Robust Deep Identification using ECG and Multimodal Biometrics for Industrial Internet of Things

Ebrahim Al Alkeem[1,2], Chan Yeob Yeun[1], Jaewoong Yun[3], Paul D. Yoo[4], Myungsu Chae[3], Arafatur Rahman[5] and A. Taufiq Asyhari[6]

[1]Center for Cyber-Physical Systems, EECS Department, Khalifa University, Abu Dhabi, UAE
[2]Security and Safety, Nawah Energy Company, Abu Dhabi, UAE
[3]Research Institute NOTA Incorporated, Daejeon, South Korea
[4]CSIS, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, United Kingdom
[4]Faculty of Computing, University Malaysia Pahang, Pahang, Malaysia
[4]CDT, Birmingham City University, Millennium Point, Birmingham, B4 7XG, United Kingdom

*Correspondence: chan.yeun@ku.ac.ae

Abstract: The use of electrocardiogram (ECG) data for personal identification in Industrial Internet of Things can achieve near-perfect accuracy in an ideal condition. However, real-life ECG data are often exposed to various types of noises and interferences. A reliable and enhanced identification method could be achieved by employing additional features from other biometric sources. This work, thus, proposes a novel robust and reliable identification technique grounded on multimodal biometrics, which utilizes deep learning to combine fingerprint, ECG and facial image data, particularly useful for identification and gender classification purposes. The multimodal approach allows the model to deal with a range of input domains removing the requirement of independent training on each modality, and inter-domain correlation can improve the model generalization capability on these tasks. In multitask learning, losses from one task help to regularize others, thus, leading to better overall performances. The proposed approach merges the embedding of multimodality by using feature-level and score level fusions. To the best of our understanding, the key concepts presented herein is a pioneering work combining multimodality, multitasking and different fusion methods. The proposed model achieves a better generalization on the benchmark dataset used while the feature-level fusion outperforms other fusion methods. The proposed model is validated on noisy and incomplete data with missing modalities and the analyses on the experimental results are provided.

Keywords: Personal identification; multimodal biometrics; deep learning; gender classification; electrocardiogram; fingerprint; face recognition; feature-level fusion

1. Introduction

Personal identification using electrocardiogram (ECG) data is a recent development in biometrics and has a great potential to be applied in Industrial Internet of Things (IIoT) environments. Such system presented with ECG data can return identifications by verifying whether or not individuals are in the database. Previous researches have focused on identification and authentication using databases of ECG profiles acquired from only a few subjects under restricted conditions [1–11]. However, ECGs captured in real-life are likely to contain noises and interferences, requiring the identification based on unseen ECG profiles to be robust towards noisy and incomplete data.

Traditional machine learning (ML) methods for personal identification include feature extraction and classification methods, but deep learning (DL) approaches can achieve better results

for several reasons. First, ML methods are generally comprised of a channel of tailored feature extraction and classification models, including k-nearest neighbors (kNN) or random forest, resulting in uncertainty of being able to extract solely informative features. Performances are determined by the selected combination of feature extraction and classification methods, which may lead to poor performances. In contrast, the DL utilizes neurons, which broaden the choice of feature extraction and learning models. This degree of freedom means that modeling between raw data and the label is trained by the gradient descent rule, which minimizes overall losses.

Second, the DL is superficial with structured raw data (e.g., audio and image data), whereas the hand-crafted features captured by the ML may also contain unrelated features. If the input is less structured (e.g., age or gender) then feature extraction is unnecessary. However, image and audio data are highly dimensional, so feature extraction should be considered carefully. DLs outperform MLs on image and audio data in various domains [12, 13].

Third, the DL is typically preferred for the analysis of ECG data [14–16]. For example, AlexNet [12] has been used as a pre-trained model to perform feature extraction from ECG profiles and predict for cardiac arrhythmia with 92.4% accuracy [14]. An ECG biometric recognition using convolutional neural networks (CNNs) has also been tested, attaining a similar error percentage of 2.26% [15] and 93.6% precision in the screening of paroxysmal atrial fibrillation [16]. However, when kNN and support vector machine (SVM) models are attached on the top of the CNN feature, the precision drops to 90.7% and 92.9%, respectively. This shows that the traditional pipeline involving the selection of feature extraction and classifier modules is outperformed by an end-to-end trainable DL model.

A multimodality is preferable to a single modality in DL models, and can involve various biometric inputs such as ECG, facial images, fingerprints, voice, ear images, and iris data. Any model using single modality appears to be easily corrupted by powerline fluctuations, surrounding noises, electromyography (EMG), movement artifacts, and electrodes, so the challenge lies on developing accurate unimodal identification algorithms. Noise reduction techniques have been tested to overcome these effects [17], but multimodality has many additional benefits. If the signal-to-noise ratio of one modality is low, offsetting the effects can be performed by another to maintain the cumulative performance. Furthermore, latent correlations between modalities can be trained to improve the performance. Finally, it is easier to train multimodal models than multiple independent models for each modality.

An appropriate fusion method should also be combined with multimodal models, particularly a feature-level fusion, because it accommodates correlations between modalities. However, the algorithms before and after the feature-level fusion are performed independently, allowing any errors that occur before the fusion can accumulate. The end-to-end training of the system minimizes such accumulated errors.

Incomplete ECG profiles can be overcome by a multitask learning. In real-world settings, ECG data are influenced by factors such as heart rate and disease. Such factors can be marginalized if the model can be trained by the multitask learning to focus on the target task and ignore other factors. When implementing the multitask learning, the loss of each task can be weighted and a suitable method to determine weightings is therefore required. In this paper, our tasks are person identification and gender classification: if the former is well trained, then the latter would also show an improvement in performance. In other words, if the identification is difficult to train, then gender classification can help to train the network. The benefit of multitask learning is that deficiencies in one task helps to regularize the other, so that the training constructively progresses regardless of incompleteness of either characteristics of personal identification or gender classification. Accordingly, one model can perform two tasks by comparing the models trained on each task.

ECG in the wild could also contain noise because it may be collected from various devices differing in precision (e.g., smart watches, smart bands). It may also contain errors, or outliers, which could be due to faulty instruments, data transmission/interference or technology compatibility issues. The condition of the subject at the time of measurement also has an influence on noise. Similarly, facial image data can be affected by the viewing angle, brightness, and blurring, while fingerprint

data can be influenced by the position and pressure of the fingers. The model architecture must therefore be robust and must consider a generalization. Therefore, the proposed model is also relevant and applicable to intelligent home systems where the access is granted using briometrics of the household members. The main contributions of the work described in this article can be summarized as follows:

- multimodal biometrics using combined facial image, fingerprint, and ECG data,
- multitask learning for personal identification and gender classification tasks,
- feature fusion methods using a deep neural network,
- end-to-end trainable architecture, and
- a robust model under noisy and/or incomplete data situations.

The remainder of the paper is structured as follows. Section II presents related work on modality, fusion methods and multitask learning. Section III describes preprocessing, architecture, and deep learning methods for user identification and gender classification. Section IV discusses our experimental results under various conditions, and Section V concludes with the provision of future work.

2. Related Work

2.1. ECG Analysis

ECG profiles provide useful biometric data because the electrical properties of the heart carry unique information suitable for personal identification or authentication [18–21]. Among various approaches that have been tested, single-lead ECG data collected across 19 subjects was fed into a Ziv-Merhav cross-parsing algorithm, which achieved 100% accuracy for a larger number of experiments [2]. Other successful methods include SVMs [3, 4], qualifying similarity or dissimilarity [5], and random forest models [6]. Personal identification has also been achieved by analyzing the frequency features of ECG signals without a fiducial point [7, 8] and classifying the shape of heart rate variability by principal component analysis and linear discriminate analysis [1, 9].

2.2. Multimodal Learning

Multimodal learning can combine multiple types of biometric data to recognize or authenticate users, and the fusion of multimodal features can increase the accuracy of verification [22, 23]. For example, combined ECG and fingerprint data achieves greater accuracy than either of the individual modes [24][25]. The uncertainty in the time domain features of ECG was 5.0% [18], but by combining ECG and facial image data this was reduced to 1.0% [23]. Finally, the accuracy of personal identification using palm prints (82.1%) or ECG (89%) was increased to 94.7% by combining them into a multimodal model [26]. In addition to multimodal learning research related to user identification using ECG, there are various multimodal learning studies related to disease diagnosis using magnetic resonance imaging (MRI) [27–29].

2.3. Handling Multimodal Databases

Multimodal databases are handled by constructing either a true or virtual database to combine the unimodal data. True multimodal databases provide all the biometric data from real people. However, this is an expensive and time-consuming process. There is a limited set of data that can be gathered, and the risk of losing personal information increases. Most studies in this area, therefore, use virtual multimodal databases, which are quicker and easier to assemble [30–34]. Given two mutually exclusive subject-based databases, a virtual multimodal database would be formed by pairing subjects in each database so that the data for each subject would be combined. Achieving this is possible by assuming that a single subject can independently be represented using a variety of biometric traits [34]. Currently, there is a lack of accurate multimodal database in the public domain possessing face, fingerprint and ECG data.

2.4. Feature Fusion Techniques

As data propagates from the input layer to the label, fusion can be achieved at five different stages, namely the input, feature, score, decision and rank.

Input-level fusion means that the data are concatenated or fused and then used as the model input. This approach is suitable if each modality has the same data type. The input-level fusion is not widely used because most modalities have different data types, which need to be extracted in an independent manner.

A feature-level fusion integrates different features from multiple biometrics, thus, constructing a dataset by concatenating each independent biometric [35]. Little progress was made for a long time due to implementational challenges, but more recent work has been successful with the feature-level fusion in three different scenarios [36] as well as the feature-level concatenation of face and fingerprint data [37].

A score-level fusion achieves a good level of compromise between the simplicity of implementation and effectiveness given that it is more accessible than input-level or feature-level fusion but can nevertheless achieve successful matches with incomplete data. Many score-level fusion approaches have therefore been reported, including those using Bayesian framework [38] and density-based methods [39]. The effectiveness of score-level fusion reflects the presentation of results as raw scores, quantized scores, or probabilities when biometric traits are combined and features are matched. However, one disadvantage of the score-level fusion is given by the diversity of scores obtained from a variety of matching strategies.

A decision-level fusion generates a binary outcome (yes or no) as typically applied in a personal identification scenario (identified/present in database or not identified/not present). The decision-level fusion is therefore used for other binary outcomes such as majority voting [40] or Boolean operations [41]. Despite the simple and intuitive outcome, the decision-level fusion is less popular than the score-level and the rank-level fusions because certain types of data cannot be adapted to this strategy naturally.

Finally, a rank-level fusion consolidates larger than two results from identification to improve the reliability of a recognition task, and is therefore popular in the domains of data mining and pattern recognition. Variations of the rank-level fusion method include top rank [42], Borda count [43], weighted Borda count [44] as well as Bayes fuse based on a Bayesian inference [45, 46]. The performance of this strategy keeps improving, as shown in the mixed group ranks method [47].

2.4. Feature Fusion Techniques

A multitask learning intends to exploit meaningful information embedded in a variety of connected tasks to enhance generalization across all the relevant tasks. It can be utilized to identify several features at once, such as the recognition of gender, age and identity using an ensemble of features subjected to decision tree and Bayes network analysis [48]. The multitask learning allows the sharing of representations between tasks, and the execution of the main task can sometimes be enhanced by utilizing knowledge from other tasks [49]. In contrast to optimizing the learning of individual tasks, multitask learning considers a set of interrelated tasks that must be solved, and the performance is enhanced by jointly executing the tasks and exchange the data representing different features [50]. To reduce learning bias, multitask learning can be combined with other learning techniques such as semi- or unsupervised, reinforcement, active, multi-view, and graphical models. Many groups have applied multitask learning to ECG profiles, and a deep multitask learning model with additional network fine-tuning has increased the accuracy of the ECG signal evaluation by 5.1% [51].

3. Methods

Our network architecture is organized into three different components, namely a feature extraction that converts the input into the embedding space, a fusion layer that combines features from each modality, and a task layer that performs the task (Fig. 1). First, the feature extraction is performed in a unique manner on each modality, given their specific characteristics. The data for fingerprint (x_p) and face (x_f) biometrics are static images, whereas the ECG data (x_e) are temporal signal traces. Classifiers trained on sizeable databases might be applied as an established feature extractor [52]. We employed Residual Network 50 (ResNet50) to extract features using x_f and x_p.

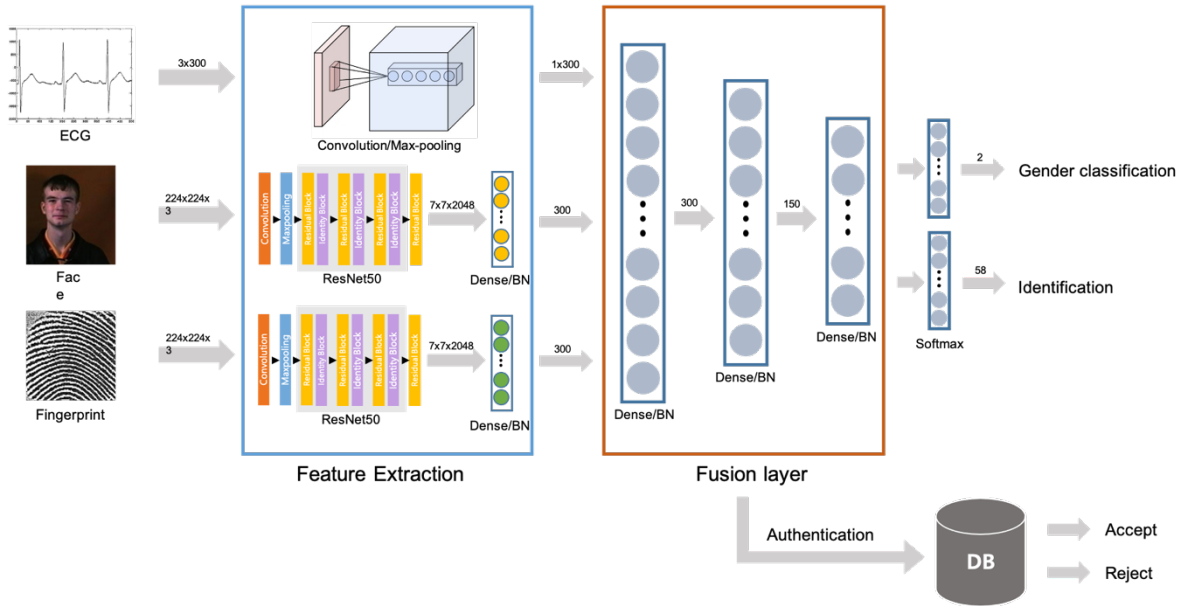

Figure 1. Comprehensive network architecture of the multimodal deep multitask learning system.

A fully connected (FC) layer was considered to align the ECG data with the dimension number. The CNN was continued by a max pooling component intended for x_e (ECG data) with the purpose of removing reliance on a temporal axis for the extracted feature. The fusion layer (neural network) considers chained features as the input from the feature extraction component. Following up the data propagation, information representing every modality is combined to reduce the overall loss. It is concluded by the task layer using the combined features from the fusion layer as its input, and classification is performed after a single layer, with the node quantity specifying the class quantity.

3.1. Preprocessing

3.1.1. Face and Fingerprint Data

As stated above, only limited data are available for user identification in multimodal datasets, thus, more facial images and fingerprints are required to implement a generalized network. The augmentation of facial images and fingerprint data was achieved by rotation, translation and cropping (the latter not for facial images, because the entire face is needed for identification). Images were rotated from –30° to 30° in 5° intervals to obtain 12 new images per original image. For translation, each image was translated from –5 to 5 pixels in 1-pixel intervals to obtain 10 new images per original image. For cropping, we created images with 60% of the original size at random positions in the original image to obtain 47–50 new images per original image. After augmentation, we used a transfer-learning technique where a pre-trained neural network, ResNet50 is adopted to extract features comprising a set pair of fingerprint and facial images.

3.1.2. ECG Data

The Pan Tompkins QRS detection algorithm is widely used to find the R peak, which is the center of the QRS complex within the ECG signal. Herein, the ECG signal is passed through low and high bandpass filters, and differential values are obtained for peak detection based on these data. During this process, we determine the threshold ECG peak value and the minimum time to generate the R peak, yielding candidate groups for actual R peaks. The detection algorithm calculates the differential values and then squares each value for detecting the peak. Values are set to zero if the peak is negative, as it is not intuitive to retain them.

The R peak detection process was implemented as shown in Fig. 2. Initially, a bandpass filter is considered to sharpen the peaks and smooth the rest of the signal. The original value is then subtracted from the present value in the filtered signal to calculate the differentiated signal. Following the replacement of any negative values with zero, the peaks are derived from candidate points that are greater than both the previous value and the next value. Finally, to validate the computed R peaks, the algorithm superimposes the peak indices with the same indices in the original ECG signal. If the overlaps are confirmed, the indices are stored. Having applied the Pan Tompkins QRS detection algorithm, we added one more step to account for any delay compared to the raw signal. The additional step ensures that, if there is a peak value larger than the detected peak value in the adjacent range, we set the larger peak value as the R peak value.

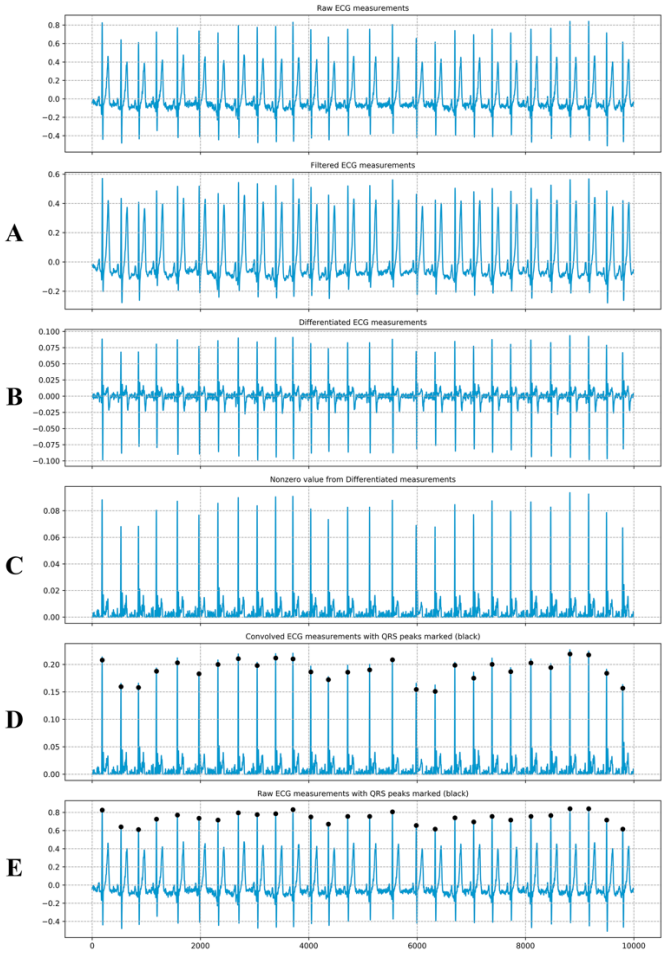

Figure 2. Process for R peak detection. (A) Application of a bandpass filter, (B) differentiation, (C) clipping of negative values to zero, (D) detection of maximum points, and (E) matching to the original signal.

We cropped a total of 300 data samples prior and post R peak detection to produce a vector representing the QRS complex. Following this mechanism, we normalize each QRS complex using a min-max method.

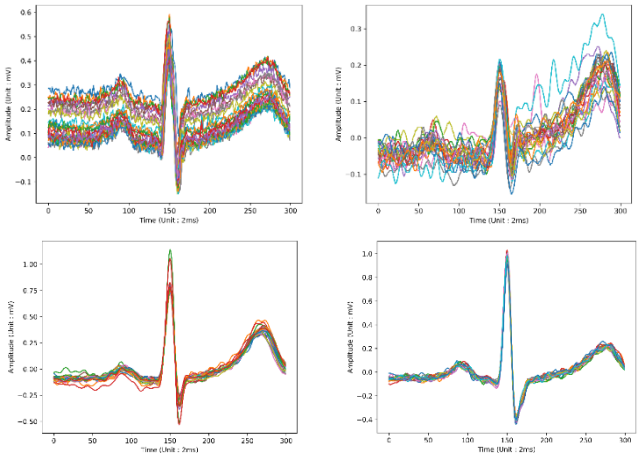

Figure 3. Four examples of QRS compositions extracted from the same record. The compositions on the top row show consistency within the record, but those on the bottom row show irregularities within the record.

Given the less importance nature of the inter-QRS complexes temporal information, compared to the complexes themselves, we extracted three QRS complexes to make one input. A single sequence therefore possesses three time-steps, indicating formation of this sequence using a group of three QRS complexes. The extracted QRS complex, based on the R peak in each recording and plotted in the same grid, is captured in Fig. 3. The consistency of the R peak wave varies depending on the records.

3.2. Feature Extraction

3.2.1. ResNet50

ResNet is a known method to perform feature extraction from image data [53]. It has been constructed via training of ImageNet, one of the comprehensive datasets for object classification. Herein our feature extraction method used ResNet50 that is applied to our face/fingerprint images, with mean pooling of features (7 × 7 × 2048). The input size of the images was set to 224 × 224. ResNet50 is a CNN with 50 layers that is trained to classify images into 1000 categories in the ImageNet database. It comprises a convolution layer (3 × 3 filter), a max-pooling layer and a residual network comprising a residual block and identity block as depicted in Fig. 4. The overall structure is given in Fig. 5. The output shape after the identity block is described at the bottom of each identity block box in Fig. 5.

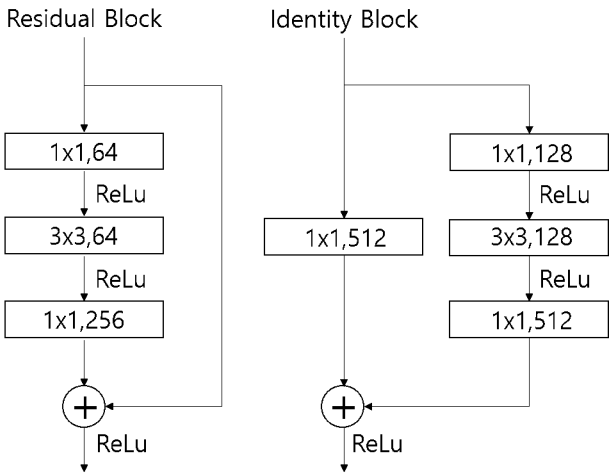

Figure 4. Detailed structure of the residual and identity blocks. The residual block allows direct connection from input to output, whereas the identity block uses an identity matrix instead of a direct connection.

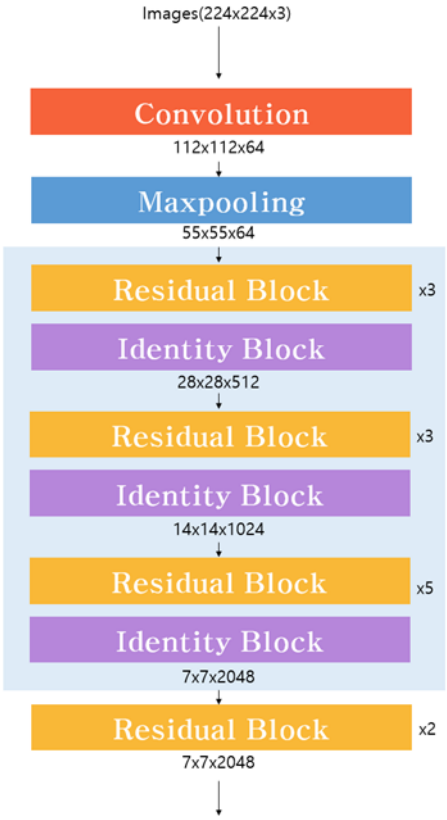

Figure 5. The architecture of ResNet, comprising a convolution and a max-pooling layers, followed by series of alternating residual and identity blocks.

3.2.2. CNN Model for ECG

Although models based on long short-term memory architecture has been widely used for identifying [54], the CNN was employed herein to extract features, seeing the less importance nature of inter-QRS temporal information, in comparison to the actual complexes. For the CNN, we have a one-dimensional convolution layer, which considers a tensor with shape (batch size, time step, 300) as an input and yields an output tensor with a similar shape characteristics. This is then followed by operating a single max-pooling task. The results of ECG feature extraction are chained with two feature vectors arising from the different modalities (Fig. 6).

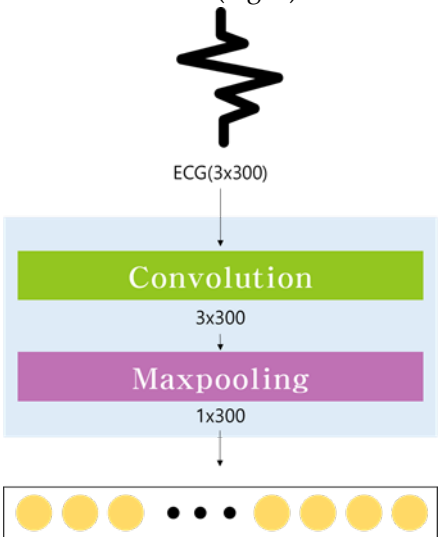

Figure 6. The one-dimensional CNN used to extract ECG features.

3.3. Feature Fusion

Input level fusion using three different modalities effectively is difficult. We thus compared the score-level fusion with feature-level fusion. For the score-level fusion, we tested three techniques, namely sum, product and max rule. For the feature-level fusion, three features were extracted from every modality for normalization as well as concatenation as the model inputs.

3.4. Classification

Our proposed system has to decide the most likely class for each test case. We therefore used the softmax activation function, which considers the class-wise outputs and their transformation into corresponding probabilities via Eq. (1):

$$\text{Softmax}(y_i) = \frac{\exp(y_i)}{\sum_j \exp(y_j)} \tag{1}$$

Herein y denotes the network output with the same dimension as the class count, and y_i is the i^{th} component of y. Note that Eq. (1) ensures that normalization of the exponential numerator by the sum of exponential terms in the denominator. The class with highest number is the target class. For identifying users and classifying genders, the cost function considered the cross-entropy loss during the model training. The function can be computed using Eq. (2):

$$\mathrm{H}(\mathrm{y}, \hat{y}) = -\sum_i y_i \log(\hat{y}_i) \tag{2}$$

Here we have y and $\hat{y}$ denoting the output and original distributions, respectively. The logarithm usage in Eq. (2) corresponds to penalty being applied for incorrect predictions, i.e., high loss with divergence of the predicted class from the real label. Eq. (2) does not represent a symmetrical function with $\mathrm{H}(\mathrm{y}, \hat{y}) \neq \mathrm{H}(\hat{y}, \mathrm{y})$ because only the logarithm of predicted probabilities is considered.

3.5. Joint Loss

Joint training of the network to simultaneously handle more than one tasks was performed by applying a joint loss that integrates cross-entropy (binary) for classifying genders (L_1) and cross-entropy (categorical) for identifying users (L_2). Two losses were weighted and summed for final loss calculation in the model training. The formula for each loss is shown in Eq. (3)-(5). The most favorable indicators for w_1 and w_2 in Eq. (5), where the sum is 1, were found experimentally.

$$L_1 = -y \log \hat{y} - (1 - \mathrm{y}) \log(1 - \hat{y}) \tag{3}$$

$$L_2 = -\sum_i y_i \log \hat{y} \tag{4}$$

$$L_{joint} = w_1 \times L_1 + w_2 \times L_2 \tag{5}$$

3.6. Optimization

For optimizing parameters, Adam, an efficient procedure for the gradient-based optimization of stochastic objective functions, was utilized relying on lower-order moments adaptive estimation [55]. This takes advantages of the two well-known methods, namely AdaGrad and RMSProp [56, 57]. Parameter optimization was achieved as shown in Eq. (6) and (7).

$$\alpha_t = \alpha \cdot \sqrt{1-\beta_2^t}/(1-\beta_1^t) \quad (6)$$

$$\theta_t \longleftarrow \theta_{t-1} - \alpha_t \cdot m_t/(\sqrt{v_t} + \hat{\epsilon}) \quad (7)$$

where α_t is the step size at time-step t and β_1, β_2 are rates characterizing exponential decrement for moment estimation. To calculate the updated parameter vector θ_t, we used updated biased moment estimates (m_t, v_t) and α_t.

4. Experiments

Nine experiments were carried out to investigate the performance of the new model. The first experiment considered multitask learning (identifying users and classifying genders) using single modality and multimodality. The second experiment compared multitask and single-task learning based on multimodal biometrics comprising ECG data from ECG-ID [58, 59], PTB [58, 60]), facial images (Face95 [61]) and fingerprints (FVC2005 [62]. The third experiment tested the user authentication using two distance metrics. The fourth experiment added noises to the data to verify the robustness of the model.

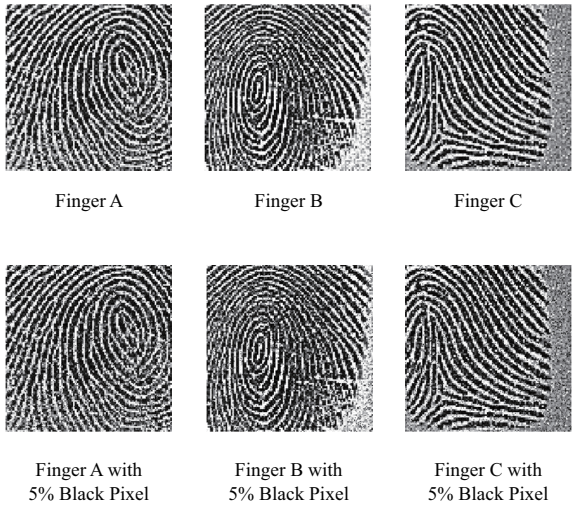

Figure 7. Noise in the fingerprint input data for the multimodal deep multitask learning system.

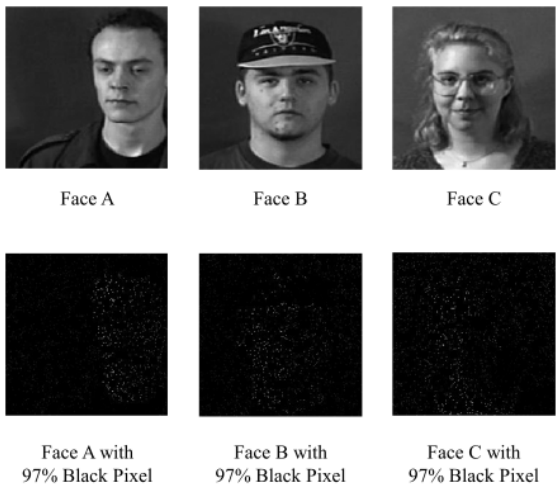

Figure 8. Noise in the face input data for the multimodal deep multitask learning system.

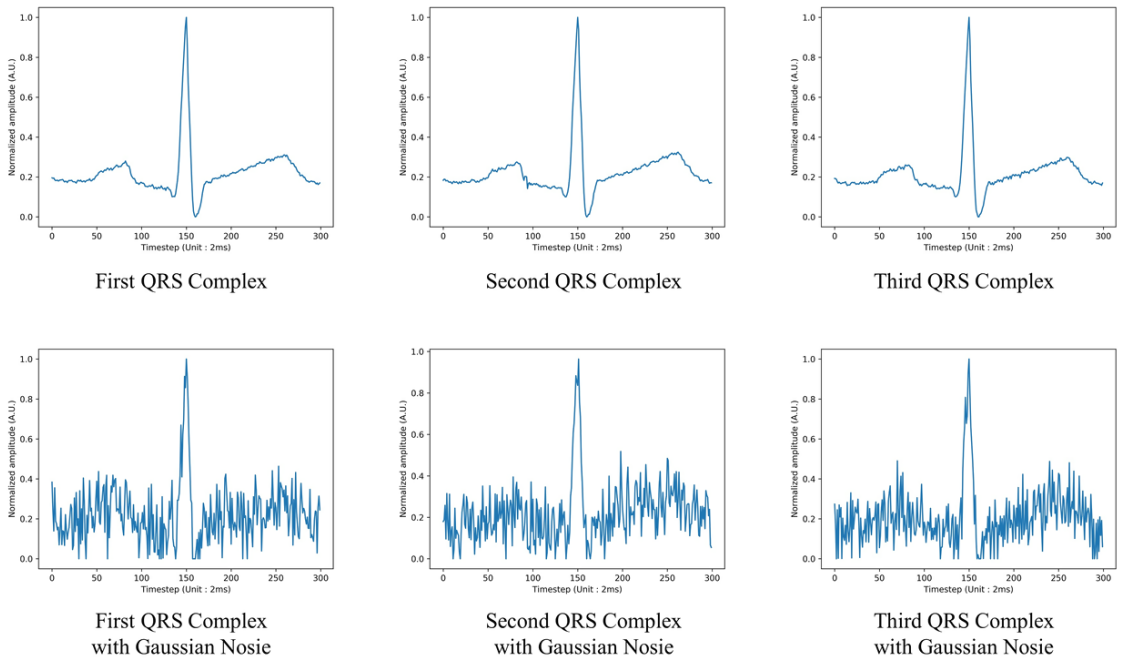

Figure 9. Noise in the ECG input data for the multimodal deep multitask learning system.

The fifth experiment investigated the impact on performance when one or two of the three biometric datasets were missing. The sixth experiment compared the performances of feature-level and score-level fusions where the latter based on sum, max or product rules. The seventh experiment modified the joint loss to give various weights. The eighth experiment investigated how data augmentation affected the performance of the multimodal/multitasking model. The final experiment changed the hyper-parameters of the fusion model to find the optimal model architecture. The number of nodes was changed to compare the accuracy of each task (user identification and gender classification).

4.1. Dataset

4.1.1. ECG Data

The ECG-ID has 310 records from 90 subjects with composition: 44 males and 46 females with ages from 13–75 years. Each single-lead trace was observed and stored for 20 s and sampled at 500 Hz at 12-bit resolution. The range of observation is ±10 mV. The dataset provides not only raw signals (EC I), but also processed signals with high-frequency and low-frequency noises filtered out (ECG I filtered). Associated data include subject age, gender and recording date.

The PTB ECG has 549 records from 290 subjects with composition: 209 males and 81 females with ages from 17–87 years. Every record comprises 15 concurrently measured signals, namely the standard 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) along with the three Frank lead ECGs (vx, vy,vz). Every signal is sampled at 1 kHz and at 16-bit resolution. The range of observation is ±16.384 mV. Associated data include a comprehensive clinical summary, gender, age and diagnostic classes. To ensure an identical sampling rate to ECG-ID, PTB ECG signals were resampled at 500 Hz.

4.1.2. Facial Images

The Faces95 database contains 1440 images with composition: 72 male and female subjects (20 per subject), primarily bachelor-level students. The images are portrait formatted and have resolution of 180 × 200 pixels. For data gathering purposes, the subjects take a single step movement approaching the camera to simulate realistic changes such as variations in head scale and lighting as

well as face position translation. The images also show variations in facial expression, but no variation in hair style.

4.1.3. Fingerprint Images

The FVC2006 database has 7200 images from 150 subjects, each image based on four sensors (1800 images per sensor) with different resolutions: electric field sensor (96 × 96 pixels), optical sensor (400 × 560 pixels), thermal sweeping sensor (400 × 500 pixels) and SFinGe v3.0 (288 × 384 pixels). The dataset is segmented into subsets. Each of the subsets DB1-A to DB4-A contains 140 subjects with 12 images per subject, giving 1680 images per subset. Each of the subsets DB1-B to DB4-B contains 10 subjects with 12 images per subject, giving 120 images per subset.

4.1.4. Virtual Dataset

A total of 58 virtual subjects were produced to decrease the variability. Because the subjects in the Face95 database are primarily bachelor-level students, it was viewed that the ages of the virtual individuals ranged from teens to thirties. The gender information was labeled by two different annotators in a complete affirmation. Utilizing criteria from the labeled gender and the age range, a valid sample was selected, matching age/sex variables from the ECG dataset with the face attributes. To ensure accuracy and fairness, we selected the scanner type used when the subjects were fingerprinted by using the fingerprint images in subset DB1-A of the FVC2006, and arranged each image randomly with the virtual subject already assigned to ECG and face data. Virtual subjects were therefore designed with three modalities (face, fingerprint and ECG) according to the gender and age labeled in the dataset.

4.2. Results

4.2.1. Comparison of the Unimodal and Multimodal Models

Like earlier models, the proposed model achieved the perfect accuracy based on the individual modalities of the ECG-ID, Face95 and FVC2006 datasets. There is no possibility to compare the performance of different models if the accuracy in a test scenario is 100%, so we added noise to achieve better generalization and discrimination (Figs 7–9). To the ECG dataset, we added Gaussian noise whose standard deviation is 0.1 for a series of three normalized QRS complexes. For the fingerprint images, we selected 5% of the pixels and changed the color to black. And for the facial images, we selected 97% of the pixels and changed the color to black.

The new model takes account of multimodality by propagating each modality to a feature extraction module, where it is represented by the fixed size of embedding (the same size for each modality). Unimodal and multimodal models were distinguished in terms of performance by the specific inclusion or exclusion of modalities. Changing the number of input modalities also changed the output of the feature extraction module. Based on the extracted feature embedding, user identification and gender classification are therefore carried on in the fusion and classification layers. The results of the unimodal and multimodal experiments are compared in Table 1.

Each experiment was carried out three times with different random seeds for initializing associated weights, and the average results in each case were reported to avoid biases caused by the parameter initialization. For the identification of virtual subjects, the accuracy was ≤85% when using a single modality, but this increased to >90% when using two or three modalities, with the highest score achieved when all three modalities were included (98.28%).

Similarly, the multimodal models achieved better gender classification results (>93% accuracy) than models based on a single modality, and the model combining all three modalities showed the highest performance (97.70%).

These results using a feature-level fusion confirm that the user identification works well even if there are noises in the input biometric data and that three modalities provide more accurate results than two. The suitability of deep and wide networks for the merging of three modalities is unclear, so the use of hyper-parameters in the fusion layer should be explored. Fig. 10 demonstrates the

consequences of changing the node quantity in each FC layer when processing noisy data from a virtual database. We did not add further layers because this increases the computational requirements and the size of the model.

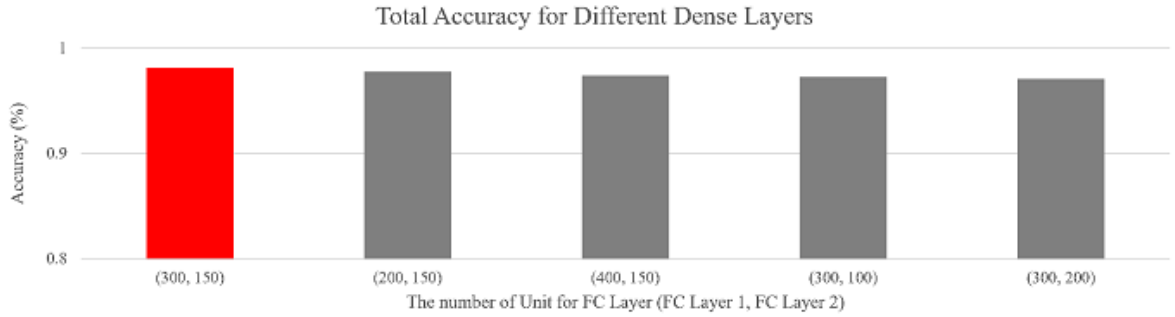

Figure 10. The total accuracy of user identification and gender classification depending on the number of nodes in each fully-connected (FC) layer.

Table 1. Multitasking accuracy for different combinations of modalities (feature-level fusion).

Modality			Task		Accuracy (%)	
ECG	Face	Finger	ID	Gender	ID	Gender
O			O		77.49	-
O				O	-	91.95
	O		O		76.44	-
	O			O	-	88.51
		O	O		83.91	-
		O		O	-	90.80
O	O		O		95.98	-
O	O			O	-	93.68
O		O	O		94.83	-
O		O		O	-	95.40
	O	O	O		96.55	-
	O	O		O	-	94.83
O	O	O	O		**98.28**	-
O	O	O		O	-	**97.70**

Table 2. Comparative performance of single- and multi-task learning models.

Task		Accuracy (%)	
ID	Gender	ID	Gender
O		98.28	-
	O	-	**97.70**
O	O	**98.97**	96.55

The number of nodes in each FC layer was initialized as 150 for the first and second layers, resulting in 900 features after feature fusion. We obtained at least 600 features from two of the three biometric datasets (ECG, facial image and fingerprint). Fig. 10 also summarizes the performance achieved in the experiment and demonstrates the benefits of FC layers with (300,150) nodes, which increased the accuracy of user identification and gender classification to 98%. Our model also outperformed the baseline approach with 300 nodes in the first FC layer and 200 nodes in the second (the poorest performance in this experiment). We therefore found that the (300,150) approach achieved the greatest accuracy in this experiment on the virtual database, and confirmed that the performance of the model degrades if the number of nodes is too large or too small.

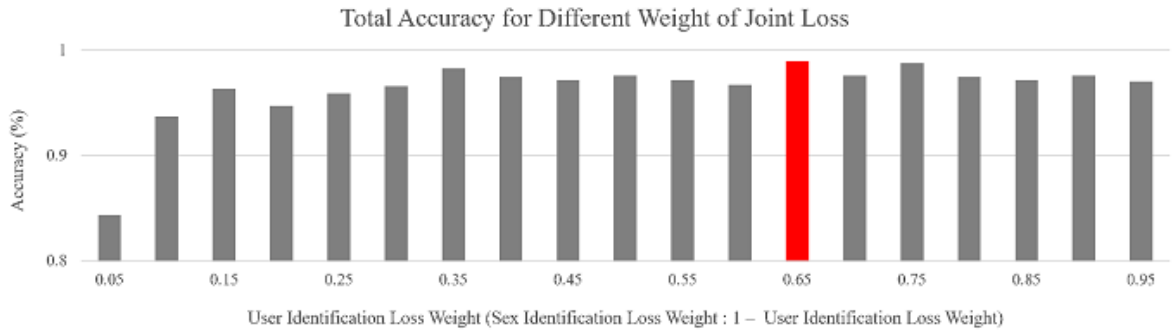

Figure 11. The total accuracy of user identification and gender classification using different sets of joint loss weights.

4.2.2. Single- and Multi-Task Learning

Single- and multi-task learning were compared by adding a loss term as necessary in each experiment. For example, when a gender classification was omitted, the loss function only contained the log likelihood of user-identity classification. However, in multitask learning experiments we included the cross-entropy of user identification and gender classification, and the balance between the two loss functions was controlled by the joint loss weight. The optimized weights were determined experimentally as follows. The performance of the multimodal and multitask model was evaluated on the virtual database with noises. The joint loss weight was varied from 0 to 1 in increments of 0.05. The parameters of the network were initialized according to different random seeds for each experiment. Also, six runs were carried out and averaged to reduce the bias. The final joint loss weights were 0.65 and 0.35 for identifying users and classifying genders, respectively. The performance of our model surpassed that of the baseline approach that applies uniform weights (0.5) for both tasks, and the improvement was greatest for the user identification task (Fig. 11). This confirmed that each task shared features with different weights to enhance the performance of multi-task in comparison to a single-task learning. As shown in Table 2, the multitask learning achieved 0.69% better score in the user identification task but the single-task model was 1.15% better for gender classification. This reflects the trade-off relationship inherent in multiple tasks. For example, adding loss weight to the identification task achieves a better performance in this task, but reduces the accuracy of gender classification. For this context, we took the mean of the identification and classification tasks as the model performance. Even if we cannot achieve the greatest accuracy for gender classification, it excels the user identification accuracy. We can also see the improve efficiency in our proposed multi-task model due to equal predictive performance at a half training time, in comparison to the single-task counterpart.

Table 3. Confusion matrix obtained during authentication.

Data Type		Threshold	FAR (%)	Accuracy (%)
Non-Augmented	Euclidean	6.02–8.56	0.0	100
	Cosine	0.14–0.32	0.0	100
Augmented	Euclidean	9.09	0.66	99.67
	Cosine	0.20–0.25	0.0	100

4.2.3. Authentication

The authentication experiment incorporated a feature vector that consolidates three different modalities (ECG, facial image and fingerprint) from the fusion layer. There is a trade-off during authentication when the model inputs are unimodal data. For example, facial image data are easy to obtain but their accuracy may be compromised by factors such as aging. Fingerprint data are highly accurate, but the quality of the source material deteriorates over time. In a previous study [63], the

equal error rate of authentication was range of 2.69 to 3.07% using fingerprint data alone. We conducted authentication experiments using multimodal data to overcome the tradeoffs inherent to unimodal data and thus improve the performance.

We evaluated the authentication performance using our virtual dataset of 58 subjects. We divided the subjects into two groups of 53 and 5, respectively, then picked 5 random subjects from the group of 53 and combined them with the first group of 5. The authentication experiment therefore involved 53 subjects in the database and 10 test subjects, allowing us to compare the similarity between the two sets. The 53 subjects in the database also have feature vectors with average pooling. Because each subject has nine features, the average pooling can represent the general characteristics of the subject. In addition, we tested authentication with augmented data to prevent over-fitting and to establish the general performance.

Similarity can be measured using various metrics, including the Euclidean distance that represents the actual distance between two points and the cosine distance which is useful when the vector size is not significant. With non-augmented data, the accuracy was 100% and the false acceptance rate (FAR) was 0% when we applied an optimal Euclidean distance threshold range of 6.02 to 8.56 and optimal cosine distance threshold range of 0.14 to 0.32 (Table 3). With augmented data, the accuracy was 99.67% with the Euclidean distance metric but 100% with the cosine distance metric (Table 3). The authentication performance of the multimodal model is therefore near 100%. The receiver operating characteristics (ROCs) for the non-augmented and augmented data are shown in Fig. 12 and Fig. 13, respectively.

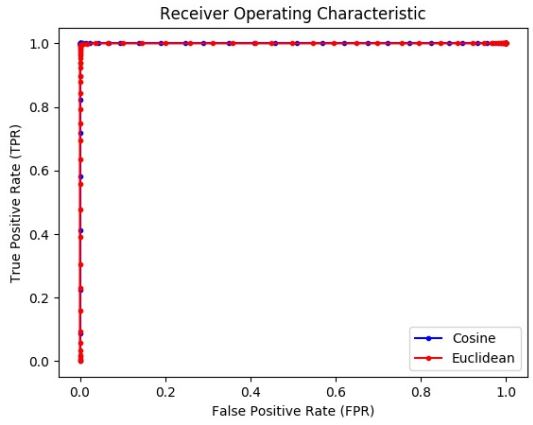

Figure 12. Receiver operating characteristics using non-augmented data.

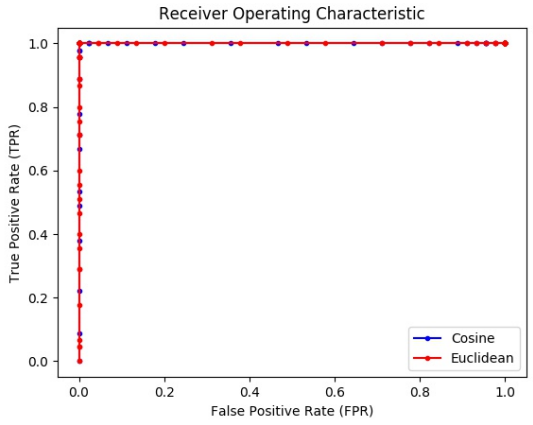

Figure 13. Receiver operating characteristics using augmented data.

4.2.4. Robustness to Noisy Data

To determine whether our model is robust when presented with noisy data, we added noises to one or more of the modalities during the user identification and gender classification tasks (Table 4). For the identification task, the highest accuracy achieved with a single modality was 84.29% (for the noisy fingerprint data). When the noises were added to two of the modalities, the highest accuracy was 95.21%. However, when the noise was added to all three modalities, the accuracy increased to 98.97%. For the gender classification task, the best-performing single modality was ECG, with an accuracy of 90.04% despite the noisy data. When the noise was added to two modalities, the highest accuracy was 95.21% (noisy ECG and fingerprint data). When the noise was added to all three modalities, the accuracy reached 96.55%. The observations herein confirm that the multimodal model performs superior to models with any single modality when noises are included in the data. This mirrors the results with clean data for user identification, where the best single modality (facial images) achieved an accuracy of 99.42% but the trimodal approach achieved an accuracy of 100%. In contrast, in the clean data experiment for gender classification, the trimodal approach achieved an accuracy of 99.43% but was outperformed by the unimodal and bimodal models (100%). Overall,

these experiments demonstrate that that the trimodal model is robust to data, with only a 1.03% drop in performance compared to the clean data for the user identification task.

Table 4. Unimodal and multimodal test for noise input.

Modality			Noise	Accuracy (%)	
ECG	Face	Finger		ID	Gender
O			O	80.08	90.04
	O		O	72.99	87.74
		O	O	84.29	88.70
O	O		O	94.83	95.02
O		O	O	93.68	95.21
	O	O	O	95.21	92.91
O	O	O	O	98.97	96.55
O				98.27	98.85
	O			99.43	100.00
		O		76.44	89.66
O	O			100.00	100.00
O		O		98.85	96.55
	O	O		100.00	98.85
O	O	O		100.00	99.43

Table 5. Accuracy of the multimodal model with incomplete biometrics.

Modality			Noise	Accuracy (%)	
ECG	Face	Finger		ID	Gender
O			O	31.61	82.18
	O		O	39.66	79.89
		O	O	52.30	83.91
O	O		O	81.61	86.21
O		O	O	**87.36**	**89.66**
	O	O	O	**87.36**	89.08
O				21.26	83.33
	O			94.25	95.98
		O		43.10	80.46
O	O			98.28	**98.28**
O		O		70.69	85.63
	O	O		**100.00**	97.70

4.2.5. Robustness to Partial Modalities

In the experiments described above, the combination of modalities in the training and testing sessions were the same. However, in real-world there may be scenarios in which one or more of the modalities would be incomplete. To prove the robustness of the multimodal model when presented with incomplete input data, we trained the model using all three modalities but evaluated its performance with one or more missing, in the presence and absence of the noise (Table 5).

In the presence of artificial noise, the accuracy of the model never reached 90% regardless of which modality or modalities were omitted, but there was a jump in performance when the number of modalities increased from one to two, with all combinations of two modalities achieving an accuracy of 80% or more for both tasks. In the absence of artificial noise, the accuracy reached 100% for user identification and 98.28% for gender classification even in the absence of ECG data. The model therefore works well with limited biometric input if the noise in each dataset is weaker than

the levels indicated in Fig. 7–9. Alternative models are not required as long as we can use at least two modalities.

4.2.6. Fusion Techniques

Our new model uses fusion techniques to concatenate biometric characteristics, thus reducing the independent characteristics of each modality for reliability improvement and accuracy maintenance [64]. We compared feature-level fusion; which was applied prior to matching to aggregate the features into a single vector, and score-level fusion; which was applied post matching to compute the similarity level utilizing particular procedures (i.e., sum, product and max in our experiments) for every output to generate the final output vector (Table 6).

When using the feature-level fusion, the accuracy of user identification was 98.97% and the accuracy of gender classification was 96.55%. When using the score-level fusion, we found that the sum rule yields the accuracy of 98.85% and 99.42% for the same tasks. This represents an improvement of 2.87% for the gender classification, reflecting the robustness of the sum rule when exposed to noisy data [65]. For the score-level fusion method, the change of weights was also experimented for each modality, assigning a weight of 0.5 to one modality and 0.25 to the others. The results indicate that achieving the greatest accuracy is noted when the ECG weight is larger than or equal to the other modalities. Herein ECG appears to make the most significant contribution among the three modalities to improve the performance.

Table 6. Comparison of feature-level and score-level fusion methods.

Fusion Level	Rule	Weight			Accuracy (%)	
		ECG	Face	Finger	ID	Gender
Feature	-	-	-	-	**98.97**	96.55
Score	Sum	0.33	0.33	0.33	98.27	**99.42**
		0.50	0.25	0.25	98.85	**99.42**
		0.25	0.50	0.25	98.85	**99.42**
		0.25	0.25	0.50	97.70	**99.42**
	Product	0.33	0.33	0.33	96.55	89.08
		0.50	0.25	0.25	95.98	89.66
		0.25	0.50	0.25	93.10	89.08
		0.25	0.25	0.50	93.68	87.36
	Max	0.33	0.33	0.33	89.66	89.66
		0.50	0.25	0.25	89.66	87.93
		0.25	0.50	0.25	89.66	86.21
		0.25	0.25	0.50	89.66	87.36

4.2.7. Data Augmentation

A data augmentation virtually increases the amount of samples in the dataset using the existing ones while playing a role of a regularizer preventing the overfitting. It is particularly helpful for improving the performance in imbalanced class problems. We applied the data augmentation techniques described in preprocessing section and generated 455 data points for each modality. The experiments dealing with two tasks simultaneously were performed for each combination of modalities, but only the database was different. We found that the difference in the performance of the model when tested on the original and augmented databases ranged from 1.9% to 24.47% depending on the individual modality. When multiple modalities were used in the augmented dataset, the accuracy was always 99% or more. We therefore confirmed that data augmentation makes the model get generality, and that it is possible to create a model that attains high accuracy even with a small amount of data in real-world settings.

Modality			Augment	Accuracy (%)	
ECG	Face	Finger		ID	Gender
O				80.08	90.04
	O			72.99	87.74
		O		84.29	88.70
O	O			94.83	95.02
O		O		93.68	95.21
	O	O		95.21	92.91
O	O	O		98.97	96.55
O			O	99.92	100.00
	O		O	97.46	98.31
		O	O	86.19	95.23
O	O		O	100.00	100.00
O		O	O	99.98	100.00
	O	O	O	99.49	99.03
O	O	O	O	100.00	100.00

5. Conclusion

We have introduced a new multimodal and multitask learning model that uses ECG, facial image and fingerprint features for user identification and gender classification. We have conducted a number of experiments in an environment that assumes extreme noises, indicating that our model is robust to noises, and have achieved greater accuracies than unimodal approaches. The proposed model has also proven to be robust against the spoof attack problem that unimodal model is vulnerable to. Our results show the desirable characteristics of our proposal. For identifying users, classifying genders and authentication, our model outperforms other approaches reported in the existing literature. The feature-level fusion ensures that the proposed model achieves similar performances despite incomplete data (losing one out of the tree), indicating its suitability for practical scenarios with a good level of security and accuracy. Future research directions that could lead to more accurate user identification, gender classification and authentication are as follows. End-to-end learning approach can be applied to the whole network with a sufficiently large database and more biometrics. A model that works well with more biometrics can achieve better performance in the missing modal situation since the remaining modals can complement the functionality of the missing modal. It should be noted that the technique used for the multimodal data fusion in the proposed model was to simply concatenate the multimodal features, and this can be improved further by using an attention model to select the most suitable modality or feature that is present in the sample.

Acknowledgement

This work is supported in part by the Center for Cyber-Physical Systems, Khalifa University, under Grant Number 8474000137-RC1-C2PS-T3. The authors declare no conflict of interest.

References

1. Wang, Y.; Agrafioti, F.; Hatzinakos, D.; Plataniotis, K.N. Analysis of human electrocardiogram for biometric recognition. *EURASIP journal on Advances in Signal Processing* **2007**, *2008*, 148658.
2. Coutinho, D.P.; Fred, A.L.; Figueiredo, M.A. One-lead ECG-based personal identification using Ziv-Merhav cross parsing. 2010 20th International Conference on Pattern Recognition. IEEE, 2010, pp. 3858–3861.

3. Lee, J.; Chee, Y.; Kim, I. Personal identification based on vectorcardiogram derived from limb leads electrocardiogram. *Journal of Applied Mathematics* **2012**, *2012*.
4. Li, M.; Narayanan, S. Robust ECG biometrics by fusing temporal and cepstral information. 2010 20th International Conference on Pattern Recognition. IEEE, 2010, pp. 1326–1329.
5. Fang, S.C.; Chan, H.L. Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space. *Pattern Recognition* **2009**, *42*, 1824–1831.
6. Kim, J.; Lee, K.B.; Hong, S.G. ECG-based biometric authentication using random forest. *Journal of the Institute of Electronics and Information Engineers* **2017**, *54*, 100–105.
7. Plataniotis, K.N.; Hatzinakos, D.; Lee, J.K. ECG biometric recognition without fiducial detection. 2006 Biometrics symposium: Special session on research at the biometric consortium conference. IEEE, 2006, pp. 1–6.
8. Gang, G.W.; Min, C.H.; Kim, T.S. Development of Single Channel ECG Signal Based Biometrics System. *Journal of the Institute of Electronics Engineers of Korea CI* **2012**, *49*, 1–7.
9. Odinaka, I.; Lai, P.H.; Kaplan, A.D.; O'Sullivan, J.A.; Sirevaag, E.J.; Rohrbaugh, J.W. ECG biometric recognition: A comparative analysis. *IEEE Transactions on Information Forensics and Security* **2012**, *7*, 1812–1824.
10. Kim, S.K.; Yeun, C.Y.; Damiani, E.; Lo, N.W. A machine learning framework for biometric authentication using electrocardiogram. *IEEE Access* **2019**, *7*, 94858–94868.
11. Al Alkeem, E.; Kim, S.K.; Yeun, C.Y.; Zemerly, M.J.; Poon, K.F.; Gianini, G.; Yoo, P.D. An enhanced electrocardiogram biometric authentication system using machine learning. *IEEE Access* **2019**, *7*, 123069–123075.
12. Sun, L.; Lu, Y.; Yang, K.; Li, S. ECG analysis using multiple instance learning for myocardial infarction detection. *IEEE transactions on biomedical engineering* **2012**, *59*, 3348–3356.
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **2017**, *60*, 84–90.
14. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.r.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; others. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine* **2012**, *29*, 82–97.
15. Isin, A.; Ozdalili, S. Cardiac arrhythmia detection using deep learning. *Procedia computer science* **2017**, *120*, 268–275.
16. Labati, R.D.; Muñoz, E.; Piuri, V.; Sassi, R.; Scotti, F. Deep-ECG: convolutional neural networks for ECG biometric recognition. *Pattern Recognition Letters* **2019**, *126*, 78–85.
17. Velayudhan, A.; Peter, S. Noise analysis and different denoising techniques of ECG signal-a survey. *IOSR journal of electronics and communication engineering* **2016**, *3*, 641–644.
18. Biel, L.; Pettersson, O.; Philipson, L.; Wide, P. ECG analysis: a new approach in human identification. *IEEE Transactions on Instrumentation and Measurement* **2001**, *50*, 808–812.
19. Israel, S.A.; Irvine, J.M.; Cheng, A.; Wiederhold, M.D.; Wiederhold, B.K. ECG to identify individuals. *Pattern recognition* **2005**, *38*, 133–142.
20. Kim, K.S.; Yoon, T.H.; Lee, J.W.; Kim, D.J.; Koo, H.S. A robust human identification by normalized time-domain features of electrocardiogram. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, 2006, pp. 1114–1117.
21. Gahi, Y.; Lamrani, M.; Zoglat, A.; Guennoun, M.; Kapralos, B.; El-Khatib, K. Biometric identification system based on electrocardiogram data. 2008 New Technologies, Mobility and Security. IEEE, 2008, pp. 1–5. Song, H.K.;
22. AlAlkeem, E.; Yun, J.; Kim, T.H.; Yoo, H.; Heo, D.; Chae, M.; Yeun, C.Y. Deep user identification model with multiple biometric data. *BMC bioinformatics* **2020**, *21*, 1–11.
23. Israel, S.A.; Scruggs, W.T.; Worek, W.J.; Irvine, J.M. Fusing face and ECG for personal identification. 32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. IEEE, 2003, pp. 226–231.
24. BE, M.; M Abhishek, A.; KR, V.; M Patnaik, L.; others. Multimodal Biometric Authentication using ECG and Fingerprint. IJCA, 2015, Vol. 111, pp. 33–39.
25. Boumbarov, O.; Velchev, Y.; Tonchev, K.; Paliy, I.; Chetty, G. Face and ECG based multi-modal biometric authentication. In *Advanced biometric technologies*; InTech, 2011.
26. Zokaee, S.; Faez, K. Human identification based on ECG and palmprint. *International Journal of Electrical and Computer Engineering* **2012**, *2*, 261.
27. Fan, J.; Cao, X.; Wang, Q.; Yap, P.T.; Shen, D. Adversarial learning for mono-or multi-modal registration. *Medical image analysis* **2019**, *58*, 101545.

28. Zhou, T.; Liu, M.; Thung, K.H.; Shen, D. Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE transactions on medical imaging* **2019**, *38*, 2411–2422.
29. Zhou, T.; Thung, K.H.; Liu, M.; Shi, F.; Zhang, C.; Shen, D. Multi-modal latent space inducing ensemble SVM classifier for early dementia diagnosis with neuroimaging data. *Medical Image Analysis* **2020**, *60*, 101630.
30. Gavrilova, M.L.; Monwar, M. *Multimodal biometrics and intelligent image processing for security systems*; Information Science Reference, 2013.
31. Chergui, O.; Bendjenna, H.; Meraoumia, A.; Chitroub, S. Combining palmprint & finger-knuckle-print for user identification. 2016 International Conference on Information Technology for Organizations Development (IT4OD). IEEE, 2016, pp. 1–5.
32. Huo, G.; Liu, Y.; Zhu, X.; Dong, H.; He, F. Face–iris multimodal biometric scheme based on feature level fusion. *Journal of Electronic Imaging* **2015**, *24*, 063020.
33. Snelick, R.; Uludag, U.; Mink, A.; Indovina, M.; Jain, A. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE transactions on pattern analysis and machine intelligence* **2005**, *27*, 450–455.
34. Ross, A.A.; Nandakumar, K.; Jain, A.K. *Handbook of multibiometrics*; Vol. 6, Springer Science & Business Media, 2006.
35. Ross, A.; Jain, A. Information fusion in biometrics. *Pattern recognition letters* **2003**, *24*, 2115–2125.
36. Ross, A.A.; Govindarajan, R. Feature level fusion of hand and face biometrics. Biometric technology for human identification II. International Society for Optics and Photonics, 2005, Vol. 5779, pp. 196–204.
37. Rattani, A.; Kisku, D.R.; Bicego, M.; Tistarelli, M. Feature level fusion of face and fingerprint biometrics. 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems. IEEE, 2007, pp. 1–6.
38. Kittler, J.; Hatef, M.; Duin, R.P.; Matas, J. On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence* **1998**, *20*, 226–239.
39. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern classification*; John Wiley & Sons, 2012.
40. Lam, L.; Suen, S. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **1997**, *27*, 553–568.
41. Daugman, J. Biometric decision landscapes. Technical report, University of Cambridge, Computer Laboratory, 2000.
42. Rautiainen, M.; Seppanen, T. Comparison of visual features and fusion techniques in automatic detection of concepts from news video. 2005 IEEE International Conference on Multimedia and Expo. IEEE, 2005, pp. 932–935.
43. Dummett, M.A.; others. *Principles of electoral reform*; Oxford University Press Oxford, 1997.
44. Ho, T.K.; Hull, J.J.; Srihari, S.N. Decision combination in multiple classifier systems. *IEEE transactions on pattern analysis and machine intelligence* **1994**, *16*, 66–75.
45. Aslam, J.A.; Montague, M. Models for metasearch. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001, pp. 276–284.
46. Aslam, J.A.; Montague, M. Bayes optimal metasearch: a probabilistic model for combining the results of multiple retrieval systems. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 2000, pp. 379–381.
47. Melnik, O.; Vardi, Y.; Zhang, C.H. Mixed group ranks: Preference and confidence in classifier combination. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2004**, *26*, 973–981.
48. Ergin, S.; Uysal, A.K.; Gunal, E.S.; Gunal, S.; Gulmezoglu, M.B. ECG based biometric authentication using ensemble of features. 2014 9th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, 2014, pp. 1–6.
49. Caruana, R. Multitask Learning: A Knowledge-Based Source of Inductive Bias ICML. *Google Scholar Google Scholar Digital Library Digital Library* **1993**.
50. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098* **2017**.
51. Ji, J.; Chen, X.; Luo, C.; Li, P. A deep multi-task learning approach for ECG data analysis. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2018, pp. 124–127.
52. Lin, Y.; Lv, F.; Zhu, S.; Yang, M.; Cour, T.; Yu, K.; Cao, L.; Huang, T. Large-scale image classification: Fast feature extraction and SVM training. CVPR 2011. IEEE, 2011, pp. 1689–1696.

53. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261* **2016**.
54. Salloum, R.; Kuo, C.C.J. ECG-based biometrics using recurrent neural networks. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017, pp. 2062–2066.
55. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
56. Duchi, J.; Hazan, E.; Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research* **2011**, *12*.
57. Hinton, G.; Srivastava, N.; Swersky, K. Neural networks for machine learning. *Coursera, video lectures* **2012**, *264*.
58. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation* **2000**, *101*, e215–e220.
59. Lugovaya, T. Biometric human identification based on electrocardiogram. *Master's thesis, Faculty of Computing Technologies and Informatics, Electrotechnical University 'LETI', Saint-Petersburg, Russian Federation* **2005**.
60. Bousseljot, R.; Kreiseler, D.; Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik/Biomedical Engineering* **1995**, *40*, 317–318.
61. Hond, D.; Spacek, L. Distinctive Descriptions for Face Processing. BMVC, 1997, number 0.2, pp. 0–4.
62. Cappelli, R.; Ferrara, M.; Franco, A.; Maltoni, D. Fingerprint verification competition 2006. *Biometric Technology Today* **2007**, *15*, 7–9.
63. Jain, A.K.; Hong, L.; Pankanti, S.; Bolle, R. An identity-authentication system using fingerprints. *Proceedings of the IEEE* **1997**, *85*, 1365–1388.
64. El-Sayed, A. Multi-biometric systems: a state of the art survey and research directions. *IJACSA) International Journal of Advanced Computer Science and Applications* **2015**, *6*.
65. Vishi, K.; Mavroeidis, V. An evaluation of score level fusion approaches for fingerprint and finger-vein biometrics. *arXiv preprint arXiv:1805.10666* **2018**.