[16] R. Ding, H. Zhong, J. Ma, X. Liu, and J. Ning, "Lightweight privacy-preserving identity-based verifiable IoT-based health storage system," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8393–8405, Oct. 2019.

[17] P. Huang, L. Guo, M. Li, and Y. Fang, "Practical privacy-preserving ECG-based authentication for IoT-based healthcare," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 9200–9210, Oct. 2019.

[18] J. Xu*et al.*, "Healthchain: A blockchain-based privacy preserving scheme for large-scale health data," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8770–8781, Oct. 2019.

[19] T. Ji and A. Pachi, "Frequency and velocity of people walking," *Struct. Eng.*, vol. 84, no. 3, pp. 36–40, 2005.

[20] P. Podder, M. Hasan, M. Islam, and M. Sayeed, "Design and implementation of Butterworth, Chebyshev-I and elliptic filter for speech signal analysis," *Int. J. Comput. Appl.*, vol. 98, no. 7, pp. 12–18, 2014

[21] Y. Chen, R. Ou, Z. Li, and K. Wu, "WiFace: Facial expression recognition using Wi-Fi signals," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2020.3001989.
[22] Y. Xun, J. Liu, and Z. Shi, "Multi-task learning assisted driver identity authentication and driving behavior evaluation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7093–7102, Oct. 2021, doi: 10.1109/TII.2020.3034276.
[23] A. Perera, N. Papamichail, N. Bârsan, U. Weimar, and S. Marco, "On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions," *IEEE Sensors J.*, vol. 6, no. 3, pp. 770–783, Jun. 2006.
[24] R. C. Luo and T. J. Hsiao, "Dynamic wireless indoor localization incorporating with an autonomous mobile robot based on an adaptive signal model fingerprinting approach," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 1940–1951, Mar. 2019.
[25] Y. Liu, G. Yang, M. Li, and H. Yin, "Variational mode decomposition denoising combined the detrended fluctuation analysis," *Signal Process.*, vol. 125, pp. 349–364, 2016.
[26] Q. Tian *et al.*, "New security mechanisms of high-reliability IoT communication based on radio frequency fingerprint," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7980–7987, Oct. 2019.
[27] J. Zhang *et al.*, "SmartSO: Chinese character and stroke order recognition with smartwatch," *IEEE Trans. Mobile Comput.*, vol. 20, no. 7, pp. 2490–2504, Jul. 2021.
[28] T. T. Um *et al.*, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. ACM Int. Conf. Multimodal Interact.*, 2017, pp. 216–220.
[29] P. Sethuraman, "A comparison of DNN, CNN and LSTM using TF/Keras," 2020. [Online]. Available: https://towardsdatascience.com/a-comparison-of-dnn-cnn-and-lstm-using-tf-keras-2191f8c77bbe
[30] Y. Xun, J. Liu, N. Kato, Y. Fang, and Y. Zhang, "Automobile driver fingerprinting: A new machine learning based authentication scheme," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1417–1426, Feb. 2020.
[31] J. He, L. Cai, C. Zhao, P. Cheng, and X. Guan, "Privacy-preserving average consensus: Privacy analysis and algorithm design," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 5, no. 1, pp. 127–138, Mar. 2019.
[32] J. Chen, J. He, L. Cai, and J. Pan, "Disclose more and risk less: Privacy preserving online social network data sharing," *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 6, pp. 1173–1187, Nov./Dec. 2018.
[33] R. C. Luo and T.-J. Hsiao, "Indoor localization system based on hybrid Wi-Fi/BLE and hierarchical topological fingerprinting approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10791–10806, Nov. 2019.
[34] J. Zhang, H. Bi, Y. Chen, M. Wang, L. Han, and L. Cai, "SmartHandwriting: Handwritten Chinese character recognition with smartwatch," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 960–970, Feb. 2020.
[35] L. Wang *et al.*, "Unlock with your heart: Heartbeat-based authentication on commercial mobile phones," *Proc. ACM Interactive Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–22, 2018.
[36] J. He, L. Cai, P. Cheng, J. Pan, and L. Shi, "Distributed privacy-preserving data aggregation against dishonest nodes in network systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1462–1470, Apr. 2019.

Hongliang Bi (Member, IEEE) received the B.S. degree from the Xi'an University of Technology, Xi'an, China, in 2013, the M.S. degree from Soochow University, Suzhou, China, in 2015, and the Ph.D. degree from Wuhan University, Wuhan, China, in 2020, all in engineering.

He is currently an Assistant Professor with the School of Cybersecurity, Northwestern Polytechnical University, Xi'an. His research interests include machine learning, IntelliSense, privacy security, Internet of Things, etc.

Jiajia Liu (Member, IEEE) received the B.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2004, the M.S. degree in computer science from Xidian University, Xi'an, China, in 2009, and the Ph.D. degree in information sciences from Tohoku University, Sendai, Japan, in 2012.

He is currently a Full Professor (Vice Dean) with the School of Cybersecurity, Northwestern Polytechnical University, Xi'an, China. He has authored or coauthored more than 100 peer-reviewed papers in many high-quality publications, including prestigious IEEE journals and conferences. His research interests include intelligent and connected vehicles, mobile/edge/cloud computing and storage, Internet of Things security, wireless and mobile ad hoc networks, and space-air-ground integrated networks.

Dr. Liu was the recipient of the IEEE VTS Early Career Award in 2019, IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2017, IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2019, Niwa Yasujiro Outstanding Paper Award in 2012, Best Paper Awards from many international conferences including IEEE flagship events, such as IEEE GLOBECOM in 2016 and 2019, IEEE WCNC in 2012 and 2014, IEEE WiMob in 2019, and IEEE IC-NIDC in 2018, and Tohoku University President Award in 2013. He has been actively joining the society activities, such as Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (since May 2018), IEEE TRANSACTIONS ON COMPUTERS (from October 2015 to June 2017), and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (since January 2016); the Editor of the IEEE NETWORK (since July 2015) and IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING (since January 2019); the Guest Editor of top ranking international journals, such as IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, IEEE VEHICULAR TECHNOLOGY MAGAZINE, IEEE INTERNET OF THINGS JOURNAL; and is on technical program committees of numerous international conferences, such as the leading symposium Co-Chair of AHSN Symposium for GLOBECOM 2017, CRN Symposium for ICC 2018, and AHSN Symposium for ICC 2019. He is the Vice-Chair of IEEE IoT-AHSN TC and is a Distinguished Lecturer of the IEEE Communications Society and Vehicular Technology Society.

Nei Kato (Fellow, IEEE) received the B.E. degree from Polytechnic University, Tokyo, Japan, in 1986, and the M.S. and Ph.D. degrees in information engineering from Tohoku University, Sendai, Japan, in 1988 and 1991, respectively.

He was a Strategic Adviser to the President with Tohoku University in 2013. From 2015 to 2019, he was the Director with the Research Organization of Electrical Communication. He is currently a Full Professor and the Dean with the Graduate School of Information Sciences, Tohoku University. He has authored or coauthored more than 450 papers in prestigious peer-reviewed journals and conferences. His research interests include computer networking, wireless mobile communications, satellite communications, ad hoc sensor and mesh networks, UAV networks, smart grid, AI, IoT, big data, and pattern recognition.

Prof. Kato has been the Vice-President (Member and Global Activities) of IEEE Communications Society, since 2018, and he has been the Editor-in-Chief of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, since 2017. He is a Clarivate Analytics Highly Cited Researcher for 2019 and 2020. He is a Fellow of The Engineering Academy of Japan and IEICE.