Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 4s (2025)

Towards Improved Biometric Security: EEG-Based Person Identification Enhanced by Deep Learning and Facial Recognition

Shalu Verma[1], Sanjeev indora[2], Rohtash Dhiman[3]

[1]Research Scholar, Department of computer science and engineering, Deenbandhu, Chhotu Ram University of Science and Technology, Murthal, Sonepat, Haryana, (India).

[2]Associate Professor, Department of computer science and engineering, Deenbandhu, Chhotu Ram University of Science and Technology, Murthal, Sonepat, Haryana, (India).

[3]Assistant Professor, Department of computer science and engineering, Deenbandhu, Chhotu Ram University of Science and Technology, Murthal, Sonepat, Haryana, (India).

Corresponding Email: 19001901010shalu@dcrustm.org

Article History:

Received: *20-09-2024*

Revised: *25-11-2024*

Accepted: *02-12-2024*

Abstract:

Due to their inherent qualities of being secretive, vivid, and unpredictable, electroencephalogram (EEG) signals are considered a valuable tool for security-related identification. However, research on using EEG signals for person identification is still in its early stages. The challenges lie in decoding these signals accurately and implementing effective EEG-based identification methods. In recent years, EEG has been at the forefront of scientific research on User Authentication (UA), leading to innovative experiments that aim to identify individuals based on their unique brain activity in specific usage scenarios. The utilization of EEG signals, which are derived from brain activity, holds great potential for addressing contemporary security concerns in conventional knowledge-based user authentication, including the vulnerability to shoulder surfing. This research investigates a new method for person identification that combines electroencephalogram (EEG) signals with facial video. A hybrid model is proposed, incorporating features from both MobileNet and a Convolutional Neural Network with Long short-term memory (LSTM-CNN) architecture giving a person identification accuracy of 99.81%. The model is trained and tested on the 'DEAP' dataset to identify individuals by leveraging unique EEG patterns and facial features, thereby improving biometric identification through the integration of these insights.

Keywords: Biometric identification, Convolutional neural network(CNN), Long short-term memory(LSTM), MobileNet architecture.

1. Introduction

In recent times, rapid advancements in biometrics, forensics, and informatics have created a conducive environment for safeguarding personal privacy and public security[1]. Among these developments, biometric technology has emerged as a vital tool for verifying individuals' identities. Biometrics involves the recognition of people based on their physiological or behavioral traits, including features like face[2], fingerprints[3], iris patterns[4], gait[5], signature[6], and voice[7]. Biometric technologies have seamlessly integrated into our daily lives, finding applications in various fields. For instance, smart phones incorporate facial[8] and fingerprint recognition for security, while airport security[9] relies on biometrics for passport control, among other uses. The widespread adoption of biometric systems depends on the assurance of reliable recognition processes and the secure management of the generated data. Therefore, ensuring privacy and security has become pivotal for the success of biometrics, alongside its high accuracy. In an era where personal information is increasingly digitized[10] and shared, maintaining the trust of individuals in biometric technology is essential. Biometrics has the potential to improve security, but it also raises important concerns about protecting sensitive data and preventing unauthorized access[11]. The evolving relationship between biometrics, privacy, and security highlights the crucial need for ethical and secure practices. Balancing individual privacy with public safety remains a key challenge in our digital world. Biometric technology plays a significant role in this landscape, serving as a tool to protect both personal freedoms and collective well-being[12].Traditional approaches to EEG-based person identification typically involve two key steps: feature representation and classifier training. Deep learning has revolutionized the field, with various neural networks taking center stage in research. Some researchers opt for an end-to-end approach, combining feature representation and classification within a single framework. Others leverage deep learning models specifically for classification, applying them to independently developed feature spaces[13]. Biometric identification is a critical aspect of security systems, and recent advancements have seen a surge in research exploring novel modalities. This paper introduces a Hybrid model that integrates EEG and facial data for person identification. The motivation behind this approach lies in the potential synergy between temporal EEG patterns and spatial facial features.

In the context of classical machine learning techniques, merely adding more features will unavoidably expand the dimension of the machine learning model. This becomes particularly pronounced when substantial redundancy exists among different features, causing the model's complexity to far exceed the actual feature dimension. As a consequence, over fitting may arise.

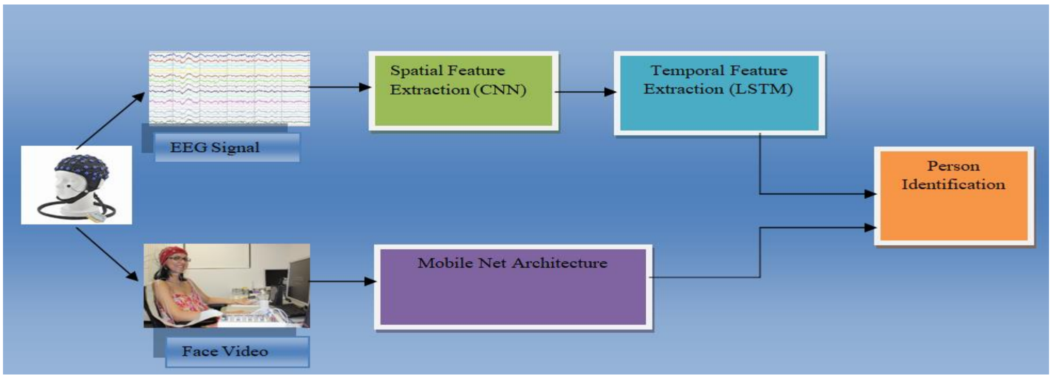

Fig.1. The deep learning-based network structure for integrated facial recognition and EEG

The study's significant contributions are summarised here.

1) This paper proposes an EEG-based person identification (PI) strategy based on deep learning with CNN-LSTM and MobileNet algorithms.

2) The proposed approach demonstrates improvement in accuracy for user identification through the fusion of biometric traits with the utilization of deep learning techniques.

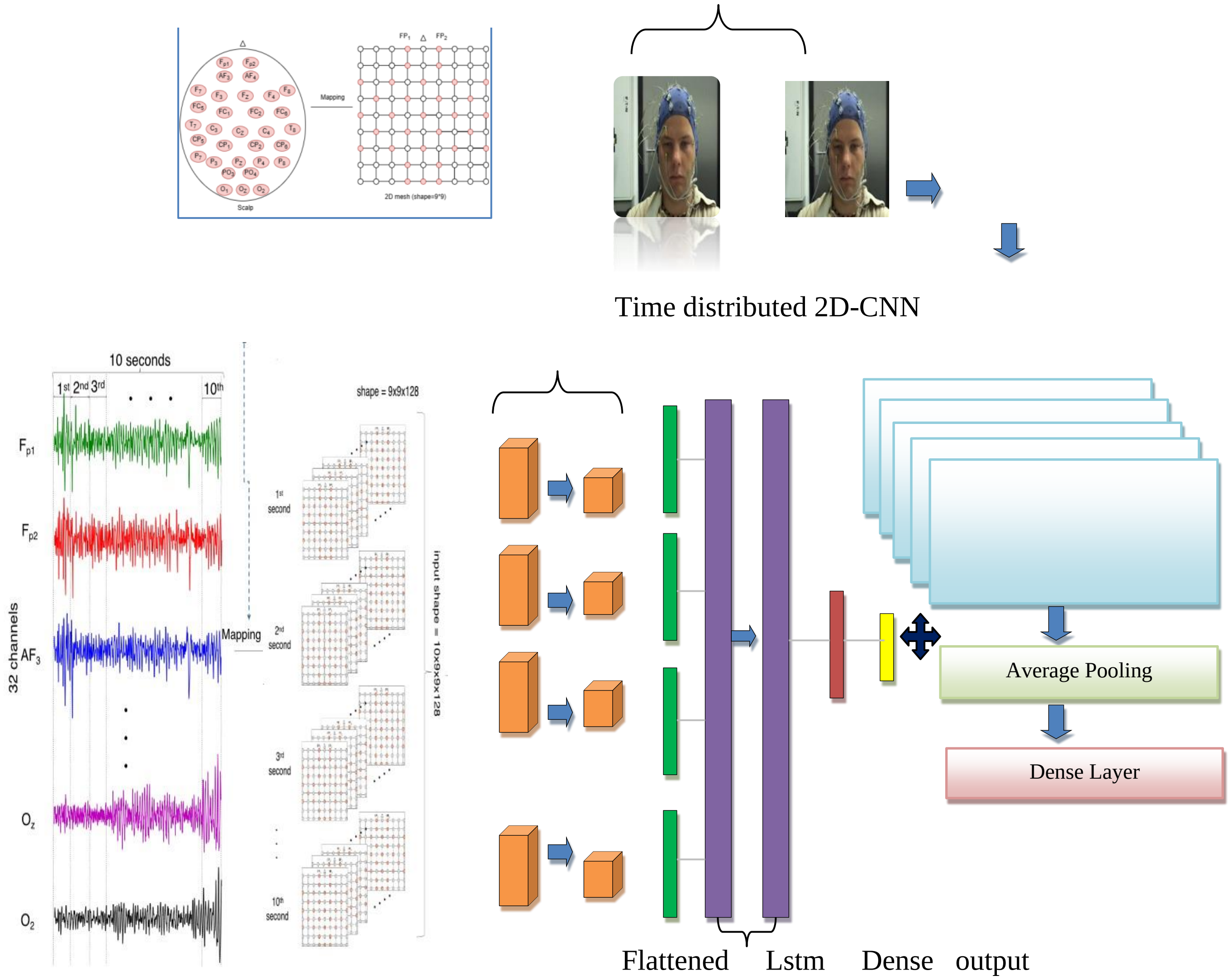

Fig. 2. Implementation of the cascade CNN-LSTM-mobileNet model[14] according to EEG& Face videos

2. Literature survey

Recently, there has been increasing interest in EEG-based Person Identification (PI)[15]. EEG-based PI methods use two types of algorithms: standard machine learning and deep learning. Machine learning techniques, including support vector machines (SVM), random forests, and k-nearest neighbours (KNN), are commonly used in EEG-based PI. Deep learning (DL) is increasingly being used for EEG-based PI, following its great success in several sectors. DL models are being increasingly incorporated into EEG-based PI research, often involving fine-tuning parameters or redesigning architectures, including CNNs) and LSTM networks. In 2019, Banee Bandana Das and colleagues[16] developed a technique to identify persons using EEG waves. The system used convolutional neural networks (CNNs) to extract spatial features from raw data, resulting in robust and informative results. The features were fed into an LSTM to manage temporal data and identify it. The method achieved outstanding accuracy rates of 99.95% and 98% for individual identification. Yingnan Sun and colleagues[17] created a one-dimensional convolutional long short-term memory neural network-

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 4s (2025)

based EEG signal-based user recognition system. During the training process, EEG data from electrodes ranging from 4, 16, 32, to 64 were input into the network. The results revealed that an optimal classification accuracy of 99.58% was attained with 16 electrodes. Emanuele Maiorana [18] presented a deep learning method that uses Siamese convolutional neural networks to analyze EEG data from 45 subjects in a multisession database. The results suggest a promising potential for applying EEG-based biometric verification across different tasks. In 2022, Alyasseri Z [19] introduced a methodology that integrates the Gray Wolf Optimizer (BGWO) with a Support Vector Machine classifier employing a Radial Basis Function kernel (SVM-RBF) for biometric human identification utilizing EEG data. This methodology attained an accuracy rate of 94.13% while utilizing just 23 sensors and 5 autoregressive coefficients.

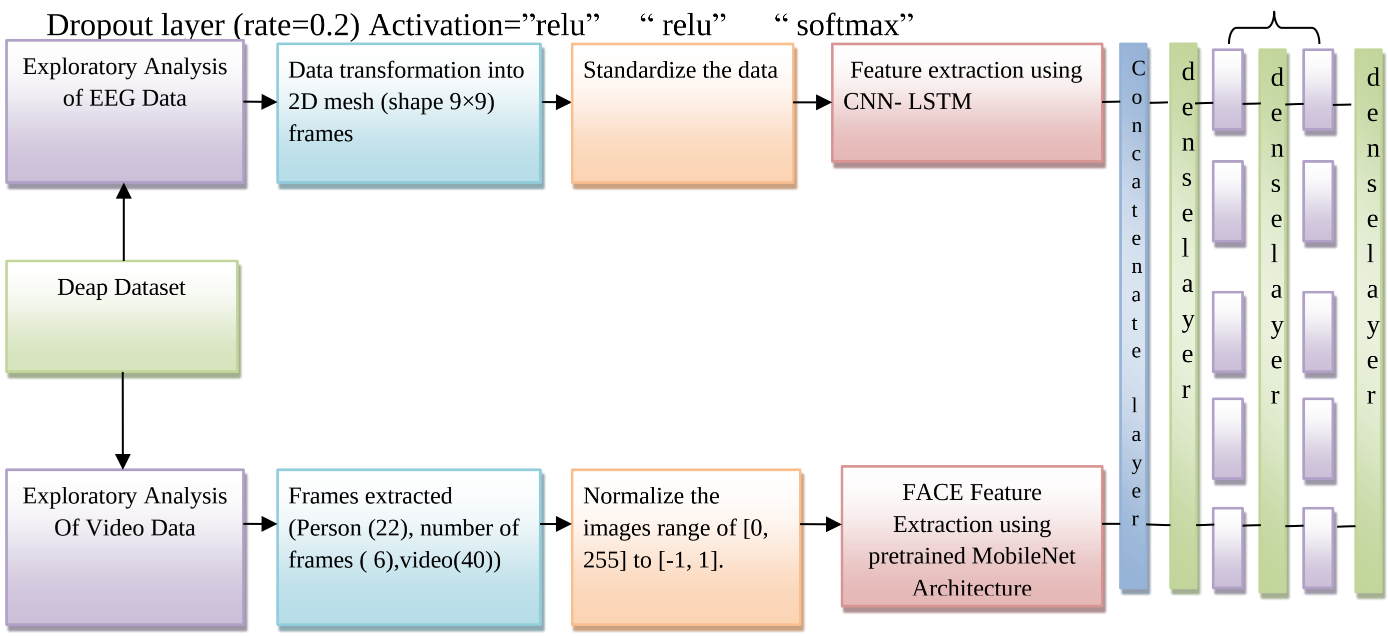

Fig. 3.Flowchart for hybrid proposed technique cnn+lstm+mobilenet

3. Deep Learning approach to EEG

In recent years, the application of deep learning (DL) techniques for classifying EEG signals—bio-potentials recorded from the scalp over time—has seen a substantial increase. Researchers frequently employ DL architectures to capture both the spatial and temporal features of these signals [16]. A typical approach involves using a combination of CNNs followed by Recurrent Neural Networks RNNs, such as Long Short-Term Memory (LSTM) networks. This layered architecture leverages the hierarchical structure of neural networks, where earlier layers extract features that are processed by later layers. CNNs are often used as the initial layers in deep learning models to capture significant patterns or features[20]. A key feature of CNNs is their use of convolution operations with small filter patches (kernels). These filters learn local patterns on their own, and when multiple CNN layers are stacked, they combine these patterns to create more complex features. Within this stack, pooling layers are often added to reduce the dimensionality by retaining only the maximum value from each small region, allowing subsequent convolutional layers to operate on a different scale. The features extracted by CNNs can then be used as input for other network architectures, aiding tasks like object detection or semantic segmentation. In the context of EEG signals, LSTMs, which take input from the local features learned by CNNs, excel at capturing temporal information.[14]Wilaiprasitporn et al focus on improving EEG-based identification using deep learning, especially while people are in different emotional states (affective EEG) and uses a mix of CNNs , RNNs to analyze both the spatial and temporal aspects of EEG signals. Utilizing biometric methods like EEG is vital for robust person recognition, addressing vulnerabilities in traditional modalities. Existing EEG approaches face limitations in scalability and generalization. Alsumari et al [21] introduce a lightweight CNN model, achieving high identification accuracy and authentication performance with minimal EEG data, promising real-world application in biometric security systems. Zhang et al.[22] attempted to use a 3-D CNN to directly capture spatiotemporal information in a single layer. However, their results were slightly inferior to those achieved with the combined CNN-LSTM model. This difference may be due to the fact that LSTMs are generally more effective at handling temporal information. LSTMs can selectively retain or discard information based on context, which is crucial for processing sequential data. As a specialized type of recurrent neural network (RNN), LSTMs are designed to manage long-range dependencies in time-series data. In this context, LSTMs use Rectified Linear Unit (ReLU) activation functions, which provide several benefits over traditional functions like tanh. The LSTM unit's output is calculated using the forget gate, input gate, and ReLU activation functions.

1. Cell State $c_t$and Gates:

The LSTM unit maintains a cell state c_t that acts as its memory.

Gates, including the forget gate f_t, input gate i_t, and output gate o_t, manage the flow of information within the LSTM:

Forget Gate (f_t): Decides which information from the previous cell statec_{t-1} should be removed.

Input Gate (i_t): Determines which new information should be added to the current cell state c_t.

Output Gate (o_t): RegRegulates which parts of the cell state contribute to the output h_t.

2. Equations for LSTM Operations with ReLU:

The cell state c_t is updated using the forget gate f_t, input gate i_t , and a candidate update $\hat{c}_t$ (computed using ReLU activation):

$$f_t = \sigma\left(w_{f.}\left[h_{t-1,}x_t\right] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(w_{i.}\left[h_{t-1,}x_t\right] + b_i\right) \tag{2}$$

$$\hat{c}_t = ReLU(w_{c.}\left[h_{t-1,}x_t\right] + b_c) \tag{3}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \hat{c}_t \tag{4}$$

where$w_{f,}w_{i,}w_{c,}$are weight matrix,$h_{t-1,}$previous hidden state and x_t with various$b_f b_i b_c$ bias terms.$\hat{c}_t$ represents candidate cell update.

3. Hidden State h_tComputation:

The output gate $o_t$modulates the cell state$c_t$before passing it through ReLU activation to compute the hidden state h_t:

$$o_t = \sigma\left(w_{o.}\left[h_{t-1,}x_t\right] + b_o\right) \quad (5)$$

$$h_t = o_t \circ ReLU(c_t) \quad (6)$$

h_t denotes the current hidden state at time step t and o_t is Output of gate output. ReLU is a Rectified Linear Unit used as an activation function. c_t represents the Current cell state,$\circ$ shows Element-wise multiplication operation.

Advantages of ReLU in LSTMs

Non-Saturating Activation: ReLU does not suffer from the vanishing gradient problem encountered with tanh or sigmoid activations, which accelerates convergence during training.

Sparse Activation: ReLU activations are sparser compared to sigmoid or tanh, promoting faster computation and improved network efficiency.

Stability and Expressiveness: The linear nature of ReLU allows LSTMs to model complex non-linear relationships more effectively, enhancing the network

3. Materials and Methods

In this segment, we initially presented the DEAP affective EEG dataset [23], which served as the foundation for our experimental investigations, alongside outlining the preprocessing procedures integral to our solution. Given that DEAP was primarily crafted for mental state classification, we detailed our approach to partitioning the data to suit the demands of the PI task. Subsequently, we elaborated on the conceptualization and execution of our proposed deep learning (DL) methodology**.**

A. Data Set:

In this study, we carried out experiments with the DEAP affective EEG dataset, which is a widely established benchmark for emotion and recognition tasks. Thirty-two individuals of sound health were recruited to partake in the experiment. Their task is to watch emotionally charged music videos and provided subjective ratings of valence and arousal for forty video clips while their EEG data was collected. A detailed summary of data is given in further steps.

- The sampling data sample frequency was decreased to 128 Hz.
- EOG (electrooculogram) artefacts were eliminated.
- Applying a bandpass filter with a range of 4.0 Hz to 45.0 Hz.
- By averaging the data to a common reference point, the data was normalised.
- After discarding the baseline 3-second trial, the data was segmented into 60-second trials. Most of the researchers have been using this dataset for human emotion classification. However, we used this dataset to study EEG-based person identification.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 4s (2025)

B. Data Organization:

- There are 32 persons (or subjects) in the dataset.
- Each person performed 40 trials, possibly in an experimental task or study.
- The EEG data was collected from 40 channels, which could represent different electrode placements on the scalp.

C.Data Segmentation:

- The continuous EEG data for each trial is segmented into fixed-size chunks.
- Each trial is segmented into a 60-second period and a 3-second pre-trial baseline.
- The baseline segment (3 seconds) serves as the reference or initial period before any stimulus or event occurs.

D. Data Sampling Rate:

The EEG data is sampled at a rate of 128 Hz, resulting in 128 data points (samples) per second for each channel.

E. Calculation of Samples per Trial:

- For each trial, the duration is 63 seconds (60 seconds trial + 3 seconds baseline).
- Since the data is recorded at 128Hz, there are 128 data points (samples) for each second of recording.
- Therefore, for each trial, there are 63 seconds x 128Hz = 8064 samples.

F. data:

- A 4-dimensional array of shapes (32, 40, 40, 8064).
- The first dimension represents the 32 persons (subjects) in the dataset.
- The second dimension represents the 40 trials performed by each person.
- The third dimension represents the 40 EEG channels.
- The fourth dimension represents the 8064 samples for each trial.

G. Data Transformation

Based on the provided information, the data transformation involves mapping EEG data to a 9x9 image plane for person identification using short-length EEG segments.

H. EEG Data Segmentation:

- Each EEG trial is originally 60 seconds long (7680 samples after removing the 3-second pre-trial baseline).
- The 60-second trial is further divided into 6 equal parts, each of 10 seconds in duration. These 10-second segments are referred to as "subsamples."

- Each 10-second subsample contains 1280 EEG data points (10 seconds × 128 Hz sampling rate).

I. Mapping to 9x9 Image Plane:

- To convert each 10-second subsample into an image representation, they are mapped to a 9x9 image plane.
- The 9x9 image plane likely represents a grid of 9 rows and 9 columns shown in Fig 4.
- Each data point of the 10-second subsample is mapped to a corresponding position in the 9x9 grid as shown in Figure 4.

J. Data and Label Structure:

- The data consists of EEG subsamples from multiple participants.
- For each participant, there are 30 subsamples (6 subsamples × 5 trials) for each affective state or condition.

The data is structured as follows:

- Data: (number of participants) × 30 subsamples × 1280 EEG data points (10 seconds at a 128 Hz sampling rate).
- Label: (number of participants) × 30 subsamples × 1 (participant's ID).

K. Person Identification:

- The analysis aims to identify individuals based on short EEG segments, each lasting 10 seconds.
- The labels used for this identification are the participant's ID, which is a unique identifier for each participant.

In summary, the data is transformed into a format suitable for person identification using short-length EEG segments. Each 60-second trial is divided into 6 equal 10-second subsamples, which are then mapped to a 9x9 image plane. The image plane representation allows for the application of various image-based machine learning or pattern recognition techniques for person identification based on EEG data. The labels used for identification are the participant's ID, and the goal is to classify individuals based on their EEG responses during the 10-second subsamples.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 4s (2025)

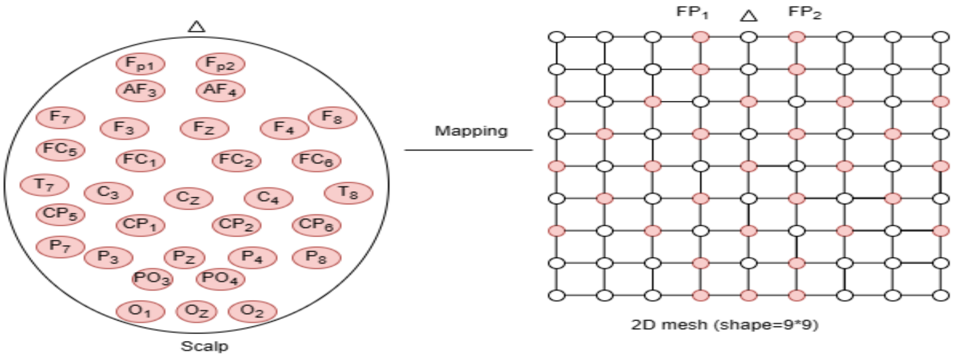

Fig.4. Multichannel EEG signals are transformed into sequences of 2-D images.

4. Proposed Approach

The LSTM-CNN model integrates Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures for sequence modeling. The LSTM part handles temporal features and long-range dependencies, while the CNN part learns local patterns or spatial features from the data. whereas MobileNet that are trained on the deep dataset video frames for facial recognition.

4.1 CNN (Convolutional Neural Network):

The CNN architecture employs a 2D convolutional layer, where the initial layer extracts low-level features using 128 filters with a 3x3 kernel size, followed by Batch Normalization and a ReLU activation function. Batch Normalization standardizes the outputs of the convolutional layer, accelerating the training process and enhancing model stability. The ReLU activation introduces non-linearity, allowing the network to capture more complex patterns. The second convolutional layer further enhances feature extraction with 64 filters, refining the low-level features identified by the first layer. This layer also includes Batch Normalization and a ReLU activation function, ensuring consistency in activation and normalization throughout the network.The third convolution layer extracts higher-level features with 32 filters. By progressively reducing the number of filters, the network hones in on more specific features. Similar to the previous layers, Batch Normalization and ReLU activation are applied. The subsequent layer, known as the Flatten layer, converts the 2D matrix outputs from the convolutional layers into 1D vectors, preparing them for input into the Dense layer. The Dense layer, consisting of 1024 units, then processes these flattened vectors, enabling the network to learn high-level representations of the image data.

4.2 LSTM (Long Short-Term Memory):

LSTM, a variant of recurrent neural network (RNN), is specifically designed to effectively process sequences and time-series data. It excels at capturing long-range dependencies in sequential data and is capable of retaining pertinent information across extended time steps.In the context of sequence modeling, LSTM is frequently employed to capture and process the temporal features of data. In This work output from the CNN is passed to first LSTM layers to capture temporal dependencies. its purpose Captures long-range dependencies in the sequence data, with 512 units. The second Lstm layer processes the sequential data, with 256 units.

4.3 Integration of LSTM and CNN in the LSTM-CNN Model:

The final model combines both the CNN and LSTM components to leverage the strengths of both architectures. The CNN layers effectively capture spatial features from each frame, while the LSTM layers model the temporal dynamics across frames. This hybrid approach is powerful for applications like video classification, where understanding both spatial and temporal patterns is essential. The concatenated features from both models are processed through additional dense layers, resulting in a comprehensive and robust model for sequence-based image classification tasks.By freezing the layers of the feature extractor and focusing on training only the top dense layers, the model leverages pre-trained knowledge while fine-tuning the classification layers for the specific task. This approach balances the need for robust feature extraction with the flexibility to adapt to new data, making it a versatile solution for various sequence classification problems. LSTM-CNN will get trained from scratch since there are no pre-trained weights. Take person ID, form an array with relative indices, and extrapolate those 22 into 22x6x40 samples according to our input data.

4.4 MobileNet

In this work, we are using the networks of MobileNet that are trained on the deep dataset video frames. MobileNet is often referred to as a lightweight convolutional neural network, making it an efficient architecture for mobile applications. One of the key advantages of MobileNet is its minimal computational requirements. Unlike standard convolutions, MobileNet employs depth-wise separable convolutions, which result in a lower number of multiplications compared to traditional convolutions. This reduction in computational complexity contributes to the overall efficiency of MobileNet. Depth-wise separable convolution comprises both depth-wise convolutions and point-wise convolutions. Unlike standard CNNs where convolution is applied to all channels simultaneously, depth-wise convolution involves applying convolution to one channel at a time. A 5D NumPy array is initialized to hold facial images with dimensions set as (number of persons, 240, 224, 224, 3). The pixel values of these images are then rescaled from their original range of [0, 255] to a new range of [-1, 1].The division by 127.5 centres the data around 0, and subtracting 1 scales it to the range [-1, 1]. Scaling pixel values to a range of [-1, 1] is a common preprocessing step in machine learning, especially for neural networks. It can improve convergence during training and help the model better handle the data. This sets up a feature extractor using the MobileNetV3Small model from Tensor Flow's Keras applications. This sets up a pre-trained MobileNetV3Small model for feature extraction. The model is configured to exclude the top layer, use ImageNet pre-trained weights, take images with dimensions (224, 224, 3) as input, and use global average pooling. All layers of the model are frozen, preventing them from being updated during subsequent training. This feature extractor can then be used as part of a larger model for tasks such as facial feature analysis or classification.

5. Hybrid Approach

The proposed model, illustrated in Figure 3 consists of extracted feature result of mobile net architecture and the Lstm-cnn model. The proposed model architecture consists of a bottleneck structure with two hidden layers having 256 and 64 neurons, with alternative dropout layer and an output layer with 22 neurons using softmax activation. The bottleneck structure with reduced dimensions (256 and 64 neurons) is often used for dimensionality reduction and feature abstraction. It

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 4s (2025)

helps in learning a compact representation of the input features. The weights of the LSTM-CNN layers and the bottleneck layers are trained together. This enables the model to learn a task-specific representation while preserving the valuable features learned by MobileNet. The output layer uses softmax activation to transform the model's raw outputs into probabilities, making it ideal for multi-class classification.Training the LSTM-CNN layers along with the bottleneck ensures that the model adapts to the specific task of classifying persons, taking into account both the facial features from MobileNet and the temporal and spatial features from the LSTM-CNN. In deep learning training, the categorical cross-entropy loss function is used to quantify the difference between the predicted class probabilities and the actual labels. This loss guides the optimization process, adjusting the model parameters to minimize the discrepancy and improve classification accuracy. During testing, the categorical cross-entropy loss function is also applied to assess the model's performance on unseen data. By comparing predicted probabilities to the actual labels, it measures the model's ability to generalize and make accurate predictions beyond the training set. In both training and testing, this loss function plays a vital role in evaluating the model's effectiveness and generalization capabilities in classification tasks, as illustrated in Figure 5.

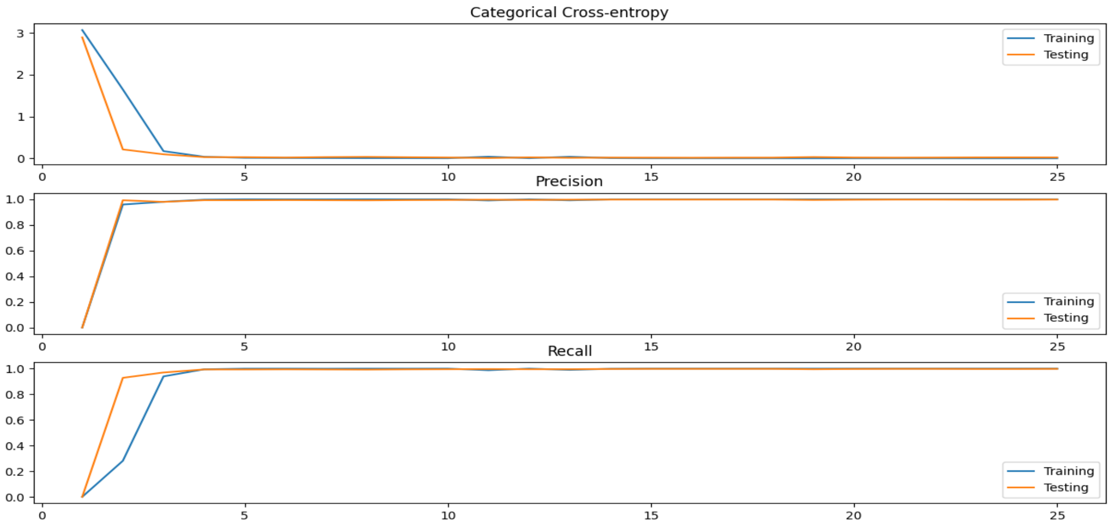

Fig. 5. training loss and testing loss against the number of epochs

6. Comparative performance analysis

In this section, a comparative analysis has been conducted with a previous study [14] that utilized deep learning for subject classification. The prior method employed two deep learning models, CNN-GRU and CNN-LSTM, achieving an impressive correct recognition rate of 99.17% by focusing on brain signals from the frontal region. Even with a reduction in EEG sensors from 32 to 5 for practicality, the approach maintained strong performance. In contrast, the current work achieved a maximum accuracy of 99.81% by integrating EEG brain wave data with a facial multimodal recognition system, further enhancing identification accuracy.

7. Results and discussion:

The electroencephalogram (EEG) and peripheral physiological data of 32 participants were captured while they watched 40 one-minute clips of music videos. This dataset, which is publicly accessible, is encouraged for use by other researchers to test their affective state estimation methods.. For each trial, the duration is 63 seconds (60 seconds trial + 3 seconds baseline). Since the data is recorded at 128Hz, there are 128 data points (samples) for each second of recording. Therefore, for each trial, there are 63 seconds x 128Hz = 8064 samples. To focus on identifying individuals using short EEG segments, we used 10-second subsamples. Each 60-second EEG trial in the DEAP dataset was divided into six subsamples, giving us 30 subsamples (6 subsamples × 5 trials) per participant for each affective state. In these experiments, the labels represent the participants' unique IDs. The data and labels are structured as follows:

Data: number of participants × 30 subsamples × 1280 EEG data points (10 seconds with 128 Hz sampling rate)

Label: number of participants × 30 subsamples × 1(ID). The analysis aims to identify a person from the short-length EEG segments (10-second subsamples).

The labels used for this identification are the participant's ID, which is a unique identifier for each participant. For each 10-second segment (sample) in the 5-dimensional array. For each channel (32 channels), interpolate the data points onto the 9x9 mesh grid. This can be done using various interpolation techniques such as bilinear, bicubic, or spline interpolation. The interpolated values for each channel will be represented on the 9x9 grid. This mesh representation allows you to observe and analyze the spatial patterns of EEG data across the 9x9 grid, which could be helpful for further analysis, visualization, or feature extraction. Performance is evaluated using three statistical metrics: classification accuracy, precision, and recall. A confusion matrix (shown in Figure 6) is used to determine true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Additionally, categorical cross-entropy, precision, and recall are plotted against the number of epochs for both training and testing loss (as seen in Figure 5). Table 1 summarizes the classification results. The classification report provides a detailed summary of precision, recall, and F1 score for each class, while the last column, "support," indicates the number of samples for each class.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} * 100 \tag{7}$$

$$Precision = \frac{TP}{TP + FP} * 100 \tag{8}$$

$$Recall = \frac{TP}{TP + FN} * 100 \tag{9}$$

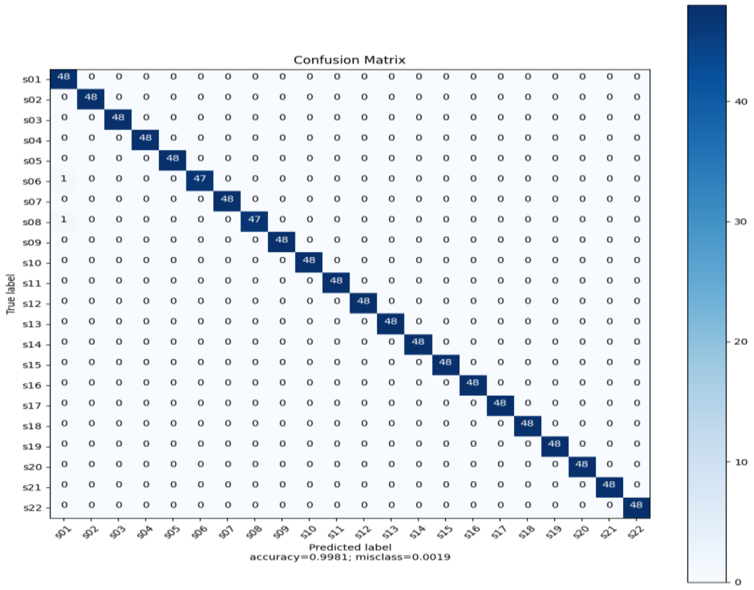

Fig.6. Confusion matrices of the combined model with predicted label accuracy

Table 1: Classification Report for the Feature-Level Fusion using cnn-lstm and mobileNet architecture

subjects	Precision	Recall	F1-score	support
S01	0.96	1.00	0.98	48
S02	1.00	1.00	1.00	48
S03	1.00	1.00	1.00	48
S04	1.00	1.00	1.00	48
S05	1.00	1.00	1.00	48
S06	1.00	0.98	0.99	48
S07	1.00	1.00	1.00	48
S08	1.00	0.98	0.99	48
S09	1.00	1.00	1.00	48
S10	1.00	1.00	1.00	48
S11	1.00	1.00	1.00	48
S12	1.00	1.00	1.00	48
S13	1.00	1.00	1.00	48
S14	1.00	1.00	1.00	48
S15	1.00	1.00	1.00	48

S16	1.00	1.00	1.00	48
S17	1.00	1.00	1.00	48
S18	1.00	1.00	1.00	48
S19	1.00	1.00	1.00	48
S20	1.00	1.00	1.00	48
S21	1.00	1.00	1.00	48
S22	1.00	1.00	1.00	48
Macro avg	1.00	1.00	1.00	1056
Weighted avg	1.00	1.00	1.00	1056

8. Conclusion

This research presents an innovative method for person identification by integrating EEG and facial data. The study seeks to enhance the effectiveness of EEG-based person identification (PI) through the use of an LSTM-CNN model, which combines Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures for a robust sequence modeling approach. The LSTM component is responsible for processing temporal features and capturing long-range dependencies, while the CNN component is used to learn local patterns or spatial features from the data. The proposed method is assessed using the DEAP dataset, a state-of-the-art benchmark for affective data. The findings reveal that the CNN-LSTM model, enhanced with MobileNet, can achieve a mean correct recognition rate of up to 99.81% for person identification (PI). This study employs the LSTM-CNN model, which integrates Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures for effective sequence modeling.The LSTM component is responsible for processing temporal features and capturing long-range dependencies, while the CNN component is used to learn local patterns or spatial features from the data. The proposed hybrid model demonstrates promising results, and the findings contribute to the evolving field of biometric identification. Future research could focus on optimizing the model architecture, investigating additional features, and enlarging the dataset to improve generalization capabilities.

Data availability

The DEAP dataset (Database for Emotion Analysis using Physiological Signals) is publicly accessible online via the dataset DOI: 10.1109/TAFFC.2012.12. It can be found at the following URL: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/ (accessed on May 13, 2023).

Declarations: Conflict of interest : The authors declare that they have no conflict of interest.

References

[1] W. Li, Y. Yi, M. Wang, B. Peng, J. Zhu, and A. Song, "A Novel Tensorial Scheme for EEG-Based Person Identification," in *IEEE Transactions on Instrumentation and Measurement*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/TIM.2022.3225016.

[2] W. Kudisthalert, K. Pasupa, A. Morales, and J. Fierrez, “SELM: Siamese extreme learning machine with application to face biometrics,” *Neural Comput. Appl.*, vol. 0123456789, 2022, doi: 10.1007/s00521-022-07100-z.
[3] S. B. Abdullahi, Z. A. Bature, P. Chophuk, and A. Muhammad, “Sequence-wise multimodal biometric fingerprint and finger-vein recognition network (STMFPFV-Net),” *Intell. Syst. with Appl.*, vol. 19, no. April, p. 200256, 2023, doi: 10.1016/j.iswa.2023.200256.
[4] Y. L. Lai, T.-Y. Chai, M. Lee, and B.-M. Goi, “Iriscode Matching Comparator to Improve Decidability of Human Iris Recognition,” in *Proceedings of the 6th International Conference on Digital Signal Processing*, in ICDSP ’22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 120–126. doi: 10.1145/3529570.3529591.
[5] X. Zhang, L. Yao, C. Huang, T. A. O. Gu, Z. Yang, and Y. Liu, “DeepKey: A Multimodal Biometric Authentication System via Deep Decoding Gaits and Brainwaves,” in *ACM Transactions on Intelligent Systems and Technology*, 2020. doi: 10.1145/3393619.
[6] D. Das Chakladar *et al.*, “A multimodal-Siamese Neural Network (mSNN) for person verification using signatures and EEG,” in *Information Fusion*, Elsevier B.V., 2021, pp. 17–27. doi: 10.1016/j.inffus.2021.01.004.
[7] C. Ding and D. Tao, “Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1002–1014, 2018, doi: 10.1109/TPAMI.2017.2700390.
[8] K. Patel, H. Han, and A. K. Jain, “Secure Face Unlock: Spoof Detection on Smartphones,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 10, pp. 2268–2283, 2016, doi: 10.1109/TIFS.2016.2578288.
[9] N. Bala, R. Gupta, and A. Kumar, “Multimodal biometric system based on fusion techniques: a review,” *Inf. Secur. J.*, vol. 31, no. 3, pp. 289–337, 2022, doi: 10.1080/19393555.2021.1974130.
[10] A. Al Abdulwahid, N. Clarke, I. Stengel, S. Furnell, and C. Reich, “Continuous and transparent multimodal authentication: reviewing the state of the art,” *Cluster Comput.*, vol. 19, no. 1, pp. 455–474, 2016, doi: 10.1007/s10586-015-0510-4.
[11] T. Fukami, Y. Abe, T. Shimada, and B. Ishikawa, “Authentication system preventing unauthorized access of a third person based on steady state visual evoked potentials,” in *International Journal of Innovative Computing, Information and Control*, ICIC International, 2018, pp. 2091–2100. doi: 10.24507/ijicic.14.06.2091.
[12] W. Deng, T. Hassner, X. Liu, and M. Pantic, “TBIOM Special Issue on Trustworthy Biometrics-Editorial,” *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 4, no. 3, pp. 301–302, 2022, doi: 10.1109/TBIOM.2022.3185447.
[13] A. Dediu, C. Martín-vide, and R. Goebel, *LNAI 7978 - Statistical Language and Speech Processing*, no. July. 2013.
[14] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, “Affective EEG-Based Person Identification Using the Deep Learning Approach,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 12, no. 3, pp. 486–496, 2020, doi: 10.1109/TCDS.2019.2924648.
[15] M. Wang, J. Hu, and H. A. Abbass, “BrainPrint: EEG biometric identification based on analyzing brain connectivity graphs,” in *Pattern Recognition*, Elsevier Ltd, 2020, p. 107381. doi: https://doi.org/10.1016/j.patcog.2020.107381.
[16] B. B. Das, P. Kumar, D. Kar, S. K. Ram, K. S. Babu, and R. K. Mohapatra, “A spatio-temporal model for EEG-based person identification,” *Multimed. Tools Appl.*, vol. 78, no. 19, pp. 28157–28177, 2019, doi: 10.1007/s11042-019-07905-6.
[17] Y. Sun, F. P.-W. Lo, and B. Lo, “EEG-based user identification system using 1D-convolutional long short-term memory neural networks,” in *Expert Systems with Applications*, 2019, pp. 259–267. doi: https://doi.org/10.1016/j.eswa.2019.01.080.
[18] E. Maiorana, “Learning deep features for task-independent EEG-based biometric verification,” *Pattern Recognit. Lett.*, vol. 143, pp. 122–129, 2021, doi: 10.1016/j.patrec.2021.01.004.
[19] Z. Abdi *et al.*, “EEG Channel Selection for Person Identification Using Binary Grey Wolf Optimizer,” vol. 10, 2022, doi: 10.1109/ACCESS.2021.3135805.
[20] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, “Affective EEG-Based Person Identification Using the Deep Learning Approach,” in *IEEE Transactions on Cognitive and Developmental Systems*, IEEE, 2020, pp. 486–496. doi: 10.1109/TCDS.2019.2924648.
[21] W. Alsumari, M. Hussain, L. Alshehri, and H. A. Aboalsamh, “EEG-Based Person Identification and Authentication Using Deep Convolutional Neural Network,” *Axioms*, vol. 12, no. 1, 2023, doi: 10.3390/axioms12010074.
[22] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, “Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain computer interface,” *32nd AAAI Conf. Artif. Intell. AAAI 2018*, no. August, pp. 1703–1710, 2018, doi: 10.1609/aaai.v32i1.11496.
[23] I. Sander Koelstra, Student Member, IEEE, Christian M¨uhl, Mohammad Soleymani, Student Member, IEEE, Jong-Seok Lee, Member, IEEE, Ashkan Yazdani, Touradj Ebrahimi, Member, IEEE, Thierry Pun, Member, IEEE, Anton Nijholt, Member, IEEE, Ioannis Patras, Member and Abstract—We, “DEAP: A Database for Emotion Analysis using Physiological Signals,” *IEEE TRANS. Affect. Comput.*, 2011.