ORIGINAL RESEARCH

A study on user recognition using 2D ECG based on ensemble of deep convolutional neural networks

Min-Gu Kim[1] · Hoon Ko[1] · Sung Bum Pan[1]

Abstract

The risk of tampering exists for conventional user recognition methods based on biometrics such as face and fingerprint. Recently, research on user recognition using biometric signals such as electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG) has been actively performed to overcome this issue. We herein propose a user recognition method applying a deep learning technique based on ensemble networks after transforming ECG signals into two-dimensional (2D) images. A preprocessing process for one-dimensional ECG signals is performed to remove noise or distortion; subsequently, they are projected onto a 2D image space and transformed into image data. For the proposed algorithm, we designed deep learning-based ensemble networks to improve the degraded performance arising from overfitting in a single network. Our experimental results demonstrate that the proposed ensemble networks exhibit an accuracy that is 1.7% higher than that of the single network. In particular, the performance of the ensemble networks is up to 13% higher compared to the single network that degrades the recognition rate by displaying similar features between classes.

Keywords Biometrics · Electrocardiogram · Ensemble networks · CNNs · User recognition

1 Introduction

With the rapid progress toward an information society and the increase in various systems and devices, research for identifying individuals have been actively performed and the results are used in real life. Recently, research on user recognition using biometric signals such as ECG, EEG, and EMG has been conducted. The major benefits of user recognition methods using biometric signals are as follows: (1) identification is difficult to forge using the signals generated inside the body compared with methods using anatomical and physical forms such as the face and fingerprints, (2) it is obtainable from all living individuals, (3) it includes information regarding the clinical or psychological state of the users, and (4) the signal waveforms do not change

significantly over time, thus rendering the easy re-identification of the users (Ogiela and Ogiela 2016).

Among the biometric signals, the ECG signal is not stimulated and is difficult to modulate. Thus, it is considered the next-generation user recognition technology. The ECG signal is outputs 12 waveform according to the attachment position of the sensor, different for each individual, depending on factors such as the location, size, and structure of the heart, and the age and sex of the individuals. Specific feature points exist for normal ECG signals. We can extract the features for user recognition by aptly using these feature points. Therefore, we can recognize the individuals using unique individual features, and recognize the users using the ECG signal regardless of the position of the users. The existing ECG-based user recognition technology used various machine-learning techniques such as SVM (Mehta and Lingayat 2008), K-NN (Zhao and Zhang 2005), and random forest (Khazaee and Zadeh 2014).

Recently, research on user recognition technology using deep learning with automatic feature extraction and learning in a learning process without an additional feature extraction process has been performed. However, performance limitations exist because a single network cannot learn all data that are difficult to recognize in the existing neural network

✉ Sung Bum Pan
  sbpan@chosun.ac.kr

  Min-Gu Kim
  happy9433@gmail.com

  Hoon Ko
  skoh21@chosun.ac.kr

[1] Chosun University, 309, Pilmun-daero, Dong-gu, Gwangju 61452, Republic of Korea

composed of a single structure network. If learning is performed on all the data that are difficult to recognize, overfitting occurs during the learning process, thus leading to performance degradation (Hawkins 2004).

We herein propose a deep-learning-based ensemble network method that relearns the excellent features output from multiple networks for data that are difficult to learn in a single network. Before transforming one-dimensional (1D) ECG signals into 2D images, the noise generated from an ECG measuring device, the muscle noise, and the baseline fluctuation noise caused by the movement of the measurers are removed. The noise-removed 1D ECG signals are projected onto a 2D image space by a periodic segmentation process including P, QRS complex, and T waves. For the proposed algorithm, we design deep learning-based ensemble networks to solve the problem of low recognition rate caused by displaying similar features between classes and overfitting occurring in a single network. Re-training is performed by fusing the excellent features output from each single network. The ECG data used in the experiment is from the MIT-BIH normal sinus rhythm database (NSRDB); we used 4500 pieces of training data, 2700 pieces of verification data, and 1800 pieces of test data. The experimental results show that the recognition rate was 97.3%, 97.9%, and 97.2%, respectively, when using three networks designed for the ensemble networks as a single network. The proposed ensemble networks show the recognition rate of 98.9%, i.e., 1.7% higher than the single network. In particular, it shows that the recognition rate is 13% higher than that of the single network where the recognition rate is degraded by displaying similar features between classes.

The structure of this paper is as follows. Section 2 describes the existing studies on user recognition using ECG. Section 3 presents the proposed ensemble networks using deep-learning-based 2D ECG images. Section 4 analyzes the performance of the proposed method. Section 5 concludes this paper and discusses the future research.

2 Related works

The risk of tampering exists for user recognition methods using biometrics such as the face and fingerprints, and a special place is required to implement a user recognition system. Recently, research on user recognition using biometric signals inside the human body such as ECG, EEG, and EMG has been performed. In particular, various types of research have been conducted on ECG-based user recognition because ECG exhibits unique individual features owing to the electrophysiological factor, location and size of the heart, and physical conditions of individuals.

Israel et al. proposed an ECG-based recognition system using temporal features. After removing the noise of the

ECG signals, they conducted research on user recognition technology by applying the features of RQ, RS, RP, RL, RP, RT, RS, RT, P width, T width, ST, PQ, PT, LQ, and ST to the linear discriminant method, as shown in Fig. 1. The experimental results using the self-obtained ECG DB from 27 people showed the individual recognition accuracy of 100% (Israel et al. 2005). Wang proposed a user recognition technique by fusing time/amplitude features with R waveform features from ECG signals. After detecting the R-peak point in the preprocessed ECG signals and setting it as the reference point, they measured the time and amplitude distance. The user recognition method showed the recognition rate of 98% for 13 subjects using principal component analysis and linear discriminant analysis for morphological features. Although the recognition rate was relatively high, the experiment was conducted with the self-obtained DB and a few subjects (Wang et al. 2007).

Chan et al., proposed a feature extraction framework using a distance measuring set that includes the wavelet transform distance. Data were collected from 50 subjects through electrodes between fingers. They achieved the recognition rate of 89% by applying the wavelet transform distance to the user recognition method. Compared to the previous experiments, they performed their experiments on a large number of subjects yet achieved a low recognition rate (Chan et al. 2008). Chiu et al., proposed an individual recognition method using wavelet and Euclidean classifiers. ECG signals were obtained from 45 subjects, and the Euclidean distance was used after extracting the features using the wavelet transform. Although they performed a relatively simple process, they achieved the low recognition rate of 90.5% (Chiu et al. 2008). Louis et al., proposed a one-dimensional multi-resolution local binary patterns (1DMRLBP) method for feature extraction after dividing periodic ECG signals. Unlike the existing LBP methods, they extracted the features transformed by LBP for the distance between the feature points and point values without
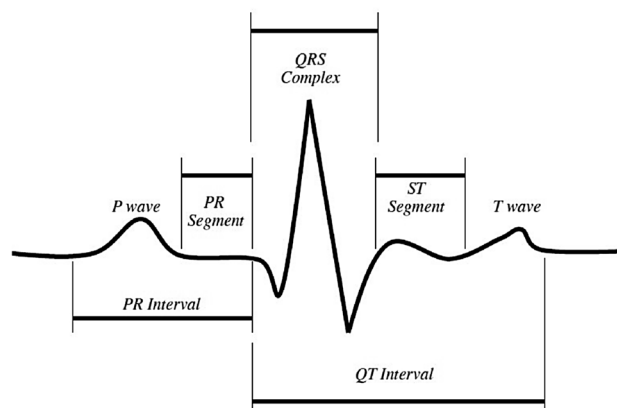

Fig. 1 Example of ECG feature waveform

Table 1 Performance analysis of previous research on user recognition using ECG

Works	Feature	Result	Number of Subject
Israel	LDA	ACC: 100%	27
Wang	PCA, LDA	ACC: 98%	13
Chan	WT	ACC: 89%	50
Chiu	WT	ACC: 90%	45
Louis	LBP, SVM	EER: 7.89%	1020
Hejazi	WT	FMR: 3.5%	52

using a fixed number of points. They achieved an EER of 7.89% for a single user recognition performance using SVM and Bagging for the classification and recognition methods (Louis et al. 2016). Hejazi et al., performed user recognition using adaptive noise cancellation methods such as least mean squares, recursive least squares, and wavelet transform threshold. They transformed high-level input data into low-level data through principal component analysis and used SVM for both nonlinear and linear data classification (Hejazi et al. 2016).

The recent research on user recognition methods using deep learning in various technology fields such as recognition, classification, and prediction has shown excellent performance Table 1. For conventional machine learning methods, extracting meaningful features was an important factor in determining performance. Therefore, experts with background knowledge are required to directly extract the features. Meanwhile, deep learning automatically extracts features without any additional feature extraction such as Fig. 2. In particular, the excellent performance of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been proven in the field of image classification and recognition, and research on user recognition using

deep learning is in progress (Buduma and Locascio 2017). However, deep running technology using ECG focuses only on analyzing a users health condition through heart rate classification (Kiranyaz et al. 2016, 2015; Mathews et al. 2018). Rajpurkar et al. (2017) developed a new network composed of 34 layers that predicts 12 arrhythmias in a single lead ECG signal. Chauhan and Vig (2015) developed a neural network structure that repeatedly used several LSTM layers to detect abnormal ECG signals. Rahhal et al. (2016) conducted research on a neural network structure composed of a feature representation layer and a softmax regression layer. The feature representation layer is learned through autoencoders that is an unsupervised learning method to remove cumulative noise from ECG signals; subsequently, ECG signals were classified through the softmax regression layer.

3 Proposed algorithms for ensemble networks based on 2D-ECG

Figure 3 is a flowchart of an ensemble-network-based user recognition system using ECG signals proposed herein. To build an ECG DB for 2D CNNs, any noise generated while capturing ECG signals is removed by frequency filtering. After detecting the peak of the R wave using the Pan and Tompkins algorithm, a periodic signal segmentation including the P, QRS, and T waves is performed. However, because the baseline fluctuation noise caused by the respiration of measurers is not removed by the filter and the median filter, ECG images are obtained by projecting signals onto a 2D space after estimating the partial baseline using the first-order regression analysis.

Finally, user recognition is processed through deep learning with automatic feature extraction and learning. However, performance limitation exists because a single network cannot learn all the data that is difficult to recognize.
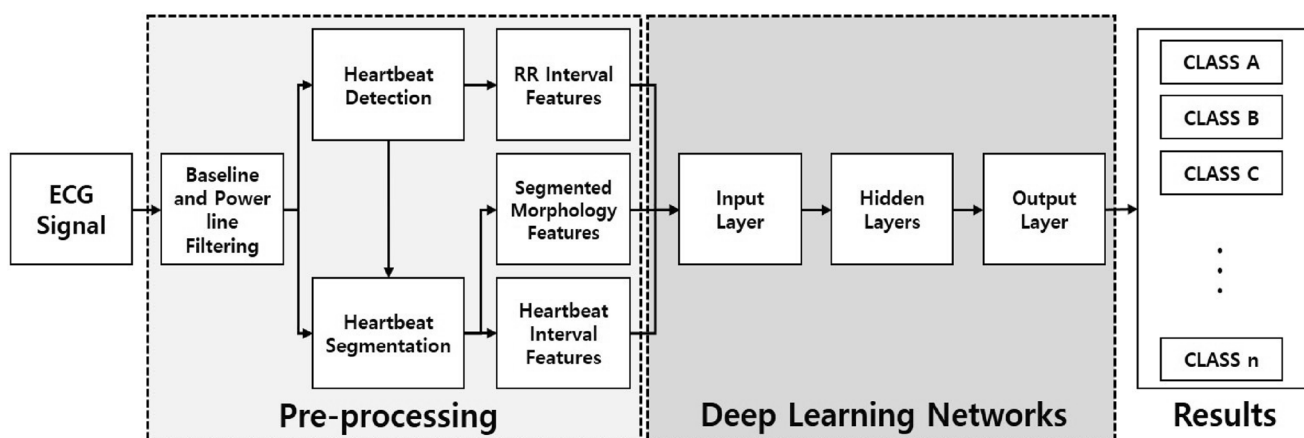
Fig. 2 Flowchart of existing deep learning for ECG classification
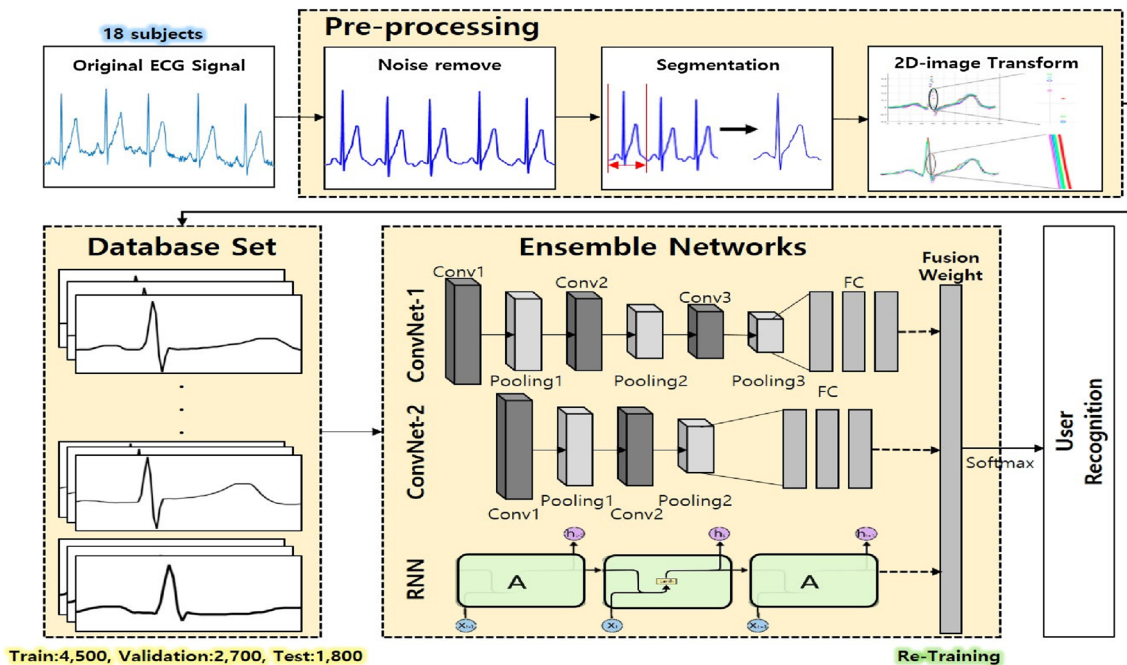
Fig. 3 Flowchart of proposed user recognition system using ECG

If learning is performed on all the data that are difficult to recognize, overfitting occurs during the learning process, thus leading to performance degradation. We herein apply ensemble networks designed with CNNs with a different number of parameters and layers, and RNNs using temporal information.

3.1 Preprocessing

Before 1D ECG signals are transformed into 2D images, a preprocessing process composed of noise removal and correction is performed. The noise from ECG signals is removed by frequency filtering, R wave detection, and median filtering, as shown in Fig. 4 (GyuHo et al. 2018). Frequency filtering uses a bandpass filter to remove power-line noise, muscle noise, and electrode contact noise generated while measuring ECG. The peak of the R wave is detected using the Pan and Tompkins algorithm for ECG signals filtered by the bandpass. Based on the detected peak of the R wave, noise is removed by applying the median filter to the remaining interval excluding the QRS complex interval that contains the unique individual physical feature information.

$$Partial\ baseline = \omega \cdot x + \beta, \omega = \frac{y_2 - y_1}{x_2 - x_1}, \beta = y_1 - \frac{y_2 - y_1}{x_2 - x_1} \cdot x_1.$$

(1)

However, a calibration process is required because the baseline fluctuation noise caused by the respiration of the users

is not removed by the frequency filter and median filter. The technique used to remove the baseline fluctuation noise estimates the partial baseline using a first-order regression analysis.

The partial baseline calculates the first-order regression analysis by applying $M_{TP}(x_1, y_1)$, i.e., the middle value of the ECG T wave and P wave, and $M_{PQ}(x_2, y_2)$, i.e., the middle value of the P wave and Q wave to Eq. (1). Finally, 1D ECG signals are transformed into 2D images through projection and a linear equation for applying to a 2D CNN. Further, 1D ECG signals are projected through Eq. (2) using the amplitude value over time.

$$PS_{ecg} = A_{ecg}(t) + |A_s| + \alpha.$$

(2)

PS_{ecg} is the pixel position of ECG, and A_{ecg} is the ECG amplitude value according to time t. When 1D ECG signals are projected onto a 2D space, data loss occurs between pixels owing to inconsistent voltage values. Therefore, 1D ECG signals are projected onto a 2D image space by minimizing data loss using a linear equation.

3.2 Ensemble architecture based on deep convolutional neural networks

Figure 5 shows the proposed ensemble networks composed of two CNN models with different numbers of layers and parameter values and one RNN that can use temporal information. The CNN learns by autonomously extracting features using a certain size filter for a continuous input
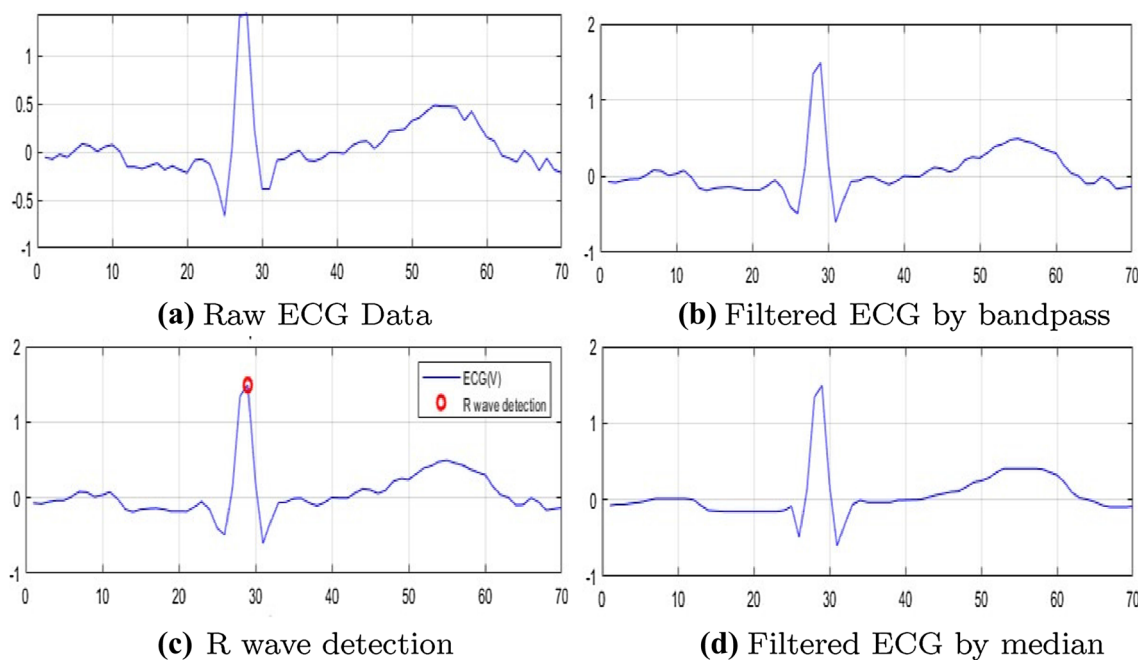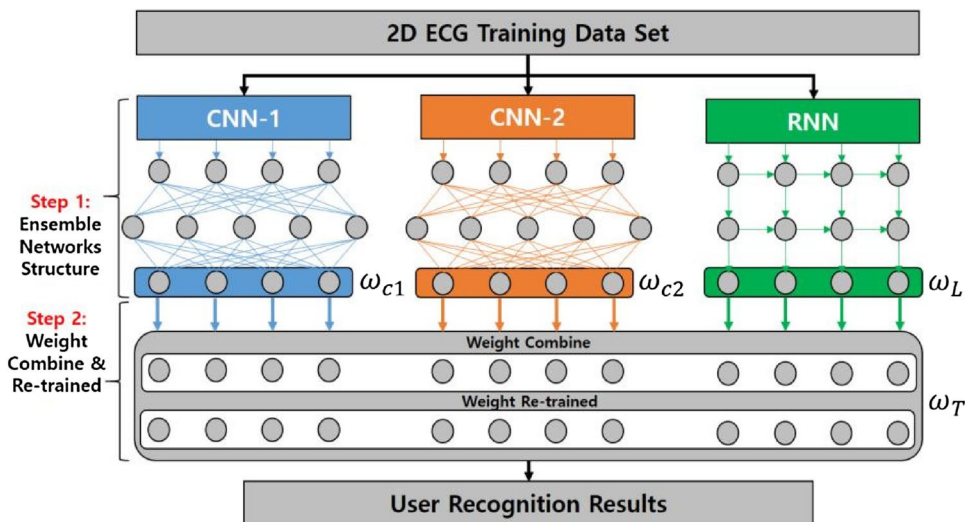
(a) Raw ECG Data

(b) Filtered ECG by bandpass

(c) R wave detection

(d) Filtered ECG by median

Fig. 4 Noise removal process

Fig. 5 Structure of proposed ensemble networks

information. When extracting features, it does not use the input information as it is but extracts only the meaningful specific information. Subsequently, using it for learning, it reduces the amount of complex computation occurring during the learning process and learns only the necessary information. Various features can be extracted by changing the filter size. Features extracted using one or several filters will form one layer. The ConvNet-1 structure used in this study comprises three convolution layers, three max-pooling layers, and three fully connected layers.

ConvNet-2 comprises two convolution layers, two max-pooling layers, and three fully connected layers.

RNN is a neural network with the additional concept of using temporal information in the general neural network. It has the advantage of using the previous information at the present time by adding a weight that returns to itself from the hidden layer. However, the RNN has limitations in storing the previous information; thus, it uses the long short-term memory (LSTM) structure to overcome this issue. As shown in Fig. 6, three gates are used to control the saving of
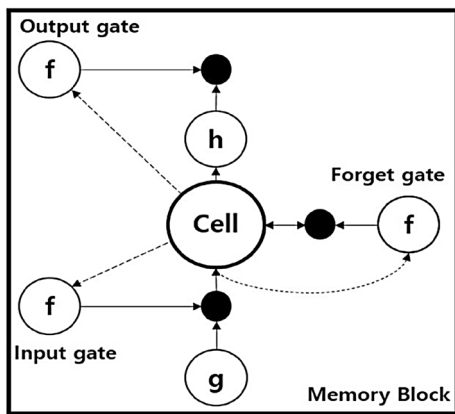
Fig. 6 Basic structure of LSTM

the previous information and one node in the basic neural network is replaced by one memory block. The input gate controls the input data at the present time, as shown in Eq. (3); the output gate controls the value at the present time at the output node, as shown in Eq. (4). Finally, the forget gate controls the saving of the current value into the cell, as shown in Eq. (5).

$$\alpha_l^t = \sum\nolimits_{i=1}^{I} \omega_{il} \times x_i^t + \sum\nolimits_{h=1}^{H} \omega_{il} \times b_h^{t-1} + \sum\nolimits_{c=1}^{C} \omega_{cl} \times S_c^{t-1}, \tag{3}$$

$$\alpha_\omega^t = \sum\nolimits_{i=1}^{I} \omega_{i\omega} \times x_i^t + \sum\nolimits_{h=1}^{H} \omega_{i\omega} \times b_h^{t-1} + \sum\nolimits_{c=1}^{C} \omega_{c\omega} \times S_c^{t-1}, \tag{4}$$

$$\alpha_\emptyset^t = \sum\nolimits_{i=1}^{I} \omega_{i\emptyset} \times x_i^t + \sum\nolimits_{h=1}^{H} \omega_{i\emptyset} \times b_h^{t-1} + \sum\nolimits_{c=1}^{C} \omega_{c\emptyset} \times S_c^{t-1}. \tag{5}$$

In this study, the LSTM structure comprises two hidden layers and one fully connected layer. As shown in Fig. 5, it fuses the weight values w_{c1}, w_{c2} output from the CNN model with the weight value w_L output from the LSTM. The fused weight value w_t is an excellent feature output from each single network model. After the Re-training process of w_t to improve the performance of user recognition, the performance of user recognition is verified using softmax regression suitable for data analysis, prediction, and identification as multiple classification is possible.

4 Discussion

4.1 Experimental results

We used the MIT-BIH NSRDB that is an ECG database measured with 128 sampling points for 18 subjects 5 men (aged 26–45 years) and 13 women (aged 20–50 years). We used 4,500 pieces of training data, 2700 pieces of verification data, and 1800 pieces of test data as the input data

for this research. We used the precision, recall, f1-score, and accuracy that were used as the performance evaluation criteria in the pattern recognition field for the performance analysis of each class. The precision, recall, f1-score, and accuracy for each class was calculated through Eqs. (6), (7), (8), and (9), respectively, with true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Figure 7 shows the results of the confusion matrix for the performance of user recognition using the proposed ensemble network-based ECG signals, and for the performance of user recognition using single network based ECG signals. They are the results of inputting 1800 pieces of experimental data into the learned neural network model. The column represents the ground truth and the row represents the number of ECG data recognized by the proposed method. For the clarification of the confusion matrix, out of 100 data, the 15th row of the O-class in Fig. 7b is divided into 89 ECG data (TP) that are correctly recognized as O-class, and the 11 ECG data (FN) that are misrecognized as F-class and Q-class. In addition, we confirmed that 1708 ECG data (TP) that are correctly recognized as classes other than O-class, and 2 ECG data (FN) that are misrecognized as O-class instead of being recognized as B-class and C-class, separately.

Table 2 shows the results of user recognition using 18 ECG signals with this analysis method.

$$Precision = \frac{TP}{TP + FP}, \tag{6}$$

$$Recall = \frac{TP}{TP + FN}, \tag{7}$$

$$F1\ Score = 2 \times \frac{Precision \times recall}{Precision + recall}, \tag{8}$$

$$Accuracy = \frac{Total\ correctly\ classified\ signals}{Total\ number\ of\ signals} \times 100. \tag{9}$$

4.2 Experimental analysis

To confirm the excellence of the proposed method, we compared the performance of a single network with that of the proposed ensemble networks. Among 1800 ECG signals, ConvNet-1, a single network, recognized 1752 pieces and misrecognized 48 pieces, thus achieving the recognition rate of 97.3%. ConvNet-2 recognized 1763 pieces and misrecognized 37 pieces, thus achieving the recognition rate of 97.9%. The RNN recognized 1750 pieces and misrecognized 50 pieces, thus showing the lowest recognition performance with the rate of 97.2%. The proposed ensemble network model recognized 1780 pieces and misrecognized 20 pieces,
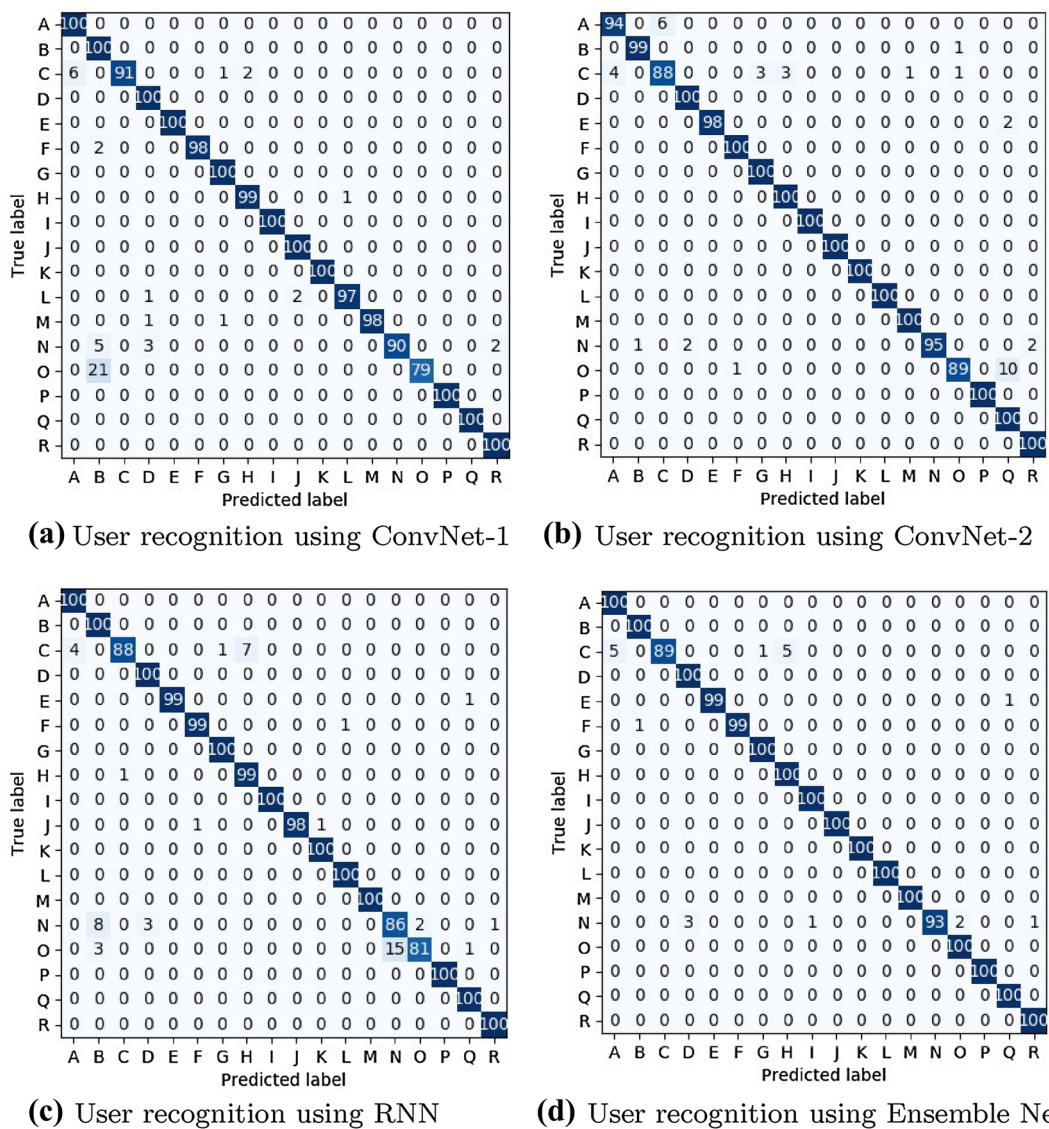
(a) User recognition using ConvNet-1

(b) User recognition using ConvNet-2

(c) User recognition using RNN

(d) User recognition using Ensemble Networks

Fig. 7 User recognition results using multi-class classification confusion matrix

	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
ConvNet-1	0.98	0.97	0.97	97.3
ConvNet-2	0.98	0.98	0.98	97.9
RNN	0.97	0.97	0.97	97.2
Proposed method	0.99	0.99	0.99	98.9

Table 2 Comparison results of user recognition

thus showing the best performance with the recognition rate of 98.9% that is 1.7% higher than that of the RNN.

In particular, when comparing the ECG data of N-class and O-class with those of other classes, they exhibit the lowest values of the variations in P, QRS, T wave intervals, and display similar signal waveforms, thus demonstrating

the low recognition rate when using a single network. However, when applied to the ensemble network model, they exhibit the highest recognition rate of 96.5%, as shown in Fig. 8, and the performance difference is 4.5% at the minimum, and 13% at the maximum when compared with a single network.

Fig. 8 User recognition accuracy rate in similar feature class type
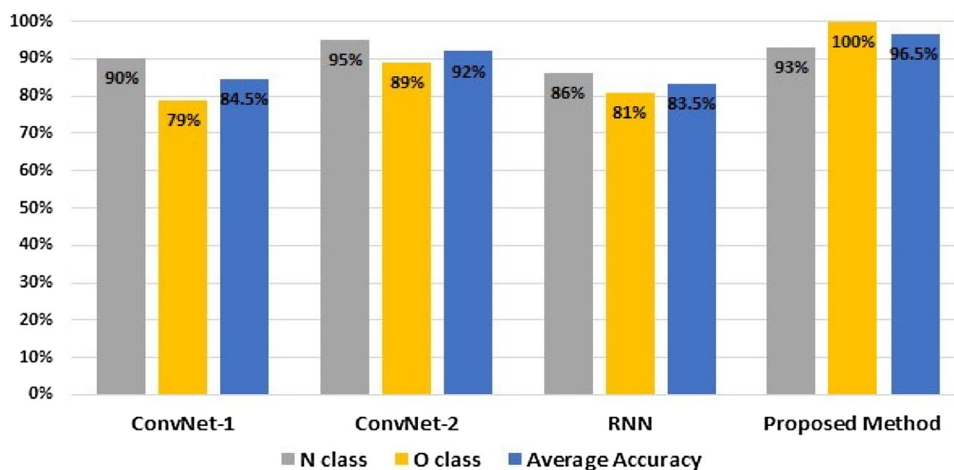
5 Conclusion

We herein proposed a user recognition method based on ensemble networks using ECG signals. To apply 1D ECG signals to the CNN to exhibit excellent performance in image recognition, classification, and prediction, we transformed them into 2D images after noise removal and a periodic segmentation process. Further, to process data that are difficult to learn in a single network, we designed an ensemble network that relearned the excellent features extracted from each single network and applied it to a user recognition system. The performance analysis results of the proposed results show that the ensemble network exhibited a higher accuracy rate of 1.7% at the maximum compared to the single network. In particular, it showed a better performance that is up to 13% higher compared to the single network for the recognition rate of the classes that display the similar features, thus solving the problems occurring in a single network.

In the future, we plan to conduct research on user recognition for real-life applications by building our own DBs according to state changes such as ECG measurement position, user exercise, sleeping, and post-drinking. We will also improve the recognition performance through various network designs and combinations according to user changes.

References

Buduma N, Locascio N (2017) Fundamentals of deep learning: designing next-generation machine intelligence algorithms. O'Reilly Media Inc, Sebastopol

Chan AD, Hamdy MM, Badre A, Badee V (2008) Wavelet distance measure for person identification using electrocardiograms. IEEE Trans Instrum Meas 57(2):248–253

Chauhan S, Vig L (2015) Anomaly detection in ECG time signals via deep long short-term memory networks. In: Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on, IEEE, pp 1–7

Chiu CC, Chuang CM, Hsu C-Y (2008) A novel personal identity verification approach using a discrete wavelet transform of the ECG signal. In: 2008 International Conference on Multimedia and Ubiquitous Engineering, IEEE, pp 201–206

GyuHo C, JaeHyo J, HaeMin M, YounTae K, SungBum P (2018) User authentication system based on basedline corrected ECG for biometrics. Intell Autom Soft Comput (will be published soon)

Hawkins MD (2004) The problem of overfitting. J Chem Inf Comput Sci 44(1):1–12

Hejazi M, Al-Haddad SAR, Singh YP, Hashim SJ, Aziz AAF (2016) ECG biometric authentication based on non-fiducial approach using kernel methods. Digit Sign Process 52:72–86

Israel SA, Irvine JM, Cheng A, Wiederhold MD, Wiederhold BK (2005) ECG to identify individuals. Pattern Recogn 38(1):133–142

Khazaee A, Zadeh AE (2014) Ecg beat classification using particle swarm optimization and support vector machine. Front Comput Sci 8(2):217–231

Kiranyaz S, Ince T, Hamila R, Gabbouj M (2015) Convolutional neural networks for patient-specific ecg classification. In: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, pp 2608–2611

Kiranyaz S, Ince T, Moncef G (2016) Real-time patient-specific ecg classification by 1-d convolutional neural networks. IEEE Trans Biomed Eng 63(3):664–675

Louis W, Komeili M, Hatzinakos Dimitrios (2016) Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ECG biometrics. IEEE Trans Inf Forensics Secur 11(12):2818–2832

Mathews SM, Kambhamettu C, Barner KE (2018) A novel application of deep learning for single-lead ECG classification. Comput Biol Med

Mehta S, Lingayat NS (2008) Svm-based algorithm for recognition of QRS complexes in electrocardiogram. IRBM 29(5):310–317

Ogiela MR, Ogiela L (2016) On using cognitive models in cryptography. In: Advanced information networking applications (AINA), 2016 IEEE 30th International Conference on, IEEE, pp 1055–1058

Rahhal MM Al, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR (2016) Deep learning approach for active classification of electrocardiogram signals. Inf Sci 345:340–354

Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY (2017) Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint arXiv:170701836

Wang Yongjin, Agrafioti Foteini, Hatzinakos Dimitrios, Plataniotis KN (2007) Analysis of human electrocardiogram for biometric recognition. EURASIP J Adv Signal Process 2008(1):148658

Zhao Q, Zhang L (2005) ECG feature extraction and classification using wavelet transform and support vector machines. In: Neural networks and brain, 2005. ICNN&B'05. International Conference on, IEEE, vol 2, pp 1089–1092