Noname manuscript No.
(will be inserted by the editor)

EEG Based Person Authentication

Aditya Sesha Sai Samineni · Praharsha Venu · Charan Sai Venkat Narayana Lolugu · Bhoomika Kotharu · Mani Chandrika Pachipulusu

Received: date / Accepted: date

Abstract Biometric authentication is pivotal in identifying individuals based on unique physiological or behavioral characteristics. This research delves into the realm of biometrics, exploring various recognition systems such as fingerprint recognition, voice recognition, iris recognition, and face recognition. Despite their widespread use, these systems possess inherent drawbacks, including susceptibility to spoofing, privacy concerns, and limitations in certain environments. To address these shortcomings, the study investigates an alternative biometric approach: Electroencephalogram (EEG) authentication. EEG involves measuring brainwave activity using electrodes and is renowned for its reliability, resistance to forgery, and inherent uniqueness, akin to fingerprints. The paper delves into the technical aspects of EEG, detailing electrode placement, signal acquisition, amplification, and digital processing. It underscores EEG's significance in liveness detection and its potential for robust biometric authentication in sensitive applications. Moreover, the research explores diverse classifier models pivotal in biometric authentication. It covers algorithms like K-Nearest Neighbors (KNN), Auto Encoders, Support Vector Machines (SVM), XGBoost, and Convolutional Neural Networks (CNNs). Each model's unique attributes, functionalities, and applications in biometric systems are analyzed, highlighting their roles in data classification, feature learning, and pattern recognition.

Keywords EEG · KNN · SVM · Machine learning · CNN · XGBoost

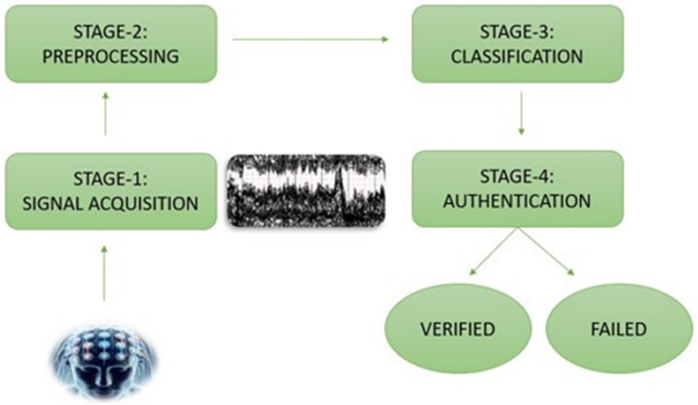

Fig. 1 Flow diagram of the proposed EEG-based person identification model.

1 Introduction

Biometrics are usually unique features related to an individual's characteristics to differentiate each person's identity. These characteristics can be sub grouped primarily into two categories – physiological and behavioural identifiers. Typically, Physiological identifiers include Iris recognition, face recognition, hand geometry identifiers, fingerprint recognition, etc whereas behavioural identifiers comprise of signatures, voice recognition, gait, handwriting. When it comes to biometrics, it involves estimating, analysing, assessing the features of individuals, on the other hand, biometric authentication comprehends a securing process that depends on the special features mentioned above to identify that respective individual. To choose a particular trait as a biometric identifier one needs to examine multiple factors such as distinctiveness, enduring, inclusivity, quantifiability and efficiency[1].To make a note of the physiological aspects furtherly, there can be several attributes such as fingerprint recognition, face recognition, Iris, EEG, etc.

1.1 Different Bio-metric Traits

Biometric authentication is a security procedure that leverages the distinctive biological traits of an individual. This technology is gaining popularity because

Aditya Sesha Sai Samineni, Praharsha Venu,Charan Sai Venkat Narayana Lolugu, Bhoomika Kotharu and Mani Chandrika Pachipulusu
Computer Science and Engineering, SRM University Andhra Pradesh, India
E-mail: adityaseshasai_s@srmap.edu.in, praharsha_venu@srmap.edu.in, lolugucharansai_v@srmap.edu.in, bhoomika_k@srmap.edu.in, manichandrika_p@srmap.edu.in

it can provide a robust level of security without causing inconvenience for the user. Typical methods of biometric authentication encompass:

1.1.1 Fingerprint recognition

This authentication process has been well recognized since more than one hundred years. Fingerprints are unique arrangements created by the raised crestlines on the fingertips, each featuring rows of interconnected pores leading to sudoriferous glands. These patterns reside throughout the life staying consistent to their typical form. With the exception of their symmetry being disrupted by any serious injury, they will sustain in their existing state.

This recognition method is one of the mostly used techniques and widely helped in person identification and thereby assisting police and courts[2].

1.1.2 Voice recognition

Voice authentication has currently evolved into an essential component in person identification and authentication. This method has replaced many weaker security options like passwords and signatures since this is comparably more secure. The process includes machines using a single voice model to identify a particular individual on the basis of their specific voice features, namely, tonal quality, sound levels, pitch, tone, speech rhythm. ASR (Automatic speech recognition) using Artificial intelligence techniques are presently being used for voice assistance. Voice recognition techniques are also playing a crucial role while working as a subsidiary tool when working with business analysing information systems and ERP systems. In accordance with demand and pertinency, we will get a measure of its prominence in our daily lives as we move forward.

1.1.3 Iris recognition

Iris recognition is the methodology of identifying an individual by inspecting the unsystematic pattern of the iris. This technical expertise combines computer vision, data pattern recognition and optical inference. These spatial regularities that are salient in the human iris are highly characteristic and remarkable to an individual. In spite of the fact that the colouration and the composition of the iris is genetically linked, the particularities of the configurations are not. So, the aim of this recognition method is to authenticate a person in real time with high performance, precision and exactness by interpreting the visual patterns evident within the iris[3].

1.1.4 Face recognition

In this identification method a face is classified as either recognised or unrecognised after examining it with the pictures of recognised persons stored in the database. The human face is exceptionally complex with everchanging

structure with aspects that can substantially and rapidly change in time. Humans can distinguish faces but too many faces sometimes being difficult to be recalled, machine learning is being upgraded to do this work. This method is comprised of two classifications: verification- The system determines if a face image matches a claimed identity template and identification- Involves scrutinizing a query face picture against all templates secured in the database[4].

Although the above-mentioned recognition methods seem feasible there are also some disadvantages in them such as fingerprints may produce false positives, can be spoofed using gels or high-resolution photographs, Voice recognition include accuracy issues with accents and its limitations in noisy environments, iris recognition encircle significant privacy preoccupations due to biometric data accumulation, facial recognition akin to deep fake manipulation and may exhibit discrepancies. So, when the machine is prone to errors like these it is easier for people to take advantage of these weaknesses. So even if the future of biometric authentication looks brilliant, that does not imply that there are not any speed bumps. Present day security methods will constantly introduce new threats, new issues, and new thoughts to task new systems. Few of these encompass physical spoofing, Deepfakes, Civil and criminal responsibility. As authentication methods are becoming greater not unusual throughout consumer and organisation systems, new styles of fraud prevention and superior biometric techniques are filling inside the hole for structures that name for better stages of reliability and security. Hence, we are considering the implementation of a more reliable techniques like EEG, ECG.

1.2 EEG Description

Researchers are exploring alternative biometric systems like authentication via brainwave recognition particularly referred to as EEG(Electroencephalogram), used to degree the activity of waves in the mind using electrodes. EEG recordings are commonly categorized into two types: non-invasive EEG recordings obtained from electrodes attached to the scalp floor and invasive EEG recordings obtained from electrodes implanted within the cranium. Implanted electrodes in invasive EEG are situated closer to the brain than scalp electrodes, allowing them to record brain alerts with better amplitudes and smaller spatial scales, starting from a single neuron. However, invasive EEG has significant clinical risks in addition to a few technical issues. Because recording electrodes are implanted inside the cortex and must work well for an extended period of time, there may be a risk of infections and other mental harm. Many studies have proven that although non-invasive EEG is less correct in assessment with invasive EEG, it nonetheless includes enough actual-time data for use as a source for distinctive programs or even in actual-time brain computer interface (BCI) machines oriented to obligations including phrase processing, net surfing or controlling a two-dimensional motion [5]. EEG serves as a dependable liveness detection approach, ensuring the real-time presence and energy of the man or woman. Moreover, EEG styles are inherently particular, akin

to fingerprints, improving identity precision. Lastly, EEG shows resistance to forgery and spoofing tries, cementing its function as a strong and comfortable preference for biometric authentication in professional and sensitive packages.

The implementation of an EEG (Electroencephalogram) system involves a chain of critical steps. It begins with the cautious placement of electrodes at the character's scalp. The configuration of electrodes can vary, but common setups consist of using a 10-20 electrode gadget or maybe more, depending on the level of element required. To set up a dependable electric connection, conductive gel or paste is typically implemented at the electrode websites. Finally, the EEG electrodes seize the extremely weak electrical signals generated by the mind. These signals, frequently measured in microvolts, are then transmitted to an amazing amplifier. The amplifier serves the twin reason of magnifying those alerts for higher detection whilst minimizing background noise. Following amplification, the EEG signals are digitized, changing the Analog electrical statistics into a digital layout for further processing and analysis. This virtual information can then be subjected to various techniques and algorithms to extract significant insights, diagnose clinical situations, or perform biometric authentication, relying at the utility of the EEG device. EEG is utilized in diverse fields and programs specifically medical analysis, neuroscience research, psychology and psychiatry, cognitive psychology, neurofeedback remedy, recognition research, sports activities science, advertising and patron research, gaming and virtual truth.

2 Related Work

The authors of [7] presents a machine learning approach for person authentication from EEG signals, demonstrating that EEG could be reliable for biometric identification and authentication across various contexts. "The study aimed to investigate the effectiveness of person authentication from brain waves in different enrolments based on various stimuli. It has tested both time-and frequency-based signal processing methods and employed feature selection and dimensionality reduction techniques before applying a random forest-based classification" [7]. The research used a public dataset that is designed for testing EEG biometric approaches. The dataset is comprised of data from 21 participants (18 males and 3 females, aged between 23 and 47 years old) and included four distinct types of stimuli, such as the standard checkerboard pattern and flashing VEP with plain colour at different frequencies. "The proposed random forest-based machine learning model achieved approximately 83.2% authentication accuracy, demonstrating the effectiveness of the approach"[7].

Alsumari et al., provides a summary of the most recent EEG-based biometric identification techniques, specifically focusing on identification and authentication scenarios. Existing EEG-based recognition methods often rely on well-engineered methods and find it difficult to apply them effectively to unidentified data. The few deep learning-based EEG-based recognition techniques have overfitting issues and necessitate learning a big number of model parameters

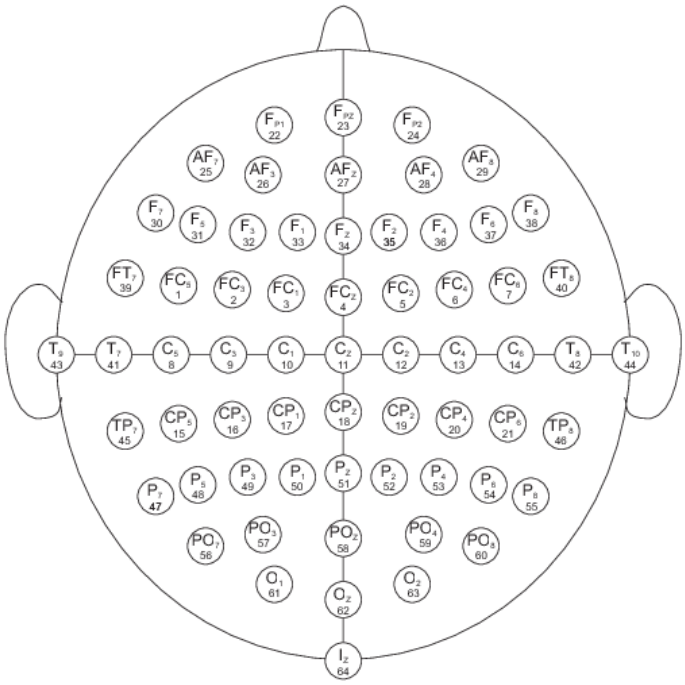

Fig. 2 Non invasive EEG electrode positioning [6].

limited EEG data. To address these issues, the paper presents a convolutional neural network (CNN) model that is lightweight and has few learnable parameters, enabling Training as well as evaluation on a small amount of available EEG data. The proposed CNN model achieves a 99% rank one identification result and a 0.187% authentication performance error rate on a public domain benchmark dataset. The system only requires two EEG channels and a signal measured over a short temporal window of 5 seconds, making it suitable for real-life settings[8].

The researcher in [9] mentions that many works have been proposed in the field of EEG biometric methods, including both shallow and deep methods, with shallow classifiers being more common in literature. Shallow models require the extraction of distinguishing features from the signals, and commonly used feature extraction methods include Autoregressive Model, Power Spectral Density, Wavelet Transform, and Fourier transform. After feature extraction, user identification/authentication is done using distance-based approaches or classification methods, with the use of Support Vector Machine being common in papers. Deep learning methods, particularly Convolutional Neural Networks (CNN), have also gained popularity in EEG biometric research, with studies showing higher accuracy compared to shallow models. Other deep learning models, such as those combining CNNs with Long Short-

Table 1 Summary of Related Work (Part 1)

Author Name & Year	Approach (Feature & Classifier)	Subjects	State & Accuracy
Mahmud Chowdhury et al., 2023	Random Forest-based machine learning model	21	State: 83.2% Accuracy, 14 electrodes
Walaa Alsumari et al., 2023	Lightweight Convolutional Neural Network (CNN) model	Not mentioned	State: 99%, Equal error rate: 0.187%, 2 EEG channels
Bidgoly et al., 2022	CNN	109	State: 98.04%, 3 channels
Wang et al., 2020	Mahalanobis Distance	109	State: 98.83%, 64 channels
Elakkiya et al., 2018	Power Spectral Density, Recurrent Neural Network, Feedforward Neural Network	25	State: 98% (spell task), 95% (read task), 5 non-invasive electrodes
Liuyin Yang et al., 2021	LSTM-based network with Bootstrap Aggregating (Bagging)	15	State: Authentication accuracy 92.6%, FAR 2.5%, FRR 5.0%, for various motor tasks
Marcel et al., 2007	SVM	9	State: 93.4%, 59 electrodes, 10-10 system
Sai Pranavi Kamaraju et al., 2023	KNN	35	State: 93.4% (Fine KNN), 89.4% (Cosine KNN), 10 channel EEG
Kritiprasanna Das et al., 2021	Linear Discriminant Analysis, MIFCSP	9	State: 83.18%, 84.44%, 3 channels for dataset1, 60 channels for dataset2
Barayeu et al., 2020	PCA-SVM	105	State: 92.09% (8 channels), 92.36% (16 channels), 95.64% (64 channels), 8, 16, 64 channels

Term Memory (LSTM) or Recursive Neural Networks (RNN), have also been explored in EEG biometric research[9].

The authors of [10] discusses the research on brain biometrics using electroencephalographic (EEG) signals, which has gained attention in recent years. It proposes a graph-based method for EEG biometric identification, consisting of a network estimation module and a graph analysis module. The network estimation module investigates seven different connectivity metrics, each characterized by a certain signal interaction mechanism. A new connectivity metric is proposed based on "the algorithmic complexity of EEG signals from an information-theoretic perspective" [10]. The graph analysis module proposes and studies six nodal features and six global features for analysing brain networks." The results demonstrate that the graph-based method improves the recognition rate and inter-state stability of EEG-based biometric identification systems"[10]. The findings provide a further understanding of the distinctiveness of humans' EEG functional connectivity and offer guidance for the design of graph-based EEG biometric systems[10].

A.Elakkiya et al., proposes an algorithm for recognizing EEG signals of individuals using a biometric authentication system. "Research on brain signals shows that each individual has a unique brain wave pattern. Electroencephalography (EEG) signals generated by mental tasks are acquired to extract

Table 2 Summary of Related Work (Part 2)

Author Name & Year	Approach (Feature & Classifier)	Subjects	State & Accuracy
Abdhulla et al., 2021	Random Forest	12	Accuracy within 1% using 32 channels, 32 channels
Hui Yap et al., 2023	Pre-trained models: GoogLeNet, Inception-V3, ResNet-50, ResNet-101, EfficientNet-B0, DenseNet-201	30	Range of 99.1-99.9%, 14 channels
Anuja Nair et al., 2023	CNN	Not mentioned, M3cv dataset	State: 99.8%, Not mentioned
Mahayar Taj Dini et al., 2023	SVM	50	State: 99.06%, 16 channels
Jin Xu 1 et al., 2023	SVM, LDA, NN, DTS, Bayesian, AdaBoost, MLP	3 datasets, one is PhysioNet	ESML1: 96%, ESML2: 99%, 3-Class task: 64 channels
Fares Yousefi et al., 2023	SVM, ANN	50 healthy people	State: 91%, 90%, 14 channels
Adhi Dharma Wibawa et al., 2022	Neural Network	43	State: 97.7%, 4 channels
Sujin Bak et al., 2023	SVM and GNB	54	SVM: 98.97%, GNB: 97.47%, 18 motor channels in the parietal region, 18 non-motor channels in the frontal and occipital regions
M Svetlakov., 2022	CalibratedClassifierCV classifier from sklearn	109, PhysioNet	9.5% Equal Error Rate (EER), 8 channels
Miao Shi et al., 2020	Pattern recognition method for optimizing SVM using the improved squirrel search algorithm (ISSA)	85.9	Not mentioned

the distinctive brain signature of an individual"[11]. The proposed algorithm uses power spectral density and both Recurrent Neural Network (RNN) and Feedforward Neural Network (FNN) for recognition of individuals. "The performance of the RNN is reported to be 98% accurate for the spell task and 95% accurate for the read task"[11]. The paper also mentions that biometric systems provide two functions: authentication and identification. Authentication confirms or denies an identity claim, while identification recognizes an individual from a group based on the claimed identity. The proposed method involves recording EEG signals from the subjects as the first stage[11].

The researcher in [12] focuses on EEG-based authentication in a real-world setting using a commercial dry-electrode EEG headset and chronic recordings on a population of 15 healthy people. The authors employed bootstrap aggregating (bagging) on an LSTM-based network to decode the EEG recordings made in reaction to performed and imagined motor tasks. identification accuracy, false rejection rate (FRR), and false acceptance rate (FAR) achieved were 92.6%, 5.0%, and 2.5% for the performed motor task; For the imagined motor task, the scores were92.5%,2.6%, and 4.9%; for the combined tasks, the scores were93.0%,1.9%, and 5.1%. The suggested approach is suggested in the paper for scenarios with limited time and data. Informed consent was obtained

from all the subjects involved in the study, and the data can be made available upon request. The study considered two authentication scenarios: one where one authentication model was developed per subject, and another where the subjects were divided into two groups to test the case where unknown individuals attempted to obtain access[12].

Kamaraju et al., proposes a framework for biometric identification using EEG signals recorded during signing, as EEG signals are difficult to replicate by another individual. "The framework utilizes multivariate variational mode decomposition (MVMD) to extract properly aligned oscillatory modes from multi-channel EEG data"[13]. Features are extracted using Fourier-Bessel series expansion-based (FBSE) entropies, specifically the multivariate FBSE-based entropy (M-FBSE-E). Machine learning-based classifiers, such as the fine K-nearest neighbor (KNN) and cosine KNN classifiers, are used to identify the EEG signals corresponding to original and forged signatures. "Experimental results show that the proposed framework achieves an average accuracy of 93.4±7.0% for subject-wise EEG-based biometric identification with the fine KNN classifier, and 89.4±1.9% average accuracy for subject-independent EEG-based biometric identification with the cosine KNN classifier"[13]. The paper highlights the uniqueness and consistency of EEG signals for each individual, making them suitable for creating a reliable biometric system[13].

The authors of [5] discusses the design of a moving average filter for MI BCI systems, which are used to analyze brain electrical activity during motor imagery movements. EEG is the most widely used technique for studying MI BCI due to its non-invasiveness, high time resolution, and mobility potential. The paper mentions that "mu (8-12 Hz) and beta (18-25 Hz) rhythms of EEG are the neurophysiological basis for MI BCI, observed in the sensory-motor cortex area of the brain"[5] . Fourier analysis has been commonly used to extract frequency bands for MI BCI, but it may not be suitable for highly nonlinear and nonstationary signals. Data-adaptive techniques like empirical mode decomposition (EMD) have been used to address the issue of subject-specific frequency bands in MI BCI. The paper validates the proposed model for MI BCI using EEG data sets from BCI competitions. The proposed method in the paper achieves an average accuracy of 83.18 for left-right (LR) MI movement, which is higher than existing state-of-the-art MI BCI algorithms[5] .

The researcher in [14] proposes an authentication system based on EEG signals recorded in response to a simple motor paradigm, using a two-stage decoder. The first stage of the decoder involves extracting EEG signals. Principal component analysis (PCA) and two deep learning neural networks—one resembling Inception and the other like VGG—are used to extract features from EEG signals. In order to authenticate the subject based on the extracted features, a support vector machine (SVM) is utilized in the second stage of binary classification. In the paper, 105 subjects' EEG motor-movement data is used to compare how well various decoders perform. For eight channels, the VGG-like NN-SVM decoder produced an overall accuracy of 88.29% with a falseacceptance rate (FAR) of 2.55%. For sixteen and sixty-four channels,

the results were similar. With an accuracy of 87.29% overall and a FAR of 4.08% for 8 channels, the Inception-like NN-SVM decoder produced comparable results for 16 and 64 channels. High accuracy results were obtained by the PCA-SVM decoder, ranging from 92.09% to 95.64% with ranging from 1.26% to 2.19% for 8, 16, and 64 channels. The authors acknowledge the need for future comparisons with other methods and have made their code publicly available [14].

Authentication methods have evolved over time, from conventional password-based approaches to more sophisticated biometric methods. EEG-based authentication is a promising approach that leverages brain signals for secure authentication. Deep learning models, such as Convolutional Neural Networks (CNNs), have shown potential in processing EEG signals and achieving better classification performance compared to traditional methods. Transfer learning, a technique that applies knowledge learned from previous tasks to new domains with limited training data, has been used in various domains but is still limited in the EEG domain. This study aims to explore the applicability of transfer learning using pre-trained CNN models for EEG-based authentication and evaluate their performance. The outcomes of the experiments demonstrated the efficiency of transfer learning in improving the performance of deep learning models for EEG-based authentication, with accuracy in the range of 99.1-99.9% [15].

Studies have explored brainwaves during resting, mental tasks, Visual Evoked Potentials (VEP), and Event Related Potentials (ERP). Spectral information obtained through frequency analysis is a popular feature used in EEG authentication. "Other features include coherence to express the phase-amplitude relationship between electrodes and mutual correlation coefficient to calculate the similarity between electrodes"[16]. Classification methods commonly used include autoregressive (AR) model, Discriminant Analysis (DA), SVM, and Neural Network (NN). Performance in biometric authentication is evaluated based on the classification rate and Equivalent Error Rate (EER). Existing studies have evaluated performance using classification rate and EER, considering both genuine and impostor data. Prior studies have used different electrode placements, ranging from three or fewer electrodes to 30 or more electrodes. This study aims to investigate the performance of EEG for authentication through an extensive analysis of combining EEG signals in terms of frequency range, physical position, and processing[16].

Mahyar TajDini et al., mentions that the genetic traits of human EEG have been studied since the early days of EEG recordings by Hans Verger in 1924. "Researchers have standardized the placement of electrodes for collecting and recording brain waves, with the 10-20 electrode system proposed by Jasper et al. in 1958 and a modification called the 10-10 system with 64 channels introduced in 1994" [17]. The paper also mentions the use of three public EEG datasets for their experiments: RSVP, Sternberg Task, and BIC2000. Different frequency bands in EEG, such as theta, alpha, beta, and gamma, correspond to different states of mind and have potential effects on experimental results.

The paper discusses the use of filtering operations for EEG denoising to ensure fair comparisons with other methods[17].

Existing biometric authentication technologies have limitations related to usability, time efficiency, and long-term viability. "Recent technological advancements have led to the development of specific devices capable of reproducing human biometrics. Utilizing alpha brainwaves is a superior option for authentication compared to other brainwave types"[18]. Deep breathing can enhance alpha waves , making it a suitable option for brainwave-based authentication. "The experiment conducted in this study demonstrated a high success rate of 91% and 90% for Support Vector Machine and Neural Network classifiers, respectively" [18]. The most dependable brain pattern for authentication was found to be deep breathing, which was corroborated by the SVM and NN classifiers' outstanding classification performance. Deep breathing promotes relaxation and enhances Alpha brainwave activity, making it an ideal choice for establishing a dependable authentication system[18].

The authors of [19] discusses the development of a biometric authentication system based on EEG signals using machine learning algorithms such as Naive Bayes, Neural Network, and Support Vector Machine (SVM). The goal of the research is to address the shortcomings of the existing authentication systems, which include their high manipulation susceptibility, low accuracy, and ease of imitation. Since EEG signals are specific to each person, they could be used as a biometric feature for identity verification. Four channels (FP1, FP2, F7, and F8) and 43 participants' EEG data collected with the OpenBCI Ultra Cortex Mark IV device are used in the study. Matlab EEGLab ToolBox was used for pre-processing stages such as Independent Component Analysis (ICA), Artifact Subspace Reconstruction (ASR), Finite Impulse Response (FIR), and Automatic Artifact Removal EOG (AAREOG). To determine the optimal EEG channel for the biometric system, the classification process is applied to each channel, with Power Spectral Density (PSD) being computed as a feature. The Neural Network algorithm achieved the F7 channel's maximum accuracy of 97.7% in contrast to Naïve Bayes and SVM. The paper also mentions the use of Naive Bayes Classifier, which assigns class labels to instances based on conditional probabilities. The dataset is divided into training and testing data, with T1 and T2 sessions used for training and T3 sessions used for testing [19].

The researchers in [20] focuses on proposing an EEG-MI methodology for user identification using optimized feature extraction methods and classifiers. "The study compares the accuracies of user recognition using support vector machine (SVM) and Gaussian Naïve Bayes (GNB) classifiers. The common spatial pattern (CSP) method is used for achieving the highest user identification accuracies of 98.97% and 97.47% using SVM and GNB, respectively. The CSP method is a spatial filtering technique that maximizes the variance between classes by simultaneous diagonalization of two covariance matrices. The study also proposes a statistical methodology for estimating a minimum dataset scale to ensure CSP performance and confirms the adequacy of the used dataset" [20]. The study contributes to the field of information secu-

rity by improving the identification accuracy in human biometrics based on EEG-MI signals [20].

Using the Holo-Hilbert spectral analysis method, S. Bak et al. suggest a subject-independent learning approach for electroencephalogram-based biometrics. Utilizing spectral maps created with the Holo-Hilbert spectral analysis method, which takes both frequency and amplitude modulation into account, is the suggested neural network architecture. For subject-independent learning, the neighbourhood components analysis loss function is employed. Achieving a 9.5% equal error rate, the architecture is tested on the PhysioNet Electroencephalogram Motor Movement/Imagery Dataset. Subject-independency and suitability for interpretation with generated spectra and the Integrated Gradients method are two benefits of the suggested approach. The integrated gradients (IG) approach and its application in the Captum framework for the creation of importance maps are also discussed in the paper. After being averaged over time, the resulting spectrum has the following form 30 frequency intervals [21].

The author in [22] proposes a pattern recognition method for optimizing the support vector machine (SVM) using the improved squirrel search algorithm (ISSA) for EEG signal classification. The ISSA algorithm is tested using benchmark functions, and the results show improved exploration ability and convergence speed. For the purpose of classifying EEG signals, the ISSA-SVM model is developed and contrasted with other widely used SVM parameter optimization models. The average classification accuracy of the ISSA-SVM model is reported to be 85.9%, which is an improvement of 2-5% over the comparison method [22]. The summary of related work is presented in Table 1 and Table 2.

3 Methodology

3.1 Preprocessing

A structured approach is employed for the preparation of electroencephalogram (EEG) data, essential for subsequent machine learning tasks. Commencing with a systematic iteration through a predefined number of Subjects, the glob library is employed to dynamically retrieve the relevant EEG files associated with each individual. This dynamic file retrieval not only enhances the code's scalability but also ensures adaptability by allowing the system to automatically locate and process EEG files without the need for hardcoded file paths.

Subsequently, essential information is extracted from the raw EEG data, including details about EEG channels and data frames. This step is crucial in understanding the characteristics of the EEG signals, laying the foundation for subsequent preprocessing steps. To enable robust model evaluation on previously unseen data during testing, the approach incorporates a train-test split mechanism. By utilizing a specified train-test split ratio, the data is par-

titioned into training and testing sets, ensuring a comprehensive evaluation of the model's generalization capabilities.

Ensuring that the model is trained on normalized data, Sklearn's StandardScaler is applied to both the training and testing datasets. This normalization process mitigates the impact of feature scale discrepancies, promoting better convergence and overall model performance. The subsequent concatenation of training and testing data frames separately contributes to the creation of unified datasets, maintaining consistency in feature dimensions for effective model training and evaluation.

In addition to data preparation, unique labels are assigned to each data point, associating them with specific subject identifiers. This subject-label association is crucial for the model to learn and establish connections between EEG patterns and subject characteristics. In essence, this methodology section outlines a systematic and comprehensive approach to preparing EEG data for machine learning, encompassing critical steps such as data loading, preprocessing, splitting, scaling, concatenation, and label generation.

This dataset included recordings from 109 subjects, and each subject had data from 64 electrode channels. Our main goal was to understand how changing the size of our dataset would affect the performance of our models. So, we looked at four different scenarios: one with 25 subjects, second with 50 subjects, another with 75 subjects and the last one with the full 109 subjects. This way, we could see how the number of subjects used for training impacted the overall effectiveness of our machine learning models.

Recognizing the significance of selecting the right electrode channels in EEG analysis, we also conducted experiments on different channel setups. We specifically explored subsets with 8, 16, 32, and 64 channels within each subject group. This detailed exploration into the number of channels helped us uncover how the spatial details in the EEG data influenced our model's performance.

4 Model and Discussion

4.1 Dataset Description

Information set includes over 1500 one and two-minute EEG recordings, received from 109 volunteers. Topics performed distinct motor/imagery duties at the same time as 64-channel EEG were recorded using the BCI2000 gadget [6]. Every subject is made to perform 14 experimental runs which included two one-minute baseline runs with eyes open and eyes closed and three two-minute runs of each of the four following tasks:

A target or a signal appears on left side or right side of the screen and the subject opens and closes the corresponding fist till the target disappears [6]. And in the next case the Subject imagines opening and closing the respective fist.

A target or a signal appears on top portion or on the bottom portion of the screen. The subject opens and closes both fists if it is on top and opens and closes both feet if it is on bottom [6]. And in the next case the Subject imagines opening and closing fists or feet.

In precise, the experiments are: eyes open and eyes closed as base cases, open and close left or right fist as the first task, imagining it is task two, open and close both fists or both feet as task three, imagining it is task 4. These tasks are repeated further for testing the individual respectively. The records are supplied right here in EDF+ layout, containing 64 EEG indicators, each sampled at a hundred and sixty samples in keeping with 2d, and an annotation channel. Each annotation consists of one in every of 3 codes (T0, T1, or T2). T0 , corresponding to rest, T1 corresponding to onset of movement (actual or imagined) of the left fist (in runs 3, four, 7, 8, eleven, and 12), Both fists (in runs five, 6, nine, 10, 13, and 14), T2 corresponding to onset of movement (real or imagined) of the right fist (in runs 3, 4, 7, eight, eleven, and 12), Both feet (in runs 5, 6, nine, 10, 13, and 14) [6]. "Within the BCI2000-layout versions of these documents, which can be available from the participants of this records set, those annotations are encoded as values of zero, 1, or 2 in the Target Code nation variable. The EEGs have been recorded from 64 electrodes as consistent with the global 10-10 system" [6]. The numbers beneath every electrode name indicate the order in which they appear within the information; be aware that alerts within the data are numbered from 0 to 63, even as the numbers within the discern range from 1 to 64 [6].

4.2 Classifier Description

4.2.1 k-NN

It is a popular and straightforward supervised machine learning approach for classification applications. It is a member of the instance-based lazy learning algorithm family. A data point in k-NN is classified according to the majority class of its k-nearest neighbours, with "k" being a user-specified parameter. Some key aspects are as follows,

- The similarity between data points is measured by k-NN using a distance metric (such as the Manhattan distance, Euclidean distance, etc.).
- Selecting "k": It is important to choose the parameter 'k' carefully. A big 'k' can tame local patterns, whereas a tiny 'k' could be noise-sensitive.
- Decision Rule: In order to classify data, the algorithm chooses the class label that the k-nearest neighbours share the most, as illustrated in Fig 3.

4.2.2 Auto Encoder

Auto encoder belongs to the style of neural networks commonly used for unsupervised learning. The fundamental objectives include proficient representation of data, usually used for dimensionality reduction, feature learning, data

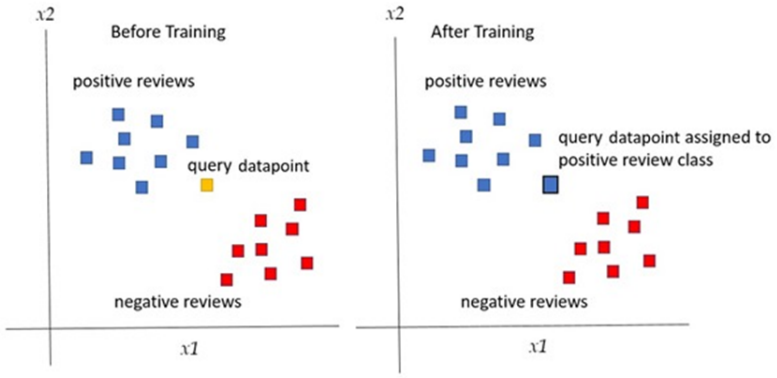

Fig. 3 Illustration for k-NN

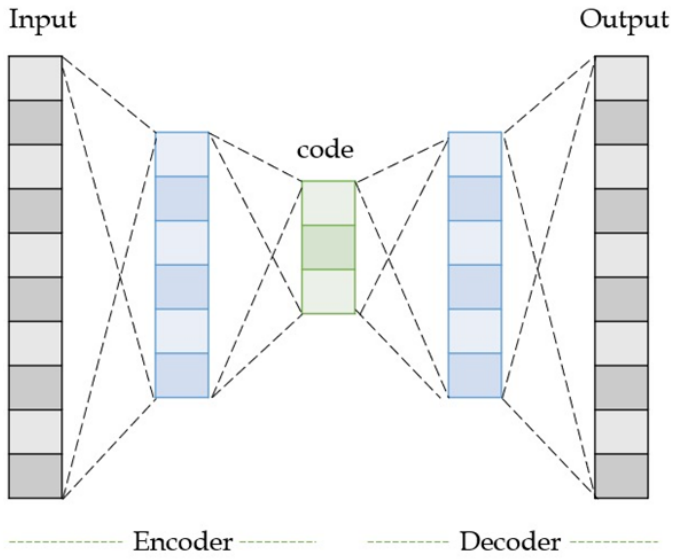

Fig. 4 Illustration for Auto Encoder

denoising, etc. This basically consists of two parts – the encoder and the decoder.

Encoder: This part of the autoencoder compresses the data into lower dimensional such that only the essential features will be extracted for efficient use of the reduced input data. Latency, as in other words, is a abridged

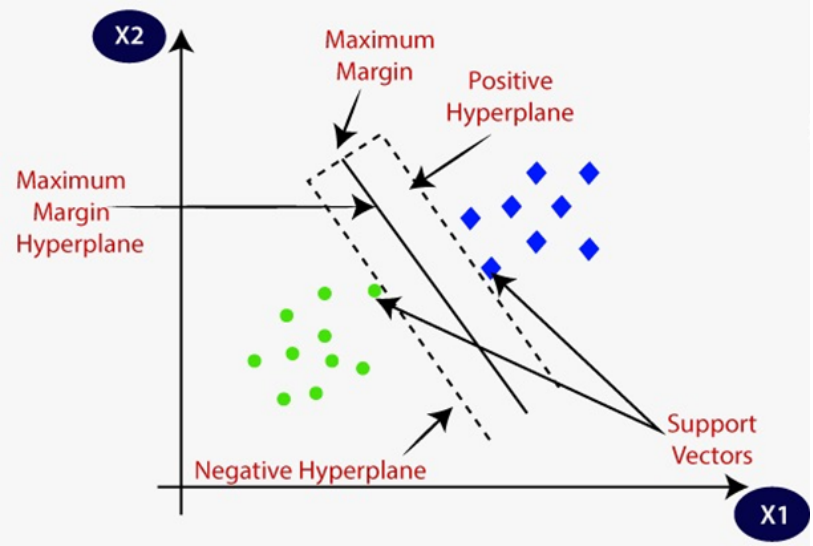

Fig. 5 Illustration for SVM

representation of the input data that optimally retains the most particular information needed for reformation.

Decoder: Decoder, as the name suggests, reconstructs the data taken from the encoder to return it back to its original form. The main purpose is to form an output that is as close as possible to input data.

The training process is to reduce the difference between the input and reconstructed output, which is done by detailing a loss function that computes the disparity between the input and the reconstructed output. Backpropagations are then used adjust weights and minimize the loss.

4.2.3 SVM

Although it is primarily known for its classification abilities, Support Vector Machine (SVM) is a supervised machine learning algorithm that is used for regression and classification tasks. As seen in Fig 5, SVM looks for the optimal hyperplane to maximize the margin between data points of different classes in order to improve generalization. Both linearly and non-linearly separable data can be handled by SVM and it transforms that data into higher-dimensional features spaces by using kernel functions. SVM is most impactful in high-dimensional spaces and is relatively less affected by overfitting, making it suitable for complex datasets. One drawback of SVM is its computational complexity, particularly with large datasets, because it necessitates solving a quadratic optimization problem. However, with the advent of efficient optimization techniques, SVM remains a popular choice in various applications like optical character recognition (OCR) systems for recognizing handwritten characters and digits, binary and multiclass classification problems, and so on.

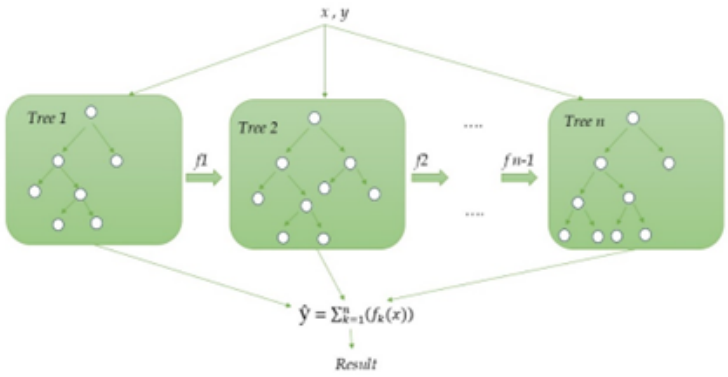

Fig. 6 Illustration for XGBoost

4.2.4 XG Boost

XGBoost, or eXtreme Gradient Boosting, stands as a potent machine learning algorithm renowned for its supremacy in supervised learning tasks like classification and regression. Operating within an ensemble learning framework, XGBoost constructs a sequence of decision trees as illustrated in Fig 6, each compensating for prior tree errors through a gradient descent approach, thus progressively refining predictions. Its appeal lies in efficient handling of missing data, regularization techniques to curb overfitting, and superior computational speed due to parallel processing. Notably, XGBoost employs tree-pruning methods to prevent unnecessary splits, optimizing predictive accuracy.

Its versatility in distributed computing and inherent cross-validation capabilities further solidify its status as a go-to choice in the realm of machine learning algorithms. This algorithm has gained immense popularity in various domains for its ability to handle large datasets, maintain model interpretability, and deliver impressive performance, often outshining other machine learning techniques.

4.2.5 CNN

Convolutional Neural Networks (CNNs), a specialized form of deep neural networks, are renowned for their intrinsic ability to automatically extract intricate features from data. These methods excel in learning intricate patterns by hierarchically arranging layers to progressively learn complex representations a process often termed as hierarchical or deep representation learning. CNNs primarily find applications in various domains, prominently excelling in image recognition tasks due to their capacity to capture spatial hierarchies and patterns within images. Comprising convolutional layers, pooling layers, and fully connected layers, these networks strategically employ convolution operations.

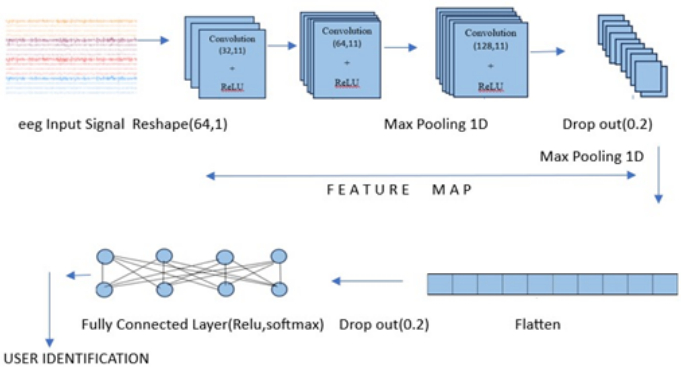

Fig. 7 Illustration for CNN

- **CONVOLUTIONAL LAYER:** This is the core building block of CNNs. The convolutional layers apply convolution operations to the input using various filters. These filters slide across the input data, extracting spatial patterns and features.
- **RELU:** It is an activating function it comes into action after the application of the convolution layer and it works on the feature map. It converts all the negative values to zero therefore achieving non linearity to the sequential model.
- **POOLING LAYER:** The pooling layer in a CNN classifier serves to lessen the feature maps' spatial dimensions, which will aid in managing computational complexity. There are three varieties: maximum pooling, average pooling, and sum pooling. Max pooling, a common technique, takes the maximum value from a specific region in each feature map, retaining essential information while reducing the size. This aids in preserving important features and in preventing Overfitting.
- **FLATTEN LAYER:** This is placed before the fully connected layers. The feature maps are flattened into a single dimensional vector. This layer is necessary because fully connected layers require a 1D input.
- **DROPOUT LAYER:** This regularized technique aids us in avoiding overfitting.
- **FULLY CONNCETED LAYER:** The dense layer in a CNN classifier is the fully connected layer responsible for making final predictions. It combines features from previous layers and applies non-linear transformations. The dense layer outputs class probabilities, enabling the classification of the input data.

Table 3 Electrode-Based Accuracy Analysis for different tasks with 25 Subjects

Classifier	Task	0-8	0-16	0-32	0-64
CNN	Task 3	18	64	80	90
	Task 4	18	58	70	85
	Task 5	16	62	80	90
	Task 6	16	54	76	86
kNN	Task 3	38	44	50	52
	Task 4	34	40	44	45
	Task 5	39	46	53	54
	Task 6	36	40	46	47
SVM	Task 3	38	73	76	84
	Task 4	36	52	69	77
Auto Encoder-KNN	Task 3	18	24	31	44
	Task 4	17	21	36	43
	Task 5	20	22	35	45
XGBoost	Task 3	45	63	83	92
	Task 4	41	59	83	92
	Task 5	46	63	86	95
	Task 6	42	50	75	89

Table 4 Electrode-Based Accuracy Analysis for different tasks with 50 Subjects

Classifier	Task	0-8	0-16	0-32	0-64
CNN	Task 3	8	49	57	80
	Task 4	9	46	61	82
	Task 5	9	52	64	82
	Task 6	8	49	58	80
kNN	Task 3	26	33	39	39
	Task 4	25	31	35	34
	Task 5	27	35	41	41
	Task 6	25	29	32	32
SVM	Task 3	-	47	67	73
	Task 4	27	43	60	68
Auto Encoder-KNN	Task 3	12	17	26	30
	Task 4	14	18	28	31
	Task 5	17	17	27	32
XGBoost	Task 3	37	53	75	85
	Task 4	35	52	76	89
	Task 5	37	55	79	88
	Task 6	33	47	66	81

5 Results and Discussion

The tables provide a comprehensive analysis of the accuracy scores achieved by different classifiers across varying tasks, each delineated by the number of subjects (25, 50, 75, and 109) and electrode configurations (0-8, 0-16, 0-32, and 0-64). Across all classifier types—CNN, kNN, SVM, Auto Encoder-KNN, and XGBoost—distinct patterns emerge. For instance, in Table 1 (25 subjects), XGBoost consistently demonstrates the highest accuracy across tasks, particularly excelling at 64 electrodes, achieving up to 95% accuracy. As the number of subjects increases in subsequent tables, trends become evident.

Table 5 Electrode-Based Accuracy Analysis for different tasks with 75 Subjects

Classifier	Task	0-8	0-16	0-32	0-64
CNN	Task 3	6	46	58	74
	Task 4	7	38	56	77
	Task 5	7	44	64	79
	Task 6	8	49	58	80
kNN	Task 3	19	24	30	29
	Task 4	19	25	30	30
	Task 5	20	25	31	30
	Task 6	25	29	32	32
SVM	Task 3	-	-	-	64
	Task 4	-	-	55	63
XGBoost	Task 3	30	45	68	79
	Task 4	26	41	69	84
	Task 5	30	45	71	83
	Task 6	27	41	65	81

Table 6 Electrode-Based Accuracy Analysis for different tasks with 109 Subjects

Classifier	Task	0-8	0-16	0-32	0-64
CNN	Task 3	4	42	53	76
	Task 4	5	38	57	76
	Task 5	5	40	58	76
	Task 6	5	39	55	77
kNN	Task 3	16	24	29	28
	Task 4	16	22	27	27
	Task 5	18	24	28	28
	Task 6	16	22	27	26
XGBoost	Task 3	26	42	66	78
	Task 4	23	39	66	80
	Task 5	26	42	66	80
	Task 6	24	40	63	77

The CNN model shows consistent performance across tasks, while kNN and auto Encoder-KNN exhibit a decreasing trend in accuracy as subject numbers rise. Moreover, SVM's accuracy improves with electrode count but displays limitations in scaling with higher subject counts. The tables exhibit a consistent trend: as electrode count rises from 8 to 64, all classifiers show increased accuracy across subject counts of 25, 50, 75, and 109. Conversely, as subject numbers increase, accuracy diminishes consistently for all electrode configurations—0-8, 0-16, 0-32, and 0-64. These trends highlight the critical role of electrode density in improving classifier accuracy while indicating the challenge of maintaining accuracy with larger subject cohorts in EEG signal analysis. For CNNs, there's a noticeable swift enhancement in accuracy as the electrode count rises. This model seems to particularly benefit from increased electrode inputs, showcasing a rapid and proportional improvement in accuracy. On the other hand, models like XGBoost and SVM exhibit a more steady and consistent increase in accuracy with additional electrodes. While they do improve, the rate of improvement is not as drastic as observed in CNNs. However, autoencoders and KNN display a different behavior. These

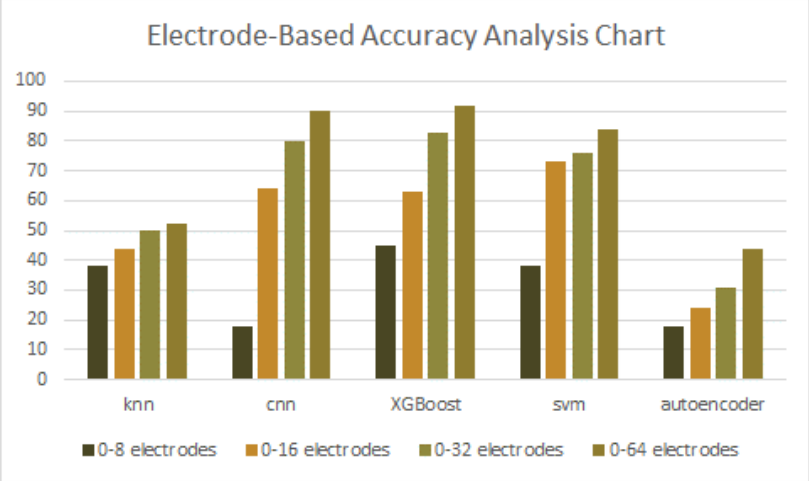

Fig. 8 Electrode-Based Accuracy Analysis Chart

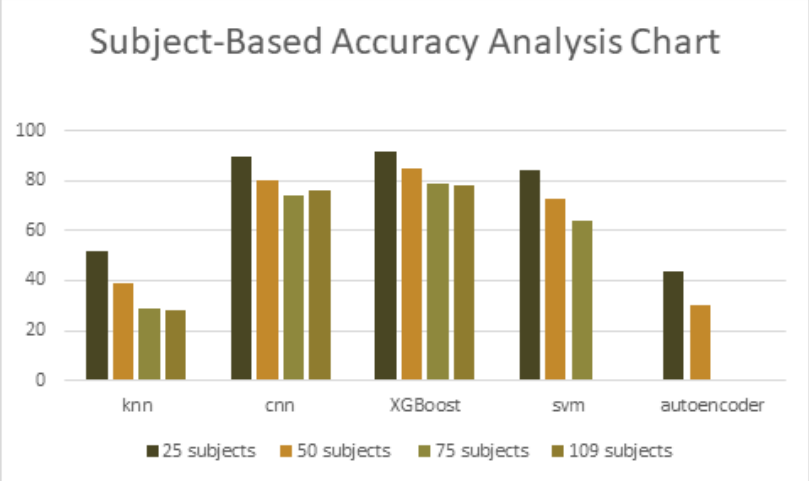

Fig. 9 Subject-Based Accuracy Analysis Chart

models exhibit slower growth in accuracy as the electrode count increases. This suggests that they might not be as inherently sensitive to the increased electrode inputs as CNNs, XGBoost, or SVM.

6 Conclusion and Discussion

This project delved into person identification using EEG data from 109 individuals, employing a diverse set of machine learning models, which includes "Convolutional Neural Network (CNN)", "Support Vector Machine (SVM)",

"XGBoost","k-Nearest Neighbors (KNN)", and an "autoencoder-based" approach.

The results highlight the exceptional performance of both the CNN and XGBoost models, with CNN emerging as the top performer, showcasing its efficacy in capturing intricate patterns within EEG data for accurate person identification. The inherent ability of CNN to extract spatial dependencies and the robustness of XGBoost contributed to their higher levels of accuracy.

SVM also demonstrated strong performance, suggesting its suitability for person identification tasks. These models exhibited competitive accuracies, emphasizing their versatility and effectiveness in handling EEG data for non-medical applications.

While KNN demonstrated reasonable accuracy, its sensitivity to noise and scalability issues may need consideration, particularly in large-scale deployments. Despite this, its simplicity and ease of implementation could make it a practical choice in specific contexts.

The autoencoder-based approach, although not surpassing the accuracy of CNN and XGBoost, revealed potential for unsupervised feature learning. Further exploration and refinement of this approach, perhaps in conjunction with other models, could enhance its effectiveness for person identification.

In conclusion, the most suitable model varies on the particular requirements of person identification applications. Both the CNN and XGBoost models stand out for their superior accuracy, but considerations such as interpretability, computational efficiency, and scalability should guide the selection process.

7 Future Work

Expanding on our current discoveries, upcoming studies might explore combining multiple classifiers using ensemble learning methods to boost accuracy in EEG-based biometric authentication. Techniques like Random Forests, Gradient Boosting, or Stacking could join forces, pooling the strengths of individual classifiers. This collaboration might help surpass the accuracy mark achieved in our study. Investigating how different sets of subjects and electrode configurations influence various classifiers could reveal new insights, refining how we authenticate users. Additionally, diving deeper into feature extraction from EEG signals—digging out more detailed temporal or spectral features—could improve classifier performance, revealing hidden information crucial for precise identification.

Another exciting avenue for future research involves customizing deep learning setups for EEG-based authentication. Crafting neural network architectures or using pre-existing models like recurrent neural networks (RNNs) tailored for sequential data could capture the subtle patterns inherent in EEG signals, potentially outperforming traditional classifiers. Also, it is crucial to consider potential security threats like adversarial attacks or noisy EEG signals, conducting robustness tests to strengthen the reliability and safety of EEG-based biometric authentication systems.

References

1. N. Ammour, Y. Bazi, and N. Alajlan. Multimodal approach for enhancing biometric authentication. *J. Imaging*, 9:168, 2023. 2
2. Nitin Kaushal and Purnima Kaushal. Human identification and fingerprints: A review. *Journal of biometrics and biostatistics*, 2, 2011. 3
3. Neha Kak, Rishi Gupta, and Sanchit Mahajan. Iris recognition system. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 1(1), 2010. 3
4. Nawaf Hazim Barnouti, Sinan Sameer Mahmood Al-Dabbagh, and Wael Esam Matti. Face recognition: A literature review. *International Journal of Applied Information Systems (IJAIS)*, 11(4), 2016. 4
5. K. Das and R. B. Pachori. Electroencephalogram-based motor imagery brain–computer interface using multivariate iterative filtering and spatial filtering. *IEEE Transactions on Cognitive and Developmental Systems*, 15(3):1408–1418, 2023. 4, 9
6. A. Goldberger and et al. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*, 101(23):e215–e220, 2000. 6, 13, 14
7. M. M. Chowdhury and M. H. Imtiaz. A machine learning approach for person authentication from eeg signals. In *2023 IEEE 32nd Microelectronics Design & Test Symposium (MDTS)*, pages 1–5, 2023. 5
8. W. Alsumari, M. Hussain, L. Alshehri, and H.A. Aboalsamh. Eeg-based person identification and authentication using deep convolutional neural network. *Axioms*, 12:74, 2023. 6
9. AJ Bidgoly, HJ Bidgoly, and Z. Arezoumand. Towards a universal and privacy preserving eeg-based authentication system. *Sci Rep*, 12(1):2531, 2022. 6, 7
10. Min Wang, Jiankun Hu, and Hussein A. Abbass. Brainprint: Eeg biometric identification based on analyzing brain connectivity graphs. *Pattern Recognition*, 105:107381, 2020. 7
11. A. Elakkiya and G. Emayavaramban. Biometric authentication system using eeg brain signature. *International Journal of Scientific Research in Science and Technology (IJSRST)*, 4(5):1797–1805, 2018. 8
12. L. Yang, A. Libert, and M.M. Van Hulle. Chronic study on brainwave authentication in a real-life setting: An lstm-based bagging approach. *Biosensors*, 11:404, 2021. 8, 9
13. Sai Pranavi Kamaraju, Kritiprasanna Das, and Ram Bilas Pachori. Eeg based biometric authentication system using multivariate fbse entropy, 2023. 9
14. U. Barayeu, N. Horlava, A. Libert, and M. Van Hulle. Robust single-trial eeg-based authentication achieved with a 2-stage classifier. *Biosensors*, 10:124, 2020. 9, 10
15. H.Y. Yap, YH. Choo, and Z.I. et al. Mohd Yusoh. An evaluation of transfer learning models in eeg-based authentication. *Brain Inf.*, 10:19, 2023. 10

16. H. Vadher, P. Patel, and A. et al. Nair. Eeg-based biometric authentication system using convolutional neural network for military applications. *Security and Privacy*, 2023. 10
17. Mahyar TajDini, Volodymyr Sokolov, Ievgeniia Kuzminykh, and Bogdan Ghita. Brainwave-based authentication using features fusion. *Computers & Security*, 129:103198, 2023. 10, 11
18. J. Xu, E. Zhou, Z. Qin, T. Bi, and Z. Qin. Electroencephalogram-based subject matching learning (esml): A deep learning framework on electroencephalogram-based biometrics and task identification. *Behav. Sci.*, 13:765, 2023. 11
19. Fares Yousefi and Hoshang Kolivand. A robust brain pattern for brain-based authentication methods using deep breath. *Computers & Security*, 135:103520, 2023. 11
20. A. D. Wibawa, B. S. Y. Mohammad, M. A. K. Fata, F. A. Nuraini, A. Prasetyo, and Y. Pamungkas. Comparison of eeg-based biometrics system using naive bayes, neural network, and support vector machine. In *2022 International Conference on Electrical and Information Technology (IEIT)*, pages 408–413, 2022. 11, 12
21. S. Bak and J. Jeong. User biometric identification methodology via eeg-based motor imagery signals. *IEEE Access*, 11:41303–41314, 2023. 12
22. M. Svetlakov, I. Hodashinsky, and K. Sarin. Representation learning for electroencephalogram-based biometrics using holo-hilbert spectral analysis. *Pattern Recognit. Image Anal.*, 32:682–688, 2022. 12