

## Article

# A Lightweight GCT-EEGNet for EEG-Based Individual Recognition Under Diverse Brain Conditions

Laila Alshehri and Muhammad Hussain \* 

Department of Computer Science, King Saud University, Riyadh 11421, Saudi Arabia;  
441204115@student.ksu.edu.sa

\* Correspondence: mhussain@ksu.edu.sa

**Abstract:** A robust biometric system is essential to mitigate various security threats. Electroencephalography (EEG) brain signals present a promising alternative to other biometric traits due to their sensitivity, non-duplicability, resistance to theft, and individual-specific dynamics. However, existing EEG-based biometric systems employ deep neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which face challenges such as high parameter complexity, limiting their practical application. Additionally, their ability to generalize across a large number of subjects remains unclear. Moreover, they have been validated on datasets collected in controlled environments, which do not accurately reflect real-world scenarios involving diverse brain conditions. To overcome these challenges, we propose a lightweight neural network model, GCT-EEGNet, which is based on the design ideas of a CNN model and incorporates an attention mechanism to pay attention to the appropriate frequency bands for extracting discriminative features relevant to the identity of a subject despite diverse brain conditions. First, a raw EEG signal is decomposed into frequency bands and then passed to GCT-EEGNet for feature extraction, which utilizes a gated channel transformation (GCT) layer to selectively emphasize informative features from the relevant frequency bands. The extracted features were used for subject recognition through a cosine similarity metric that measured the similarity between feature vectors of different EEG trials to identify individuals. The proposed method was evaluated on a large dataset comprising 263 subjects. The experimental results demonstrated that the method achieved a correct recognition rate (CRR) of 99.23% and an equal error rate (EER) of 0.0014, corroborating its robustness against different brain conditions. The proposed model maintains low parameter complexity while keeping the expressiveness of representations, even with unseen subjects.

**Keywords:** EEG brain signals; biometric recognition; convolutional neural network; deep learning

**MSC:** 68T07



**Citation:** Alshehri, L.; Hussain, M. A. Lightweight GCT-EEGNet for EEG-Based Individual Recognition Under Diverse Brain Conditions. *Mathematics* **2024**, *12*, 3286. <https://doi.org/10.3390/math12203286>

Academic Editors: Grigoreta-Sofia Cojocar, Adriana-Mihaela Guran and Laura-Silvia Dioşan

Received: 22 September 2024

Revised: 16 October 2024

Accepted: 16 October 2024

Published: 20 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Identity recognition is crucial for verifying users and preventing imposters in various biometric applications. Traditional methods such as cards, keys, and passwords are widely used, but they are susceptible to loss or theft. Biometric traits such as fingerprints, iris appearance, voice, and gait offer promising alternatives, though each has its limitations [1,2]. For instance, biometrics involving the eyes, fingers, or faces cannot be easily replaced once compromised. To address these security concerns, EEG-based brain biometrics have emerged as a viable solution [3]. The EEG has been extensively used in medicine to assess brain health and in brain-computer interface systems, and is gaining acceptance as a biometric method due to its user-friendliness, the availability of portable headsets, and its non-invasive nature [4]. The EEG records electrical activity across the scalp using electrodes, offering advantages such as cost-effectiveness, temporal precision, resistance to theft, and the ability to verify live subjects. Each individual's EEG is unique, exhibiting

minimal intra-subject and significant inter-subject variability [5]. The primary challenge lies in developing a reliable EEG-based recognition system that recognizes individuals despite their brain activity variability. Numerous EEG-based biometric methods have evolved from those based on hand-engineered features using conventional machine learning to more advanced modern techniques such as convolutional neural networks (CNNs) [6–10] and recurrent neural networks (RNNs) [11–14]. Traditional methods rely on hand-engineered features, and they often preprocess EEG recordings to remove unwanted artifacts such as power supply noise, eye blinking, or muscle activity [15]. After preprocessing, features are extracted using methods such as auto-regressive (AR) models [16,17] and power spectral density (PSD) [18–20]. They are often difficult to tune and time-consuming, and they usually require expert knowledge. The designs of methods based on white-box models, such as auto-regressive models, assume simple and linear relationships in the data, making them less effective in capturing intricate patterns in EEG signals [4,21]. Therefore, simpler models often miss essential details, whereas EEG signals vary between subjects and brain states, and intricate patterns play a key role in discriminating subjects' identities. In contrast, a deep learning model automatically learns intricate patterns from the data hierarchically, making it better suited to capture discriminative features relevant to the identity of subjects from their EEG signals. There are many research works based on deep learning for EEG-based recognition [6–14]. However, they are not generalizable because their designs are based on small datasets that were collected during specific tasks with fewer than 60 participants. This limits their applicability to real-world scenarios. Further research is needed to improve the generalizability and applicability of EEG-based biometric systems by developing task-independent feature extraction methods and ensuring low time and space complexity in the model design. This study proposes a solution to tackle these issues through a compact and efficient deep learning model that automatically captures discriminative information for individual identification, thereby enhancing the system's usability and applicability in real-world scenarios. The key contributions of this research include the following:

- A lightweight deep neural network model based on the design ideas of CNN models and an attention mechanism to selectively focus on salient frequency bands for extracting discriminative features relevant to the identity of a subject from an EEG trial under various brain conditions.
- A robust EEG-based system for identification and authentication that is agnostic to various brain conditions, i.e., resting states, emotions, alcoholism, etc., and one that uses a short EEG trial of one second to reveal or authenticate the identity of a subject.
- A thorough evaluation for validating the proposed EEG-based system using a large dataset of 263 subjects who underwent EEG trials that were captured in various brain states.

The remainder of this study is organized as follows: Section 2 presents an overview of the existing works and Section 3 describes the proposed method. Section 4 explains the evaluation method and Section 5 describes the detailed experiments with discussions. Finally, the findings are summed up and future research is discussed in Section 6.

## 2. Related Work

The use of EEG-based biometrics has been explored since the 1980s, leveraging distinct electrical activity patterns for individual identification [22]. Over the years, research efforts have increasingly focused on extracting discriminative information from EEG recordings. Maiorana [23] highlighted the effectiveness of deep learning techniques, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in extracting unique features from various EEG representations and architectures. This section reviews the current literature on deep learning methods for EEG biometrics, covering both identification and verification approaches.

In the identification task, CNNs gained increasing attention due to their exceptional feature learning and classification abilities. Das et al. [11] applied a CNN–LSTM model

to identify 109 subjects from PhysioNet dataset, achieving 99.9 for eyes-closed (EC) and 98% for eyes-open (EO) tasks in trials of 12 s. Similarly, Jijomon and Vinod [12] developed a CNN–LSTM model that was applied to a private dataset consisting of 20 subjects performing auditory tasks (AEPs), reaching a 99.5% CRR with trials of 2 s. Wilaiprasitporn et al. [14] also employed CNN–LSTM and CNN–GRU for 2D meshes on the DEAP dataset, which involved 32 subjects performing emotion-related tasks, achieving a CRR of over 99% with a 10 s trial length. Jin et al. [24] proposed the CTNN model, which was employed on a private dataset of 20 subjects performing different brain tasks, achieving a CRR of 99.9%.

Different CNN-based models have also been used for verification tasks, such as spatial–temporal convolutions [7], depth-wise separable convolutions [25], and Siamese networks [9]. Chen et al. [7] used a CNN with global spatial and local temporal kernels on multiple datasets with different brain tasks, achieving an EER of 2.94. Debie et al. [25] applied a depth-wise separable convolution technique to a CNN on two public datasets with fewer than 54 subjects, performing different kinds of tasks, achieving a false acceptance rate (FAR) and false rejection rate (FRR) of less than 2%. In [13], a CNN–LSTM model was applied to the PhysioNet dataset, achieving an EER of 0.41. Seha and Hatzinakos [10] used 3D tensors with a CNN encoder, and the features were classified using an SVM on a private dataset of 13 subjects performing AEPs, achieving EERs between 3 and 7.5%.

Some previous methods treated identification and verification tasks as classification problems, making them impractical in real-world applications. In contrast, other studies [26–28] treated these tasks as matching problems using CNNs and focusing on specific protocols such as eyes-open (EO) and eyes-closed (EC) tasks or time-locked brain protocols (e.g., event-related potentials, ERPs). Alsumari et al. [27] and Bidgoly et al. [26] used the PhysioNet dataset, achieving correct recognition rates (CRRs) of 99.05% and 98.04%, with error rates of 0.187 and 1.96, respectively. In [28], ERPs were extracted from two datasets with 40 and 41 subjects, achieving CRRs of 95.63% and 99.92% and equal error rates (EERs) of 1.37% and 0.14%, respectively.

Although EEG-based biometric systems have made great progress over the years, research in this area still faces significant challenges. First, some methods [7,11,14,19,24] stack layers to CNNs or apply an RNN on top of a CNN in an end-to-end model, leading to parameter explosion as the number of subjects increases. To lower the number of parameters, the recognition problem should be treated as a matching problem. Additionally, most research works rely on private datasets or datasets with fewer than 100 subjects, making them less generalizable. These systems often need repeated stimuli in controlled environments, requiring subjects' cooperation to create the same brain state each time, which is not always possible because brain states are dynamic and not constantly at rest. In addition, external factors such as fatigue, mood, and alcohol use are not considered in many studies such that most research focuses on datasets such as PhysioNet, DEAP, and private datasets, which are limited to specific tasks such as motor imagery (MI), visual evoked potentials (VEPs), and auditory evoked potentials (AEPs).

Further, although some studies such as [11,14] achieved high accuracies, they used long trial lengths of 10 and 12 s, respectively. Jijomon and Vinod [12] also achieved high identification with only two electrodes, but their study involved a small number of subjects, limiting its applicability to real-world scenarios. End-to-end models come with a high cost in terms of time and space, as they need to be retrained every time a new subject registers. With a large number of individuals, the number of parameters can grow rapidly. While some techniques, such as depth-wise separable convolutions [24], help reduce spatial complexity, the overall computational demands and the need to retrain models for new subjects further limit their scalability. All these problems make the models less efficient for real-time applications.

Despite advancements in EEG-based biometric systems, ensuring effective performance in the presence of varying mental and physical activities remains a challenge, as EEG signals are influenced by factors such as movement, artifacts, fatigue, and emotions. These issues necessitate the development of a model that focuses on learning intricate

intrinsic and discriminative features from EEG trials across diverse brain states, including often neglected psychological factors such as fatigue. To address these issues, inspired by the design of EEGNet [29] and the limitations of the existing deep models for EEG-based recognition, we designed a model that incorporates depth-wise separable convolutions and attention mechanisms to focus on the most important EEG features, improving accuracy while reducing complexity.

### 3. The Proposed Method

We address the recognition problem using EEG brain signals as a biometric modality. First, we define and formulate the problem. The challenging part of the solution to this problem is the extraction of discriminative features from EEG signals. We present the details of a lightweight deep neural network model for feature extraction from EEG trials.

#### 3.1. Problem Specification and Formulation

In biometric recognition, there are two primary tasks. Given an EEG trial (a query trial) of a subject, the aim is to reveal (identification problem) or authenticate (verification problem) the identity  $l$  of the subject. In the identification task, the system determines the identity of an unknown subject by matching the query EEG trial  $x$  against the trials of all subjects in the gallery set; this task is formulated as a one-to-many matching problem. Identification can be with a closed set, where the trials of the query subject are known to be in the gallery set, or an open set, where the query subject may not be in the gallery set; this trial design is more challenging. In the verification task, the system verifies the claimed identity of a subject by comparing the query EEG trial  $x$  with the EEG trials of the same subject in the gallery set; this task is formulated as a one-to-one matching problem.

We represent an EEG trial or epoch as a matrix of size  $C \times T$ , i.e.,  $x \in R^{C \times T}$ , where  $C$  is the number of channels used to capture the brain's electrical activity over different locations on the scalp, and  $T$  is the number of timestamps recorded within a fixed time interval. Let  $X = \{X_1, X_2, \dots, X_N\}$  be the collection of EEG trials acquired from  $N$  subjects, such that  $X_i = \{x_1^i, x_2^i, \dots, x_{n_i}^i\}$  is the set of trials from the  $i$ th subject; for simplicity, we write this as  $X_i = \{x_1, x_2, \dots, x_{n_i}\}$ . In addition, let  $L = \{1, 2, \dots, N\}$  be the set of subject labels/IDs, and  $l \in L$  be the ID or label of the  $l$ th subject. Let  $V = \{V_1, V_2, \dots, V_N\}$ , where  $V_i = \{v_1^i, v_2^i, \dots, v_{n_i}^i\}$  is the set of feature vectors extracted from the EEG trials  $x_1, x_2, \dots, x_{n_i}$  corresponding to the  $i$ th subject; for simplicity, we write this as  $V_i = \{v_1, v_2, \dots, v_{n_i}\}$ . The crucial part of the design of the recognition system is the extraction of discriminative features  $v_i$  from EEG trials  $x_i$ . Inspired by the outstanding performance of deep learning models in automatic feature learning in various applications and, specifically, EEG-based applications [6,7], we design a lightweight deep model  $f$  for feature extraction in such a way that  $f(x; \theta) = v$ , where  $x$  is the input EEG trial,  $v$  is the feature vector extracted by  $f$ , and  $\theta$  represents the learnable parameters of  $f$ . The complexity of the model depends on the learnable parameters  $\theta$ ;  $f$  must be designed so that this complexity is low to avoid overfitting.

For the design of an identification or verification system, we divide the available data of each subject into query and gallery sets,  $V_i^q$  and  $V_i^g$ , respectively. Let  $v_q$  and  $v_g$  be the feature vectors extracted from a query and a gallery trial, respectively, i.e.,  $v_q \in V_i^q$  and  $v_g \in V_i^g$ . We compute the matching score  $s_{q,g} \in [0, 1]$  of  $v_q$  and  $v_g$  using the metric  $d(v_q, v_g)$  and let  $t$  be a predefined threshold. In case of an open-set identification problem, let  $s_i^q = \max \{0, s_{q,g} \mid s_{q,g} \geq t, v_g \in V_i^g\}$  be the maximum matching score of the query vector  $v_q$  from each of the gallery vectors  $v_g \in V_i^g$  of the  $i$ th subject. The predicted label of the query trail  $x_q$  is  $l_q$ , where  $l_q = \max_{1 \leq i \leq N} \{s_i^q \mid i = 1, 2, \dots, N\}$ . If each  $s_i^q$  is zero, the subject does not exist. In case of verification, let  $s_i^q = \max \{0, s_{q,g} \mid s_{q,g} \geq t, v_g \in V_i^g\}$  be the maximum matching score of the query vector  $v_q$  from each of the gallery vectors  $v_g \in V_i^g$  of the query subject. The attempt is genuine if  $s_i^q \neq 0$ ; otherwise, it is an impostor.

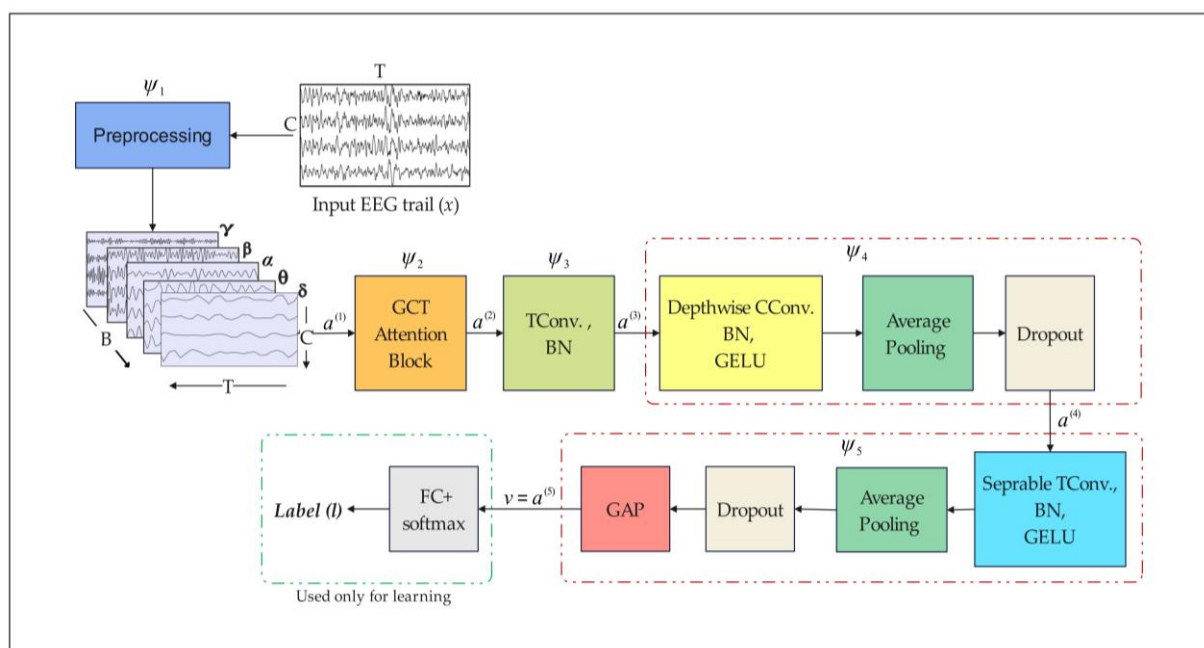
We examine metrics such as Euclidean, Manhattan, and cosine similarity to compute the similarity between two trials.

### 3.2. Deep-Learning-Based Feature Extractor

This section presents the details of a lightweight and task-independent deep model, GCT-EEGNet. Its architecture is inspired by EEGNet's success and excellent generalizability in various BCI paradigms [29]. It is designed as a feature extractor  $f$  to extract discriminative features relevant to the identity of a subject from their EEG trial  $x$ . Figure 1 provides an overview of the model architecture, and Table 1 gives its complete specification. First, an EEG trial  $x$  is preprocessed using the mapping  $\psi_1$ , which normalizes the trial and then decomposes it into frequency sub-bands (rhythms) using the discrete wavelet transform (DWT). Then, mapping  $\psi_2$  assigns weights to each rhythm according to its contribution to learning discriminative features; it is implemented as an attention block that helps to pay attention to the significant rhythms. It follows the mappings  $\psi_3$  and  $\psi_4$ , which learn low-level spectral-spatio-temporal features using temporal convolution (TConv), depth-wise channel convolution (DCCov), and average pooling. Finally, the mapping  $\psi_5$  learns high-level spectral-spatio-temporal features using separable temporal convolution (STCov) and global average pooling (GAP) blocks. The output of GAP is the feature vector ( $v$ ) used for identification and verification. Mathematically,  $f$  is a composition of the following five mappings:

$$f(x; \theta) = \psi_5 \circ \psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1(x) \quad (1)$$

where  $\theta = \{ \theta_2, \theta_3, \theta_4, \theta_5 \}$  and  $\theta_2, \theta_3, \theta_4, \theta_5$  are learnable parameters of  $\psi_2, \psi_3, \psi_4$ , and  $\psi_5$ , respectively. The details of each mapping are given in the following paragraphs. The GCT-EEGNet is trained as an end-to-end classification model; for this purpose, an FC layer with the softmax function is added after GAP during the training time. After training, the classification layer is removed, and the model is used as a standalone feature extractor.



**Figure 1.** The architecture of the attention-based EEGNet model (GCT-EEGNet), where  $B$  is the number of frequency bands,  $T$  is the time points,  $C$  is the number of channels of the EEG signal,  $\alpha, \beta, \gamma, \theta$ , and  $\delta$  are the alpha, beta, gamma, theta, and delta frequency bands, respectively. The TConv, CConv, BN, and GAP represent temporal and channel convolutions, batch normalization, and global average pooling layers, respectively, and  $v$  is the learned feature vector and  $l$  is the subject label.



**Table 1.** The specifications of the architecture of GCT-EEGNet, where C and D are hyperparameters, C is the number of channels, and D is a depth multiplier that specifies the number of spatial filters for each feature channel of the input feature map.

Transformation Mapping	Block Layers	#Kernel/Size	Output	Options	Learnable Parameters
	Input	-	$32 \times 128$		0
$\psi_1$	Preprocessing	-	$5 \times 32 \times 128$		0
$\psi_2$	GCT	-	$5 \times 32 \times 128$		15
$\psi_3$	Conv2D	$64/1 \times 64$	$64 \times 32 \times 128$	Padding = same	20,480
	BatchNorm	-	$64 \times 32 \times 128$		512
$\psi_4$	Depthwise Conv2D	$D \times 64/C \times 1$	$64 \times 1 \times 128$	D = 1 C = 32	2048
	BatchNorm	-	$64 \times 1 \times 128$		512
	GELU	-	$64 \times 1 \times 128$		0
	Average Pooling2D	$1 \times 4$	$64 \times 1 \times 32$		0
	Dropout	0.5	$64 \times 1 \times 32$		0
$\psi_5$	Separable Conv.	$128/1 \times 16$	$128 \times 1 \times 32$		9216
	BatchNorm	-	$128 \times 1 \times 32$		128
	GELU	-	$128 \times 1 \times 32$		0
	Average Pooling 2D	$1 \times 8$	$128 \times 1 \times 4$		0
	Dropout	0.5	$128 \times 1 \times 4$		0
	GAP Layer	-	128		0
	FC + Softmax	-	236		30,444
	Total Parameters				62,764

### 3.2.1. Data Preprocessing

An EEG trial  $x$  is preprocessed with the mapping  $\psi_1$ , which is composed of the following two functions:

$$\psi_1(x) = \mathcal{F} \circ \aleph(x) \quad (2)$$

where the function  $\aleph$  normalizes the input EEG trial  $x$ , and then the function  $\mathcal{F}$  decomposes it into rhythms. The function  $\aleph$  is defined using Z-score normalization [13] as follows:

$$x'_{c,t} = \frac{x_{c,t} - \mu_c}{\sigma_c}, \quad c = 1, 2, \dots, C = 1, 2, \dots, T \quad (3)$$

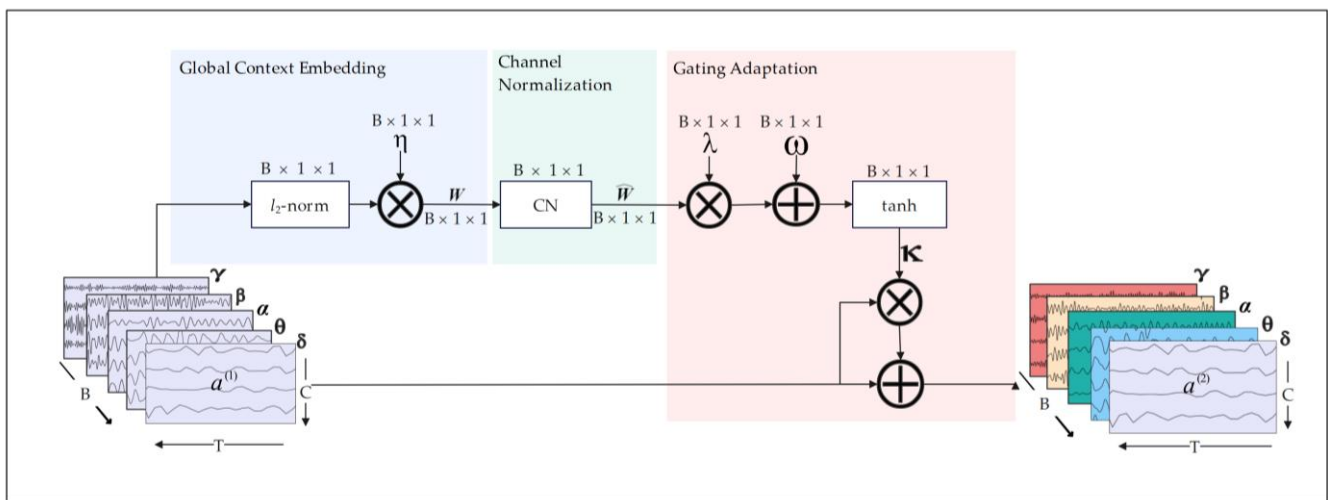
where  $c$ ,  $t$ ,  $\mu_c$ , and  $\sigma_c$  refer to the channel identifier, the signal value at a specific time point, and the mean and standard deviation of the  $c$ th channel, respectively. Note that the function is applied on each channel individually to address differences in the feature unit and scale while improving the convergence speed. Instead of utilizing the entire frequency spectrum, which is rarely employed in biometrics, specific frequency bands or rhythms known to be more discriminative are used [30]. The study in [15] indicated that EEG bands below 50 Hz have higher energy for biometric identification. Consequently, the function  $\mathcal{F}$  decomposes each EEG trial into frequency bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–16 Hz), beta (16–32 Hz), and gamma (32–50 Hz)—to assess their significance in the recognition process. Following a previous study [31], the function  $\mathcal{F}$  is based on the DWT due to the nonstationary rapid fluctuations in EEG [32]. The DWT with the fourth-order Daubechies mother wavelet (db4) is used to decompose an EEG segment into the A5 (low-frequency) and D1–D5 (high-frequency) bands, where A5 is the delta ( $\delta$ ) band, while D2 to D5 are the theta ( $\theta$ ), alpha ( $\alpha$ ), beta ( $\beta$ ), and low gamma ( $\gamma$ ) bands, respectively. The choice of the DWT, specifically the db4 wavelet, is well suited for EEG analysis because its morphology is similar to that of EEG data [33,34]. Finally, the mapping  $\psi_1$  transforms the input EEG trial  $x \in R^{C \times T}$  into a tensor  $a^{(1)} \in R^{B \times C \times T}$  of  $B$  bands,  $C$  channels, and  $T$  time samples, as depicted in Figure 1.

### 3.2.2. GCT Attention Block

For subject recognition, not all brain rhythms and channels are equally important. Therefore, identifying the most significant ones without extensive experiments is crucial. To address this issue, we employ an attention mechanism using the GCT block with mapping  $\psi_2$ , which is a composition of the following mappings:

$$\psi_2(a^{(1)}; \theta_2) = \mathcal{X}_4 \circ \mathcal{X}_3 \circ \mathcal{X}_2 \circ \mathcal{X}_1(a^{(1)}; \eta) \quad (4)$$

where  $\theta_2 = \{\eta, \lambda, \omega\}$  are learnable parameters of  $\mathcal{X}_1$  and  $\mathcal{X}_3$ , respectively. The GCT [35] is a simple and effective attention module that simulates channel interactions without extra parameters. It helps prioritize and emphasize key rhythms and channels for the recognition task. It consists of the following three main components: global context embedding, channel normalization, and gating mechanism, as depicted in Figure 2.



**Figure 2.** Gate channel transformation (GCT) module, where  $B$  is the number of frequency bands,  $T$  is the time points,  $C$  is the number of channels of the EEG signal,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\theta$ , and  $\delta$  are the alpha, beta, gamma, theta, and delta frequency bands, respectively.  $\eta$  denotes the trainable embedding weights,  $W$  represents the global context information,  $\lambda$  and  $\omega$  represent the gating weights and biases, and  $\kappa$  is the output of tanh function. Different colors in the output  $a^{(1)}$  indicate the varying significance assigned to each band.

Initially, global contextual information is captured by mapping  $\mathcal{X}_1$  using the  $l_2$ -norm from the EEG trail  $a^{(1)} \in R^{B \times C \times T}$  as follows:

$$\mathcal{X}_1(a^{(1)}; \eta) = W = [w_1, w_2, \dots, w_b]^t, \quad W \in R^{B \times 1 \times 1} \quad (5)$$

$$w_b = \eta_b \sqrt{\left[ \sum_{i=1}^C \sum_{j=1}^T (a^{(1)}_b)^{i,j} \right]^2} + \varepsilon, \quad b = 1, 2, \dots, B \quad (6)$$

where  $W = [w_1, w_2, \dots, w_b]^t$  represents the globally collected contextual information along each frequency band dimension  $b \in B$ ,  $\varepsilon$  serves as a small constant to avoid the derivation problem at zero,  $\eta$  denotes the trainable embedding weights for controlling and emphasizing each frequency band's significance, and  $C$  and  $T$  refer to the number of channels and time points, respectively. Then, channel normalization (CN) is applied using mapping  $\mathcal{X}_2$  by normalizing each component  $w_b$  of  $W$ , as shown in the following Equation (7):

$$\mathcal{X}_2(W) = \hat{W} = (\hat{w}_1, \hat{w}_2, \dots, \hat{w}_b)^t, \quad \hat{W} \in R^{B \times 1 \times 1} \quad (7)$$

where

$$\hat{w}_b = \frac{\sqrt{B} \cdot w_b}{\sqrt{\left(\sum_{b=1}^B w_b^2\right) + \varepsilon}} \quad (8)$$

and the scaler  $\sqrt{B}$  is used to adjust the scale of  $\hat{W} = (\hat{w}_1, \hat{w}_2, \dots, \hat{w}_b)$ . This adjustment helps prevent  $\hat{w}_b$  from becoming too small when the number of frequency bands is large. Channel normalization encourages channel interactions, whereas the  $l_2$ -norm operates across channels. It permits larger responses for some frequency band coefficients and suppresses others with smaller feedback. Finally, using the normalized vector  $\hat{W}$  and the frequency bands  $a^{(1)}$ , gating adaptation takes place using mappings  $\mathcal{X}_3$  and  $\mathcal{X}_4$ , as shown in the following:

$$\mathcal{X}_3(\hat{W}; \lambda, \omega) = \kappa = \tanh(\lambda \hat{W} + \omega), \quad \kappa \in R^{B \times 1 \times 1}, \quad (9)$$

$$a^{(2)} = \mathcal{X}_4(a^{(1)}, \kappa) = a^{(1)} \oplus a^{(1)} \otimes \kappa; \quad a^{(2)} \in R^{B \times C \times T} \quad (10)$$

where  $\lambda$  and  $\omega$  represent the gating weights and biases that control the activation of features, while  $a^{(1)}$ ,  $a^{(2)}$  denote input and output features for the gating mechanism module, respectively. The gating mechanism boosts competition and cooperation between frequency bands during the training process. To enhance feature extraction and classification performance, we employed convolutional layers to extract both temporal and spatial features.

### 3.2.3. Temporal Convolution Block

This block is designed to capture time-dependent features within an EEG trial, enabling the model to learn important temporal relationships that are crucial for distinguishing between different subjects. It operates along the time axis via a standard 2D convolutional layer  $\varphi_1$ , transforming the tensor  $a^{(2)} \in R^{B \times C \times T}$  into the tensor  $a^{(3)} \in R^{k_1 \times C \times T}$  through the mapping  $\psi_3$  as defined below:

$$\psi_3(a^{(2)}; \theta_3) = BN \circ \varphi_1(a^{(2)}; \theta_3) \quad (11)$$

where  $k_1$  is the number of kernels used for temporal convolution with a size of  $1 \times 64$  to detect the temporal features for each frequency band. In our experiments, we set  $k_1$  to 64. To enhance neural network performance and achieve faster training convergence, each convolutional layer is followed by a subsequent batch normalization (BN) layer [36].

### 3.2.4. Depth-Wise Channel Convolution Block

To reduce the model's computational complexity while extracting spatial features, a depth-wise channel convolution layer  $\varphi_2$  is applied. Similar to the approaches used in Xception [37] and MobileNet [38], this layer applies a single filter per input channel, effectively isolating channel-specific features without the overhead of traditional convolution operations. Using  $\varphi_2$ , the mapping  $\psi_4$  transforms  $a^{(3)}$  into  $a^{(4)} \in R^{k_2 \times 1 \times \frac{T}{4}}$  as follows:

$$\psi_4(a^{(3)}; \theta_4) = \mathcal{P}_a \circ g \circ BN \circ \varphi_2(a^{(3)}; \theta_4) \quad (12)$$

where  $k_2$  is the number of kernels, each of size  $C \times 1$ , and  $C$  is the number of channels. These kernels are applied along the spatial (channel) axis, enabling the network to learn  $D$  spatial kernels, with each kernel being dedicated to a distinct feature map. The result is an output feature map of an extended dimension  $D \times k_2$ . This approach provides the following two key advantages: first, it serves as a spatial cross-channel feature learner, enhancing the global feature extraction, especially in multi-channel EEG data. Second, it reduces the number of learnable parameters, as this layer is not connected to all outputs from the preceding layer. Then, it is followed by BN. Unlike the original EEGNet activation function  $g$ , which utilizes exponential linear units (ELUs), we employ the Gaussian error linear unit (GELU) [39], which was motivated by its success in vision transformers



(ViTs) [40]. It is applied in the second and third convolutional layers, combining the benefits of dropout [41] and randomly removing neurons during the training process. To further reduce dimensionality, an average pooling layer  $\mathcal{P}_a$  with a window of size  $1 \times 4$  is employed after the GELU layers. All convolution layers are applied with a stride of one.

### 3.2.5. Separable Temporal Convolution Block

Finally, the separable convolution layer  $\varphi_3$  integrates depth-wise and pointwise convolutions to decompose the convolution process further, thereby enabling the model to process spatial and temporal features independently. This approach not only improves the model's ability to capture complex patterns in EEG data but also enhances its computational efficiency, leading to more robust and accurate classification output. Specifically, this layer employs a  $1 \times 16$  kernel to aggregate individual feature maps. Then, a pointwise convolution with 128 kernels, each of size  $1 \times 1$ , is employed to combine these feature maps optimally. This setup effectively exploits temporal and spatial features for individual recognition. Employing  $\varphi_3$ , the transformation map  $\psi_5$  that converts the tensor  $a^{(4)} \in R^{k_2 \times 1 \times \frac{T}{4}}$  into  $a^{(5)} \in R^{k_2 \times 1 \times \frac{T}{4 \times 8}}$  is defined as follows:

$$\psi_5(a^{(4)}; \theta_5) = \mathcal{P}_g \circ \mathcal{P}_a \circ g \circ BN \circ \varphi_3(a^{(4)}; \theta_5) \quad (13)$$

where  $BN$ ,  $g$ , and  $\mathcal{P}_a$  denote the batch normalization, GELU, and average pooling with a window of size  $1 \times 8$ , respectively. Instead of incorporating a fully connected layer, which would increase the model complexity, a GAP layer  $\mathcal{P}_g$  is utilized, where  $\mathcal{P}_g(a^{(5)}) = v$ . This serves later as a feature extraction layer. This GAP layer reduces feature dimensionality and model parameters for efficient feature extraction. The resulting feature vector  $v$  is then fed into a softmax classifier with  $N$  units, corresponding to the total number of subjects that the model is trained on.

### 3.2.6. Training of GCT-EEGNet

The network was trained as an end-to-end model using a categorical cross-entropy loss for 100 epochs with a batch size of 500. The AdamW optimizer [42] with its default parameters was used for training. The training stopped if validation loss did not improve for three consecutive epochs, and an early stopping technique [43] was employed to prevent overfitting.

For model evaluation, a stratified 10-fold cross-validation was applied based on the subjects. In each fold, the subjects were divided into the following two groups: 90% were used for training, and the remaining 10% were used for testing. This format ensured that each fold used distinct subjects for testing [44]. After training, the model was utilized as a feature extractor, and the identification and authentication performances were assessed using 10% of the subjects reserved for testing. All results are reported as the average performance across folds.

## 4. Evaluation Protocol

This section first describes the datasets used to evaluate the proposed method. Then, it provides an overview of the performance metrics used for evaluation. For evaluation, 10-fold cross-validation was used, as described in Section 3.2.6.

### 4.1. Datasets

To validate the generalization of the proposed model across diverse brain activations, three publicly available EEG benchmark datasets were combined to create a larger dataset with a large number of subjects—263 in total—encompassing diverse human states. The EEG signals in each dataset were downsampled to 128 Hz, and the same channels were selected from all datasets.

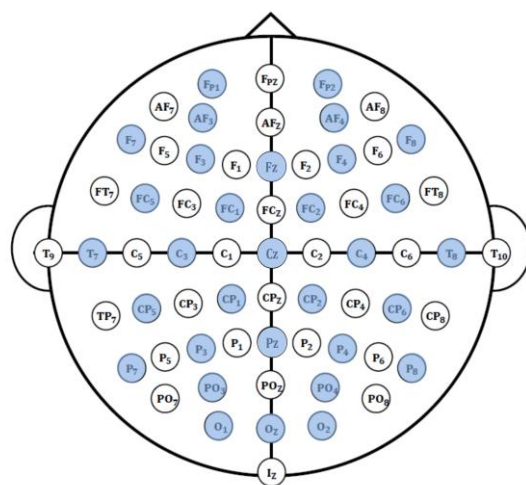
The DEAP [45] was developed to analyze human affective states. It was recorded from 32 individuals using a BioSemi headset with 32 EEG channels placed on the skull based on

a 10–20 system and a 512 Hz sampling rate. The subjects watched 40 one-minute music videos that corresponded to different emotional states, i.e., valence, arousal, dominance, and liking. For a fair comparison, we used the preprocessed version.

Physionet motor/imagination [46] is a well-known and widely used EEG dataset with 64 channels; 160 Hz EEG recordings were captured from 109 healthy subjects. The international 10–10 system for the placement of electrodes was utilized. Each participant's EEG was recorded over 14 tests. One baseline run lasted one minute in both “eyes-open” (EO) and “eyes-closed” (EC) conditions, and the last four minutes contained four motor/imagination (MI) activities.

The EEG UCI dataset was produced for alcoholism-related genetic studies and involves recordings from 122 patients, including 45 controls and 77 alcoholics, with each completing 120 one-second trials. It was collected using a 10–20 system with 64 electrodes at 256 Hz; subjects viewed black-and-white photos [47] for 300 ms, with a separation of 1.6 s. Subjects were asked to determine if the two photos were identical.

The combined dataset (CD) integrated data from all subjects across the *DEAP*, *UCI*, and *PhysioNet* datasets. Due to differences in EEG data collection equipment, 32 channels of each EEG trial were selected according to the 10–20 electrode placement system (see Figure 3). The EEG records were then segmented into non-overlapping one-second epochs (EEG trials), resulting in EEG trials of  $32 \times 128$ , where 32 was the number of channels, and 128 was the number of time samples. This dataset included EEG trials from a total of 263 subjects, with each having different brain activations. Subjects were numbered sequentially for training purposes, starting with those from *DEAP*, followed by *PhysioNet* and then *UCI*.



**Figure 3.** Channel positions of all 64 electrodes (channels) using a 10–20 system where the highlighted channels were used in experiments [24].

#### 4.2. Performance Metrics

Commonly used metrics for identification and verification tasks were employed to evaluate the efficacy of GCT-EEGNet. The study considered the correct recognition rate (CRR) and cumulative match characteristic (CMC) curve for identification and the equal error rate (EER) and detection error trade-off (DET) curve for verification. A lower EER signifies a better performance in authentication scenarios.

### 5. Experimental Results and Discussion

This section presents the details of the experiments conducted to validate the performance of the method and discusses the results obtained. All experiments were performed on a computer with 128 GB of RAM and an NVIDIA Quadro RTX 6000 GPU. The model was implemented using Python v3.7, Pytorch Lightning v1.7, and torch v1.13.1.

### 5.1. Ablation Study

In this section, we discuss ablation experiments that were performed to assess the impact of each model component using 10-fold cross-validation. Initially, for each fold, we trained the model using all 32 channels and used 90% of the subjects for training; e.g., in the DEAP dataset, 28 subjects were used for training, and the remaining 10% were used for testing. We utilized the same hyperparameters in GCT-EEGNet that are common in EEGNet [29], the baseline network, which was trained for 100 epochs per fold. The impacts of various factors and hyperparameters on the performance of the model are shown in Table 2. The model configuration that gave the highest average validation accuracy over 10-fold cross-validation on the datasets was considered for further improvement.

#### 5.1.1. Input Configuration and Optimizers

We compared the results obtained with the 3D input shape [ $5 \times 32 \times 128$ ] with those obtained with the original 2D shape [ $32 \times 128$ ] while keeping all hyperparameters fixed in GCT-EEGNet as in the baseline EEGNet network. Besides the network's architectural design, the training method affects the model's performance [48]. The vision transformer [49,50] introduced a new collection of modules and new training methods (e.g., the AdamW optimizer). As shown in Table 2, the 3D input shape using the same optimizer achieved better performance for non-preprocessed datasets using the Adam optimizer, particularly the UCI dataset, with a difference of almost 13%. In addition, we observed that the AdamW optimizer yielded better outcomes for three datasets, with a slight decrease in validation accuracy of 0.05% for DEAP.

#### 5.1.2. Number of Kernels and Activation Functions

Table 2 also shows the impact of altering the number of kernels from (8, 16) to (32, 64) and (64, 128) for the first and second convolution layers, respectively, in GCT-EEGNet while keeping the ELU activation function fixed. It is evident that using 64 and 128 kernels produced the best results. A larger number of kernels in the first layer resulted in better accuracy than a smaller number of kernels, with a slight improvement in the DEAP dataset, since it was preprocessed, and there was a significant improvement in the other two datasets in comparison with the small number of kernels (8, 16). In addition, the most often used activation functions were assessed, and the results indicated that the GELU activation function achieved the best results, with a modest improvement over ReLU and ELU; the SiLU activation function showed a slight improvement for some datasets, but its long training time is a major drawback.

**Table 2.** Ablation study of the performance of GCT-EEGNet; the performance is reported as the mean validation performance  $\pm$  standard deviation using 10-fold cross-validation; *ELU* is the exponential linear unit, *Avg* is the average pooling layer, *ReLU* is the rectified linear unit, *GELU* is the Gaussian error linear units, *SiLU* is the sigmoid linear unit, *GAP* is global average pooling, *SE* is squeeze and excitation, *GCT* is gated channel transformation, and *RMSProp* is root mean squared propagation.

Experiment	Choices	Datasets			
		DEAP	PhysioNet	EEG UCI	Combined
Raw 2D input without DWT decomposition ( $32 \times 128$ )					
Optimizers, kernels 8, 16	Adam	$99.87 \pm 0.08$	$74.51 \pm 2.75$	$49.10 \pm 7.29$	$69.80 \pm 3.02$
	AdamW	$99.92 \pm 0.06$	$73.72 \pm 2.14$	$50.50 \pm 5.26$	$69.81 \pm 3.05$
Raw 3D input with DWT decomposition ( $5 \times 32 \times 128$ )					
Optimizers, kernels 8, 16	Adam	$99.88 \pm 0.05$	$76.90 \pm 1.63$	$62.57 \pm 2.74$	$74.56 \pm 1.35$
	AdamW	<b><math>99.87 \pm 0.11</math></b>	$77.57 \pm 0.87$	$64.22 \pm 3.80$	$74.69 \pm 1.41$

Table 2. Cont.

Experiment	Choices	Datasets			
		DEAP	PhysioNet	EEG UCI	Combined
Number of kernels	32, 64	$99.99 \pm 0.02$	$99.13 \pm 0.15$	$95.58 \pm 0.64$	$98.69 \pm 0.19$
	64, 128	$100 \pm 0.01$	$99.75 \pm 0.05$	$97.41 \pm 0.94$	$99.54 \pm 0.08$
Activation functions	ReLU [51]	$97.75 \pm 0.77$	$99.67 \pm 0.07$	$97.75 \pm 0.77$	$99.39 \pm 0.09$
	SiLU [52]	$97.84 \pm 0.54$	$99.77 \pm 0.07$	$97.84 \pm 0.54$	$99.53 \pm 0.04$
	GeLU [39]	$100 \pm 0.01$	$99.79 \pm 0.06$	$97.90 \pm 0.52$	$99.50 \pm 0.03$
Pooling Layer	Max	$100 \pm 0.01$	$99.68 \pm 0.08$	$97.50 \pm 0.61$	$99.38 \pm 0.09$
GAP layer	GAP	$100 \pm 0.00$	$99.80 \pm 0.06$	$98.51 \pm 0.40$	$99.58 \pm 0.08$
Attention Layer	SE	$100 \pm 0.00$	$99.68 \pm 0.06$	$98.73 \pm 0.36$	$99.54 \pm 0.10$
	GCT	$100 \pm 0.00$	$99.84 \pm 0.05$	$98.87 \pm 0.33$	$99.66 \pm 0.04$
Dropout	0.5	-	-	-	$99.66 \pm 0.04$
	0.25	-	-	-	$99.63 \pm 0.06$
	Without dropout	-	-	-	$99.24 \pm 0.19$

### 5.1.3. Pooling Layer

To reduce the feature map dimensionality, pooling layers were employed. We compared the following two most popular types of pooling in this experiment: average and maximum pooling. Because the average pooling layer was used from the beginning of the first test, all previous results included the average pooling layer. Table 2 shows that the average pooling achieved good accuracy compared with max pooling.

### 5.1.4. GAP and Attention Layer

Instead of simply flattening or adding an FC layer, we used a GAP layer. The GAP layer averaged spatial information to strengthen the input against spatial translations. The results in Table 2 show that the validation accuracy improved with a notable reduction in the learnable parameters. In addition, we evaluated two attention approaches to capture the channel importance. The GCT layer was shown to be better than the squeeze and excitation (SE) block, with a small difference.

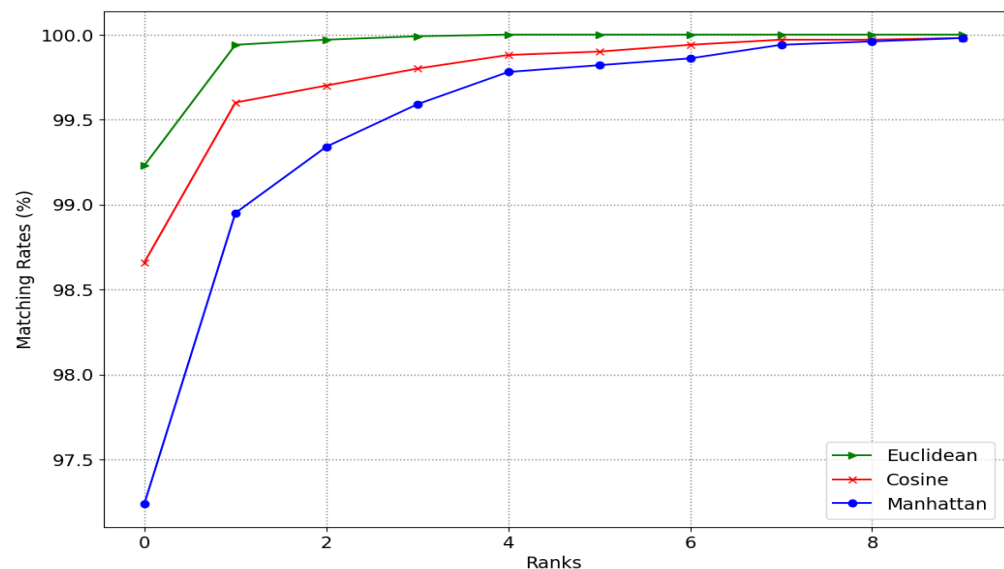
### 5.1.5. The Effect of Employing GELU with Dropout

As the GELU activation function incorporated a dropout functionality, we needed to guarantee that this layer had a positive effect on the presence of GELU or that the validation accuracy would be reduced. As a result, we found that the dropout was beneficial for this application, as removing this layer would result in a modest decline in outcomes (see Table 2). Note that all previous experiments included dropout with a 50% rate. This experiment applied only to the combined dataset.

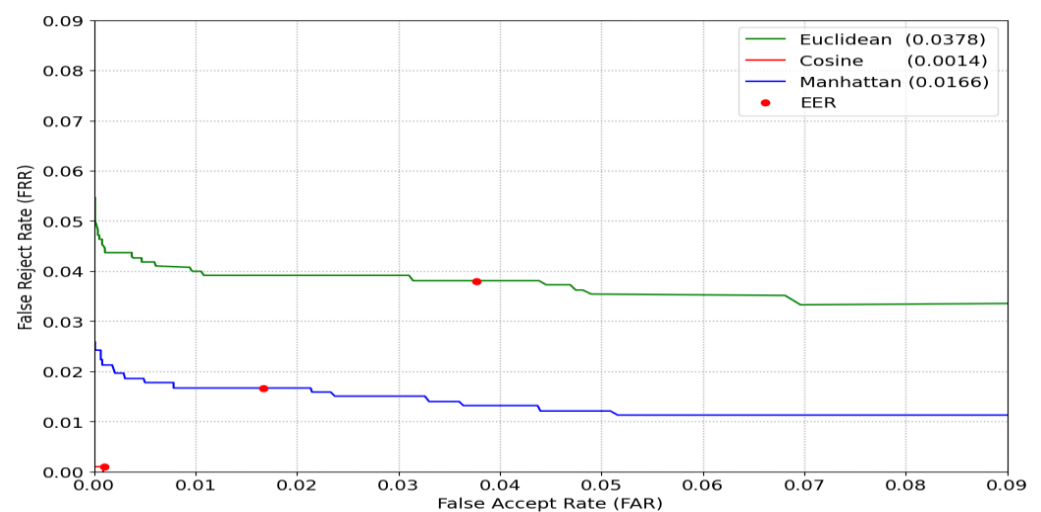
## 5.2. The Identification and Verification Results

The ablation study helped to find the best configuration of GCT-EEGNet. Using its best configuration, we extracted the feature vector of each EEG trial as the output of the GAP layer. Then, we matched pairs of EEG trials by determining the similarity between their feature vectors. We explored several similarity metrics for matching, including Euclidean, Manhattan, and cosine similarity. The choice of a similarity metric can significantly impact the results, and we aimed to identify the most effective one. Our findings revealed that the cosine similarity measure (red line) consistently outperformed the others in both the identification and verification scenarios (see Figures 4 and 5). Figure 4 shows the CMC curves for the identification scenario, illustrating the top ten ranks for the combined dataset. The best results were achieved using the cosine similarity measure with a CRR of 99.23%. For the verification scenario, we considered genuine pairs (within a class) and impostor (between classes) pairs, with 1080 and 28,080, respectively. The DET curves

depicted in Figure 5 show that the cosine distance measure resulted in the lowest EER of 0.0014%. The EER represents the threshold at which the false acceptance rate (FAR) equals the false rejection rate (FRR). The results indicate that the cosine distance is the best similarity measure, and this outcome was also confirmed by the authors of [26] using the PhysioNet dataset.



**Figure 4.** Performance in identification: CMC curves for the combined dataset.



**Figure 5.** Performance in verification: DET curves for the combined dataset.

### 5.3. Robustness to Diverse Brain States

To replicate real-world scenarios and demonstrate the robustness of GCT-EEGNet across diverse brain states, we extracted epochs (EEG trials) from EEG signals, regardless of the onsets or offsets of cognitive tasks, and performed two experiments. This approach ensured the generalization of GCT-EEGNet across diverse cognitive states, as shown in Table 3. In the first experiment, the model was trained on cognitive states different from those employed during the testing stage. The results indicated that the model achieved a good CRR and EER on the DEAP dataset, where nearly equal samples were sampled from each cognitive state. Additionally, the results from the PhysioNet and UCI datasets highlighted the impact of the difference in training and testing sample sizes on the model performance. For example, in the PhysioNet dataset, the performance decreased by nearly 9% from 97.83% to 88.72% when the model trained on a small number of samples (13,195)



for EO and EC conditions and tested on a much larger sample size of 161,647 for PHY and IMA. Conversely, when trained on the larger PHY and IMA sample, the model's performance improved, yielding a CRR of 97.83% and an EER of 0.0047. The performance was particularly good for the UCI dataset when trained on the alcoholic state, which was likely due to the larger sample size of 6989 for 77 subjects compared with the 4015 samples from 45 subjects in the non-alcoholic training phase. These findings emphasize the critical influence of the training data size on model performance.

**Table 3.** Test results of experiments with the CRR and EER (average  $\pm$  standard deviation), where *HH* is high valence, high arousal; *HL* is high valence, low arousal; *LH* is low valence, high arousal; *LL* is low valence, low arousal; *EO* is eyes open; *EC* is eyes closed; *PHY* is motor physical activity; and *MI* or *IMA* is motor imagination activity.

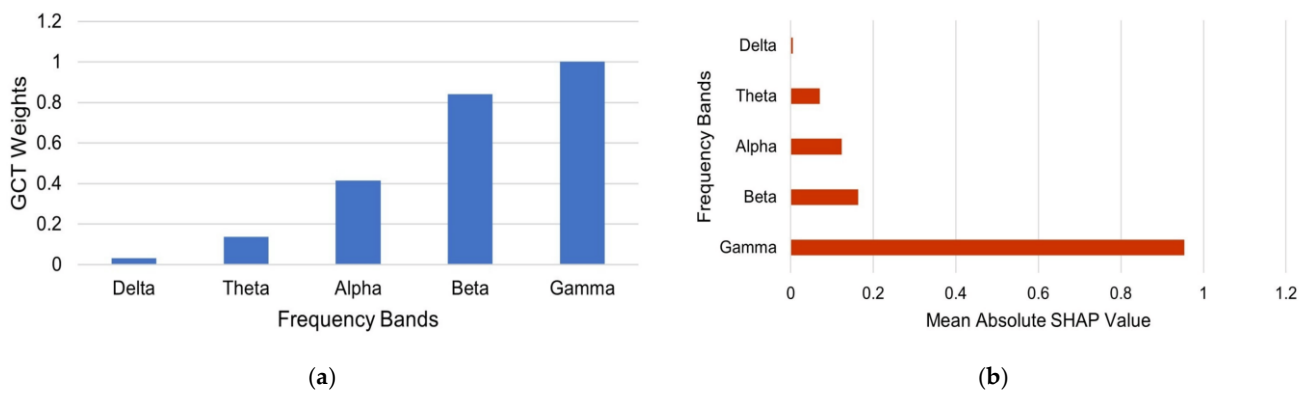
Experiment # 1				
Dataset	Training States	Testing States	CRR	EER
DEAP	LL, HH	LH, HL	99.99 $\pm$ 0.04	0.0215 $\pm$ 0.0183
	LL, HL	LH, HH	99.98 $\pm$ 0.05	0.0272 $\pm$ 0.0253
	LL, LH	HL, HH	99.96 $\pm$ 0.06	0.0283 $\pm$ 0.0106
	HH, HL	LL, LH	99.93 $\pm$ 0.09	0.1079 $\pm$ 0.0216
	HH, LH	LL, HL	99.98 $\pm$ 0.05	0.0860 $\pm$ 0.0597
	LH, HL	LL, HH	100.00 $\pm$ 0.00	0.0523 $\pm$ 0.0061
PhysioNet	EO, EC	PHY, IMA	88.72 $\pm$ 1.12	0.0514 $\pm$ 0.0105
	PHY, IMA	EO, EC	97.83 $\pm$ 1.66	0.0047 $\pm$ 0.0023
EEG UCI	Alcoholic	Non-Alcoholic	84.25 $\pm$ 0.83	0.0087 $\pm$ 0.0015
	Non-Alcoholic	Alcoholic	77.47 $\pm$ 0.56	0.0041 $\pm$ 0.0008
Experiment # 2				
Datasets			CRR	EER
DEAP			100.00 $\pm$ 0.00	0.0004 $\pm$ 0.0008
PhysioNet			98.90 $\pm$ 0.48	0.0043 $\pm$ 0.0014
EEG UCI			99.25 $\pm$ 0.91	0.0009 $\pm$ 0.0016
Combined			99.23 $\pm$ 0.50	0.0014 $\pm$ 0.0008

In addition, binding a system to a specific mental state during registration is often impractical in real-world biometric applications. To address this variation, the model was trained with diverse states, enabling it to adapt to subject variability. In the second experiment, data from various brain states were merged for both training and testing. This approach demonstrated that the proposed model achieved better identification results (more than 98% for all datasets) and verification results of less than 0.004. This indicated that the model generalized well over diverse cognitive states, even with new subjects that were never introduced during the training procedure. This adaptability to intra-person EEG variability makes the model a promising candidate for real-world biometric applications.

#### 5.4. The Effects of Different Frequency Bands

This section examines how various frequency bands, including the delta, theta, alpha, beta, and gamma bands, impacted the brainprints derived from EEG-generated spontaneous brain activity. The GCT attention block within the model played a crucial role in determining the most contributive frequency bands for recognition. Attention weights were computed per frequency band generated in the testing samples and then averaged to account for subject variability, as the different subjects generated distinct attention patterns. Figure 6a shows that the beta (14–32 Hz) and gamma (32–50 Hz) bands dominated the combined dataset. These findings suggest that lower frequencies correspond to common brain activities, while higher frequencies are associated with individual distinctiveness. To reveal the significance of frequency bands, we applied the deepSHAP technique [53]. Figure 6b presents a global interpretation of the model's decisions, highlighting the prominence of

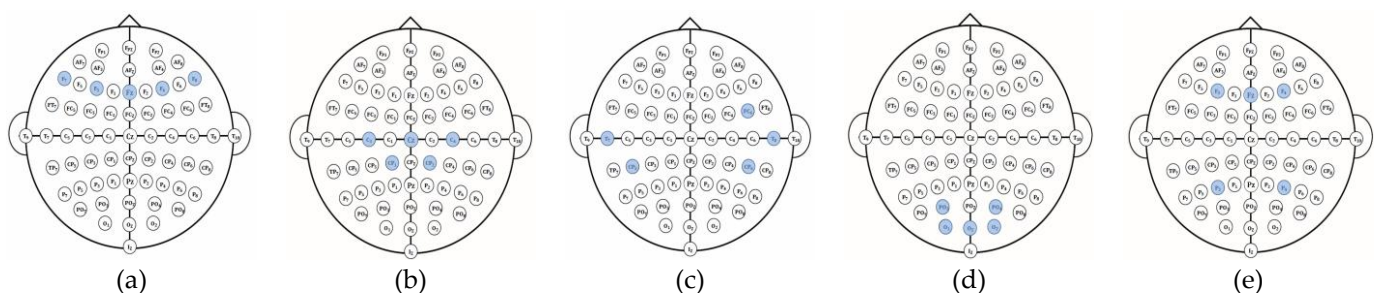
the gamma band across all 32 channels, which aligns with the results observed in the GCT layer. This suggests that the attention layer of our model could provide valuable insights into identifying the frequency band with the greatest contribution.



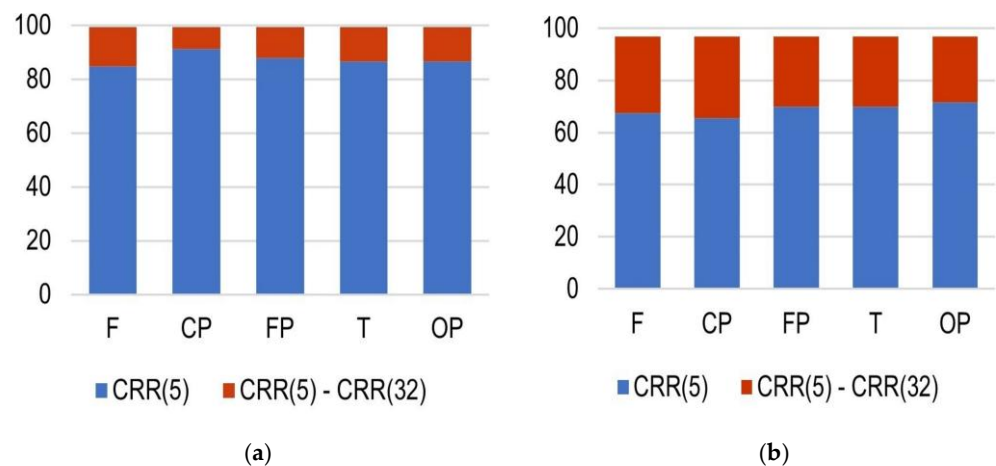
**Figure 6.** The effect of frequency bands on the combined dataset. (a) The GCT attention mechanism weights. (b) The respective mean SHAP values.

### 5.5. The Effect of Channel Reduction

In this experiment, we analyzed the effect of reducing the number of EEG channels on the model performance. To improve the system's user-friendliness, it was necessary to minimize the number of electrodes while maintaining satisfactory performance. Figure 7a–e displays five sets of EEG channels defined by Wilaiprasitporn et al. [14], with each covering the following distinct regions of the scalp: frontal (F), central and parietal (CP), temporal (T), occipital and parietal (OP), and frontal and parietal (FP). The results depicted in Figure 8a illustrate the performance of these channel subsets using all frequency bands on the combined dataset. The blue color represents the performance of five distinct channel sets, while the red color indicates the performance difference between these subsets and the full 32-channel configurations. The model performance was degraded as the number of electrodes decreased. Moreover, the channels from the CP region exhibited the best performance, exceeding 90%. Figure 8b shows the results when only the gamma band was used; though there was a slight decrease in the performance, the gamma band played key role in identification, as identified in the previous section.



**Figure 7.** Five different channel configurations, each highlighting different regions of the scalp: (a) frontal (F), (b) central and parietal (CP), (c) temporal (T), (d) occipital and parietal (OP), (e) frontal and parietal (FP).



**Figure 8.** Performance of the proposed method among five different sets of channels. (a) All frequency bands, (b) gamma band, where CRR (5) denotes the performance of the five distinct channel sets, while CRR (5)–CRR (32) indicate the performance differences among the five channel subsets and the 32 channels.

#### 5.6. Comparison with the State of the Art

To demonstrate the effectiveness of the proposed method, we compared its performance with that of state-of-the-art EEG-based deep learning biometric techniques on public domain datasets, including DEAP, PhysioNet, UCI, and a combined dataset. It is to be noted that many state-of-the-art methods also used DEAP and PhysioNet for evaluation; the comparison of these datasets is given in Table 4.

**Table 4.** Comparison with state-of-the-art EEG-based biometric systems according to the number of subjects (# Sub.), trial length (TL), number of channels (# Chan.), network (NW), Euclidean distance (L2), and Manhattan distance (L1).

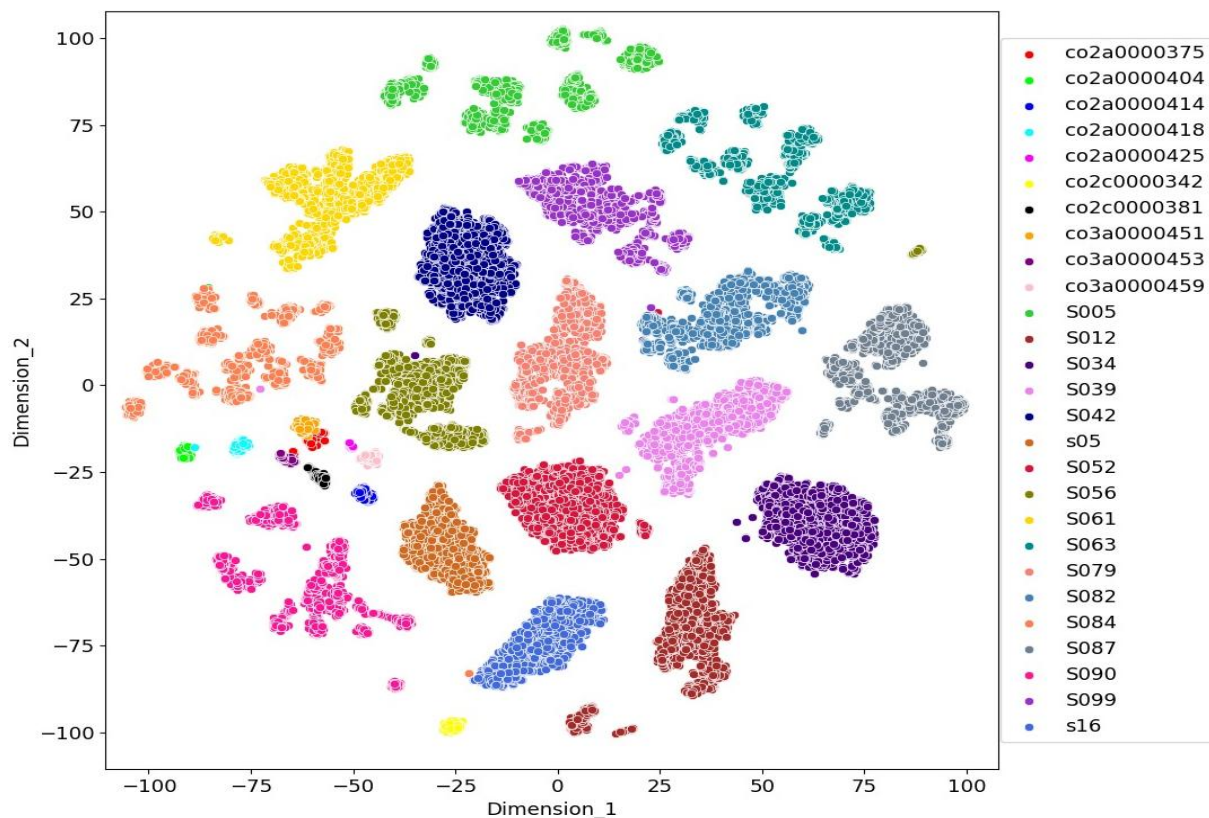
	Dataset	# Sub	Method	# Ch	TL (sec.)	CRR (%)	EER (%)	Parameters
Sun et al. [13]—2019	PhysioNet	109	CNN, LSTM	16	1	99.58	0.41	505,281,566
Wilaiprasitporn et al. [14]—2019	DEAP	32	CNN, LSTM CNN, GRU	5	10	>99	-	324,032 496,384
Jin et al. [24]—2020	MTED	20	CTNN	7	1	99	0.1	4600
Bidgoly et al. [26] 2022	PhysioNet	109	CNN, Cosine	3	1	98.04	1.96	NA
Alsumari et al. [27]—2023	PhysioNet	109	CNN, L1	2	5	99.05	0.187	74,071
Fallahi et al. [28]—2023	ERP CORE	40	Siamese NW, L2	30	0.10	95.63	1.37	NA
	Brain Invaders	41		32		99.92	0.14	
Proposed approach	DEAP	32	GCT-EEGNET, Cosine	32	1	100.00	0.0004	35,900
	PhysioNet	109				98.90	0.0043	45,000
	UCI	122				99.25	0.0009	62,100
	Combined	263				99.23	0.0014	62,800

Most of the state-of-the-art techniques were trained and evaluated on a single dataset involving a small number of tasks, limiting their performance evaluation to specific scenarios and often involving smaller groups of subjects. For instance, Sun et al. [13] evaluated their CNN–LSTM model on the PhysioNet dataset (109 subjects, 16 channels), achieving a high CRR of 99.58%. However, their model incorporated LSTM layers, increasing its complexity to over 505 million parameters, which raised the risk of overfitting, especially

when trained on a dataset with only 109 subjects. The large number of parameters also makes the model computationally expensive and difficult to deploy in real-world systems, unlike the proposed method, which uses only 62,800 parameters while maintaining competitive performance. Similarly, Bidgoly et al. [26] utilized PhysioNet with three channels, achieving a CRR of 98.04% by stacking CNN layers. However, this method uses only three channels, thus it lacks the spatial information of EEG trials, leading to a relatively higher EER of 1.96%. In addition, Alsumari et al. [27] employed the PhysioNet dataset with only two channels, achieving a CRR of 99.05% but at the cost of higher error rates (EER of 0.187%). Their model's simplicity raises concerns about its robustness against diverse brain conditions. In contrast, the proposed model captures richer spatial-temporal information while maintaining low complexity and achieving a much lower EER of 0.0043% on the same dataset. Wilaiprasitporn et al. [14] employed the DEAP dataset (32 subjects, 5 EEG channels) with CNN-LSTM and CNN-GRU networks, yielding a CRR of more than 99%, but these models were tested on a relatively small number of subjects and brain conditions. This limitation of dataset size could affect the generalization of the model to larger populations. Similarly, the proposed model was applied to the DEAP dataset, gaining 100% with a small number of parameters of 35,900. Jin et al. [24] used the MTED dataset (20 subjects, 7 channels), which resulted in a CRR of 99% and an EER of 0.1%. Although these metrics are impressive, the small dataset size (20 subjects) raises concerns about the model's applicability to broader real-world conditions. In contrast, the proposed method demonstrated a much wider generalization by achieving an EER of 0.0014% on a dataset of 263 subjects in diverse brain states. Fallahi et al. [28] used the ERP CORE and Brain Invaders datasets with 40 and 41 subjects, respectively, achieving a CRR of 99.92%. However, their method relies on a Siamese network, which, while effective for specific tasks, introduces a relatively higher EER of 1.37%. Moreover, these datasets focus on specific cognitive tasks, limiting their applicability across broader EEG conditions. In contrast, the proposed method was designed to perform well across multiple brain states and cognitive conditions, as evidenced by its consistently low EER on the DEAP, PhysioNet, UCI, and combined datasets. The method was tested on a large combined dataset created from DEAP, PhysioNet, and UCI with a larger number of subjects, which helped to validate its broader generalization. Despite using a low-complexity architecture with fewer parameters, the method achieved competitive results. Specifically, it attained a CRR of 99.23% and an EER of 0.0014% across diverse brain states and short temporal intervals of one second. This indicated that the proposed model can handle the variability in real-world EEG-based biometric input more effectively than more complex models that are tuned for specific tasks or datasets. In conclusion, although the datasets vary in their characteristics, the proposed method offers a balanced solution with good accuracy, lower complexity, and greater flexibility across different EEG datasets. This highlights its practicality for real-world EEG-based biometric systems, particularly in scenarios that require adaptability across diverse brain conditions and subjects.

### 5.7. Visualization of the Features Learned by the Model from EEG Segments

To verify the effectiveness of the model, we employed the t-distributed stochastic neighbor embedding (t-SNE) [54] method to visualize the learned features in lower-dimensional 2D space from the GAP layer. This technique helped us evaluate whether the model had effectively learned features that distinguished individuals. Figure 9 shows the results for the combined dataset, with each color representing a different subject. The visualization shows the GAP layer's remarkable ability to classify the testing subjects into distinct groups. Although most subjects were well separated, there were a few outliers (i.e., S039, S034, and S099), which appeared to be incorrectly grouped with other subjects. Overall, this visualization indicated that our approach effectively extracted distinctive features from EEG data for each individual, achieving this with just two layers.



**Figure 9.** The t-SNE visualization for high-dimensional features of the GAP layer.

### 5.8. Discussion

This study presents an EEG-based biometric system based on GCT-EEGNET with a large number of individuals (263) and diverse mental states. While several models in [11–14] achieved high performance by integrating CNN and RNN layers to exploit both spatial and temporal features, their robustness and generalization to a large number of subjects are questionable due to increased complexity. To address this problem, our approach utilized depth-wise separable convolution layers within a CNN architecture. This design efficiently captured both spatial and temporal features while significantly reducing parameter complexity to just 62,764 parameters. This reduction enhanced the model's efficiency and generalization, even with a larger number of subjects. An ablation study on the model hyperparameter choices was discussed in Section 5.1.

Additionally, the method automatically selected optimal frequency bands through the analysis of the GCT layer attention scores, reducing the need for costly experiments (see Section 5.4). The results showed that the gamma and beta bands were the most significant frequency bands, which was consistent with the prior findings in [24,27,55,56]. This indicated that distinctive human features prevail in higher-frequency bands. Channel reduction simplified the system's equipment and applicability. We observed a decline in the model's performance when employing fewer channels compared with utilizing 32 channels. This decline may be attributed to the correlation between channels of an EEG segment. Figure 9 confirms the model's ability to discriminate against unseen subjects. Overall, the system has potential benefits for individuals with disabilities and security concerns.

The proposed EEG-based biometric system shows promising results for recognizing individuals with diverse brain states. However, several limitations need to be addressed for real-world adoption. Reducing the number of EEG channels decreases its performance, complicating practical use and requiring simpler setups. In addition, the limited availability of large, multi-session datasets spanning long time intervals consisting of a large number of subjects may affect the system's ability to generalize across different brain conditions. While the model performs well under diverse mental states, EEG signals exhibit high



variability across sessions, even for the same individual. This variability could affect the system's long-term reliability in real-world applications, where EEG data may be collected over weeks or months. Addressing these issues will be key for real-world adoption, with future work focusing on improving robustness and acceptance by considering channel reduction, handling cross-session variability, and reducing computational requirements for broader applicability.

## 6. Conclusions

This study introduced an efficient, lightweight GCT-EEGNet model for EEG-based biometric recognition by leveraging attention mechanisms and advanced convolutional layers. Our model captures both temporal and spatial features from EEG signals, utilizing diverse cognitive states. The results demonstrated the model's effectiveness, achieving a high CRR of 99.23 and a low EER of 0.0014 with a short one-second temporal window while utilizing 32 electrodes on the combined dataset. The system required only a short one-second temporal window for identification and verification. The integrated GCT layer emphasized the significance of higher-frequency bands, particularly the beta and gamma bands, for individual distinction. A depth-wise separable convolution layer was employed to avoid excessive growth in the number of trainable parameters as the number of subjects increased. Furthermore, comparisons with state-of-the-art methods showed GCT-EEGNet's ability to balance high performance with minimal computational complexity, making it a strong candidate for scalable EEG-based biometric recognition. Future research could explore alternative attention mechanisms for automated channel selection and further investigate the system performance on multi-session datasets to enhance the system's real-world applicability and long-term usability.

**Author Contributions:** Conceptualization, L.A. and M.H.; Data curation, L.A.; Formal analysis, L.A. and M.H.; Funding acquisition, M.H.; Methodology, L.A. and M.H.; Project administration, M.H.; Resources, M.H.; Software, L.A.; Supervision, M.H.; Validation, L.A.; Visualization, L.A.; Writing—original draft, L.A.; Writing—review and editing, L.A. and M.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported under the Researchers Supporting Project, number (RSP2024R109), King Saud University, Riyadh, Saudi Arabia.

**Data Availability Statement:** Public-domain datasets were used for the experiments. The DEAP dataset is available at <https://www.eecs.qmul.ac.uk/mmv/datasets/deap/download.html> (accessed on 16 December 2023). The PhysioNet dataset is available at <https://physionet.org/content/eegmmidb/1.0.0/> (accessed on 16 December 2023). The EEG UCI dataset is available at <https://archive.ics.uci.edu/dataset/121/eeg+database> (accessed on 16 December 2023).

**Conflicts of Interest:** The authors declare no competing interests.

## References

1. Zhang, D.D. *Automated Biometrics: Technologies and Systems*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7.
2. Jain, A.K.; Ross, A.; Prabhakar, S. An Introduction to Biometric Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 4–20. [CrossRef]
3. Poulos, M.; Rangoussi, M.; Chrissikopoulos, V.; Evangelou, A. Parametric Person Identification from the EEG Using Computational Geometry. In Proceedings of the ICECS'99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 99EX357), Paphos, Cyprus, 5–8 September 1999; pp. 1005–1008.
4. Gui, Q.; Ruiz-Blondet, M.V.; Laszlo, S.; Jin, Z. A Survey on Brain Biometrics. *ACM Comput. Surv.* **2019**, *51*, 1–38. [CrossRef]
5. Van Dis, H.; Corner, M.; Dapper, R.; Hanewald, G.; Kok, H. Individual Differences in the Human Electroencephalogram during Quiet Wakefulness. *Electroencephalogr. Clin. Neurophysiol.* **1979**, *47*, 87–94. [CrossRef] [PubMed]
6. Zhang, X.; Yao, L.; Wang, X.; Zhang, W.; Zhang, S.; Liu, Y. Know Your Mind: Adaptive Cognitive Activity Recognition with Reinforced CNN. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; pp. 896–905.
7. Chen, J.X.; Mao, Z.J.; Yao, W.X.; Huang, Y.F. EEG-Based Biometric Identification with Convolutional Neural Network. *Multimed. Tools Appl.* **2019**, *79*, 1–21. [CrossRef]

8. Xu, T.; Wang, H.; Lu, G.; Wan, F.; Deng, M.; Qi, P.; Bezerianos, A.; Guan, C.; Sun, Y. E-Key: An EEG-Based Biometric Authentication and Driving Fatigue Detection System. *IEEE Trans. Affect. Comput.* **2021**, *14*, 864–877. [\[CrossRef\]](#)
9. Maiorana, E. Learning Deep Features for Task-Independent EEG-Based Biometric Verification. *Pattern Recognit. Lett.* **2021**, *143*, 122–129. [\[CrossRef\]](#)
10. Seha, S.N.A.; Hatzinakos, D. Longitudinal Assessment of EEG Biometrics under Auditory Stimulation: A Deep Learning Approach. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; pp. 1386–1390.
11. Das, B.B.; Kumar, P.; Kar, D.; Ram, S.K.; Babu, K.S.; Mohapatra, R.K. A Spatio-Temporal Model for EEG-Based Person Identification. *Multimed. Tools Appl.* **2019**, *78*, 28157–28177. [\[CrossRef\]](#)
12. Jijomon, C.M.; Vinod, A.P. Person-Identification Using Familiar-Name Auditory Evoked Potentials from Frontal EEG Electrodes. *Biomed. Signal Process. Control.* **2021**, *68*, 102739. [\[CrossRef\]](#)
13. Sun, Y.; Lo, F.P.-W.; Lo, B. EEG-Based User Identification System Using 1D-Convolutional Long Short-Term Memory Neural Networks. *Expert Syst. Appl.* **2019**, *125*, 259–267. [\[CrossRef\]](#)
14. Wilaiprasitporn, T.; Dittthapron, A.; Matchaparn, K.; Tongbuasirilai, T.; Banluesombatkul, N.; Chuangsuwanich, E. Affective EEG-Based Person Identification Using the Deep Learning Approach. *IEEE Trans. Cogn. Dev. Syst.* **2019**, *12*, 486–496. [\[CrossRef\]](#)
15. Yang, S.; Deravi, F. On the Usability of Electroencephalographic Signals for Biometric Recognition: A Survey. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 958–969. [\[CrossRef\]](#)
16. Maiorana, E.; La Rocca, D.; Campisi, P. EEG-Based Biometric Recognition Using EigenBrains. In Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 29 June–5 July 2015; pp. 1–6.
17. Rodrigues, D.; Silva, G.F.; Papa, J.P.; Marana, A.N.; Yang, X.-S. EEG-Based Person Identification through Binary Flower Pollination Algorithm. *Expert Syst. Appl.* **2016**, *62*, 81–90. [\[CrossRef\]](#)
18. Thomas, K.P.; Vinod, A.P. EEG-Based Biometric Authentication Using Gamma Band Power during Rest State. *Circuits Syst. Signal Process.* **2018**, *37*, 277–289. [\[CrossRef\]](#)
19. Jijomon, C.M.; Vinod, A.P. EEG-Based Biometric Identification Using Frequently Occurring Maximum Power Spectral Features. In Proceedings of the 2018 IEEE Applied Signal Processing Conference (ASPCON), Kolkata, India, 7–9 December 2018; pp. 249–252.
20. Nakamura, T.; Goverdovsky, V.; Mandic, D.P. In-Ear EEG Biometrics for Feasible and Readily Collectable Real-World Person Authentication. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 648–661. [\[CrossRef\]](#)
21. Zhang, S.; Sun, L.; Mao, X.; Hu, C.; Liu, P. Review on EEG-Based Authentication Technology. *Comput. Intell. Neurosci.* **2021**, *2021*, 5229576. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Stassen, H.H. Computerized Recognition of Persons by EEG Spectral Patterns. *Electroencephalogr. Clin. Neurophysiol.* **1980**, *49*, 190–194. [\[CrossRef\]](#)
23. Maiorana, E. Deep Learning for EEG-Based Biometric Recognition. *Neurocomputing* **2020**, *410*, 374–386. [\[CrossRef\]](#)
24. Jin, X.; Tang, J.; Kong, X.; Peng, Y.; Cao, J.; Zhao, Q.; Kong, W. CTNN: A Convolutional Tensor-Train Neural Network for Multi-Task Brainprint Recognition. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *29*, 103–112. [\[CrossRef\]](#)
25. Debie, E.; Moustafa, N.; Vasilakos, A. Session Invariant EEG Signatures Using Elicitation Protocol Fusion and Convolutional Neural Network. *IEEE Trans. Dependable Secur. Comput.* **2021**, *9*, 2488–2500. [\[CrossRef\]](#)
26. Bidgoly, A.J.; Bidgoly, H.J.; Arezoumand, Z. Towards a Universal and Privacy Preserving EEG-Based Authentication System. *Sci. Rep.* **2022**, *12*, 1–12. [\[CrossRef\]](#)
27. Alsumari, W.; Hussain, M.; Alshehri, L.; Aboalsamh, H.A. EEG-Based Person Identification and Authentication Using Deep Convolutional Neural Network. *Axioms* **2023**, *12*, 74. [\[CrossRef\]](#)
28. Fallahi, M.; Strufe, T.; Arias-Cabarcos, P. BrainNet: Improving Brainwave-Based Biometric Recognition with Siamese Networks. In Proceedings of the 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom), Atlanta, GA, USA, 13–17 March 2023; pp. 53–60.
29. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain–Computer Interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Fraschini, M.; Hillebrand, A.; Demuru, M.; Didaci, L.; Marcialis, G.L. An EEG-Based Biometric System Using Eigenvector Centrality in Resting State Brain Networks. *IEEE Signal Process. Lett.* **2014**, *22*, 666–670. [\[CrossRef\]](#)
31. Kaur, B.; Singh, D.; Roy, P.P. A Novel Framework of EEG-Based User Identification by Analyzing Music-Listening Behavior. *Multimed. Tools Appl.* **2017**, *76*, 25581–25602. [\[CrossRef\]](#)
32. Kawabata, N. A Nonstationary Analysis of the Electroencephalogram. *IEEE Trans. Biomed. Eng.* **1973**, 444–452. [\[CrossRef\]](#)
33. Kumari, P.; Vaish, A. Brainwave Based User Identification System: A Pilot Study in Robotics Environment. *Robot. Auton. Syst.* **2015**, *65*, 15–23. [\[CrossRef\]](#)
34. Ting, W.; Guo-Zheng, Y.; Bang-Hua, Y.; Hong, S. EEG Feature Extraction Based on Wavelet Packet Decomposition for Brain Computer Interface. *Measurement* **2008**, *41*, 618–625. [\[CrossRef\]](#)
35. Yang, Z.; Zhu, L.; Wu, Y.; Yang, Y. Gated Channel Transformation for Visual Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13 June 2020; pp. 11794–11803.
36. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.

37. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21 July 2017; pp. 1251–1258.
38. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [[CrossRef](#)]
39. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (Gelu). *arXiv* **2016**, arXiv:1606.08415. [[CrossRef](#)]
40. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929. [[CrossRef](#)]
41. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
42. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2017**, arXiv:1711.05101. [[CrossRef](#)]
43. Yao, Y.; Rosasco, L.; Caponnetto, A. On Early Stopping in Gradient Descent Learning. *Constr. Approx.* **2007**, *26*, 289–315. [[CrossRef](#)]
44. Lin, C.; Kumar, A. A CNN-Based Framework for Comparison of Contactless to Contact-Based Fingerprints. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 662–676. [[CrossRef](#)]
45. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [[CrossRef](#)]
46. 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. [[CrossRef](#)]
47. Snodgrass, J.G.; Vanderwart, M. A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *J. Exp. Psychol. Hum. Learn. Mem.* **1980**, *6*, 174. [[CrossRef](#)]
48. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A Convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18 June 2022; pp. 11976–11986.
49. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training Data-Efficient Image Transformers & Distillation through Attention. In Proceedings of the International Conference on Machine Learning, Online, 18–24 July 2021; pp. 10347–10357.
50. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11 October 2021; pp. 10012–10022.
51. Agarap, A.F. Deep Learning Using Rectified Linear Units (Relu). *arXiv* **2018**, arXiv:1803.08375. [[CrossRef](#)]
52. Elfving, S.; Uchibe, E.; Doya, K. Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning. *Neural Netw.* **2018**, *107*, 3–11. [[CrossRef](#)]
53. Cui, J.; Yuan, L.; Wang, Z.; Li, R.; Jiang, T. Towards Best Practice of Interpreting Deep Learning Models for EEG-Based Brain Computer Interfaces. *arXiv* **2022**, arXiv:2202.06948. [[CrossRef](#)] [[PubMed](#)]
54. Van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
55. Wang, M.; El-Fiqi, H.; Hu, J.; Abbass, H.A. Convolutional Neural Networks Using Dynamic Functional Connectivity for EEG-Based Person Identification in Diverse Human States. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 3259–3272. [[CrossRef](#)]
56. Fraschini, M.; Pani, S.M.; Didaci, L.; Marcialis, G.L. Robustness of Functional Connectivity Metrics for EEG-Based Personal Identification over Task-Induced Intra-Class and Inter-Class Variations. *Pattern Recognit. Lett.* **2019**, *125*, 49–54. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.