SignEEG v1.0 : Multimodal Electroencephalography and Signature Database for Biometric Systems

Ashish Ranjan Mishra[1,2,*], Rakesh Kumar[2], Vibha Gupta[3], Sameer Prabhu[3], Richa Upadhyay[3], Prakash Chandra Chhipa[3], Sumit Rakesh[3], Hamam Mokayed[3], Marcus Liwicki[3], Foteini Simistira Liwicki[3], and Rajkumar Saini[3]

[1]Rajkiya Engineering College Sonbhadra, UP, India
[2]Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India
[3]Luleå University of Technology, Department of Computer Science, Electrical and Space Engineering, Embedded Intelligent Systems LAB, Sweden

ABSTRACT

Noninvasive electroencephalography (EEG) is a method for measuring electrical brain activity from the surface of the scalp. Recent developments in artificial intelligence accelerate the automatic recognition of brain patterns, allowing more reliable and increasingly faster Brain-Computer interfaces, including biometric applications. Biometric research is also focusing on multimodal systems using EEG along with other modalities. This paper presents a new multimodal *SignEEG v1.0* dataset based on EEG and hand-drawn signatures from 70 subjects. EEG signals and hand-drawn signatures have been collected with *Emotiv Insight* and *Wacom One* sensors, respectively. The multimodal data consists of three paradigms with increasing brain functioning: (i) visualizing a signature image, (ii) doing a signature in mind, and (iii) physically drawing a signature. Extensive experiments have been done in order to provide a solid baseline with machine learning classifiers. We release the raw, pre-processed data and easy-to-follow implementation details.

1 Introduction

This study delves into a relatively unexplored realm of biometrics, where the fusion of physiological signals and behavioral aspects is investigated to uncover their potential for developing efficient biometric systems. Specifically, the study integrates brain signals (EEG) and hand-drawn signatures within a multimodal biometric framework. In pursuit of the study's goal, two primary objectives have been defined: user identification and verification. To evaluate these objectives, a series of four tasks were designed (Section 2.6).

Essentially, the study proposes an intriguing future plan to expand the investigation by examining the long-term stability of EEG signals. This will be accomplished by progressively accumulating a dataset from the same group of subjects over the course of multiple years. By adopting this longitudinal approach, the research seeks to delve into the temporal variations inherent in EEG signals, allowing for a comprehensive understanding of their dynamic nature over time. This exploration of EEG dynamics holds promise for enhancing the modeling and interpretation of EEG-based biometric systems.

User identification refers to establishing unique credentials or attributes within a system, while user verification confirms the claimed identity through interaction with the system. In terms of biometrics, user identification involves establishing the identity of an individual from a larger dataset, while user verification focuses on confirming the claimed identity of an individual through a one-to-one match. Various physical traits, such as retina, iris, sclera, ear, hand, palm geometry, fingerprints, and face, along with behavioral characteristics like keystroke pattern, voice, signature, and gait, have been utilized in the design of biometric systems[1]. Signature-based biometric systems[2,3] are preferred due to their convenience, non-invasiveness, cost-effectiveness, wide adoption, cancelable nature, and legal recognition. However, forgery of signatures is a recognized limitation in signature-based systems, as skilled individuals can imitate someone else's signature with varying levels of accuracy. EEG-based biometric systems provide a high level of security against forgery due to the inherent difficulty of replicating an individual's unique brain activity patterns[4]. Unlike signatures or physical biometrics, EEG signals are challenging to forge, making them a reliable and robust method for identity verification.

While unique to individuals, signatures can still be forged, necessitating a more robust authentication system. By combining EEG signals with signature traits, the system aims to enhance identification and verification performance, ensuring information confidentiality and minimizing the possibility of impersonation. The interconnected nature of hand-drawn strokes and EEG signals are leveraged in this study for user identification and verification purposes.

The existing research[5,6] on integrating hand-drawn signatures and EEG signals is scarce and constrained to a few subjects.

Table 1. Related Work: EEG and Signatures

Paper	Participants (genuine + forged)	Total Samples	Mental Imagery	Motor Imagery	Physical Execution	Availability
[5]	33 + 25	1650	-	-	✓	-
[6]	70 + 70	1400	-	-	✓	-
SignEEG v1.0	70 + 70	2043 × 3	✓	✓	✓	✓

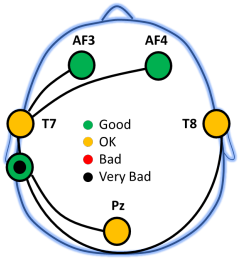


Figure 1. Electrode placement according to the 10-20 international system[8]. The green and orange colors indicate the good and moderate (OK) quality of the signal, respectively. While red and black colors indicate bad and very bad signal quality, respectively.

Furthermore, the unavailability of publicly accessible data from these studies hinders researchers from further investigating this direction. Some of the EEG and signature-based multimodal works are summarized in Table 1.

SignEEG v1.0 consists of three paradigms, unlike others where only one paradigm is used. The proposed dataset gives the opportunity to develop multimodal biometric systems as it covers different paradigms and hand-drawn signatures; paradigms are discussed in Section 2.4. By incorporating these three paradigms, a multimodal approach to user authentication can be established, combining the strengths of mental imagery, motor imagery, and physical execution. This comprehensive approach enhances the security and reliability of the biometric system by capturing multiple facets of the individual's unique signature-related brain activity and behavior. To conclude, our key contributions can be summarized as:

- A multimodal biometric dataset (SignEEG v1.0) with EEG and hand-drawn signatures from 70 healthy participants collected over several months. Along with the raw dataset, we provide preprocessed data and code for preprocessing, feature extraction, and machine learning.
- A benchmark for biometrics from EEG signals consisting of three paradigms (tasks) with increasing difficulty.
- A comprehensive experiment conducted to establish the initial performance on the proposed benchmark.

The rest of the paper is organized as follows. Section 2 describes the data acquisition process, including details of participants, tools used, recording setup, experimental paradigms, data acquisition protocol, and objectives, followed by data pre-processing. Sections 3, 4, and 5 discuss the physical structure of the data, results, and conclusion, respectively.

2 Data Acquisition

In this section, we describe in detail the data acquisition process, including devices used, participants (subjects), the experimental paradigm, recording setup, and acquisition protocol.

2.1 Participants

The dataset development involved 70 enrolled participants (anonymized). All 70 participants were healthy individuals, comprising five females and sixty-five males, aged between seventeen and thirty-six years, and without any neurological disorders or brain-related medications/surgery. None of the participants had prior experience with EEG. Participants were instructed not to consume alcohol or tobacco for two days prior to the data collection. After the data collection, the participants independently rated their satisfaction, boringness, horribleness, and calmness regarding their experience, and the average scores (out of 10) were 9.56, 1.43, 1.17, and 9.50, respectively. Written informed consent was obtained from all participants before they participated in the experiment; a sample consent form may be provided as supplementary material if required.

2.2 Device and software Details

The **Emotiv Insight headset**[7] is a wireless EEG device designed for research and personal use. It has five channels *(AF3, AF4, T7, T8, Pz)* for recording EEG signals and two references *(CMS, DRL)*, with a sampling rate of 128 Hz and a resolution of 14 bits. The electrode placement according to the 10-20 international system[8] is shown in Figure 1; colors indicate the quality of the electrical signal passing through the sensors and the reference.

The headset is lightweight, comfortable, and easy to set up, making it suitable for BCI (brain-computer interfaces) research. The Emotiv Insight requires an **EmotivPRO** software toolkit[9]; it allows researchers to record, visualize, and analyze EEG data

collected from the headset. The software provides a user-friendly interface for easy data analysis and visualization. Additionally, the toolkit is quite useful in the development of custom applications and experiments using the Emotiv headset.

The **Wacom One Display Pen Tablet**[10] has a 13.3-inch Full HD display with a resolution of 1920 × 1080 pixels. It supports up to 4096 levels of pressure sensitivity and has a battery-free pen with tilt support. The tablet has a viewing angle of 170 degrees and a color gamut of 72% NTSC (National Television System Committee). It has a durable and scratch-resistant surface and comes with an adjustable stand for comfortable use. The tablet also features HDMI and USB Type-C ports for connectivity. Online signatures drawn on the tablet were captured using the Wacom **SignatureScope** software tool[11]. The software captures real-time signature data and provides trajectory information, including coordinates (x and y), timestamp, stroke, pressure, azimuth, altitude of the pen, and other related information. This information is captured at a sampling rate of 200 frames per second.

2.3 Recording Setup

To ensure a controlled environment for data collection, a physical recording studio was set up, which is illustrated in Figure 2b. The studio is designed with acoustic walls to minimize external noise and is maintained at a constant temperature of 20 °C. The subjects are seated comfortably and provided with a monitor screen and a Wacom tablet with a stylus for online signature recording. Audio instructions are provided through a speaker to ensure consistency across all subjects. The dataset curator is responsible for monitoring and recording the EEG and online signature data, while an assistant helps the subjects with the EEG headset placement. Both the curator and assistant maintain a safe distance from the subject to avoid any interference with the data collection process. In addition, the studio is thoroughly cleaned and disinfected to ensure hygiene, and the power supply is checked to ensure a stable recording environment before the start of each recording session.

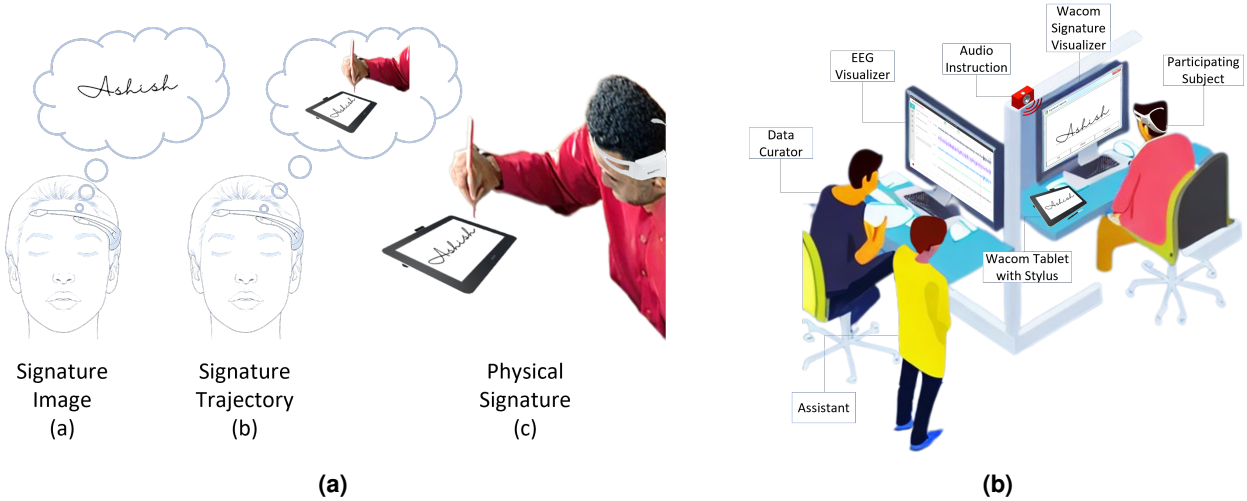


Figure 2. (i) Experimental Paradigms for the multimodal biometric research, (ii) Physical setup of the data collection lab.

2.4 Experimental Paradigms

The multimodal dataset SignEEG v1.0 consists of three paradigms with increasing brain functioning: (i) visualizing a signature image, (ii) doing a signature in mind, and (iii) physically drawing a signature. These three paradigms are of interest to neuroscience, BCI, and machine learning research. The paradigms are discussed below (also depicted in Figure 2a).

1. Closed eyes, thinking of the signature image: This paradigm involves mental imagery[12–14] or the creation of a sensory experience in the absence of external stimuli. When a person closes their eyes and thinks of an object, they are engaging in mental imagery. This is the simplest of the three paradigms, as it requires minimal motor activity and cognitive effort as compared to others. The main focus is on generating a visual image in the mind's eye.

2. Closed eyes, drawing the signature mentally: This paradigm also involves mental imagery[12,13]; however, it requires more complex cognitive processes such as spatial reasoning and motor planning. The person needs to create a mental representation of the object and then simulate the motor movements needed to draw it. This can activate additional brain regions involved in motor planning and execution. This paradigm may also be seen as a motor imagery task from the conventional definition of motor imagery[14]. The detailed discussion can be found in[14,15].

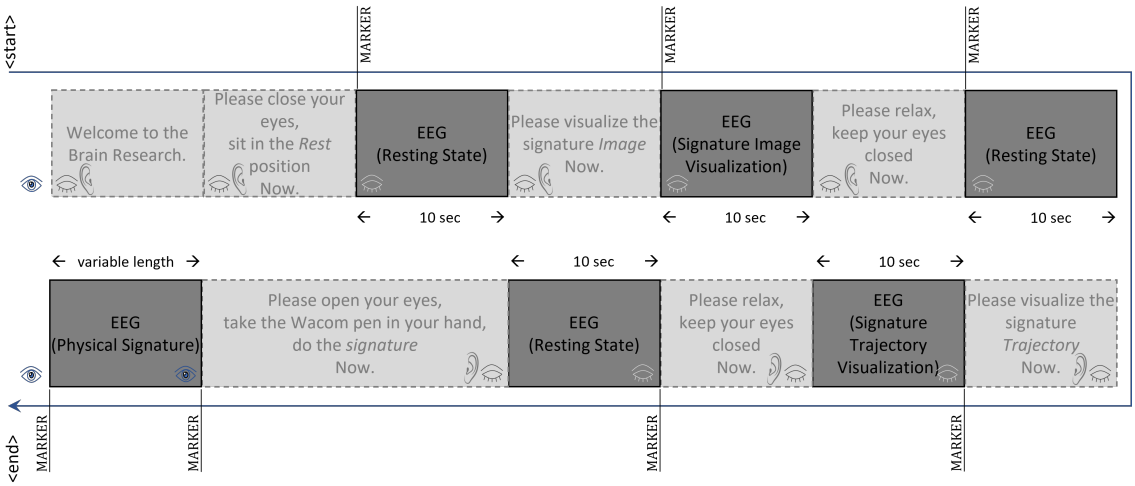


Figure 3. Acquisition protocol for the data collection. Markers (value=4) are inserted into recordings to fetch the EEG signals related to 'signature' or 'rest' activities.

3. Open eyes, drawing an object physically: This paradigm involves actual motor execution[16,17] and visual perception[18]. The person needs to visually perceive the object and then use motor planning and execution to draw it. This paradigm can activate additional brain regions involved in sensorimotor integration, feedback control, and visual perception. This paradigm requires coordination between the motor and visual systems and engages higher-level cognitive processes such as planning, attention, and working memory.

Different regions of the brain may activate differently in different paradigms; therefore, these paradigms can be analyzed individually or combined.

2.5 Acquisition Protocol

The acquisition protocol is designed based on an increasing workload on the brain. The workload on the brain increases from mental imagery alone to mental imagery with motor simulation and finally to the combined demands of motor execution, visual perception, and cognitive processes. The protocol employed for recording both EEG and online signature data is illustrated in Figure 3. Initially, subjects sit in a calm/resting state; the subject's head is fitted with the EEG headset, and electrodes are placed on the scalp according to the 10-20 system[8]. After the channel and EEG signal quality indicators reach a stable state (indicated by a green color in the Emotiv Pro software), the EEG data acquisition is initiated following the procedure outlined in Figure 3.

During the EEG data collection, except for the period when the subjects are providing their physical signatures, their eyes are closed. The subjects are instructed via audio commands throughout the data collection process, with the audio instructions playing simultaneously with the EEG recording. At the start of the data collection, the audio instructs the subjects to 'Welcome to the Brain Research. Please close your eyes, sit in the Rest position Now'. Following this audio instruction, a marker with a value of 4 is inserted into the EEG recording, signifying the beginning of the rest period. During this rest period, which lasts for 10 seconds, the subjects are instructed to maintain a resting position with their eyes closed.

Subsequently, the audio command instructs the subjects to 'Please visualize the signature Image Now,' and a second marker is inserted into the EEG recording. During the following 10 seconds, the subjects are instructed to visualize the image of their signature in their minds' eye, and if the image disappears, they are to recall it in their minds again. After the 10 seconds have elapsed, the audio instruction advises the subjects to 'Please relax, keep your eyes closed Now.' Following this instruction, a third marker is inserted into the EEG recording, and the subjects are instructed to remain in a resting position for an additional 10 seconds.

Subsequently, the audio instruction directs the subjects to 'Please visualize the signature Trajectory Now.' For the signature trajectory task, the subjects are instructed beforehand to simulate the act of signing on paper in their minds. They are required to envision the process of producing the signature from start to finish as if they were signing on a physical sheet of paper. Following this instruction, a fourth marker is inserted into the EEG recording, and the subjects simulate the signature as many times as they can within the next 10 seconds. After the 10 seconds have passed, the audio command advises the subjects to 'Please relax, keep your eyes closed Now.' Following this instruction, a fifth marker is inserted into the EEG recording, and the subjects are required to remain in a resting position for a further 10 seconds.

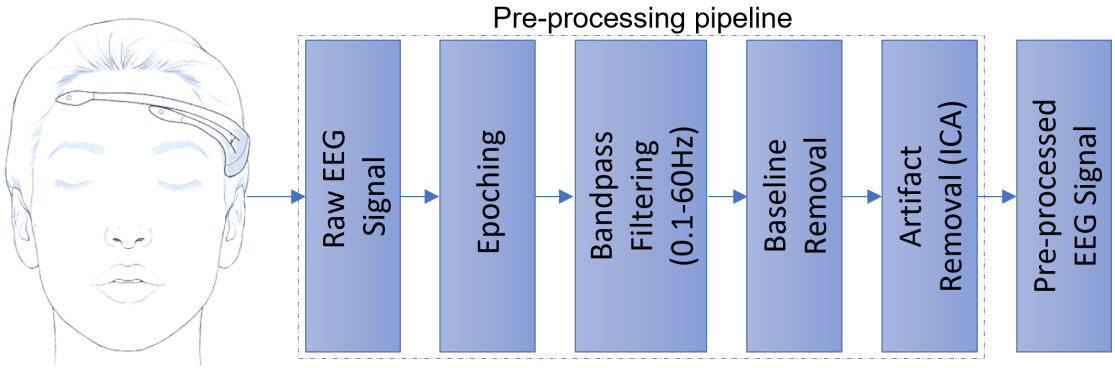


Figure 4. EEG data pre-processing pipeline.

Following the resting period, the audio instruction prompts the subjects to 'Please open your eyes, take the Wacom pen in your hand, do the signature Now.' After the audio instruction, the subjects open their eyes and pick up the Wacom stylus in their hand. Once they are ready to perform the signature on the Wacom tablet, a sixth marker is inserted into the EEG recording. The subjects proceed to sign on the Wacom tablet, and the final marker (7th time) is inserted into the EEG recording, indicating that the signature process has been completed. At this point, the dataset curator saves both the EEG and online signature data into permanent storage.

All markers inserted into the EEG recordings, except for the last one (7th marker), indicate the beginning of the EEG activities, whether it is the resting period or related to the signature tasks. It is noteworthy that all markers in the EEG recordings are represented by the digit 4. Additionally, in an EEG recording, each of the EEG activities, except for the one corresponding to the physical signature task, has a duration of 10 seconds. The final EEG activity associated with the physical signature has a duration equivalent to the time spent performing the signature on the Wacom tablet.

2.6 Objectives and Tasks

The complete data acquisition process intends to explore two major objectives: user identification and verification.

User Identification: User identification involves establishing the identity of an individual from a larger dataset. To identify who the subject is from the pool, given the input data.

User Verification: User verification focuses on confirming the claimed identity of an individual through a one-to-one match. Given input data, it aims to verify if the subject is *Genuine* or *Forged*.

Multiple tasks are designed to evaluate the above-mentioned objectives for different experimental paradigms explained in Section 2.4. The tasks are as follows.

Task 1: This task aims to develop a biometric system based on paradigm one that corresponds to the EEG signals of mental imagery of the signatures.

Task 2: This task aims to develop a biometric system based on paradigm two that corresponds to the EEG signals of motor imagery of the signatures.

Task 3: This task aims to develop a biometric system based on paradigm three that corresponds to the EEG signals of the physical execution of the signatures.

Task 4: This task aims to develop a biometric system combining paradigm three and Wacom signatures.

2.7 Data Pre-processing and Validation

The pre-processing of the EEG data was done with EEGLAB[19] (a MATLAB-based toolbox) following the EEG data pre-processing pipeline as depicted in Figure 4. Epoching refers to segmenting the continuous EEG data into smaller segments for further analysis. The EEG data obtained using the acquisition protocol discussed in Section 2.5 underwent further processing before being subjected to signal processing. For each subject, there were instances of genuine and forgery attempts (approximately 30 in total) recorded. EEG activities corresponding to rest and signatures were extracted by identifying inserted markers (marker value = 4). Each of the EEG activities related to the first six markers had a duration of 10 seconds, equivalent to 1280 frames starting from the marker's position, given that the EEG headset had a sampling rate of 128 frames per second. The duration of the last EEG activity varied depending on when the subject completed the signature on the Wacom tablet.

It is recommended that the epoch length be consistent for a specific type of EEG activity. Therefore, the duration of the last EEG activity, which corresponds to physical signatures, is constant in all instances (genuine and forged) for a given subject. The instances corresponding to the last EEG activity for each subject were interpolated[20] to match the length of the longest

genuine instance for that subject. This means that all instances (genuine and forged) for a particular subject have the same length (epoch length = $5 \times 10 \times 128 + k$), where k is the duration of the last EEG activity. The value of k varies among subjects. Once the EEG data has been segmented using epoching, a bandpass filter is applied to the data, with a frequency range of 0.1-60 Hz, followed by the baseline removal. The purpose of baseline removal is to eliminate non-neural activity in the EEG signal by subtracting the mean of the signal from its data points. This process results in a more refined signal that reflects the neural activity under investigation. Following epoching and filtering, Independent Component Analysis (ICA) is utilized to remove artifacts resulting from various sources such as eye movements, muscle activities, and environmental noise. By applying ICA, the components associated with the artifacts as well as the brain activity, are identified. The components with a substantial amount of artifacts are eliminated, while the components linked with the brain activity are kept to produce the final pre-processed signals.

3 SignEEG v1.0 Physical Structure

Following Section 2, 2047 ($\times$3, *edf, csv, mat*) instances have been recorded and pre-processed. Each subject is given a unique number (e.g., *200894*). The directory structure '*200894*\ *Genuine*' and '*200894*\ *Forged*' consists of instances of genuine and forgery attempts for the subject *200894*, respectively. The file naming convention is as follows.

EDF: SubjectID_{'*Genuine**Forged*'}_SignerID_*d*.edf. This is a recorded EEG datum in standard EDF format. The digit *d* represents the instance number. A signer (*SignerID*) performs the EEG and signatures for the subject (*SubjectID*). *SubjectID* and *SignerID* are the same for a *Genuine* attempt, and they differ for a forgery (*Forged*) attempt. The EDF file contains the EEG signals corresponding to the channels in the Emotiv Insight headset, along with other EEG-related information.

CSV: SubjectID_ '*Genuine**Forged*'}_SignerID_*d*.csv. This is a recorded Wacom signature datum in standard CSV (comma-separated value) format. It follows the same naming convention as EDF except for the extension. It consists of the pen data (Index, Stroke, Btn, coordinate X, coordinate Y, Timestamp (T), Pressure, Azimuth, Altitude) per frame.

MAT: SubjectID_{'*Genuine**Forged*'}_SignerID_*d*_epochLength_diff.mat. This is a MATLAB structure array (unordered *struct*, with MATLAB variable name *subject*) consisting of data related to EEG and Wacom signature after pre-processing. It follows the same naming convention as EDF; however, it has additional information in the filename. *epochLength* is the epoch length as discussed in Section 2.5, and *diff* is an averaged absolute difference of EEG signals before and after applying ICA. The fields of the MATLAB structure and their description is given in Table 2.

Table 2. MATLAB structure of pre-processed data

Struct Fields	Description	Example
SubjectID	Subject ID	'001810111230006'
Gen_Forge	*Genuine/Forged*	'Genuine'
SignerID	Signer ID	'001810111230006'
EEG	5 $\times$EpochLength. EEG data after Epoching.	5 $\times$7347
ICA_EEG	5 $\times$EpochLength. EEG data after applying ICA.	5 $\times$7347
EpochLength	Epoch length as per EEG pre-processing.	7347
Diff	Averaged absolute difference between EEG signal before and after applying ICA.	2.954159527
SignWacom	Signature data from the Wacom tab with 8 attributes as in WacomHeader.	1190x8 double
EEGHeader	The name and order of the channels in EEG and ICA_EEG.	AF3, T7, Pz, T8, AF4'
WacomHeader	The name and order of the Wacom signature attributes in SignWacom.	Index, Stroke, Btn, X, Y, Pressure, Azimuth, Altitude'
Note	Explanation of number of Frames in EEG and ICA_EEG.	There are 1280 EEG frames each for Rest1, EEG1 (Image), Rest2, and EEG2 (Trajectory), as well as Rest3. These are followed by ad...' <Preview truncated at 128 characters>
Path1	Path of corresponding raw EEG recording.	.\001810111230006\Genuine\001810111230006_Genuine_001810111230006_1.edf
Path2	Path of corresponding raw Wacom signature recording.	.\001810111230006\Genuine\001810111230006_Genuine_001810111230006_1.csv

Out of 70, the data of 5 subjects is kept aside for robustness analysis in anticipated future research.

4 Baseline Results and Discussion

4.1 Models and features

Table 3 summarizes the performance of several models on two independent objectives, namely identification, and verification, as mentioned in Section 1.

For verification, Table 3 represents the combined accuracy across all the subjects, i.e., combining all the true positives, true negatives, and so on. The models include Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN), Linear Support Vector

Table 3. Baseline results on various machine learning and deep learning models for identification and verification. The accuracy score in mean and standard deviation for five trials of each experiment are shown. This is a performance analysis on the data of Task-4 in Section 2.6, i.e., EEG collected during physically drawing the signature on the Wacom tablet. (GPU: single V100 NVidia)

Identification						
Models	EEG		Wacom	EEG + Wacom		
	Statistical features	Wavelet features	Wacom features	Statistical + Wacom	Wavelet + Wacom	Statistical+Wavelet Wacom
Gaussian NB	93.4 ± 0	90.24 ± 0	21.90 ± 0	96.31 ± 0	96.51 ± 0	96.57 ± 0
KNN	91.82 ± 0	51.98 ± 0	25.33 ± 0	93.14 ± 0	51.98 ± 0	59.10 ± 0
Linear SVM	91.82 ± 0	89.71 ± 0	10.82 ± 0	96.83 ± 0	91.29 ± 0	93.93 ± 0
Random Forest	91.66 ± 0.0043	91.08 ± 0.0045	55.25 ± 0.0125	97.63 ± 0.006	97.68 ± 0.0019	98.94 ± 0.0033
Gradient Boost	92.88 ± 0.0013	90.82 ± 0.002	38.58 ± 0.0112	96.73 ± 0.0013	94.78 ± 0.0011	94.35 ± 0.0027
XGBoost	93.93 ± 0	92.08 ± 0	51.19 ± 0	98.68 ± 0	97.89 ± 0	97.89 ± 0
	EEG		Wacom ((x,y) $\rightarrow$ image)	EEG + Wacom (feature fusion[21])		
EEGNet[22]	4.77 ± 0.8848					
ResNet50[23]	21.48 ± 2.8664		96.59 ± 1.0630			
EEGNet + ResNet50					1.33 ± 0.3280	
ResNet50 + ResNet50					2.37 ± 0.9977	
Verification						
Gaussian NB	67.28 ± 0	63.14 ± 0	92.78 ± 0	92.52 ± 0	92.52 ± 0	91.49 ± 0
KNN	68.04 ± 0	68.04 ± 0	88.91 ± 0	74.48 ± 0	68.04 ± 0	68.29 ± 0
Linear SVM	73.40 ± 0	60.30 ± 0	84.02 ± 0	86.64 ± 0	68.04 ± 0	70.10 ± 0
Random Forest	75.30 ± 0.0078	69.38 ± 0.0034	95.05 ± 0.0022	94.84 ± 0.0036	94.69 ± 0.0029	94.89 ± 0.0056
Gradient Boost	68.50 ± 0.0067	65.30 ± 0.005	93.81 ± 0.0026	93.55 ± 0.0045	93.50 ± 0.0042	93.40 ± 0.0029
XGBoost	68.04 ± 0	63.14 ± 0	91.49 ± 0	92.01 ± 0	91.49 ± 0	92.26 ± 0

Machine (SVM), Random Forest, Gradient Boost, XGBoost, EEGNet, and ResNet50 similar to[24]. These models were evaluated using different sets of features derived from electroencephalogram (EEG) data (statistical and wavelet)[25], Wacom tablet data[26], or a combination of both. Feature extraction is a process in which relevant information is extracted from raw data to represent it in a more compact and meaningful form. Feature extraction refers to the process of identifying and quantifying distinctive patterns or characteristics that are informative for both identification and verification objectives. Statistical measures like energy, variance, skewness, and kurtosis are computed from the EEG signal in the time domain. They provide information about the signal's amplitude, variability, asymmetry, and distribution. The wavelet features are capable of capturing both temporal and spectral characteristics of the EEG signal during signature analysis. Due to the high dimensionality of the wavelet features, statistical information on these features is utilized in these experiments. Based on five features from the Wacom tablet (*X, Y, pressure, azimuth,* and *altitude* as given in Table 2), various features like the mean, standard deviation, and max of writing pressure, speed, and acceleration are computed.

The data and codes are distributed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

4.2 Results and Discussion

For the statistical features derived from EEG, the Gaussian (NB), (KNN), Linear SVM, and random forest models exhibit similar performance with accuracy in the low 90% range. Gradient boost and XGBoost also perform well, with accuracies of 92.88% and 93.93%, respectively. Among these, XGBoost performs the best. While on the wavelet features, the performance of KNN is unsatisfactory. Of all the rest of the models, gradient boost and XGBoost perform relatively better, with an accuracy of around 90%. Compared to EEG signal-based features, the model performs less optimally for features from the Wacom signal. Random forest gives the highest accuracy at 55.25%. The combination of statistical and Wacom features drastically improves the performance of all models. XGBoost shows the highest accuracy at 98.68%, followed closely by Random Forest at 97.63%. All other models also exhibit significant improvement in their accuracies. Similar improvements in the performance can be observed when wavelet features and Wacom features are taken into account. However, KNN does not seem to benefit significantly from this combination, maintaining a low accuracy of 51.98%. The combination of EEG statistical, EEG wavelet, and Wacom features gives most models the highest accuracy, with random forest delivering the best results (98.94%). The only exception is KNN, which only achieves 59.10% accuracy.

In general, it has been noted that the accuracy of KNN is suboptimal compared to alternative models. This may be attributed to the selection of the parameter K, which denotes the number of neighbors to be taken into account during the prediction process. In order to ensure fair evaluation, a consistent value of K (i.e., 5) has been utilized across all features. However, this

may prove inadequate for wavelet and Wacom features, resulting in a decline in performance. On the other hand, XGBoost and random forest excel for most of the experiments, whether statistical or wavelet features. XGBoost offers several advantages that may explain its superior performance in handling various types of features, such as in-built L1 and L2 regularization and cross-validation, tree pruning, capability to handle missing as well as sparse data, and many more. The random forest ensemble learning method shows a strong performance due to its ability to mitigate overfitting through the aggregation of results from multiple decision trees.

For identification, two deep learning models are considered, EEGNet and ResNet50. The EEGNet model, which is specifically designed for EEG data, has been applied to the 5-channel EEG signal. It achieves an accuracy of 4.77%, while when pretrained ResNet50 is used, the performance improves to 21.48%. In order to apply ResNet50 to the Wacom dataset, the X and Y coordinates are utilized to generate an image representation of the signature; this image array is subsequently inputted into the model, resulting in an accuracy rate of 96.59%. The multi-modal experiments involving the fusion of EEG and Wacom features[21] yielded unsatisfactory results for both combinations of EEGNet (for EEG data) and ResNet50 (for Wacom image data), as well as ResNet50 (for EEG data) and ResNet50 (for wacom image data). A plausible explanation for the suboptimal performance of deep learning models, particularly on EEG data, may be attributed to the paucity of available data. The provided data pertaining to each subject is inadequate for acquiring the ability to differentiate among the subjects using deep convolutional models. It should be noted that since these models accept inputs of the same size, the EEG data collected during task-3, which varied in size across subjects, was interpolated to a fixed length of 1280. This length corresponds to the size of the EEG signals recorded during task-1 and task-2, as detailed in section 2.7.

The pattern in the performance across different models and features seen during identification also appears during verification. Random forest and gradient boost both achieve the highest accuracy when using statistical, wavelet, and Wacom features combined, with accuracies of 94.89% and 93.40%, respectively. For verification, deep learning models were not employed because, for training subject-specific models, more EEG data per subject is required.

5 Conclusion and Future Work

In this paper, we propose a multimodal SignEEG v1.0 dataset for biometric systems. The dataset consists of data based on three paradigms and four tasks. The paradigms are related to mental imagery, motor imagery, visual perception, and motor execution. The tasks are defined based on the paradigms. The data collected from 70 participants has been anonymized. The development of a multimodal dataset, particularly with EEG, is a complex and resource-intensive endeavor. The presented database provides a valuable opportunity for comprehensive biometric research, enabling the examination of physiological (EEG) and behavioral (signatures) aspects individually and in combination. By integrating these modalities, the study aims to advance our understanding of biometric systems and their practical applications. Additionally, future investigations will focus on exploring the long-term stability of EEG signals through continuous data collection from the same subjects over an extended duration. This longitudinal approach will offer valuable insights into the temporal dynamics and enduring properties of EEG signals.

The insufficiency of data available for training deep learning models is a primary factor contributing to their suboptimal performance on EEG data. To address this issue, a potential avenue for future research could involve the utilization of advanced deep learning techniques, such as few-shot learning and self-supervised learning, which have demonstrated efficacy in scenarios involving limited data.

References

1. Jain, A., Hong, L. & Pankanti, S. Biometric identification. *Commun. ACM* **43**, 90–98 (2000).
2. Tolosana, R. *et al.* Icdar 2021 competition on on-line signature verification. In *Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV 16*, 723–737 (Springer, 2021).
3. Tolosana, R. *et al.* Svc-ongoing: Signature verification competition. *Pattern Recognit.* **127**, 108609 (2022).
4. Palaniappan, R. Two-stage biometric authentication method using thought activity brain waves. *Int. journal neural systems* **18**, 59–66 (2008).
5. Kumar, P., Saini, R., Kaur, B., Roy, P. P. & Scheme, E. Fusion of neuro-signals and dynamic signatures for person authentication. *Sensors* **19**, 4641 (2019).
6. Saini, R. *et al.* Don't just sign use brain too: A novel multimodal approach for user identification and verification. *Inf. Sci.* **430**, 163–178 (2018).
7. Inc, E. Insight User Manual — emotiv.gitbook.io. https://emotiv.gitbook.io/insight-manual/ (2020). [Accessed 06-Jun-2023].

8. Sharbrough, F. *et al.* American electroencephalographic society guidelines for standard electrode position nomenclature. *Clin. Neurophysiol.* **8**, 200–202 (1991).

9. Inc, E. Insight User Manual — emotiv.gitbook.io. https://emotiv.gitbook.io/emotivpro-v3/ (2023). [Accessed 06-Jun-2023].

10. Wacom. Wacom User Help (DTC133) — 101.wacom.com. http://101.wacom.com/UserHelp/en/TOC/DTC133.html (2020). [Accessed 06-Jun-2023].

11. Wacom. Signature scope – developer support. https://developer-support.wacom.com/hc/en-us/sections/9304704351895-Signature-Scope (2020). (Accessed on 06/06/2023). The tool is available on request.

12. Kosslyn, S. M., Thompson, W. L. & Ganis, G. *The case for mental imagery.* (Oxford University Press, New York, NY, US, 2006).

13. Ganis, G., Thompson, W. L. & Kosslyn, S. M. Brain areas underlying visual mental imagery and visual perception: an fmri study. *Cogn. Brain Res.* **20**, 226–241 (2004).

14. Nanay, B. Mental Imagery (Stanford Encyclopedia of Philosophy) — plato.stanford.edu. https://plato.stanford.edu/entries/mental-imagery/#MentImagVsMotoImag (1997). [Accessed 09-May-2023].

15. Jeannerod, M. The representing brain: Neural correlates of motor intention and imagery. *Behav. Brain sciences* **17**, 187–202 (1994).

16. Frenkel-Toledo, S., Bentin, S., Perry, A., Liebermann, D. G. & Soroker, N. Dynamics of the eeg power in the frequency and spatial domains during observation and execution of manual movements. *Brain research* **1509**, 43–57 (2013).

17. Rimbert, S., Al-Chwa, R., Zaepffel, M. & Bougrain, L. Electroencephalographic modulations during an open-or closed-eyes motor task. *PeerJ* **6**, e4492 (2018).

18. Taylor, P. C. J. & Thut, G. Brain activity underlying visual perception and attention as inferred from tms–eeg: A review. *Brain stimulation* **5**, 124–129 (2012).

19. Delorme, A. & Makeig, S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. neuroscience methods* **134**, 9–21 (2004).

20. Gribbon, K. T. & Bailey, D. G. A novel approach to real-time bilinear interpolation. In *Proceedings. DELTA 2004. Second IEEE international workshop on electronic design, test and applications*, 126–131 (IEEE, 2004).

21. Nguyen, D., Nguyen, K., Sridharan, S., Dean, D. & Fookes, C. Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition. *Comput. Vis. Image Underst.* **174**, 33–42, https://doi.org/10.1016/j.cviu.2018.06.005 (2018).

22. Lawhern, V. J. *et al.* Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *J. Neural Eng.* **15**, 056013, 10.1088/1741-2552/aace8c (2018).

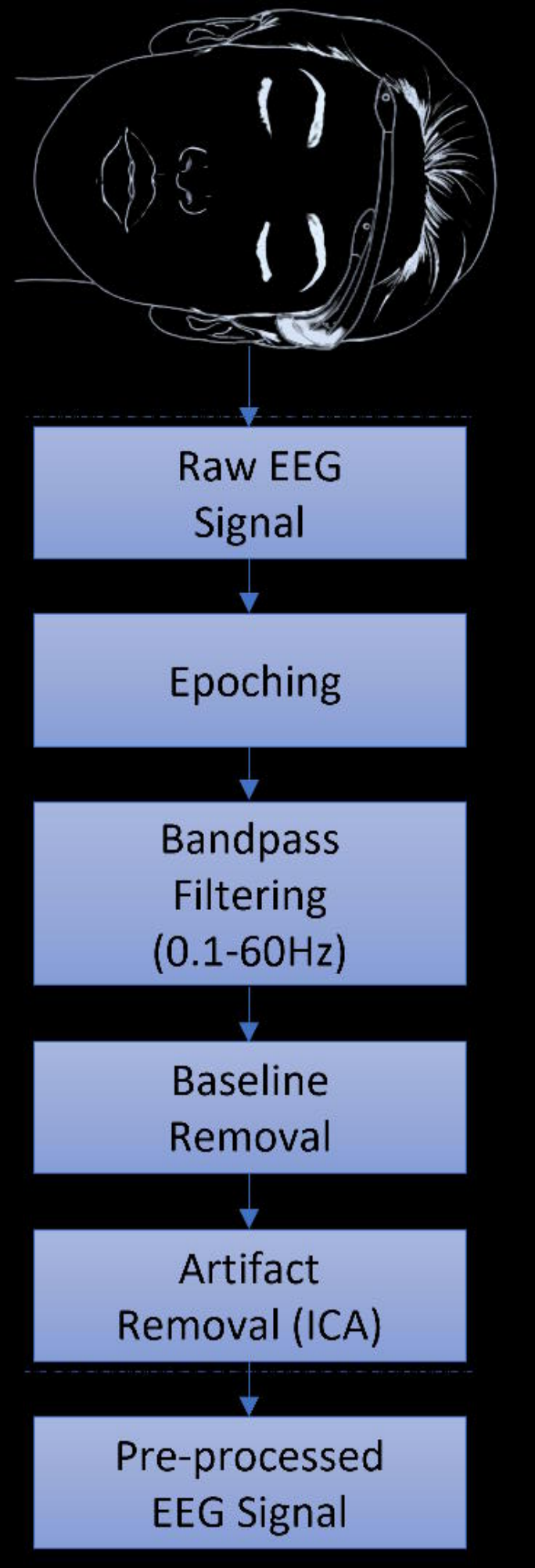
23. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, 10.1109/CVPR.2016.90 (2016).

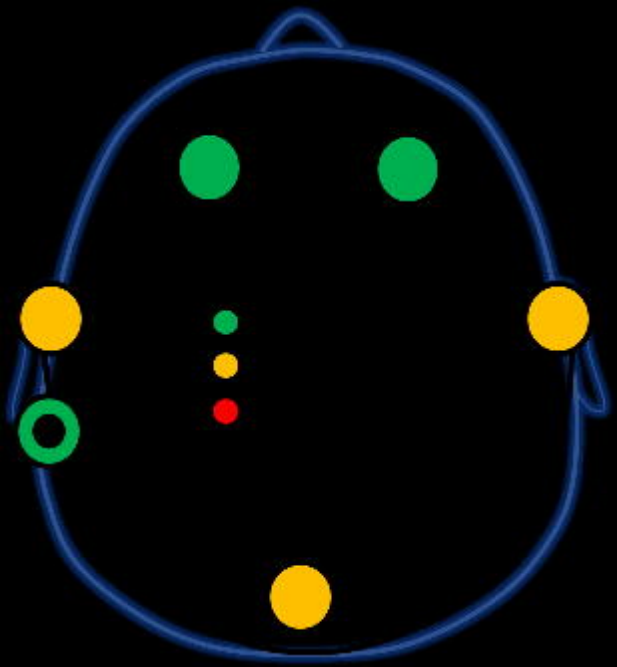
24. Kastrati, A. *et al.* Eegeyenet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. *ArXiv* **abs/2111.05100** (2021).

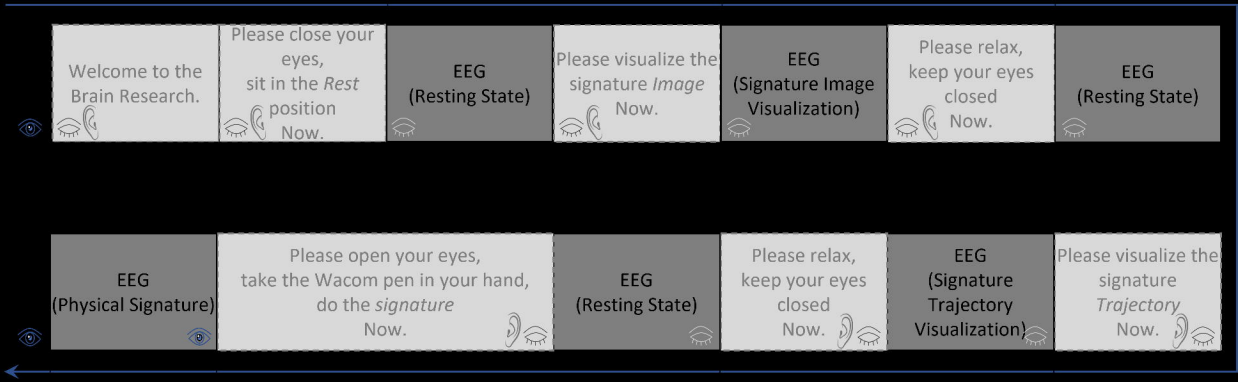
25. Saini, R. *et al.* Imagined object recognition using eeg-based neurological brain signals. In *Recent Trends in Image Processing and Pattern Recognition*, 305–319 (Springer International Publishing, 2022).

26. Hasan, M. A. M., Shin, J. & Maniruzzaman, M. Online kanji characters based writer identification using sequential forward floating selection and support vector machine. *Appl. Sci.* **12**, 10.3390/app122010249 (2022).

Wacom Signature Visualizer
Audio Instruction
EEG Visualizer
Participating Subject
Data Curator
Wacom Tablet with Stylus
Assistant


Raw EEG Signal
Epoching
Bandpass Filtering (0.1-60Hz)
Baseline Removal
Artifact Removal (ICA)
Pre-processed EEG Signal




Welcome to the Brain Research.
Please close your eyes, sit in the *Rest* position Now.
EEG (Resting State)
Please visualize the signature *Image* Now.
EEG (Signature Image Visualization)
Please relax, keep your eyes closed Now.
EEG (Resting State)
EEG (Physical Signature)
Please open your eyes, take the Wacom pen in your hand, do the *signature* Now.
EEG (Resting State)
Please relax, keep your eyes closed Now.
EEG (Signature Trajectory Visualization)
Please visualize the signature *Trajectory* Now.