A multimodal-Siamese Neural Network (mSNN) for Person Verification using Signatures and EEG

Debashis Das Chakladar[a], Pradeep Kumar[b,*], Partha Pratim Roy[a], Debi Prosad Dogra[c], Erik Scheme[b], Victor Chang[d]

[a]*Department of Computer Science & Engineering, Indian Institute of Technology Roorkee, Roorkee, Pin code-247667, India*
[b]*Institute of Biomedical Engineering, University of New Brunswick, Canada*
[c]*School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Odisha, Pin code- 752050, India*
[d]*School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK*

Abstract

Signatures have long been considered to be one of the most accepted and practical means of user verification, despite being vulnerable to skilled forgers. In contrast, EEG signals have more recently been shown to be more difficult to replicate, and to provide better biometric information in response to known a stimulus. In this paper, we propose combining these two biometrics using a multimodal Siamese Neural Network (mSNN) for improved user verification. The proposed mSNN network learns discriminative temporal and spatial features from the EEG signals using an EEG encoder and from the offline signatures using an image encoder. Features of the two encoders are fused into a common feature space for further processing. A Siamese network then employs a distance metric based on the similarity and dissimilarity of the input features to produce the verification results. The proposed model is evaluated on a dataset of 70 users, comprised of 1400 unique samples. The novel mSNN model achieves a 98.57% classification accuracy with a 99.29% True Positive Rate (TPR) and False Acceptance Rate (FAR) of 2.14%, outperforming the current state-of-the-

*Corresponding author
Email addresses: ddaschakladar@gmail.com (Debashis Das Chakladar), pkumar1@unb.ca (Pradeep Kumar), proy.fcs@iitr.ac.in (Partha Pratim Roy), dpdogra@iitbbs.ac.in (Debi Prosad Dogra), escheme@unb.ca (Erik Scheme), victorchang.research111@gmail.com (Victor Chang)

art by 12.86% (in absolute terms). This proposed network architecture may also be applicable to the fusion of other neurological data sources to build robust biometric verification or diagnostic systems with limited data size.

Keywords: User verification, Multimodal, EEG, Siamese Neural Network, LSTM, CNN

1. Introduction

User identification, such as through video surveillance or photo identification, is becoming increasingly pervasive [1, 2, 3]. This has been at least partially motivated because these modalities generally require no explicit compliance from the target subject. User verification, however, is typically used in higher security applications to provide additional assurance that a person is who they say they are. As such, it generally involves direct interaction with authentication systems to gain access to physical locations, online transactions, or for remote access control. Even with increased digitization and the availability of more advanced biometric traits such as facial characteristics, the retina or iris, and finger or palm prints [4], signatures remain a commonly used hallmark for verifying identity. A typical pen-paper-based signature elicits discriminative information through the pen's pressure, the shape of loops, the speed and care of writing, and the up-down motion of the pen. However, because these are behavioural traits, once the shape and stroke of the original signature are known, a trained imposter can learn them to exploit vulnerabilities.

A wide variety of machine learning approaches have been applied to the verification of signatures. Fang *et al.* [5], for example, discussed signature verification via tracking of features and pen-stroke positions, but reported a False Acceptance Rate (FAR) of 16.7%. Alaei *et al.* [6] developed a signature verification system using interval symbolic representation of images of signatures (termed as *offline*) and a fuzzy similarity measure. Ferrer *et al.* [7] classified geometric features of signatures using Support Vector Machine (SVM), Hidden Markov Model (HMM), and Euclidean classifiers. Several other works have

explored variations of features and classifiers with varying degrees of success, typically with Average Error Rates above 10% [8, 9, 10]. More recently, deep learning tools, such as Convolutional Neural Networks (CNN), have been used to automatically learn features from offline signatures to further prevent forgery attacks [11].

Brain-Computer-Interface (BCI) approaches have also been widely explored in recent years due to inherent advantages over other techniques. In particular, the capture non-invasive measurement of Electroencephalogram (EEG) has become increasingly used to capture brain-related signals. The use of physiological signals in biometrics are much more difficult to spoof, because they are either independent of, or at least not solely dependent on, behaviour. Studies have shown that EEG-based user verification can indeed outperform offline signature-based systems [12, 13]. Shannon entropy-based EEG features of various EEG wavebands (alpha, beta, and gamma), for example, have been used for user verification, yielding classification accuracies of 97.1% [14]. Spatial and statistical features (mean, standard deviation, kurtosis, skewness) of EEG have also been used effectively across several works [15, 16]. In [17], Pham *et al.*developed an EEG-based multifactor (age, gender) person verification system using the Power Spectral Density (PSD) feature of EEG with a SVM classifier. They achieved 97.1% and 96.7% classification accuracies with gender and age factors considered, respectively. Mu *et al.* [18] developed a person authentication system using fuzzy entropy features of EEG and a Artificial Neural Network (ANN) classifier. Recently, Kumar *et al.* [19] implemented a framework using statistical features of EEG to secure mobile devices, and reported a 25% global Half of Total Error Rate (HTER) and 2.01% local HTER with 50 users.

Because performance improves when EEG are measured in response to a stimulus (termed the Evoked-Response Potential,or ERP) [20], researchers have explored various recording scenarios including during rest [21, 22], while listening to music [23], or while entering password patterns on a mobile device [19, 24].

Without a known and consistent stimulus, however performance may degrade. For example, a user may lose focus while listening to music or watching

a video. Moreover, the emotions and facial expression generated in response to such stimuli may add confounding noise to the EEG, degrading system performance [23]. Therefore, the selection of a robust stimulus that can be easily integrated with EEG is of utmost importance when designing a biometric system.

The literature has also shown that unimodal biometric systems yield limited effectiveness and remain vulnerable to attacks and forgeries [25, 26]. Consequently, in recent years, multimodal biometric systems have gained traction, taking advantage of mixed physical and behavioral information sources. In these systems, an intruder must break more than one biometric trait, and often concurrently. In multimodal biometrics fusion, features from different biometric traits can be combined at various levels, including based on sensor data, matching score, or decision level (e.g. based on the classification performance of unimodal systems) [27]. For example, Galdi *et al.* [28] implemented a multimodal authentication framework fusing iris recognition with recognition of the image sensor. Chang *et al.* [29] combined images of the face and ear using Principal Component Analysis (PCA) to achieve a person identification rate of 90.9%. Frontal face images have been combined with text-dependent voice biometrics [30], and face has been fused with fingerprint [31], the latter of which achieved a 3.9% False Rejection Rate (FRR).

Saini *et al.* [32] proposed a multimodal user verification system using EEG and offline signatures. They achieved 98.24% user identification accuracies with FAR of 15% using Pyramid Histogram of Oriented Gradients (PHOG) and Discrete Wavelet Transform (DWT) features. As with passwords, signature-based verification systems also have the advantage of being changeable. That is, a user could, technically, change their signature if attacked by imposters/forgers. Moreover, the creation of a repeatable signature involves a series of complex neuromuscular processes requiring years of repetition and motor learning [33, 34]. This not only adds an additional layer of authentication, but also makes signatures a perfect stimuli against which to evaluate the neural response. Such a scenario is represented in Figure 1, which shows the offline signature and

corresponding EEG of a true user, Figure 1(a), and forged sample, Figure 1(b). It can be seen from the figure that the forger was able to mimic the signature quite well, but was unable to mimic the brain signals. Moreover, it can also be seen that the forger failed to mimic the signature duration, as noted by the shorter duration of corresponding EEG signals for the genuine user (350 vs 500 samples).

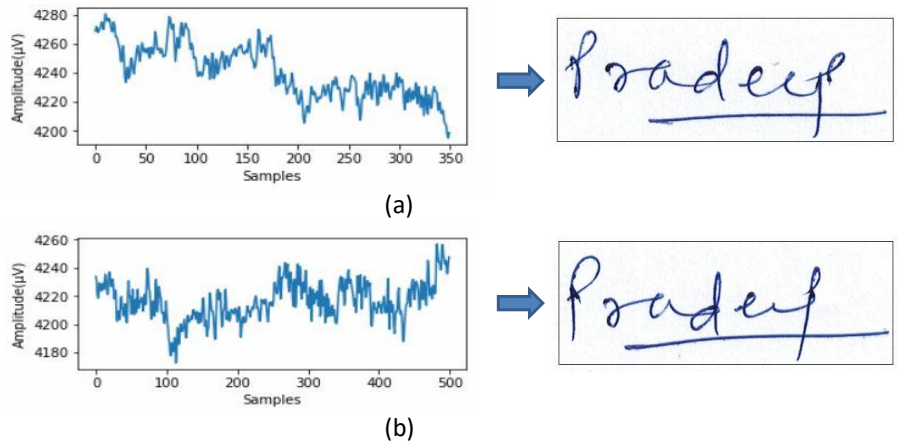

Figure 1: Example of (a) an EEG signal with original signature. (b) an EEG signal of a corresponding imposter signature. The EEG signals was taken from the F7 channel for visualization purposes only.

Deep learning architectures leveraging Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) techniques have become increasingly favored for the automatic extraction of features and classification tasks. These techniques, however, are widely accepted to require large amounts of labeled data, which may not be possible (or preferable) in biometric applications nor with short sessions of the EEG recorded while signing. Moreover, it is not always feasible or practical to retrain biometric models whenever a new user is added to the dataset. Consequently, a different approach is needed that can be trained on fewer data samples and can extend to new users without retraining the model. One potential approach is one-shot learning, which can be implemented using a Siamese network comprised of twin sister-networks that share the same weights.

Here, we propose a novel multimodal Siamese Neural Network (mSNN) for person verification using combined EEG and offline signature data. The network accepts a combination of EEG and offline signature pairs from genuine and forged users to model the biometric data, as depicted in Figure 2. The signatures are processed using CNNs, whereas the EEG is modeled using LSTMs. The resulting encodings are fused to find structural commonalities in the multimodal data. Finally, the encoding of input pairs is optimized based on a loss function to facilitate person verification.

This network configuration could also be beneficial for fusing other modalities for other applications, such as to improve the diagnosis of neurological or psychiatric diseases. This ability to discriminate using fewer samples could be particularly attractive as it would reduce the burden of collecting multiple samples of neuroimaging data.

This study therefore presents the following novel contributions:

1. A multimodal Siamese Neural Network (mSNN) is proposed that fuses two different biometric traits, namely EEG and offline signature, for person verification.
2. The proposed framework leverages a spatio-temporal architecture to generate multimodal encodings of the EEG time-series and the spatial signature data.
3. The proposed multimodal user verification system is evaluated on a large dataset, demonstrating its superiority over traditional approaches.

2. Methodology

Siamese networks, a unique kind of neural network architecture, consist of two similar types of subnetworks with the same configuration (identical parameters and shared weights) [35, 36]. These two subnetworks are connected by a loss function, which calculates a similarity score between two input samples based on the feature representation of the two subnetworks. Here, Siamese Networks are extended for use with two different modalities - behavioural (signature) and

electrophysiological (EEG), enabling a multimodal redundancy in verification. The proposed mSNN takes the combination of EEG signal and offline signature pair from two users (e.g., subject 1 and subject2) and produces an output based on the similarity between the two pairs, as depicted in Figure 2. For consistency, the EEG and signature of the genuine subject 1 are passed to the first subnetwork. Conversely, the EEG from subject 2 produced while attempting a forgery of subject 1's signature, are passed to the second subnetwork.The EEG signal is processed using an LSTM to learn temporal relationships in the data sequences. In contrast, the offline signature is processed using a CNN to find spatial features from the signature image. Before passing the EEG and signature into the respective networks, however, both inputs are pre-processed. Because the size of raw signature images may vary (from 304× 240 to 798×482 in the current dataset), all images were resized to 155×220 using binary interpolation [37]. To normalize the images, each image pixel was divided by the standard deviation of the pixels across all images in the dataset. EEG data were preprocessed using a 2^{nd} order Bandpass filter within the alpha range of 8-12 Hz [38, 39].

2.1. Multimodal Feature Learning

The proposed model takes a pair of inputs that consist of a signature image and a time-series of EEG signal. Because the structure of these inputs is entirely different, finding a common pattern or mapping between them is non-trivial. Consequently, each input was transformed to a shared space by maximizing the similarity between two embeddings of each input representation. To achieve this, a Siamese Network was designed that 1) learns the joint embeddings between the EEG and images using deep encoders and 2) maximizes the similarity measure between the two modalities [40].

For the purpose of explanation, let a multimodal dataset of sample pairs, be given by X=$\{e_k, i_k\}$, where $k = 1, 2, 3, ...N$, and the EEG signal and signature image of k^{th} sample are represented as (e_k) and (i_k), respectively. Assuming Γ and δ represent the space of EEG samples and images, respectively, the objective is to build two encoders that convert the EEG and signature image into

a common space (η). The EEG encoder (β) and image encoder (α) are then represented as: β: $\Gamma \rightarrow \eta$ and α: $\delta \rightarrow \eta$, respectively. Here, the EEG encoder (β) that maps the EEG signal into the common space was built using LSTM modules of different sizes to facilitate learning of temporal features from the EEG. A linear layer was then applied to project the learned features into a one-dimensional feature vector in the common space (η). Similarly, the image encoder (α) mapped the input image into the common space (η) using Convolutional Neural Networks (CNN) with different sizes of convolutional layers. Again, a linear layer was applied to produce a one-dimensional feature vector in the common space (η). The output of two encoders (α and β) was then concatenated using a compatibility function (F). This compatibility function (1) is used to measure the similarity between two sample pair embeddings in the common space (η). Finally, the output of F from two subnetworks is passed to a loss function to generate the final output of the model, as summarized in Figure 2.

$$F(e, i) = [\beta(e), \alpha(i)] \tag{1}$$

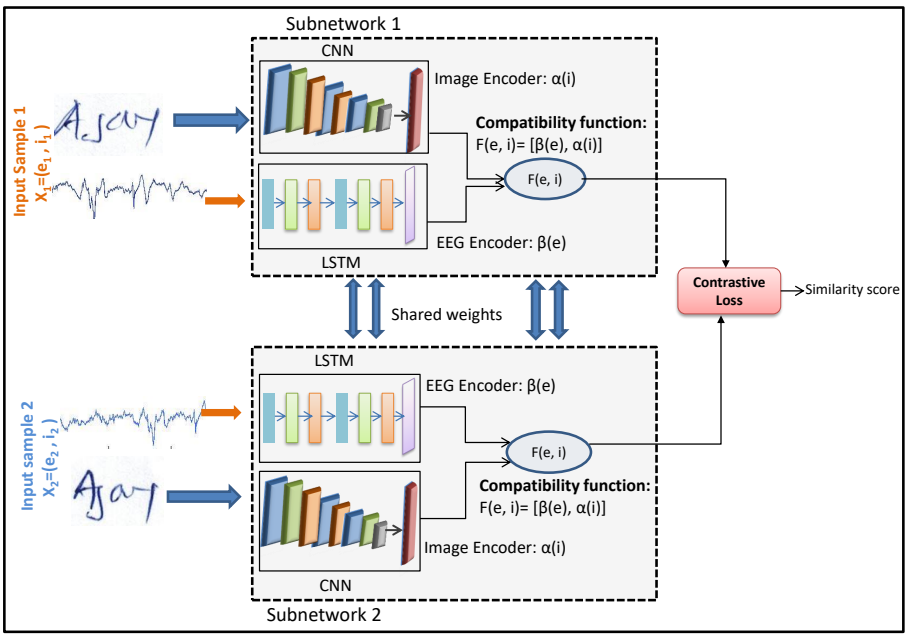

Figure 2: Block diagram of the proposed multimodal Siamese Neural Network (mSNN) for user verification using EEG signals and offline signatures.

The proposed model compares two input samples (x_1, x_2), where x_1=(e_1, i_1) and x_2=(e_2, i_2) using the distance between them in the common space calculated according to a contrastive loss function [41]. The loss is computed between pairs of training data known as positive-positive (similar) pairs and positive-negative (dissimilar) pairs. The objective is to learn representations with a small distance (D) between them for similar pairs, and greater distance than some margin value (m) for dissimilar pairs. In this case, a value of zero means the instances are similar, and a value of one indicates the instances are dissimilar. Unlike other loss functions, such as Cross-Entropy Loss (a classification loss function whose objective is to learn to predict class probabilities independently for each sample), contrastive loss is a metric learning loss, which operates on the data points produced by the network and their positions relative to each other. The objective of contrastive loss is to predict relative distances between inputs so that a

threshold can be used as a tradeoff between the genuine user and an imposter. The contrastive loss is defined in (2), where m is the margin. Here, we set the value of m to be 1. The binary indicator function (I) denotes whether the two samples (x_1, x_2) belongs to the same class or not.

$$L(x_1, x_2, I) = I\ D^2 + (1 - I)\ max(0, m - D)^2 \tag{2}$$

where D represents the Euclidean distance between two learned embeddings $F(x_1)$ and $F(x_2)$ from two subnetworks (3) .

$$D = \|F(x_1) - F(x_2)\|_2 \tag{3}$$

A detailed diagram of the model architecture for person verification, along with all associated parameters, is shown in Figure 3. Additional details about the unique processing of EEG and signature images are given as follows.

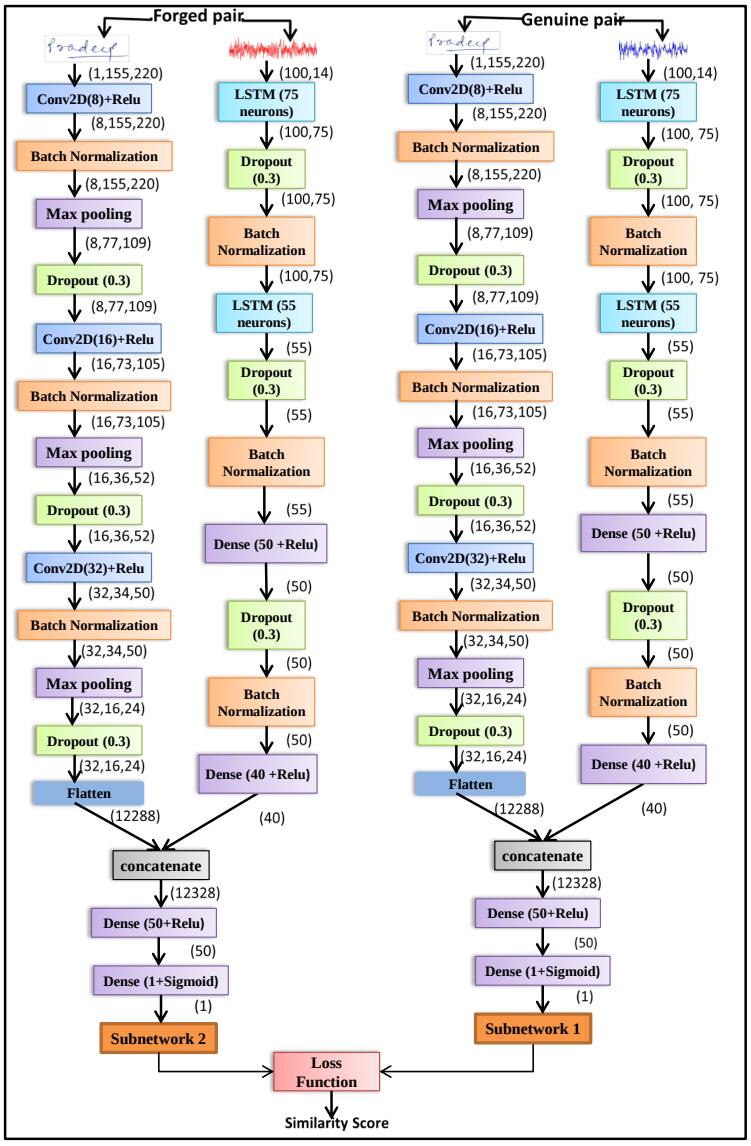

Figure 3: Proposed deep spatio-temporal siamese network for user verification using offline signature and EEG signal.

2.2. EEG Encoder

The EEG encoder (β) converts the neural signal (EEG) into a common representation. To capture the longer dynamics in the temporal dimension of

the EEG signal, we have used two consecutive LSTM layers. The EEG encoder consists of two LSTM layers; the first LSTM layer consists of 75 neurons, and the second one with 55 neurons. Each LSTM layers are followed by a dropout and batch normalization layer. Next, a dense layer of 50 neurons, followed by a dropout and batch normalization layer, is connected to another dense layer with 40 neurons. Informative features of the input EEG have been extracted using a series of LSTM layers. The EEG feature maps (in the embedding space) generated by the EEG encoder is shown in Figure 4. Each LSTM layer extracts significant temporal information from the EEG signal. LSTM layers are used to extracts deep temporal dynamics from the input data. A series of dense layers have been applied to make the network deeper, which leads to better performance.

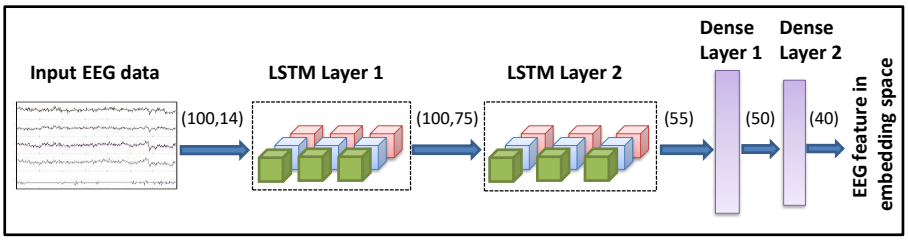

Figure 4: A block representation of the EEG encoder. EEG encodings are marked with different colour in each LSTM layer. The output dimension of each layer has been mentioned beside the block.

2.3. Signature Image Encoder

The image encoder (α) consists of a series of convolution layers, where each convolution layer further followed by batch normalization, max pooling, and dropout layer. Here, we use a 2D CNN for extracting the spatial information from the input signature. The first convolutional layer takes the signature image (size: 155×220) as input and performs the filter operation with 8 filters of size 11×11. The outputs of the first convolutional layer are passed to the second convolutional layer (16 filters with a size of 5×5). Next, the outputs of the second convolutional layer are passed to the third convolutional layer (32 filters

with a size of 3×3). Finally, the summary of all the spatial features of a signature image is passed to the flatten layer to produce a 1D feature vector.

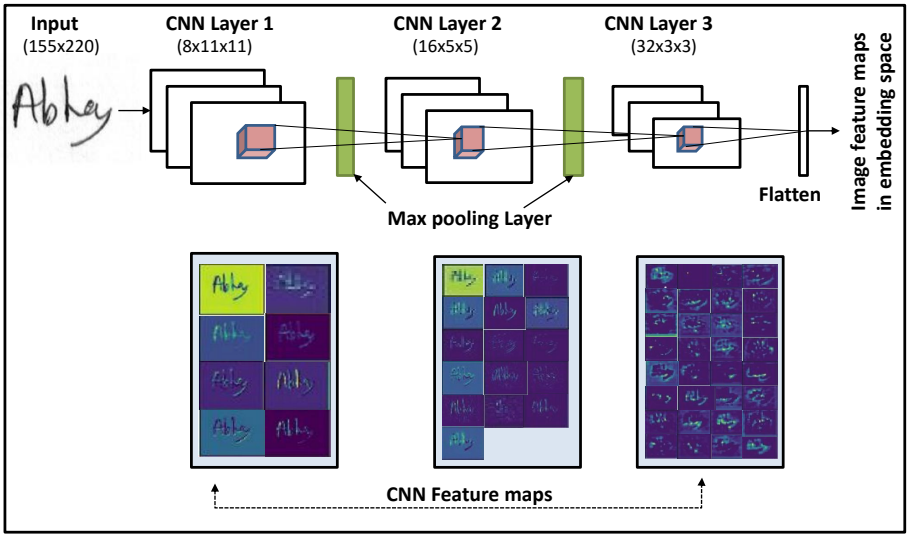

Figure 5: Image encoder with layer-wise feature maps. Initial convolution layers (layer 1, layer 2) extracts low-level visual features (color contrast, edge, shape, texture) of the signatures. The final layer (layer 3) extracts context-specific features in embedding feature space.

Feature maps can effectively show how each convolution layer (within the image encoder) extracts the signature related pixels from the input signature image. The layer-wise feature maps along with the detailed architecture of image encoder are displayed in the Figure 5. It can be noted that the feature maps close to the input detect low level detail of the input, whereas the feature maps close to the output produce more specific features.

Next, the one-dimensional feature vector of EEG and image encoder is passed to the 'concatenate' layer to produce the output for each subnetwork (1). The same configuration and weights of EEG and image encoder are applied for both the sister networks of mSNN. Finally, the loss function takes the output vectors of two sister networks and computes the distance-based similarity measure as the output.

2.4. User Verification

For a given pair of identity and input data, where the input data refers to the pair of signatures and EEG of the siamese network, the proposed system can efficiently determine whether the input data provides a genuine match against the identity or not. The proposed model gives the similarity measure as output, and based on that similarity measure; we perform the verification task. If the similarity measure is lower than the selected threshold, the user request is accepted, otherwise rejected. The similarity measure between two input pairs in the common space is computed using the 'Euclidean distance' function. The verification decision for all the users is performed using (4).

$$Decision(X|S_m) = \begin{cases} Genuine, & \text{if } S_m < t_h \\ Imposter, & \text{otherwise} \end{cases} \tag{4}$$

where X and t_h represent the query user, and threshold value, respectively. If the similarity measure (S_m) between two pairs (EEG and signature) is less than t_h, the user is treated as genuine else rejected. The verification performance is recorded in terms of FAR and FRR using Eq. (5) and (6), respectively. Here, the terms FP, TP and FN are false-positive, true-positive and false-negative, respectively.

$$FAR = FP/(FP + TN) \tag{5}$$

$$FRR = FN/(TP + FN) \tag{6}$$

2.5. Dataset

In this work, we adopted the dataset provided by Saini *et al.* [32]. The dataset consists of 70 subjects with 10 signatures each. An additional ten impersonated signatures were also collected per person (with corresponding EEG signal) to facilitate the evaluation of forgery detection. Consequently, 700 genuine signatures (70×10=700) and 700 forged signatures (70×10=700) were collected, and a total of 700+700=1400 signatures were used in this work. The signatures were collected by asking subjects to sign a sheet of paper 10 times

using a ballpoint pen, while EEG was recorded simultaneously. Subject age and gender information were also recorded for context during data collection. The EEG signals were captured using a 14-channel (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) portable EEG headset, with a sampling frequency of 128 Hz. A description of the dataset is described in Table 1, with subjects and gender ratios subdivided based on age.

Table 1: Description of Subjects and Gender Ratios Based on Age

Groups	**Age Distribution**	**#Subjects**	**Male/female**
G1	15-25	50	27/23
G2	26-44	15	10/5
G3	45-55	5	3/2

3. Results

3.1. Analysis of person verification

In this section, we perform the experimental analysis of the proposed model. This section is divided into the following subsections: person verification using unimodal & multimodal systems (Subsection 3.1.1), user verification by varying multimodal data (Subsection 3.1.2), effects of different brain lobes in user verification (Subsection 3.1.3), and comparison with machine learning classifiers in user verification (Subsection 3.1.4).

Twelve pairs each of genuine-genuine and genuine-forged signatures were created randomly for every user. These pairs were then divided into training (6 pair/user), validation (2 pair/user) and testing (4 pair/user). For training, the models were run for 150 epochs with a batch size of 100. An Adam optimizer was used with an initial learning rate of 1e-03. Running on a 2GHz CPU with 64GB RAM, each epoch took approximately 350 seconds or 5.8 minutes, resulting in a total training time of 870 minutes (150× 5.8 minutes). The details of the training parameters are listed in Table 2.

Table 2: List of Training Parameters

Parameter	**Value**
Batch Size	100
Activation	Relu
Maximum Iteration	150
Optimizer	Adam
Initial Learning Rate	1e-03
Regularization Parameter	L2 (0.01)
Epoch Time	5.8 minutes
Training Time	870 minutes

3.1.1. Person Verification

Here, we present the results of person verification using both unimodal approaches and the proposed multimodal system. For unimodal systems, two Siamese Neural Networks (SNNs) were trained independently for EEG and signature data. The distance between each pair of samples in the unimodal system (EEG or signature) and the multimodal system (a combination of EEG and signature) was calculated using the Euclidean distance. The distribution of distances for matching and non-matching pairs are shown in Figure 6 for EEG alone (a), signature alone (b), and for the EEG+Signature proposed system (c). For matching pairs, the pair-wise distance should be lower, whereas the distance should be higher for an imposter/non-matching pairs. From Figure 6, it can be seen that the multimodal system outperforms both unimodal systems based on the reduced overlap in distributions.

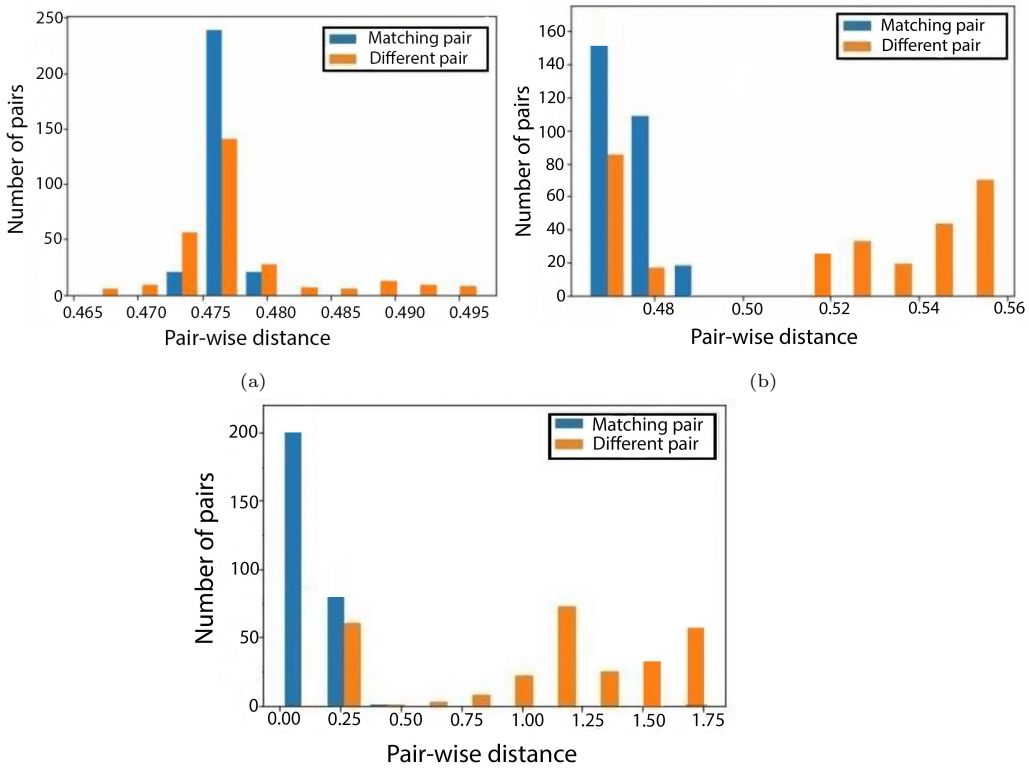

Figure 6: Distribution of matching and non-matching pair-wise distances for the (a) EEG-based unimodal, (b) Signature-based unimodal, (c) (EEG + signature)-based multimodal systems.

The ROC curves shown in Figure 7 was used to evaluate the performance of the unimodal and proposed multimodal systems across different threshold values. In this work, a threshold value (t_h) of 0.44 was chosen for the proposed model, corresponding to the point where the true positive rate (TPR) of the proposed model reached 100%. That is, no legitimate users were rejected at this point. For the EEG and signature-alone models, values of 0.48 and 0.49, respectively were chosen for the same reason. It can be seen that the multimodal system yields a much higher area under the curve (99% vs 76% and 86% for EEG and Signature, respectively).

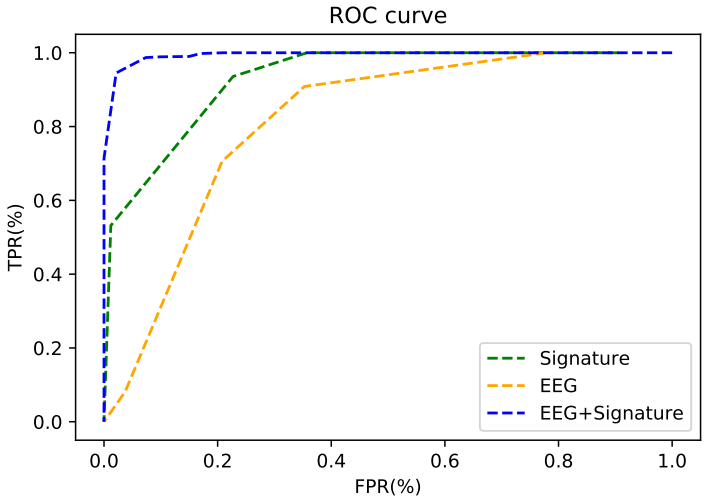

Figure 7: ROC curve of the multimodal and unimodal verification systems.

Using the best thresholds found for each system (based on the ROC curves), the accuracy of each system was computed, as shown in Table 3.

Table 3: Unimodal vs. Multimodal (mSNN) verification system. Accuracy results shown for Training/Validation/Testing.

Modality	**Features**	**Accuracy(%)**	**FAR(%)**	**FRR(%)**
Unimodal	EEG	91.93/87.45/81.78	29.28	7.14
Unimodal	Signature	93.29/86.33/81.42	30.71	6.42
Multimodal	EEG + Signature	99.80/98.60/98.57	2.14	0.71

Figure 8 shows examples of cases when the unimodal systems failed to recognize a forgery, but the proposed system does not. The signature-based Siamese Neural Network in Figure 8 (a) takes the genuine signature of the user as an input to its subnetwork 1 and the forged signature attempt as an input to its subnetwork 2, but is unable to discriminate between the skilled forgery and the original. Similarly, the EEG-based Siamese Neural Network in Figure 8 (b)

takes the EEG from two different users as inputs to its subnetworks. Driven by the stochastic nature of EEG, in some cases, it returns the wrong results due to similar peaks, amplitude, or variations in the input signals. The proposed multimodal system, however, makes its decision based on the combined pair of signature and EEG and therefore tends to produce a more robust biometric.

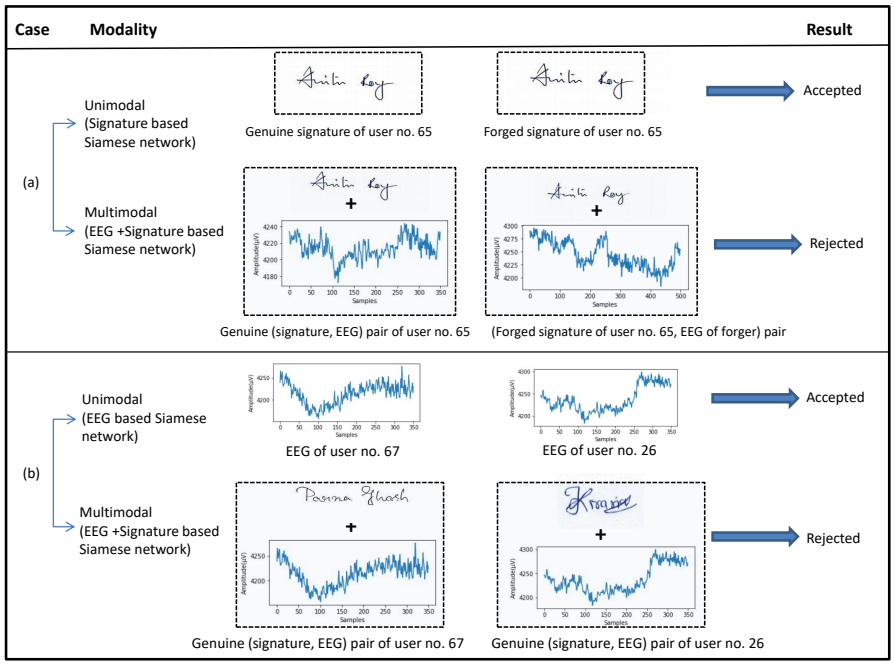

Figure 8: Comparative Example: Unimodal system (signature, EEG) vs. multimodal (EEG + signature) system for user verification. The EEG data corresponding to the P8 electrode of each user is shown for visualization purposes. In both examples (a) and (b), the proposed multimodal system rejects the imposter attempt and correctly identifies the genuine user.

3.1.2. Impact of Training Samples

To evaluate the impact of the amount of training exemplars, the number of training samples was varied from 4 to 7 signature pairs for each user. One pair of genuine and forged signatures was used for each of validation and testing, to maintain consistency. The results of this analysis are presented in Table 4, and show that the performance plateaus after 6 training pairs were included.

Table 4: Impact of the number of training exemplars on model performance

# Training Pairs (Genuine, Forged)	Accuracy(%) (Train/Val/Test)	FAR(%)	FRR(%)
(4,4)	92.83/89.25/90.02	14.29	5.71
(5,5)	96.65/93.46/94.23	3.57	2.14
(6,6)	99.30/98.60/98.57	1.42	1.42
(7,7)	99.75/98.89/98.53	1.42	1.42

3.1.3. Effect of EEG Sensor Location

The classification performance of the proposed model was also evaluated for different electrode locations over the brain. EEG data were collected from all 14 channels, but the activity from different lobes of the brain (frontal, temporal, parietal, and occipital) has been shown to be related to different kinds of tasks. Table 5, therefore, presents the results when using channels only from specific lobes, and all lobes combined. It can be noted that, although the occipital region performs well, the best result is achieved when using all sites.

Table 5: Effect of EEG lobes in user verification

Brain Lobes	Channels	Accuracy(%) (Train/Val/Test)	FAR(%)	FRR(%)
Frontal	F7, F8, F3, F4	93.29/91.45/90.32	12.14	7.14
Temporal	T7, T8	97.02/94.89/93.25	8.90	4.60
Occipital	O1, O2	98.25/97.10/96.67	4.28	2.50
Parietal	P7, P8	89.56/84.10/82.77	21.42	13.21
All lobes	All channels	99.30/98.08/98.57	2.14	0.71

3.1.4. Comparison with Classification

The performance of the SNN and mSNN approaches was compared with two conventional machine learning classifiers, SVM and Random Forest (RF), for both the unimodal and multimodal cases. For this, the commonly used

histogram of the oriented gradients (HOG) features were extracted from the signature images and three statistical features (the mean, standard deviation and root mean square) were computed from the EEG filtered data [32]. For the unimodal cases, classification was performed using the SVM or RF classifiers and either the signature or EEG features, separately. A CNN-based architecture was used for the signature-alone SNN, and an LSTM-based SNN was used for the EEG-alone data. For the multimodal system, the HOG and EEG features were combined for use with the SVM and RF classifiers for comparison with the mSNN. The results of this comparison are shown in Figure 9 for signature alone (a), EEG alone (b), and the multimodal case (c). It can be noted that the proposed mSNN-based user verification system vastly outperforms these conventional classification approaches in both the unimodal and multimodal cases.

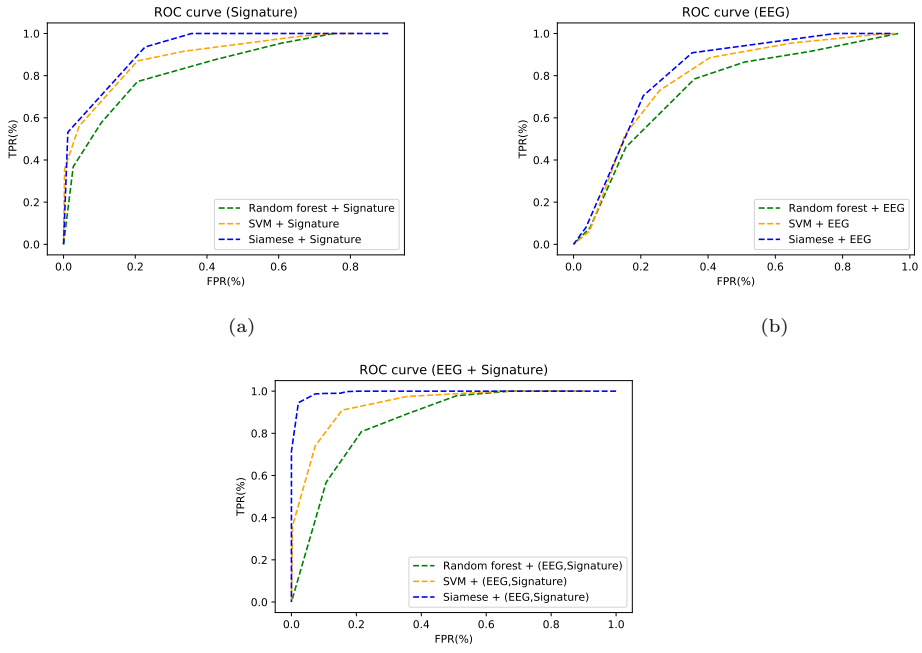

Figure 9: ROC curves comparing the unimodal SNN and multimodal mSNN with conventional SVM and Random Forest classifiers. (a) Signature-based unimodal system, (b) EEG-based unimodal system, (c) EEG & signature-based multimodal system.

3.2. Analysis of Failure Cases

Out of the 560 testing pairs (280 genuine, 280 forged), the proposed mSNN model incorrectly classified 8 samples. This yielded a 2.14% FAR (6 samples out of 280) and 0.71% FRR (2 samples out of 280). Figure 10 shows a comparison of the training and testing samples for these 8 misclassified cases.

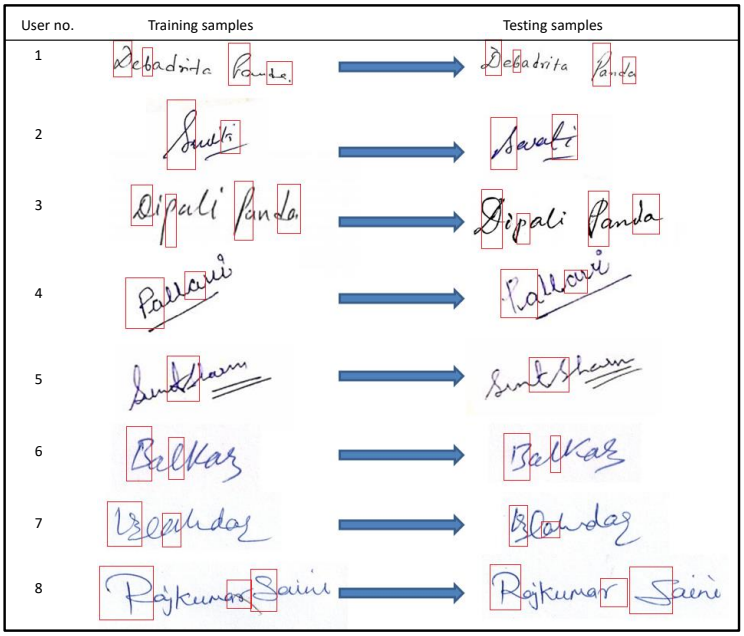

Figure 10: Training and testing samples of the misclassified (wrongly classified) users by our model. The red blocks denote the difference of characters between training and testing samples. User 1 to 5 are female subjects, and user 6 to 8 are male subjects.

The EEG signals from some of the misclassified users also exhibited the presence of head and facial muscle movements during signing. This may have led to the inclusion of motion artifacts or electromyogram (EMG) corruption in the acquired EEG signal [42], as shown in Figure 11.

4. Conclusion

In this study, we proposed a multimodal Siamese Neural Network to combine two different biometric traits, EEG and offline signatures, for user verification.

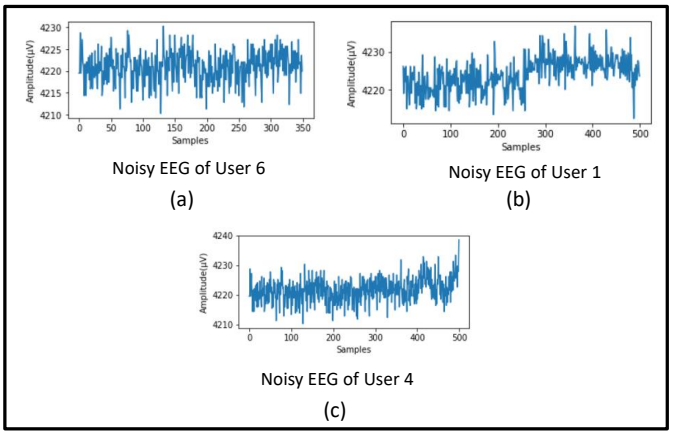

Figure 11: EEG samples from three users likely corrupted by EMG and/or motion artifact

The mSNN was designed to learn spatial and temporal features from signature images and EEG using both CNNs and LSTMs encoders. Multimodal feature fusion was performed in the embedding space and a similarity between two pairs a computed for verification. The comparison between the performance of EEG or signature-only unimodal systems and the proposed multimodal system was analyzed using ROC curves, showing that mSNN outperformed EEG based authentication by reducing FAR from 29.28% to only 2.14%. From this, it can be seen that the two traits complement each other when combined, helping to greatly reduce the success of forgery attempts. Moreover, a skilled forgery can be learned to copy offline signatures, however, the features of the corresponding EEG signal are unique for each user, making it especially difficult to forge both at the same time.

From the experiments conducted, the proposed deep spatio-temporal mSNN achieved a 98.57% classification accuracy with 2.14% FAR. This represents an absolute improvement of 12.86% in FAR compared to previous results using a Hidden Markov Models on the same dataset [32]. Some caution should be taken in interpreting this comparison, however, as the training/testing protocols may differ between the works due to the training strategy of the Siamese networks (compared to conventional classification). Under a direct and controlled com-

parison within the current work, however, while unimodal SNNs only moderately outperformed conventional SVM and RF classifiers, the multimodal mSNN vastly outperformed their multimodal equivalents.

The training approach of the mSNN offers another key advantage. It was found that only 6 pairs of multimodal data were sufficient to secure a user's identity. This is attractive as it reduces the registration burden users in potential future implementations. Furthermore, the entire model doesn't have to be retrained when adding a new user, as with conventional classifiers. This could substantially reduce the computational burden as the database of known users is increased. Future analyses should investigate, in more detail, the impact of the selection and quality of specific genuine and forgery attempts on the overall system performance. Together, these results provide a compelling argument for the robustness and applicability of the proposed mSNN. While initial internal piloting suggests that the SNN architectures translate well to other individual modalities, additional multimodal datasets are required to explore how well the proposed mSNN models generalize to new modalities.

Acknowledgment

Prof Chang's research is partly supported by VC Research (number: VCR 0000050). Dr. Scheme's research is partly supported by the New Brunswick Innovation Foundation.

References

[1] X. Zhu, X.-Y. Jing, X. You, X. Zhang, T. Zhang, Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics, IEEE Transactions on Image Processing 27 (11) (2018) 5683–5695. 2

[2] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, Q. Tian, Pose-driven deep convolutional model for person re-identification, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 3960–3969. 2

[3] H. Li, J. Brandt, Z. Lin, X. Shen, G. Hua, A multi-level contextual model for person recognition in photo albums, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1297–1305. 2

[4] A. Jain, L. Hong, S. Pankanti, Biometric identification, Communications of the ACM 43 (2) (2000) 90–98. 2

[5] B. Fang, C. H. Leung, Y. Y. Tang, K. Tse, P. C. Kwok, Y. Wong, Off-line signature verification by the tracking of feature and stroke positions, Pattern recognition 36 (1) (2003) 91–101. 2

[6] A. Alaei, S. Pal, U. Pal, M. Blumenstein, An efficient signature verification method based on an interval symbolic representation and a fuzzy similarity measure, IEEE Transactions on Information Forensics and Security 12 (10) (2017) 2360–2372. 2

[7] M. A. Ferrer, J. B. Alonso, C. M. Travieso, Offline geometric parameters for automatic signature verification using fixed-point arithmetic, IEEE transactions on pattern analysis and machine intelligence 27 (6) (2005) 993–997. 2

[8] S. Dey, A. Dutta, J. I. Toledo, S. K. Ghosh, J. Lladós, U. Pal, Signet: Convolutional siamese network for writer independent offline signature verification, arXiv preprint arXiv:1707.02131. 3

[9] A. Hamadene, Y. Chibani, One-class writer-independent offline signature verification using feature dissimilarity thresholding, IEEE Transactions on Information Forensics and Security 11 (6) (2016) 1226–1238. 3

[10] Y. Guerbai, Y. Chibani, B. Hadjadji, The effective use of the one-class svm classifier for handwritten signature verification based on writer-independent parameters, Pattern Recognition 48 (1) (2015) 103–113. 3

[11] L. G. Hafemann, R. Sabourin, L. S. Oliveira, Offline handwritten signature verification—literature review, in: Seventh International Conference

on Image Processing Theory, Tools and Applications, IEEE, 2017, pp. 1–8. 3

[12] P. Kumari, A. Vaish, Brainwave based user identification system: A pilot study in robotics environment, Robotics and Autonomous Systems 65 (2015) 15–23. 3

[13] R. Palaniappan, K. Ravi, A new method to identify individuals using signals from the brain, in: Fourth International Conference on Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia., Vol. 3, IEEE, 2003, pp. 1442–1445. 3

[14] D. Q. Phung, D. Tran, W. Ma, P. Nguyen, T. Pham, Using shannon entropy as EEG signal feature for fast person identification., in: ESANN, Vol. 4, 2014, pp. 413–418. 3

[15] I. Jayarathne, M. Cohen, S. Amarakeerthi, Brainid: Development of an EEG-based biometric authentication system, in: 7th Annual Information Technology, Electronics and Mobile Communication Conference, IEEE, 2016, pp. 1–6. 3

[16] M. Phothisonothai, An investigation of using ssvep for EEG-based user authentication system, in: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, IEEE, 2015, pp. 923–926. 3

[17] T. Pham, W. Ma, D. Tran, P. Nguyen, D. Phung, Multi-factor EEG-based user authentication, in: International Joint Conference on Neural Networks, IEEE, 2014, pp. 4029–4034. 3

[18] Z. Mu, J. Hu, J. Min, EEG-based person authentication using a fuzzy entropy-related approach with two electrodes, Entropy 18 (12) (2016) 432. 3

[19] P. Kumar, R. Saini, P. P. Roy, D. P. Dogra, A bio-signal based framework to secure mobile devices, Journal of Network and Computer Applications 89 (2017) 62–71. 3

[20] T. Koike-Akino, R. Mahajan, T. K. Marks, Y. Wang, S. Watanabe, O. Tuzel, P. Orlik, High-accuracy user identification using eeg biometrics, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2016, pp. 854–858. 3

[21] K. P. Thomas, A. Vinod, Toward EEG-based biometric systems: The great potential of brain-wave-based biometrics, IEEE Systems, Man, and Cybernetics Magazine 3 (4) (2017) 6–15. 3

[22] M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci, G. L. Marcialis, An EEG-based biometric system using eigenvector centrality in resting state brain networks, IEEE Signal Processing Letters 22 (6) (2014) 666–670. 3

[23] B. Kaur, D. Singh, P. P. Roy, A novel framework of EEG-based user identification by analyzing music-listening behavior, Multimedia tools and applications 76 (24) (2017) 25581–25602. 3, 4

[24] K. A. Sidek, V. Mai, I. Khalil, Data mining in mobile ECG based biometric identification, Journal of Network and Computer Applications 44 (2014) 83–91. 3

[25] I. Bhardwaj, N. D. Londhe, S. K. Kopparapu, A spoof resistant multibiometric system based on the physiological and behavioral characteristics of fingerprint, Pattern Recognition 62 (2017) 214–224. 4

[26] Z. Luo, Q. Gu, G. Qi, S. Liu, Y. Zhu, Z. Bai, A robust single-sensor face and iris biometric identification system based on multimodal feature extraction network, in: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), IEEE, 2019, pp. 1237–1244. 4

[27] A. Lumini, L. Nanni, Overview of the combination of biometric matchers, Information Fusion 33 (2017) 71–85. 4

[28] C. Galdi, M. Nappi, J.-L. Dugelay, Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity, Pattern Recognition Letters 82 (2016) 144–153. 4

[29] K. Chang, K. W. Bowyer, S. Sarkar, B. Victor, Comparison and combination of ear and face images in appearance-based biometrics, IEEE Transactions on pattern analysis and machine intelligence 25 (9) (2003) 1160–1165. 4

[30] N. Poh, J. Korczak, Hybrid biometric person authentication using face and voice features, in: International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, 2001, pp. 348–353. 4

[31] L. Hong, A. Jain, Integrating faces and fingerprints for personal identification, IEEE transactions on pattern analysis and machine intelligence 20 (12) (1998) 1295–1307. 4

[32] R. Saini, B. Kaur, P. Singh, P. Kumar, P. P. Roy, B. Raman, D. Singh, Don't just sign use brain too: A novel multimodal approach for user identification and verification, Information Sciences 430 (2018) 163–178. 4, 14, 21, 23

[33] R. Plamondon, A kinematic theory of rapid human movements: Part iii. kinetic outcomes, Biological Cybernetics 78 (2). 4

[34] A. Fischer, R. Plamondon, Signature verification based on the kinematic theory of rapid human movements, IEEE Transactions on Human-Machine Systems 47 (2) (2016) 169–180. 4

[35] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a" siamese" time delay neural network, in: Advances in neural information processing systems, 1994, pp. 737–744. 6

[36] G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, in: ICML deep learning workshop, Vol. 2, Lille, 2015. 6

[37] M. Hu, J. Tan, F. H. Xue, A new approach to the image resizing using interpolating rational-linear splines by continued fractions?, 2005. 7

[38] M. Del Pozo-Banos, J. B. Alonso, J. R. Ticay-Rivas, C. M. Travieso, Electroencephalogram subject identification: A review, Expert Systems with Applications 41 (15) (2014) 6537–6554. 7

[39] I. Nakanishi, S. Baba, C. Miyamoto, Eeg based biometric authentication using new spectral features, in: 2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), IEEE, 2009, pp. 651–654. 7

[40] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, M. Shah, Decoding brain representations by multimodal learning of neural activity and visual features, arXiv preprint arXiv:1810.10974. 7

[41] S. Chopra, R. Hadsell, Y. LeCun, et al., Learning a similarity metric discriminatively, with application to face verification, in: Computer Vision and Pattern Recognition (1), 2005, pp. 539–546. 9

[42] X. Jiang, G.-B. Bian, Z. Tian, Removal of artifacts from eeg signals: a review, Sensors 19 (5) (2019) 987. 22