Article

Personal Identification Using an Ensemble Approach of 1D-LSTM and 2D-CNN with Electrocardiogram Signals

Jin-A Lee and Keun-Chang Kwak *

Department of Electronic Engineering IT-Bio Convergence System, Chosun University, Gwangju 61452, Korea; kwwlsdk@naver.com

* Correspondence: kwak@chosun.ac.kr; Tel.: +82-062-230-6086

Abstract: Conventional personal identification methods (ID, password, authorization certificate, etc.) entail various issues, including forgery or loss. Technological advances and the diffusion across industries have enhanced convenience; however, privacy risks due to security attacks are increasing. Hence, personal identification based on biometrics such as the face, iris, fingerprints, and veins has been used widely. However, biometric information including faces and fingerprints is difficult to apply in industries requiring high-level security, owing to tampering or forgery risks and recognition errors. This paper proposes a personal identification technique based on an ensemble of long short-term memory (LSTM) and convolutional neural network (CNN) that uses electrocardiograms (ECGs). An ECG uses internal biometric information, representing the heart rate in signals using microcurrents and thereby including noises during measurements. This noise is removed using filters in a preprocessing step, and the signals are divided into cycles with respect to R-peaks for extracting features. LSTM is used to perform personal identification using ECG signals; 1D ECG signals are transformed into the time–frequency domain using STFT, scalogram, FSST, and WSST; and a 2D-CNN is used to perform personal identification. This ensemble of two models is used to attain higher performances than LSTM or 2D-CNN. Results reveal a performance improvement of 1.06–3.75%.

Keywords: personal identification; electrocardiogram signals; deep learning; long short-term memory; convolutional neural network; ensemble approach

Citation: Lee, J.-A.; Kwak, K.-C. Personal Identification Using an Ensemble Approach of 1D-LSTM and 2D-CNN with Electrocardiogram Signals. *Appl. Sci.* **2022**, *12*, 2692. <https://doi.org/10.3390/app12052692>

Academic Editors: Robert Hanus, Dariusz Czerwinski and Volodymyr Mosorov

Received: 17 February 2022

Accepted: 1 March 2022

Published: 4 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid advancement of artificial intelligence (AI) has achieved significant progress and, thereby, has gained substantial attention worldwide. Since the emergence of deep learning (a core technology of AI), the applications of AI technology have been expanded to various fields for increased convenience in daily human life. In addition, it has improved the quality of life through its applications in the medical, agricultural, financial, and autonomous vehicle fields. Notwithstanding the positive influences of technology advancement and diffusion, certain risks are imposed on humans as intelligent cyber-attacks increase [1]. An array of personal authentication technologies has been studied to defend against similar threats.

Personal information has been protected using passwords or OTPs. However, recently, technologies using biometrics for protecting personal information have been employed increasingly to address the risk of loss or theft. These technologies can conveniently and safely manage personal information as well as verify the identity of users. As a consequence, personal authentication technology using biometric information such as voice, gender, face, and behavior is being actively developed and used [2,3]. Seven characteristics are required for biometric recognition. Fundamental characteristics such as universality (indicates whether every individual has the information), uniqueness (indicates the distinctiveness of each individual's information), permanence (indicates whether the information remains unaltered over time and is unmodifiable), and collectability (indicates whether the biometric

can be acquired conveniently using sensors) are required. The following characteristics are required for trusting and using biometrics: accuracy including the speed and precision of receiving and processing information, accessibility that does not induce reluctance to biometric measurement, and safety against inappropriate use to achieve deception security [4].

Biometric recognition involves behavioral characteristics such as voice, gait, electrocardiogram (ECG), and brain waves as well as physical characteristics including the face, fingerprints, and the iris. The advantage of face recognition is that facial features can be identified rapidly and conveniently. However, recognition may be ineffective when facial expressions or lighting vary. Fingerprint recognition is used widely. Here, application sensors are used to capture fingerprint images. Although fingerprint recognition is simple, it is sensitive to skin wounds and impurities and can be forged or tampered with. However, recently, face or fingerprint recognition has been difficult to implement in public places during the pandemic because of the need for everyone to wear facemasks. Although voice recognition is a convenient and safe technology because it uses features extracted from voices, recognition may be difficult because of the use of a recorded file or the presence of noise. Gait recognition is a technology that involves an analysis of a user's gait characteristics, which is considerably influenced by the surrounding environment. The application of biometric information expressed outside the body is encountering severe problems owing to personal damage caused by user recognition error as well as forgery and tampering. Accordingly, research is being actively conducted on personal identification using internal signals of the human body, such as electroencephalogram (EEG), electromyogram (EMG), and ECG.

EEG is a test that measures an electrical signal at the scalp. It is utilized in brain-computer interfaces (BCIs) [5] and medical diagnosis. However, humans are reluctant to attach sensors to their head for signal measurement. Furthermore, signals may be distorted as these pass through the cranial bone. EMG records the signals generated by muscles. It is utilized in motion recognition, medical diagnosis, and rehabilitation treatment. However, signals must be set for each motion. Moreover, the sensors need to be accurately attached to the muscle position. ECG shows the electrical signals of the heart's rhythm through a micro-current, which consists of the P wave (atrial depolarization), QRS complex wave (ventricular depolarization), and T wave (ventricular repolarization). The ECG signal is unique to each individual and depends on the size and position of the heart, gender, and age. The 12-lead ECG involves the measurement of signals from both the wrists. It is utilized for medical diagnosis as well as personal identification. Because ECG measures a signal generated inside the human body, it is difficult to replicate or tamper and is unaffected by environmental transformations. Furthermore, all living humans have these signals [6]. Identification based on physiological characteristics such as EEG, EMG, and ECG has been researched extensively for security methods based on personal signals because of its various advantages. However, this study conducted personal identification using ECG, which displays all the seven characteristics of biometric recognition.

In recent years, research on personal identification using deep learning has been conducted actively. The first study on user recognition using ECG was conducted by Biel [7]. It can be classified into handcrafted and non-handcrafted fiducial-based methods. Handcrafted feature extraction methods are based on the characteristics of ECG, namely, the ECG signal amplitude, time interval (ST interval, PQ interval, and QS interval), peak (minimum and maximum peaks of P, Q, R, S, and T), and angle [8,9]. This method is a type of machine learning where features are extracted directly based on the learning performed. Israel [10] extracted 15 time-interval features by determining the maximum local value based on P, R, and S peaks, tracking the slope, and identifying the position of the minimum curve radius for personal identification using ECG. Jahiruzzaman [11] used the continuous wavelet transform (CWT) for extracting features as a representation of the time domain used for signal processing tasks such as image compression and pattern recognition. Personal identification was performed on the MIT/BIH arrhythmia database using ID matching

by applying the encryption technology to ECG signals through CWT. Zhao [12] used ensemble empirical mode decomposition (EEMD) and Welch spectrum analysis to extract intrinsic mode function (IMF) spectral features for obtaining the morphological and spectral information of signals.

However, handcrafted feature extraction methods entail problems because the detection of peaks causes high variability in signals [13]. Furthermore, these methods display reduced performance while removing noises or extracting features. Consequently, the emergence of deep learning resulted in the increased use of long short-term memory (LSTM). LSTM improves the performance of time-series and convolutional neural networks (CNNs), which display remarkable image classification performance. Deep learning extracts features through learning. Therefore, features need not be extracted directly as in handcrafted feature extraction methods. Non-handcrafted fiducial-based methods need to detect the R-peak for the signal division because characteristic points are not used based on the overall form of ECG signals. For personal identification based on ECG, Labati [14] proposed CNN-based deep-ECG using the PTB database. They performed preprocessing, CNN feature extraction, and identification in that order. A notch filter, infinite impulse response (IIR) filter, and third-order high-pass filter are used in the preprocessing step. The CNN consisting of six convolutional layers, a dropout layer, a fully-connected layer, and a softmax layer performed personal identification. Abdeldayem [15] proposed five approaches for personal identification using ECG. First, signals are distinguished using the cyclic characteristics of ECG signals. Second, the method of dividing and using the blind with constant duration of each segment, which is the period fixed to the ECG signal, can lower the complex calculation and improve the performance. This approach should incorporate the cyclic characteristic of the ECG mentioned above. Third, a noise removal step is not applied because the noise is not applied with a circulatory stop. Fourth, a 2D-CNN is used by transforming it into a power spectral density, which is the frequency domain of signals. Finally, eight open databases are combined into a database. Ciocoiu [16] removed noises through a band-pass filter in a preprocessing step for ECG signals and divided the cycle by a constant time with respect to the R-peak. After transforming the data into images using four types of spatial representation (namely, CWT, Gramian angular field (GAF), phase-space trajectories, and recurrence plots), a CNN consisting of three convolutional layers, activation function ReLU, a max pooling layer, a fully-connected layer, and a softmax layer was applied for a comparative analysis of the accuracy and equal error rate (EER) of ECG-based biometric recognition. Y. H. Byeon [17] used a CNN model of transfer learning for various time–frequency representations to examine the performance of ECG biometric recognition. Four transfer learning models using MFCC, spectrogram, log spectrogram, mel spectrogram, and scalogram were employed as a time–frequency representation method. G. H. Choi [18] proposed a personal identification method where multidimensional features are extracted by adjusting the bi-cubic 2D size for maintaining the data values and converting the ECG signals into a spectrogram. Noises are removed through preprocessing, and the signals are divided into a cycle consisting of a P wave, QRS complex wave, and T wave for personal identification. The divided signals are converted into a spectrogram to reduce the image size by 1/2 and 1/4 for identifying users. D. Jyotishi [19] proposed a method of classification by adding the output of LSTM cells for personal identification using ECG signals. In the proposed model, the variations in bit could be observed because the signals were divided into smaller units considering the fluctuations in bit. Moreover, personal identification was performed according to various window lengths. J. S. Kim [20] proposed a personal identification method based on a 2D coupling image using the cycle information of ECG signals. A 2D coupling image uses a CNN consisting of 12 convolutional layers and six max pooling layers for ECG-based user recognition. M. Hammad [21] proposed an end-to-end deep neural network (DNN) for ECG-based authentication. The first model was designed as a 1D-CNN consisting of four convolutional layers, two max pooling layers, two fully-connected layers, and one max pooling layer. The convolution product of a CNN can efficiently extract morphological features from time-series or image

data. The second model was designed as ResNet-Attention. It combines the output of the first class consisting of two convolutional layers, a normalization layer, a ReLU layer, and a dropout layer, with that of the second class consisting of two normalization layers, two ReLU layers, two dropout layers, and two convolutional layers to be used as the input of Attention. Attention inspects the user authentication performance through two dense layers, a ReLU layer and a softmax layer.

In this study, personal identification is carried out based on the ensemble of LSTM and CNN by using ECG. The CU-ECG database constructed at Chosun University is used in the study. Non-handcrafted fiducial-based methods involve the detection of the R-peak of signals and division at certain intervals. Furthermore, short-time Fourier transform (STFT), scalogram, Fourier synchrosqueezed transform (FSST), and wavelet synchrosqueezed transform (WSST) are used as time–frequency representations to convert these into images. For classifying 1D time-series signals, LSTM as well as GoogleNet, VGG-19, and ResNet-101 (which are CNN transfer learning models with remarkable image classification performance) are used to inspect the performance. In addition, the improvement in performance is examined by the ensemble method.

2. Deep Learning Model

2.1. LSTM

LSTM is the architecture of a recurrent neural network (RNN). An RNN is a neural network having a recurrent structure of output and input. Figure 1 shows the basic structure of an RNN. When a sequence with a large number of time steps is used in an RNN, the initial values decrease by the chain rule. This is because the values between -1 and 1 are multiplied by the hyperbolic tangent function ($\tanh$) in the back propagation through time (BPTT) used for training as the network becomes deeper. Therefore, an RNN involves the problem of information loss because the initial input data do not influence the output results owing to the vanishing gradient problem.

Figure 1. Structure of RNN.

An LSTM with a structure more complex than that of an RNN was proposed to solve the long-term dependency problem of an RNN. An LSTM consists of an input gate, forget gate, and output gate for preventing information loss. The sigmoid activation function outputs a value between zero and one to determine the amount of information based on the output value. Thus, it can add or remove the information of the cell state. The sigmoid and hyperbolic tangent functions are used as the activation functions of an LSTM. The input gate determines whether new information is saved in the cell state, whereas the forget gate determines whether past information is deleted from the cell state. Meanwhile, the output gate determines which information is to be output from the cell state. Figure 2 shows the structure of an LSTM.

Figure 2. Structure of LSTM.

Equations (1)–(6) show the process of updating the cell state and the output values of each gate by using the LSTM calculations. h_{t-1} represents the previous state, x_t represents the cell input, and h_t represents the cell output. w and a represent the weight and bias, respectively.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + a_f) \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + a_i) \quad (2)$$

$$\tilde{c}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + a_c) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + a_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

2.2. CNN

Deep learning is a type of machine learning technique. It is a neural network designed to have a structure similar to that of a neuron of the human brain. It refers to a DNN consisting of multiple layers including input layer, hidden layer, and output layer. A DNN has at least two hidden layers. Earlier, shallow neural networks could not perform complex computations, and vanishing gradient or overfitting occurred during the learning process. However, a DNN enables learning and yields high performance by solving similar problems.

A CNN is a type of deep learning architecture. It is most widely used for image and time-series data. A CNN is a highly appropriate architecture for analyzing and processing 2D data because features are extracted from input data through convolution products. A CNN consists of a repeating convolutional layer, ReLU activation function layer, and pooling layer. Figure 3 shows the basic structure of a CNN.

Figure 3. Structure of CNN.

A convolutional layer is for extracting features from input data through the convolution product. The computation function outputs values by adding and multiplying each element of a moving filter and the filter size image. Padding is the process of filling the surrounding values of input data with zero. It prevents the size of input data from decreasing by the convolution product computation for adjusting the output size. Stride is the process of performing the convolution product when the filter moves according to the stride value by the interval in which the filter is applied to the input image. An activation function is a non-linear function positioned between a convolutional layer and pooling layer. It includes sigmoid, ReLU, step function, hyperbolic tangent, and softmax functions. ReLU is mostly used as an activation function. The ReLU function is expressed as zero and one. Herein, a negative value is output as zero, and any value higher than zero would be output directly as the input. Equation (7) presents the ReLU function.

$$R_x = \max(0, x), R_x = \begin{cases} 0 & (x < 0) \\ x & (x \geq 0) \end{cases} \quad (7)$$

A pooling layer reduces the dimensions while maintaining important features of an image. Pooling layers are of several types such as max pooling, average pooling, and L2-norm pooling. A max pooling layer is most commonly used where the maximum values of each domain are expressed for the target domain. A fully-connected layer is used for classifying images in 1D form. In this layer, a neuron of a previous layer is connected with one of the next layer. A softmax layer shows the final classification result as a probability where the sum of output values is always one. Accordingly, a CNN demonstrates remarkable performance in image classification by adding a convolutional layer and pooling layer to a conventional neural network.

3. Proposed Personal Identification Based on Ensemble of LSTM and 2D-CNN

3.1. LSTM

An LSTM neural network is used for identifying the sequential information to analyze 1D time-series or sequence signals. An LSTM consists of a sequence input layer for entering time-series or sequence data, an LSTM layer for training long-term dependency between time-steps of a sequence, a fully-connected layer for classifying class labels, a softmax layer, and a classification layer. Because the classification accuracy improves as units consisting of the numbers of hidden layers and three cells increase, the deep structure of an LSTM neural network can be expanded by adding LSTM layers. Figures 4 and 5 show examples of one LSTM layer and two LSTM layers, respectively.

Figure 4. One LSTM layer.

Figure 5. Two LSTM layers.

3.2. Time–Frequency Transform

Because physiological signals are affected considerably by noises, the data are transformed into the time–frequency domain and expressed in a 2D image for signal analysis [22]. ECG signals use 2D images transformed by STFT, scalogram, and FSST through a time–frequency transform. The images expressed through a time–frequency transform are classified with a CNN, which displays remarkable image-classification performance.

3.2.1. STFT

A Fourier transform is a frequency representation where time-series signals are decomposed into a frequency. Frequencies observed in signals can be analyzed. However, the variations over time are not considered. The conventional Fourier transform is insufficient to analyze images because the position of each frequency with respect to time cannot be identified [23]. Therefore, STFT and DTFT have been researched to overcome these drawbacks.

STFT divides a long signal that varies over time into shorter lengths to apply the Fourier transform. The presence of a specific frequency at a specific time can be identified when signals are divided into shorter time-lengths, and whether a specific frequency is present within a specific time can be identified when signals are divided into longer time-lengths. That is, a smaller window width is more advantageous for time resolution, whereas a larger one is more advantageous for frequency resolution. Figure 6 shows the image of the time–frequency representation by STFT.

Equation (8) shows the process of dividing the signals of STFT, which is expressed in terms of signals and a moving window function. Equation (9) is the Fourier transform computation of STFT in terms of the signal $x(t)$ and window function $w(t)$ with respect to time t .

$$x(a, t) = x(t) \times w(t - a) \quad (8)$$

$$x(a, v) = \int_{-\infty}^{\infty} x(t) w(t - a) e^{-it} \quad (9)$$

3.2.2. Scalogram

A scalogram is the absolute value of the continuous wavelet transform (CWT) with respect to signals. A wavelet is explained below because the CWT is computed for expressing a scalogram. STFT complements the drawback of the conventional Fourier transform (namely that whereas detailed information on either of time and frequency according

to the window length can be obtained, it is difficult to obtain the information of both time and frequency owing to a fixed window length). A wavelet transform has been recommended to overcome this limitation of STFT. In addition, wavelet transform can be time, and frequency can be simultaneously identified in the CWT domain. A wavelet transform increases the time resolution in the signals of the high-frequency domain while lowering the frequency resolution and increases the frequency resolution in the signals of the low-frequency domain while decreasing the time resolution. Thereby, both time and frequency information can be identified simultaneously. Thus, it is efficient for analyzing discontinuous signals. The conventional Fourier transform uses an infinite sine function in the time domain, whereas a wavelet transform uses a mother wavelet function that is limited in the time domain. This enables signals to be analyzed in time and frequency domains through the scaling and shifting of signals [24]. Equation (10) shows the wavelet transforms (CWT). m and n mean resizing the mother wavelet and shifting of the mother wavelet, respectively. $h(t)$ is the input signal, and ψ is the mother wavelet.

$$G(m, n) = \frac{1}{\sqrt{m}} \int_{-\infty}^{\infty} h(t) \psi\left(\frac{t-n}{m}\right) dt \quad (10)$$

Figure 6. Time–frequency representation by STFT.

There are various types of mother wavelets, and different analysis results are expressed depending on the wavelet type. Therefore, an appropriate type of a mother wavelet must be used for each analysis type. The mother wavelets used for the CWT include Morse, Morlet, and bump wavelets. Morse wavelets are suitable for analyzing signals having time, frequency, and amplitude [25]. Equation (11) represents the Fourier transform of the Morse wavelet. Here, $P(\omega)$ is the unit time; $h_{T,\mu}$ is the normalization constant; and μ and T^2 are the products of parameters for representing the symmetry of the Morse wavelet, time, and bandwidth. α is a damping or compressing parameter rather than the product of time and bandwidth. Here, Equation (12) shows the Morse wavelet equation in terms of the parameterization of α and μ . Two parameters can be adjusted as required to represent the Morse wavelet [26]. T^2 is the product of time and bandwidth and is proportional to the wavelet duration, which varies over time. In addition, duration is the frequency at which the maximum peak frequency is positioned in the center of the window. Here, Equation (13) expresses the maximum peak frequency. μ controls the symmetry of a wavelet, which varies over time.

$$\psi_{T,\mu}(\omega) = P(\omega) h_{T,\mu} \omega^{\frac{T^2}{\mu}} e^{-\omega^\mu} \quad (11)$$

$$\psi_{\alpha,\mu}(\omega) = P(x) h_{\alpha,\mu} \omega^\alpha e^{-\omega^\mu} \quad (12)$$

$$\left(\frac{T^2}{\mu}\right)^{\frac{1}{\mu}} \quad (13)$$

Figure 7 shows the Morse wavelet according to T^2 when $\mu = 3$. Here, Figure 7a shows the Morse wavelet transform result when $T^2 = 10$, whereas Figure 7b displays the Morse wavelet transform result when $T^2 = 60$. Based on Morse wavelet (3.10) and Morse wavelet (3.60), it can be concluded that Morse wavelet (3.60) has a higher frequency resolution than Morse wavelet (3.10).

Figure 7. Morse wavelet according to T^2 . (a) $T^2 = 10$. (b) $T^2 = 60$.

3.2.3. FSST

Signals in a vibration mode, such as physiological signals and voice, can be expressed as overlapped amplitude or frequency modulation. Time–frequency (TF) analysis is expressed as the sum of analysis signals as shown in Equation (14). $f(x)$ and $\varnothing_k(x)$ are the time–frequency amplitude and phase, respectively, of the analyzed signal $X_k(x)$; K is the number of analyzed signals; and $j = \sqrt{-1}$. FSST clearly expresses a time–frequency representation based on STFT used in a spectral function. It is used as a transform technique for maintaining the time resolution at a level similar to that of the original signals [27]. Equations (15)–(17) show the process of calculating FSST. Figure 8 shows the image of the time–frequency representation by FSST.

$$f(x) = \sum_{k=1}^K f_k(x) = \sum_{k=1}^K X_k(x) e^{j2\pi\varnothing_k(x)} \quad (14)$$

$$V_h f(x, \mu) = \int_{-\infty}^{\infty} f(t) h(t-x) e^{-j2\pi\mu(t-x)} dt \quad (15)$$

$$R_h f(x, \omega) = \int_{-\infty}^{\infty} V_h f(x, \mu) \delta(\omega - \Omega_h f(x, \mu)) d\mu \quad (16)$$

$$\Omega_h f(x, \mu) = \frac{1}{j2\pi} \frac{\partial}{\partial x} V_h f(x, \mu) = \mu - \frac{1}{j2\pi} \frac{V_{\partial^2/\partial^2} f(x, \mu)}{V_g f(x, \mu)} \quad (17)$$

3.2.4. WSST

A time–frequency representation wherein signal energy is reallocated from frequency compensates for the scattering effect caused by the mother wavelet. Unlike other methods where energy is reallocated for time–frequency representation, synchrosqueezing maintains the time resolution, and energy is reallocated in the direction of frequency [28]. In addition, synchrosqueezing uses the first derivative for the CWT. Furthermore, signal reconstruction is feasible because the CWT inherits reversibility, whereas the synchrosqueezing transform inherits the properties of the CWT. The algorithm of WSST is as follows:

[Step 1] Determine the CWT of the input signal as shown in Equation (18).

$$G(m, n) = \frac{1}{\sqrt{m}} \int_{-\infty}^{\infty} h(t) \psi\left(\frac{t-n}{m}\right) dt \quad (18)$$

[Step 2] Extract the instantaneous frequency information from the output of CWT for expressing in synchrosqueezing. The equation represents 19. Meanwhile, a and b are scaling and shifting parameters, respectively.

$$h_g = -i \frac{\frac{dG(m,n)}{dn}}{G(m,n)} \quad (19)$$

[Step 3] A phase transform compresses the CWT in a certain domain so that the instantaneous frequency value is reallocated as an individual value. Accordingly, WSST results in an output with a high resolution.

Figure 8. Time–frequency representation by FSST.

Figure 9 shows the image of time–frequency representation by WSST.

Figure 9. Time–frequency representation by WSST.

3.3. 2D Transform-Based CNN

Images that are represented through time–frequency transforms such as STFT, scalogram, FSST, and WSST use a CNN, which is highly capable of image classification. A

CNN can be designed directly to examine its performance. Alternatively, transfer learning including a CNN model can be used to inspect the performance. A substantial amount of data is required for training a CNN-based deep learning model. In addition, the training time is long for a complex model, where the number of layers and hyperparameters need to be adjusted according to data. Therefore, transfer learning (a CNN model) is used. Transfer learning such as AlexNet, GoogleNet, VGG, ResNet, and SqueezeNet involves training new data using a previously developed model and is applicable for cases with a marginal amount of data. GoogleNet, VGG-19, and ResNet-101 are used as a 2D transform-based CNN. As shown in Figure 10, GoogleNet is a DNN with 22 trainable layers and 9 inception models. Furthermore, GoogleNet consists of parallel convolution filters with outputs connected from the inception modules [29]. Figure 11 shows the inception module with four types of operations: 1×1 convolution product, 1×1 convolution product + 3×3 convolution product, 1×1 convolution product + 5×5 convolution product, and 3×3 max pooling + 1×1 convolution product. These operations are used to reduce the number of operations required by decreasing the number of parameters and adjusting the number of channels. The loss values generated during the training are added to prevent the vanishing gradient problem in DNNs such as GoogleNet.

Figure 10. Structure of GoogleNet.

Figure 11. Inception module.

VGG-19 consists of 5 blocks with 19 layers including a convolutional layer and max pooling layer. Furthermore, it uses the smallest 3×3 filter for the convolution product [30]. Figure 12 shows the structure of VGG-19.

ResNet-101 includes 104 convolutional layers and consists of 33 hierarchical blocks. A block consists of a 1×1 convolution product, 3×3 convolution product, and 1×1 convolution product. Here, the operation can be reduced by a bottleneck layer. A residual connection is also added to solve the vanishing gradient problem. x is expressed as $P(x) = F(x) + x$ because it is added to the output through residual connection. Thus, the vanishing gradient problem can be resolved, and a more remarkable performance can be

achieved through deeper layers of a neural network [31]. Figure 13 shows the structure of ResNet-101, whereas Figure 14 shows the residual connection.

Figure 12. Structure of VGG-19.

Figure 13. Structure of ResNet-101.

Figure 14. Residual Connection.

3.4. Proposed Ensemble-Based Personal Identification

The performance of various models can be compared by training these and selecting those with a higher performance. However, when an individual model with remarkable performance is used, the performance improvement in other models can be omitted. Thus, the performance can be improved further by combining different models. An ensemble is a technique for improving performance by combining different models. It can demonstrate performance higher than that of the individual models. Computation, representation, and statistics can be improved by deep learning models and ensembles [32]. An ensemble uses the voting method, average, maximum, and multiplication for the output values of each model to predict the final results.

In this study, the deep learning models LSTM and 2D-CNN are combined for personal identification. The numbers of hidden layers and units are increased to enhance the classification accuracy, and LSTM layers are added to deepen the LSTM. Furthermore, 1D ECG signals are transformed to 2D data by using time–frequency representation methods. Three pre-trained CNN models (namely, GoogleNet, VGG-19, and ResNet-101) are used. An ensemble method for combining the output values of two models is used to enhance the performance of an individual model. An ensemble performs personal identification by determining the final prediction results using the output values of each model for an identical input. Personal identification using ECG can be divided into three steps: signal preprocessing, feature extraction and learning, and personal identification (see Figure 15).

Figure 15. Flowchart of proposed ensemble-based personal identification.

In the first step of signal preprocessing, noises need to be removed for maintaining the shape of ECG signals. This is because the signals can be distorted by various noises during measurement. In an ECG, breathing, friction between skin and electrode, noise from muscles, and noise from contact between electrode and power line are included during measurements. In this paper, the noises of ECG signals were removed using a low-pass filter. A low pass filter is a filter that passes through a frequency signal below the cutoff frequency. This filter is used to remove high frequency components such as muscle noise, 60 Hz power line noise, and electrode contact noise and to soften the signal. These noises are removed through a low-pass filter as shown in Figure 16a,b. In Figure 16b, where noises have been removed, the baseline of the signal shifts above or below the x -axis of the signal. Fluctuations in the baseline are low-frequency vibrations generated by the breathing, sweat, or movement of a subject. These cause variations in the impedance between electrode and skin. The baseline functions as the criteria for detecting characteristic points of ECG signals. Noises must be removed from the baseline because the morphological characteristics of ECG signals cannot be identified if the fluctuations in the baseline occur. Figure 16c shows the signals wherein the baseline has been calibrated with respect to zero, and Figure 16d shows the standardized signals.

ECG signals consist of a P wave, QRS complex wave, and T wave and include various cycles (see Figure 17a). Signals are divided to extract features of the ECG signals. As shown in Figure 17b, the R-peak is detected to divide the signal into a cycle with respect to the detected R-peak.

The second feature extraction and learning stage is an important step that affects classification performance results. For this purpose, electrocardiogram data is applied to LSTM and 2D-CNN to extract and learn features, respectively. In order to use 2D-CNN, one-dimensional ECG signals need to be converted into two-dimensional images. Therefore, it is converted through four time-frequency representations such as STFT, Scalogram, FSST, and WSST, which are transformation methods, and a pre-trained model, GoogleNet, VGG-19, ResNet-101 is used. The size of each image is reduced to 224×224 , and Adam is applied as a learning method. The minimum value of a loss function can be determined because Adam maintains both the moving average and momentum. The initial learning rate, epoch, and mini-batch size are set appropriately for the model to examine the personal identification performance.

In the third step of personal identification, an ensemble that can display performance higher than that of an individual model is used by combining various models based on the personal identification results in the previous step of feature extraction and learning. An ensemble uses the multiplication of model outputs to determine the results. Figure 18 shows the flowchart of the proposed ensemble-based personal identification using STFT and GoogleNet among the four transform methods and three models used. Two databases classify the subjects by combining the output values of LSTM and CNN through preprocessing.

Figure 16. ECG through preprocessing. (a) Original signal; (b) Filtering; (c) Adjusting the baseline; (d) Standardization.

Figure 17. Detected R-peak and signal divided into one cycle. (a) Detected R-peak; (b) one cycle.

Figure 18. Flowchart of proposed ensemble-based personal identification.

4. Experiment and Results Analysis

The CU-ECG database constructed at Chosun University was used for carrying out personal identification using ECG in this study.

4.1. Database

The CU-ECG database was constructed by Chosun University and includes a total of 100 subjects (89 male and 11 female) aged between 23 and 34. Each subject was seated comfortably in a chair for a 1-lead ECG, which is the potential difference between the right arm and left arm. The signals were recorded for 10 s, once per subject, and a total of 60 times, obtained consecutively. The sample speed of the acquired signal was 500 kHz. An analog–digital converter Keysight MSO9104 was used for acquiring the ECG data. Here, ATmega8 was the processor, and wet electrodes were attached [33]. The number of data points for each subject is 60. The ECG signals of multiple cycles were divided uniformly with respect to the R-peak point. The number of data points after the division was 16,930. Of these, 80% (13,546 data points) were used as training data whereas 20% (3384 data points) were used as validation data.

4.2. Experimental Method and Results

In this section, the performance of personal identification using LSTM and 2D-CNN from ECG signals is analyzed. LSTM was used for analyzing 1D ECG signals, and the performances of LSTM layer and two LSTM layers were analyzed comparatively. The initial learning rate of the experiment was 0.01. Adam was used as an optimization function for minimizing the error between the predicted and actual values. Epochs of 30, 50, 60, and 100 and a mini-batch size of 128 were applied repeatedly to examine the performance (see Figure 19). In addition, 100 hidden layers were used for LSTM. The LSTM-based personal identification accuracy for the CU-ECG database was highest (95.12%) when the epoch and mini-batch size were set to 100 and 128, respectively.

Because physiological signals such as ECG signals are affected by various types of noises, the signals were transformed into 2D images to be applied for personal identification using CNNs. Figure 20 shows the images transformed by STFT, scalogram, FSST, and WSST for a subject in the CU-ECG database.

Figure 19. Accuracy of LSTM-based personal identification.

Figure 20. Time–frequency representation of a subject with regard the CU-ECG database signal.

The personal identification accuracy for the transformed image was examined using a 2D-CNN. The transfer learning models GoogleNet, VGG-19, and ResNet-101 were used for the 2D-CNN. The data were converted to 224×224 to be used as input. The following were the settings for GoogleNet to conduct an experiment using the CU-ECG database: an initial learning rate of 1×10^{-4} , Adam as an optimization function, an epoch of 30, and mini-batch size of 64. Meanwhile, the following were the settings for VGG-19: an initial learning rate of 1×10^{-4} , Adam as an optimization function, an epoch of 20, and a mini-batch size of 32. ResNet-101 was applied in a similar manner as VGG-19. Furthermore, the accuracy of personal identification using ECG signals based on the proposed ensemble was examined rather than separately examining the performance of LSTM and of CNN. Table 1 presents the accuracy of personal identification based on the ensemble of LSTM and 2D-CNN and that based on 2D-CNN for the CU-ECG database. GoogleNet demonstrated the highest performance of 96.25% in FSST, whereas VGG-19 demonstrated the highest performance of 95.12%. ResNet-101 demonstrated the highest performance of 97.67% in STFT. The accuracy

of personal identification using an ensemble of LSTM and 2D-CNN was examined using GoogleNet, VGG-19, and ResNet-101 for STFT, scalogram, FSST, and WSST expressed with a time–frequency representation. The ECG signal was converted into the two-dimensional time–frequency domain through LSTM and four transformation methods, and it can be seen that the personal identification performance results using GoogleNet, VGG-19, and ResNet-101 are all excellent. However, it can be confirmed that the individual identification performance through the ensemble method using the maximum value by multiplying the score values of each model proposed in this paper is superior to the performance of the single model. GoogleNet demonstrated the highest performance in FSST, where the result of an ensemble showed an improvement of 2.33%. Furthermore, VGG-19 demonstrated the highest performance in WSST with an improvement of 3.4% from the individual models. Meanwhile, ResNet-101 demonstrated the highest performance in STFT, with an improvement of 1.06% from the individual models.

Table 1. Accuracy of LSTM and 2D-CNN-based ensemble.

CU-ECG	Time–Frequency Representation	2D-CNN Validation Accuracy	LSTM–2D-CNN Ensemble Validation Accuracy
GoogleNet	STFT	95.09%	98.37%
	Scalogram	95.69%	97.87%
	FSST	96.25%	98.58%
	WSST	95.48%	98.73%
VGG-19	STFT	94.86%	97.90%
	Scalogram	94.18%	97.93%
	FSST	94.52%	97.93%
	WSST	95.12%	98.52%
ResNet-101	STFT	97.67%	98.73%
	Scalogram	97.25%	98.73%
	FSST	96.04%	98.43%
	WSST	95.24%	98.49%

5. Conclusions

This study performed personal identification based on the ensemble of LSTM and 2D-CNN with ECG signals. ECG-based personal identification is based on a comparison of the ECG of a user with that of registered users. ECG uses unique signals of each person, which vary depending on the position and size of the heart, gender, and age. Thus, individuals can be identified with over 90% accuracy based on the characteristics of ECG signals. In addition to personal identification, ECG is being widely used in the medical field for predicting and diagnosing heart-related diseases. Therefore, personal identification using ECG signals as well as ECG-based health monitoring technology that enables remote examination of heart diseases such as cardiac arrest or arrhythmia are likely to be developed.

ECG signals are accompanied by different types of noises because these are physiological signals measured through microcurrents. Therefore, distorted signals are removed through filters to enable accurate assessment or diagnosis. Because the adjusted baseline is the reference for detecting the characteristic points of ECG signals, it becomes difficult to identify the morphological characteristics of ECG signals if the baseline is not calibrated to zero. Accordingly, noise removal and baseline fluctuation adjustment were performed as a preprocessing step. In order to classify the one-dimensional ECG signal with noise removed, two LSTM layers with higher accuracies were used as a result of comparing and analyzing one LSTM layer and two LSTM layers. As such, the more layered the classifica-

tion performance accuracy can be improved, but there is a problem that the structure can be complicated. In this paper, we propose an ensemble, so we use two layers without building more layers of LSTM to avoid complex structures. In addition, the 1D ECG signal is represented by the image using the Short-Time Fourier Transform (STFT), Scalogram, Fourier Synchrosqueezed Transform (FSST), and Wavelet Synchrosqueezed Transform (WSST) as a time–frequency representation. For performance results, the performance of each model of the composite multi-neural network with excellent image classification performance was confirmed using GoogleNet, VGG-19, and ResNet-101, and the performance of each model of the LSTM and 2D-CNN was improved through the ensemble method. To conduct the experiment, the CU-ECG database constructed by Chosun University used data containing 100 subjects, 89 men, and 11 women, who were in comfortable postures. The results of two LSTM neural networks showed that the highest performance was 95.12% when epoch was set at 100 and the minibatch size at 128, and the performance of 2D-CNN was the highest at 97.67% for ResNet-101. Finally, the performance of each model of LSTM and 2D-CNN was improved by an ensemble method, with a personal identification performance of at least 1.06% to a maximum of 3.75% compared to that of a single model.

Author Contributions: Conceptualization, J.-A.L. and K.-C.K.; Methodology, J.-A.L. and K.-C.K.; Software, J.-A.L. and K.-C.K.; Validation, J.-A.L. and K.-C.K.; Formal Analysis, J.-A.L. and K.-C.K.; Investigation, J.-A.L. and K.-C.K.; Resources, K.-C.K.; Data Curation, J.-A.L.; Writing—Original Draft Preparation, J.-A.L.; Writing—Review and Editing, K.-C.K.; Visualization, J.-A.L. and K.-C.K.; Supervision, K.-C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2017R1A6A1A03015496).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Jeong, D. Artificial intelligence security threat, crime, and forensics: Taxonomy and open issues. *IEEE Access* **2020**, *8*, 184560–184574. [\[CrossRef\]](#)
- Kim, J.; Rhee, P. Image recognition based on adaptive deep learning. *Inst. Internet Broadcast. Commun.* **2018**, *18*, 113–117.
- Lu, L.; Mao, J.; Wang, W.; Ding, G.; Zhang, Z. A study of personal recognition method based on emg signal. *IEEE Trans. Biomed. Circuits Syst.* **2020**, *14*, 681–691. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rabuzin, K.; Baca, M.; Sajko, M. E-learning: Biometrics as a security factor. In Proceedings of the International Multi-Conference on Computing in the Global Information Technology, Bucharest, Romania, 1–3 August 2006; p. 64.
- Barros, A.; Rosário, D.; Resque, P.; Cerqueira, E. Heart of IoT: ECG as biometric sign for authentication and identification. In Proceedings of the 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; pp. 307–312.
- Jiang, X.; Xu, K.; Liu, X.; Dai, C.; Clifton, A.D.; Clancy, A.E.; Akay, M.; Chen, W. Cancelable HD-sEMG-based biometrics for cross-application discrepant personal identification. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1070–1079. [\[CrossRef\]](#)
- Biel, L.; Pettersson, O.; Philipson, L.; Wide, P. ECG analysis: A new approach in human identification. *IEEE Trans. Instrum. Meas.* **2001**, *50*, 808–812. [\[CrossRef\]](#)
- Ingale, M.; Cordeiro, R.; Thenttu, S.; Park, Y.; Karimian, N. ECG biometric authentication: A comparative analysis. *IEEE Access* **2020**, *8*, 117853–117866. [\[CrossRef\]](#)
- Zhang, Q.; Zhou, D.; Zeng, X. HeartID: A multiresolution convolutional neural network for ECG-Based biometric human identification in smart health applications. *IEEE Access* **2017**, *5*, 11805–11816. [\[CrossRef\]](#)
- Israel, S.A.; Irvine, J.M.; Cheng, A.; Wiederhold, M.D.; Wiederhold, B.K. ECG to identify individuals. *Pattern Recognit.* **2005**, *38*, 113–142. [\[CrossRef\]](#)
- Jahiruzzaman, M.; Hossain, A.B.M.A. ECG based biometric human identification using chaotic encryption. In Proceedings of the International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Savar, Bangladesh, 21–23 May 2015; pp. 1–5.
- Zhao, Z.; Yang, L.; Chen, D.; Luo, Y. A human ECG identification system based on ensemble empirical mode decomposition. *Sensors* **2013**, *13*, 6832–6864. [\[CrossRef\]](#)

13. Wieclaw, L.; Khoma, Y.; Fałat, P.; Sabodashko, D.; Herasymenko, V. Biometric identification from raw ECG signal using deep learning techniques. In Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, Romania, 21–23 September 2017; pp. 129–133.
14. Labati, R.D.; Muñoz, E.; Piuri, V.; Sassi, R.; Scotti, F. Deep-ECG: Convolutional neural networks for ECG biometric recognition. *Pattern Recognit. Lett.* **2019**, *126*, 78–85. [\[CrossRef\]](#)
15. Abdeldayem, S.S.; Bourlai, T. A novel approach for ECG-Based human identification using spectral correlation and deep learning. *IEEE Trans. Biom. Behav. Identity Sci.* **2020**, *2*, 1–14. [\[CrossRef\]](#)
16. Ciocoiu, I.B.; Cleju, N. Off-Person ECG biometrics using spatial representations and convolutional neural networks. *IEEE Access* **2020**, *8*, 218966–218981. [\[CrossRef\]](#)
17. Byeon, Y.H.; Kwak, K.C. Pre-Configured Deep Convolutional Neural Networks with Various Time-Frequency Representations for Biometrics from ECG Signals. *Appl. Sci.* **2019**, *9*, 4810. [\[CrossRef\]](#)
18. Choi, G.H.; Bak, E.S.; Pan, S.B. User identification system using 2D resized spectrogram features of ECG. *IEEE Access* **2019**, *7*, 34862–34873. [\[CrossRef\]](#)
19. Jyotishi, D.; Dandapat, S. An LSTM-based model for person identification using ECG Signal. *IEEE Sens. Lett.* **2020**, *4*, 1–4. [\[CrossRef\]](#)
20. Kim, J.S.; Kim, S.G.; Pan, S.B. Personal recognition using convolutional neural network with ECG coupling image. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 1923–1932. [\[CrossRef\]](#)
21. Hammad, M.; Plawiak, P.; Wang, K.; Acharya, R.U. ResNet-attention model for human authentication using ECG signals. *Expert Syst.* **2021**, *38*, 6. [\[CrossRef\]](#)
22. Byeon, Y.H.; Pan, S.B.; Kwak, K.C. Intelligent deep models based on scalograms of electrocardiogram signals for biometrics. *Sensors* **2019**, *19*, 935. [\[CrossRef\]](#)
23. Chikkerur, S.; Cartwright, A.N.; Govindaraju, V. Fingerprint enhancement using STFT analysis. *Pattern Recognit.* **2007**, *40*, 198–211. [\[CrossRef\]](#)
24. Lee, J.W.; Lee, H.W.; Yoo, C.S. Selection of mother wavelet for bivariate wavelet analysis. *J. Korea Water Resour. Assoc.* **2019**, *52*, 905–916.
25. Olhede, S.C.; Walden, A.T. Generalized morse wavelets. *IEEE Trans. Signal Process.* **2002**, *50*, 2661–2670. [\[CrossRef\]](#)
26. Lilly, J.M.; Olhede, S.C. Higher-order properties of analytic wavelets. *IEEE Trans. Signal Process.* **2009**, *57*, 146–160. [\[CrossRef\]](#)
27. Oberlin, T.; Meignen, S.; Perrier, V. The fourier-based synchrosqueezing transform. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 315–319.
28. Kumar, A.; Gandhi, C.P.; Zhou, Y.; Vashishtha, G.; Kumar, R.; Xiang, J. Improved CNN for the diagnosis of engine defects of 2-wheeler vehicle using wavelet synchro-squeezed transform (WSST). *Knowl. Based Syst.* **2020**, *208*, 106453. [\[CrossRef\]](#)
29. Lumini, A.; Nanni, L. Deep learning and transfer learning features for plankton classification. *Ecol. Inform.* **2019**, *51*, 33–43. [\[CrossRef\]](#)
30. Habib, N.; Hasan, M.M.; Reza, M.M.; Rahman, M.M. Ensemble of cheXNet and VGG-19 feature extractor with random forest classifier for pediatric pneumonia detection. *SN Comput. Sci.* **2020**, *1*, 6. [\[CrossRef\]](#)
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Haralabopoulos, G.; Anagnostopoulos, I.; McAuley, D. Ensemble deep learning for multilabel binary classification of user-generated content. *Algorithms* **2020**, *13*, 83. [\[CrossRef\]](#)
33. Byeon, Y.H.; Lee, J.N.; Pan, S.B.; Kwak, K.C. Multilinear eigenECGs and fisherECGs for individual identification from information obtained by ad electrocardiogram sensor. *Symmetry* **2018**, *10*, 487. [\[CrossRef\]](#)