EEG based Password Authentication System
Biomedical Signal Processing
First Review

Group 6 Team Members

Vikash J	CB.EN.U4AIE22156
Nithish Ariyha K	CB.EN.U4AIE22140
Eshwanth Karti T R	CB.EN.U4AIE22118
Yeshwanth Balaji A P	CB.EN.U4AIE22102

Problem Statement

Traditional authentication methods are prone to security breaches. Developing a reliable EEG-based biometric system is challenging due to complex brainwave patterns and data variability. This project aims to tackle these issues using Recurrent Neural Networks (RNN) and Kolmogorov-Arnold Networks (KAN) for secure and accurate authentication.

Abstract

Electroencephalogram (EEG) signals present a promising avenue for safe and reliable biometric authentication due to their unique and non-replicable nature. This project aims to develop an advanced EEG-based authentication system that leverages cognitive passwords, where specific questions act as keys for verifying identity. The system utilizes Recurrent Neural Networks (RNNs) enhanced with Attention layers and RNNs integrated with Kolmogorov Arnold Networks (KAN) to capture intricate temporal dependencies and non-linear patterns within EEG signals. A comparative analysis of these models is conducted to evaluate their performance in terms of accuracy, robustness, and real-time application suitability. The Attention layer enhances feature extraction by focusing on relevant EEG signal patterns, while the KAN model further captures complex cognitive processes. This research highlights the potential of cognitive password-based EEG authentication as a secure alternative to traditional methods, paving the way for innovative applications in personal security and access control systems

Literature Review on EEG-Based Authentication Systems

We plan to place the user in a medium-light room and assign them tasks at 30-second intervals. These tasks include:

1. **Mental Arithmetic**: Multiply 233 by 22 without using pen or paper.
2. **Imaginary Conversation**: Imagine convincing your class CR to postpone the BMSP exam by a few days.
3. **Selective Attention**: A video is played to measure the user's attention span and capability.
4. **Image Comparison**: Identify the difference between two visual images.
5. **Resting State**: The user remains in a resting state.

A newly published paper proposes replacing the hidden vector calculation in RNN layers with a multilayer perceptron (MLP) model. The authors compare the performance of a linear layer against a 2x1 perceptron. One of the key challenges with KAN (Kolmogorov-Arnold Network) is its limited implementation in real-world tasks. Most current research has only explored KAN's potential on toy problems or simple datasets. Few studies investigate KAN's ability to process noisy, complex data such as EEG signals. Additionally, the impact of KAN on memory and time complexity, compared to traditional architectures, needs further exploration to increase the significance of our research.

We plan to compare a 2x1 MLP perceptron inside an RNN with a 2x1 KAN layer since KAN has limited research applications. In our dataset, we observe several outliers—one prominent source of noise is eye interference, which produces signals ranging from 100μV to 200μV for about 250ms. To manage this, we have decided to apply a threshold of 100μV. Another issue in this domain is the lack of systematic applications, such as automated door opening, which is significant despite being overlooked. It's worth noting that about 50% of locks can be picked using brute force, like trying different combinations or exploiting vulnerabilities with SQL injections, which is why we will be incorporating vector databases for enhanced security.

Lastly, we are focusing on P300 signals, which are event-related potentials (ERPs) triggered by rare events. The downside of traditional shallow feature extraction techniques, such as PSD (Power Spectral Density) and fuzzy entropy, is that they can capture information from a single EEG channel but struggle to explicitly capture interactions between channels.

Authentication processes traditionally rely on knowledge, ownership, and biometric factors, where EEG falls into the latter category due to its unique physiological characteristics. EEG-based authentication leverages brain signals that are inherently complex, dynamic, and difficult to replicate, making it a promising candidate for secure biometric authentication systems.

Methods of Data Collection and Preprocessing

Data collection methods across the reviewed papers emphasize the diversity of tasks designed to elicit distinct EEG patterns. The studies typically involve motor imagery, visual stimuli, acoustic stimuli, and mental tasks, all intended to generate unique brainwave signatures for each individual.

Preprocessing is a crucial step in EEG signal analysis. Common methods include band-pass filtering to isolate relevant brain wave frequencies, such as using Butterworth and Chebyshev filters, and spatial filtering techniques like Common Average Referencing (CAR) and Independent Component Analysis (ICA) to reduce noise and enhance signal quality. These preprocessing techniques are fundamental for extracting reliable features from EEG signals that can be used for authentication purposes.

Feature Extraction Techniques

Feature extraction is pivotal in distinguishing between individuals based on EEG data. The studies utilize a variety of methods:

- **Autoregressive Coefficients (ARs)**, **Power Spectral Density (PSD)**, and **Spectral Power (SP)** are commonly used to capture signal characteristics that reflect individual brain patterns.
- **Spatial and Temporal Features**: Some models, such as the 1D-Convolutional LSTM networks, extract both spatial and temporal features, leveraging the LSTM's capability to retain memory of temporal signal dynamics, thus improving user identification accuracy.
- **Chaotic Measures**: Features like correlation dimension, Lyapunov Exponent, and entropy are used to characterize the brain's dynamic behaviour, as explored in deep RNN-based models.

Security and Privacy Concerns

Security issues are inherent to EEG-based authentication, with challenges such as signal spoofing, imitation, and signal sniffing posing significant risks. A notable advancement discussed is the use of privacy-preserving techniques that store hashed EEG fingerprints instead of raw signals. This approach protects user data and reduces the risk of sensitive information leakage.

One study highlights a novel method employing the Gram-Schmidt orthogonalization process, which successfully reduces the number of EEG channels required for authentication while maintaining high accuracy, demonstrating the potential to simplify EEG systems without compromising performance.

Classification and Authentication Models

Several machine learning and deep learning models are employed across the studies for classification and authentication:

- **RNN and Deep Learning Models**: These models use recurrent structures to learn and classify EEG patterns based on mental tasks, such as visual counting or mental figure rotation. The deep RNN approach identifies EEG sub-band features and classifies them into corresponding mental tasks, which can be associated with the user's identity.
- **1D-Convolutional LSTM Networks**: These networks excel at learning the sequential patterns in EEG data, integrating both spatial and temporal characteristics, achieving high accuracy even with a reduced number of channels.

Challenges and Future Directions

Despite significant advancements, EEG-based authentication still faces several challenges:

- **Signal Variability**: The variability of EEG signals due to environmental factors, user states, or noise necessitates robust preprocessing and feature extraction methods to maintain consistency and reliability.
- **Data Quality and Channel Reduction**: Reducing the number of channels without losing accuracy is critical for making EEG systems more practical and less cumbersome. Current research is focused on optimizing feature extraction and classification techniques to achieve this goal.
- **Privacy and Security**: Privacy-preserving techniques are emerging as essential components of EEG-based systems, with ongoing research into new algorithms that secure user data against spoofing and unauthorized access.

Literature Review Table:

Title	Authors	Inference
A survey on methods and challenges in EEG based authentication (Jalaly Bidgoly et al., 2020)	AmirJalaly Bidgoly , Hamed Jalaly Bidgoly , Zeynab Arezoumand	This review includes a number of aspects such as the various tasks that the user required to perform and understand in EEG authentication. This paper reviews on various dataset and data collection methods. This also gives an insight into the challenges faced in the process of building EEG authentication. Various shallow preprocessing methods has also been discussed.
Low-cost electroencephalogram (EEG) based authentication (Ashby et al., 2011)	Corey Ashby ; Amit Bhatia, Francesco Tenore, Jacob Vogelstein	This publication reviews on a low-cost consumer grade EEG authentication system using shallow methods and preprocessing methods like Autoregressive Coefficients and PSD. For better results they have used a group wise voting system, instead of one overall output.
Towards a universal and privacy preserving EEG-based authentication system(Bidgoly et al., 2022)	AmirJalaly Bidgoly , Hamed Jalaly Bidgoly , Zeynab Arezoumand	This paper reviews on idea of processing EEG signals similar to simple password systems. This paper reviews on using Gram-Schmidt orthogonalization for channel reduction techniques. This paper also reviews on the idea of using neural network for feature extraction and preprocessing.
Convolutional Neural Networks Using Dynamic Functional Connectivity for EEG-Based Person Identification in Diverse Human States (Wang et al., 2019)	Min Wang Heba El-Fiqi, Jiankun Hu, Hussein A. Abbass	This paper reviews on representing EEG signals as a graph based on within-frequency and cross-frequency functional connectivity estimates, and the use of graph convolutional neural network (GCNN) to automatically capture deep intrinsic structural representations from the EEG graphs for person identification. Apart from GCNN, they also

		review on using signals collected during tasks instead of only rest state.
EEG-based user identification system using 1D-convolutional long short-term memory neural networks (Sun et al., 2019)	Yingnan Sun , Frank P.-W. LO , Benny Lo	This review approaches on idea of using a 1D convolutional LSTM for user identification. The combined use of this system helps the model to capture spatiotemporal features and reduce the number of channels used. This system involves an enrolment phase, where users' EEG biometrics are learned and stored in a 1D-Convolutional LSTM network, and an identification phase, where the network identifies users from 1-second EEG recordings. The signals pre-processed, normalized and segmented before fed into network.
Deep RNN learning for EEG based functional brain state inference (Patnaik et al., 2017)	Suprava Patnaik, Lalita Moharkar, Amogh Chaudhari	This review deals with extraction of EEG sub-band features by using wavelet transform followed through classification by means of recurrent neural network (RNN) and deep learning to identify the task. This helped us in quantitating the questionnaire for our project.
Managing EEG studies: How to prepare and what to do once data collection has begun (Boudewyn et al., 2023)	Megan A. Boudewyn Molly A. Erickson, Kurt Winsler, John Daniel Ragland, Andrew Yonelinas, Michael Frank, Steven M. Silverstein, Jim Gold, Angus W. MacDonald Ill, Cameron S. Carter, Deanna M. Barch, Steven J. Luck	This paper acted as an instruction for us on how to collect data and steps to be followed in every steps. This gives us an idea of issues that could be faced and avoided in the data collection.
Cognitive biometrics based on EEG signal (Chatra, 2015)	Ashwini S Chatra	The process of biometric signal processing uses EEG, ECG, and EDG. But this review takes only EEG due to the latter having a higher frequency response. This paper reviews more about capturing the signals and settings used.

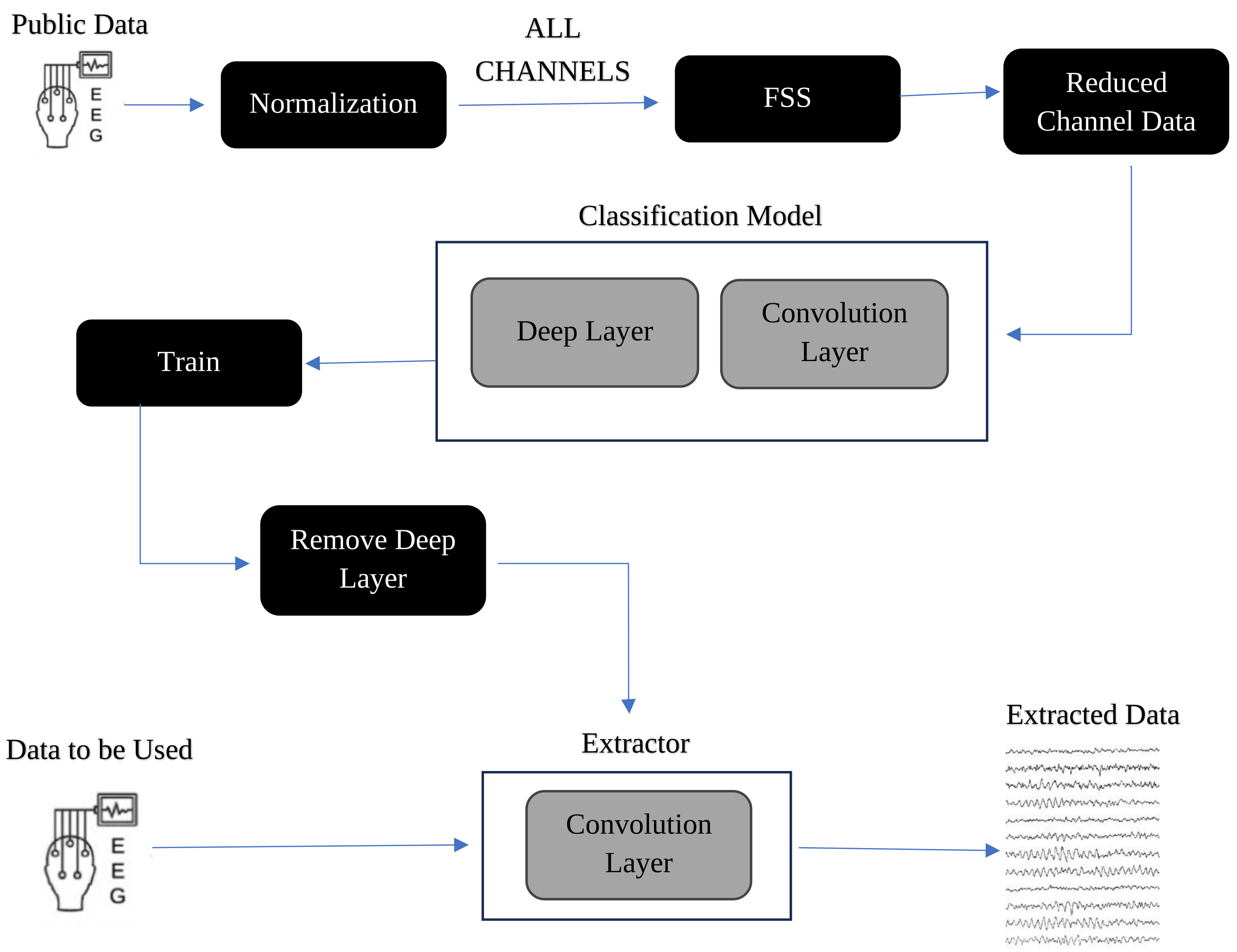
EEG-Based User Authentication in Multilevel Security Systems (Pham Tienand Ma, 2013)	Tien Pham, Wanli Ma, Dat Tran, Phuoc Nguyen & Dinh Phung	Usage EEG signals for authentication in multi-level security systems, showcasing their ability to provide secure, real-time verification and enhance overall security.
Learning to (Learn at Test Time): RNNs with Expressive Hidden States (Sun et al., 2024)	Yu Sun, Xinhao , Karan Dalar, Jiarui Xu, Arjun Vikram , Genghan Zhang, Yann Dubois, Xinlei Chenf, Xiaolong Wangi , Sanmi Koyejo, Tatsunori Hashimoto, Carlos Guestrin	Self-attention performs well in long context but has quadratic complexity. Existing RNN layers have linear complexity, but their performance in long context is limited by the expressive power of their hidden state. They propose a new class of sequence modeling layers with linear complexity and an expressive hidden state. The key idea is to make the hidden state a machine learning model itself, and the update rule a step of self-supervised learning
EEG-based biometric authentication system using convolutional neural network for military application (Vadher et al., 2024)	Himanshu Vadher, Pal Patel, Anuja Nairn Tarjni Vyas, Shivani Desai, Lata Gohil, Sudeep Tanwar, Deepak Garg, Anupam Singh	CNNs for automatic EEG feature extraction, emphasizing robustness in noisy environments, high security, and real-time processing.
Affective EEG-Based Person Identification Using the Deep Learning Approach (Wilaiprasitporn et al., 2018)	Theerawit Wilaiprasitporn, Apiwat Ditthapron, Karis Matchaparn, Tanaboon Tongbuasirilai, Nannapas Banluesombatkul and Ekapol Chuangsuwanich	Person Identification → based on emotion recognition like valance, Arousal and etc using DEAP Dataset
KAN: Kolmogorov–Arnold Networks (Liu et al., 2024)	Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljaéiél', Thomas Y. Hou, Max Tegmarkl	KANs have learnable activation functions on edges (“weights”). KANs have no linear weights at all – every weight parameter is replaced by a univariate function parametrized as a spline. For accuracy, smaller KANs can achieve comparable or better accuracy than larger MLPs in function fitting tasks. Theoretically and empirically, KANs possess faster neural scaling laws than MLPs.

Selecting methods for a modular EEG pre-processing pipeline: An objective comparison (Coelli et al., 2024)	Stefania Coelli, Alessandra Calcagno, Chiara Maria Cassani, Pierluigi Reali , Roberto Gatti , Manuela Galli , Anna Maria Bianchi, Federico Temporiti	This paper discusses EEG signal processing before feeding into a model. Techniques like: bandpass filter, low pass filter, highpass filter can be a preliminary processing. Latter for better processing techniques like, ICA and ReReferencing can be use. **ICA** – BSS, SOBI **ReReferencing** : CAR, rCAR, REST.
Biosignals for User Authentication - Towards Cognitive Biometrics? (Revett et al., 2010)	Kenneth Revett ,Farzin Deravi and Konstantinos Sirlantzis	Contains the The types of signals used and their claimed effectiveness is presented and compared challenges facing the deployment of cognitive biometrics, including sensor design issues and the need to extract information-rich and robust features.
Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation (Marcel & R. Millan, 2007)	Se´bastien Marcel and Jose´ del R. Milla´n	propose the use of a statistical framework based on Gaussian Mixture Models and Maximum A Posteriori model adaptation, successfully applied to speaker and face authentication

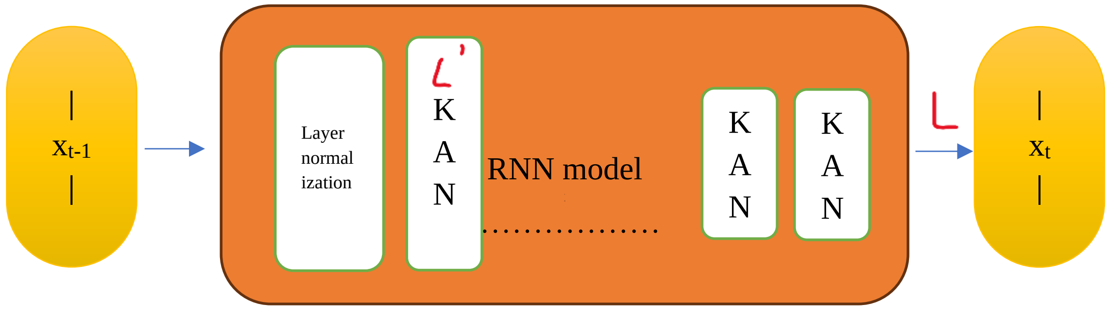
Research Gap

- Existing studies often focus on single mental tasks and person recognition rather than authentication.
- There is a lack of research on using combined mental tasks for improved accuracy and security in authentication.
- In 2020, the OpenAI scaling law paper (Kaplan et. al) showed that LSTMs (a type of RNN) could not scale similarly to Transformers or effectively use long context. Now, with modern RNNs and best practices, we try to implement this new model for our use case.
- Lack of questionnaire investigation used in this field.
- Current models use many EEG channels. Future work should aim to reduce channels without sacrificing accuracy.

Data Collection and Extraction



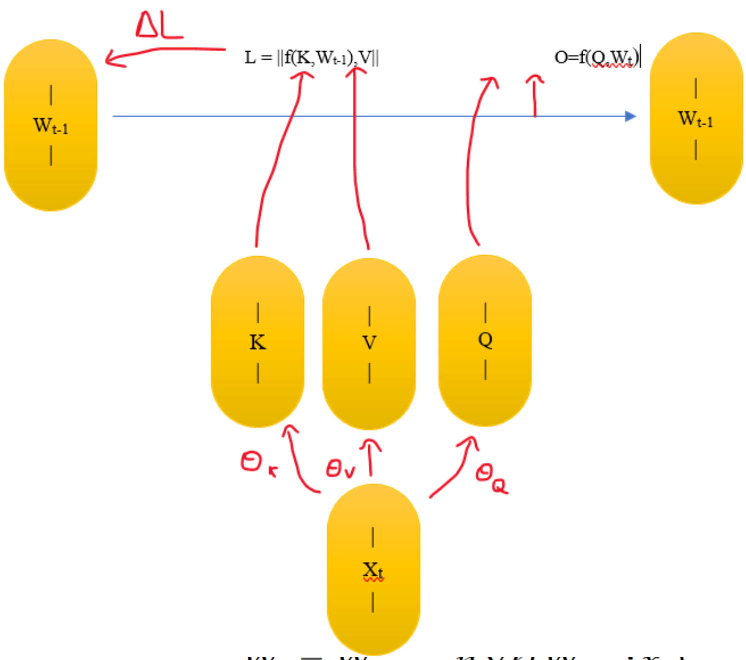
RNN WITH MLP ATTENTION



Two losses **L** and **L'**

MLP tries to minimize L' which updates rule in a step of self-supervised learning method. key idea is to use self-supervised learning to compress the historic context x1,..., xt into a hidden state st , by making the context an unlabeled dataset and the state a model. Concretely, the hidden state st is now equivalent to Wt , the weights of a model f , which can be a linear model,a small neural network, or anything else. The output rule is simply:

$$Z_t = f\,(x_t; W_t)$$



$$W_t = W_{t-1} - \eta \, \nabla \ell(W_{t-1}; x_t),$$

Since this model uses attention principle we can Parrlaleize this and run in a batch format instead of rnn's known drawback of sequential processing

RNN WITH KAN IMPLEMENTATION

Kolmogrov Arnold representation theorem states that if f is a multivariate continous function on a bounded domain , then it can be written as a finite composistion of continous functions of a single variable and the binary operation of addition. More precisely, for a smooth

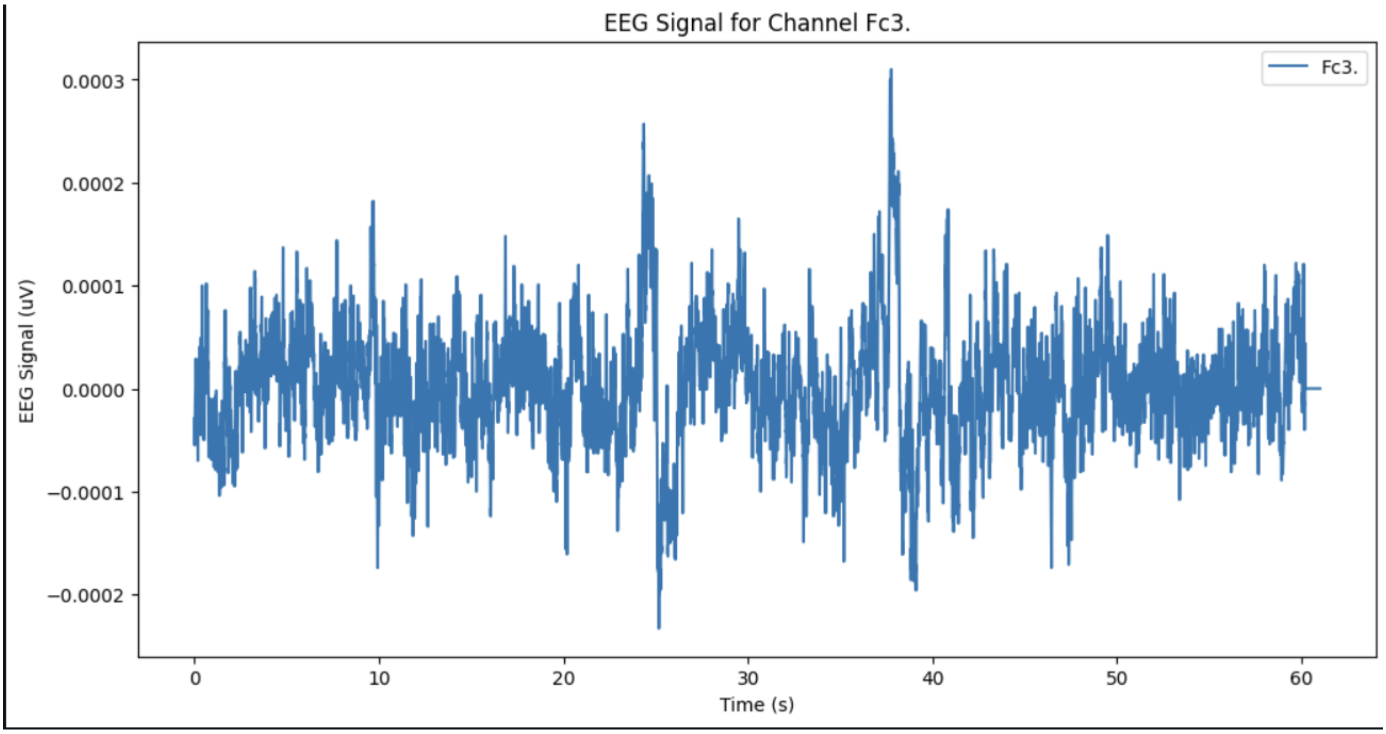
$f : [0,1]^n \to \mathbb{R}$,

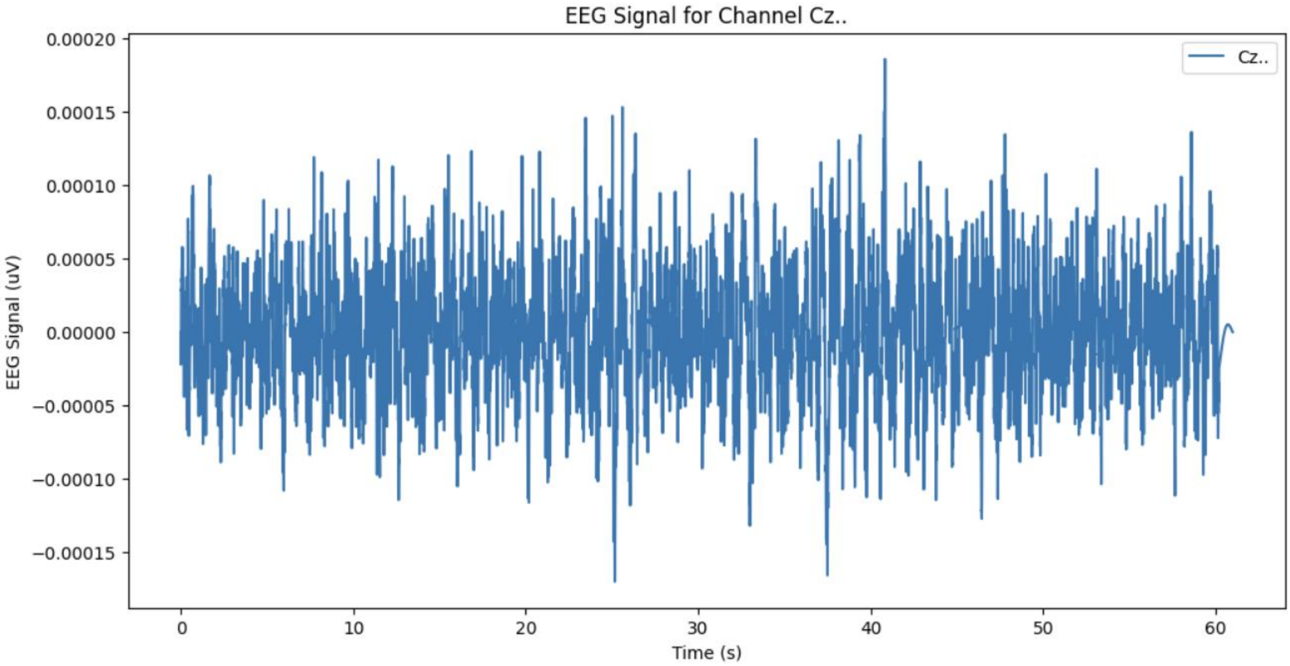
$$f(x) = f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q(\sum_{p=1}^{n} \phi_{q,p}(x_p))$$

Sample code with which we worked with toy data like emotion classification like neutral, Sad, Happy
we got around 95% CCR(Correct Classific
ation

PRELIMINARY RESULTS:

```
Reading EEG signals and events for subject: files
Reading EEG signals and events for subject: S001
Loaded file: C:/Users/nithi/College/5TH SEMESTER/BIOMEDICAL SIGNAL PROCESSING/PROJECT/eeg_motor/S001\S001R01.edf
Number of channels: 64
Channel names: ['Fc5.', 'Fc3.', 'Fc1.', 'Fcz.', 'Fc2.', 'Fc4.', 'Fc6.', 'C5..', 'C3..', 'C1..', 'Cz..', 'C2..', 'C4..', 'C6..', 'Cp5.', 'Cp3.', 'Cp1.', 'Cpz.', 'Cp2.', 'Cp4.', 'Cp6.', 'Fp1.', 'Fpz.', 'Fp2.', 'Af7.', 'Af3.', 'Af
EEG data shape: (64, 9760)
Sample times: [0.      0.00625 0.0125  0.01875 0.025   0.03125 0.0375  0.04375 0.05
 0.05625]...
Loaded events from: C:/Users/nithi/College/5TH SEMESTER/BIOMEDICAL SIGNAL PROCESSING/PROJECT/eeg_motor/S001\S001R01.edf.event
Number of events: 0
First few events: []
Events loaded successfully.
Loaded file: C:/Users/nithi/College/5TH SEMESTER/BIOMEDICAL SIGNAL PROCESSING/PROJECT/eeg_motor/S001\S001R02.edf
Number of channels: 64
Channel names: ['Fc5.', 'Fc3.', 'Fc1.', 'Fcz.', 'Fc2.', 'Fc4.', 'Fc6.', 'C5..', 'C3..', 'C1..', 'Cz..', 'C2..', 'C4..', 'C6..', 'Cp5.', 'Cp3.', 'Cp1.', 'Cpz.', 'Cp2.', 'Cp4.', 'Cp6.', 'Fp1.', 'Fpz.', 'Fp2.', 'Af7.', 'Af3.', 'Af
EEG data shape: (64, 9760)
Sample times: [0.      0.00625 0.0125  0.01875 0.025   0.03125 0.0375  0.04375 0.05
 0.05625]...
Loaded events from: C:/Users/nithi/College/5TH SEMESTER/BIOMEDICAL SIGNAL PROCESSING/PROJECT/eeg_motor/S001\S001R02.edf.event
Number of events: 0
First few events: []
Events loaded successfully.
Loaded file: C:/Users/nithi/College/5TH SEMESTER/BIOMEDICAL SIGNAL PROCESSING/PROJECT/eeg_motor/S001\S001R03.edf
Number of channels: 64
Channel names: ['Fc5.', 'Fc3.', 'Fc1.', 'Fcz.', 'Fc2.', 'Fc4.', 'Fc6.', 'C5..', 'C3..', 'C1..', 'Cz..', 'C2..', 'C4..', 'C6..', 'Cp5.', 'Cp3.', 'Cp1.', 'Cpz.', 'Cp2.', 'Cp4.', 'Cp6.', 'Fp1.', 'Fpz.', 'Fp2.', 'Af7.', 'Af3.', 'Af
EEG data shape: (64, 20000)
Sample times: [0.      0.00625 0.0125  0.01875 0.025   0.03125 0.0375  0.04375 0.05
 0.05625]...
```






```
Creating RawArray with float64 data, n_channels=64, n_times=9760
    Range : 0 ... 9759 =      0.000 ...    60.994 secs
Ready.
Filtering raw data in 1 contiguous segment
Setting up high-pass filter at 1 Hz

FIR filter parameters
---------------------
Designing a one-pass, zero-phase, non-causal highpass filter:
- Windowed time-domain design (firwin) method
- Hamming window with 0.0194 passband ripple and 53 dB stopband attenuation
- Lower passband edge: 1.00
- Lower transition bandwidth: 1.00 Hz (-6 dB cutoff frequency: 0.50 Hz)
- Filter length: 529 samples (3.306 s)

Fitting ICA to data using 64 channels (please be patient, this may take a while)
Selecting by number: 64 components
[Parallel(n_jobs=1)]: Done  17 tasks      | elapsed:    0.0s
Fitting ICA took 1.6s.
```

REFERENCES

Ashby, C., Bhatia, A., Tenore, F., & Vogelstein, J. (2011). Low-cost electroencephalogram (EEG) based authentication. *2011 5th International IEEE/EMBS Conference on Neural Engineering*, 442–445. https://doi.org/10.1109/NER.2011.5910581

Bidgoly, A. J., Bidgoly, H. J., & Arezoumand, Z. (2022). Towards a universal and privacy preserving EEG-based authentication system. *Scientific Reports*, *12*(1), 2531. https://doi.org/10.1038/s41598-022-06527-7

Boudewyn, M. A., Erickson, M. A., Winsler, K., Ragland, J. D., Yonelinas, A., Frank, M., Silverstein, S. M., Gold, J., MacDonald, A. W., Carter, C. S., Barch, D. M., & Luck, S. J. (2023). Managing EEG studies: How to prepare and what to do once data collection has begun. *Psychophysiology*, *60*(11). https://doi.org/10.1111/psyp.14365

Chatra, A. (2015). *Cognitive biometrics based on EEG signal*. 374–376. https://doi.org/10.1109/IC3I.2014.7019605

Coelli, S., Calcagno, A., Cassani, C. M., Temporiti, F., Reali, P., Gatti, R., Galli, M., & Bianchi, A. M. (2024). Selecting methods for a modular EEG pre-processing pipeline: An objective comparison. *Biomedical Signal Processing and Control*, *90*, 105830. https://doi.org/10.1016/j.bspc.2023.105830

Jalaly Bidgoly, A., Jalaly Bidgoly, H., & Arezoumand, Z. (2020). A survey on methods and challenges in EEG based authentication. *Computers & Security*, *93*, 101788. https://doi.org/https://doi.org/10.1016/j.cose.2020.101788

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). *KAN: Kolmogorov-Arnold Networks*.

Marcel, S., & R. Millan, J. Del. (2007). Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(4), 743–752. https://doi.org/10.1109/TPAMI.2007.1012

Patnaik, S., Moharkar, L., & Chaudhari, A. (2017). Deep RNN learning for EEG based functional brain state inference. *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)*, 1–6. https://doi.org/10.1109/ICAC3.2017.8318753

Pham Tien and Ma, W. and T. D. and N. P. and P. D. (2013). EEG-Based User Authentication in Multilevel Security Systems. In Z. and C. L. and Z. O. and Y. M. and W. W. Motoda Hiroshi and Wu (Ed.), *Advanced Data Mining and Applications* (pp. 513–523). Springer Berlin Heidelberg.

Revett, K., Deravi, F., & Sirlantzis, K. (2010). Biosignals for User Authentication - Towards Cognitive Biometrics? *2010 International Conference on Emerging Security Technologies*, 71–76. https://doi.org/10.1109/EST.2010.32

Sun, Y., Li, X., Dalal, K., Xu, J., Vikram, A., Zhang, G., Dubois, Y., Chen, X., Wang, X., Koyejo, S., Hashimoto, T., & Guestrin, C. (2024). *Learning to (Learn at Test Time): RNNs with Expressive Hidden States*. https://arxiv.org/abs/2407.04620

Sun, Y., Lo, F. P.-W., & Lo, B. (2019). EEG-based user identification system using 1D-convolutional long short-term memory neural networks. *Expert Systems with Applications*, *125*, 259–267. https://doi.org/https://doi.org/10.1016/j.eswa.2019.01.080

Vadher, H., Patel, P., Nair, A., Vyas, T., Desai, S., Gohil, L., Tanwar, S., Garg, D., & Singh, A. (2024). EEG-based biometric authentication system using convolutional neural network for military applications. *SECURITY AND PRIVACY*, *7*(2). https://doi.org/10.1002/spy2.345

Wang, M., El-Fiqi, H., Hu, J., & Abbass, H. A. (2019). Convolutional Neural Networks Using Dynamic Functional Connectivity for EEG-Based Person Identification in Diverse Human States. *IEEE Transactions on Information Forensics and Security*, *14*(12), 3259–3272. https://doi.org/10.1109/TIFS.2019.2916403

Wilaiprasitporn, T., Ditthapron, A., Matchaparn, K., Tongbuasirilai, T., Banluesombatkul, N., & Chuangsuwanich, E. (2018). *Affective EEG-Based Person Identification Using the Deep Learning Approach*. https://doi.org/10.1109/TCDS.2019.2924648

Time line