MASTER IN ARTIFICIAL INTELLIGENCE
M1 INTERNSHIP REPORT

JACOBO RUIZ OCAMPO

Biometric Authentication with EEG signals using Machine Learning

Referent Professor
M. Caio Corro, Paris-Saclay University

Internship Tutors:
M. Alain Destexhe, NeuroPSI
M. Sabir Jacquir, NeuroPSI

Educational institution :
Paris-Saclay University - Computer Science Department

Internship host laboratory :
The Computational Neuroscience Laboratory at Paris-Saclay Neuroscience Insitute - NeuroPSI - CNRS

31/08/2023

Abstract—Electroencephalogram (EEG) signals from the brain have become an area of interest for biometric authentication because of concerns about the reliability of traditional methods like fingerprint and retina scans. This study looks into the use of EEG signals, focusing on beta waves from the frontal brain region, for biometric authentication. We reviewed current research and found a range of methods for EEG-based biometrics. We then chose a specific model that was previously designed for a wider frequency range (4-34Hz) and used 10 electrodes. For our tests, we changed it to only focus on the beta frequency band (12-32Hz) and reduced the electrodes to the 3 located in the frontal cortex of the 10-10 EEG system. Our results showed that using only these frontal electrodes did not work well for reliable identification. However, beta waves showed promise in other settings. This highlights the importance of selecting the right frequency bands and electrode locations for EEG-based biometrics. Our findings help understand the potential and challenges of using EEG for biometric authentication.

Index Terms—EEG, Biometrics, Authentication, Deep Learning, Signal Processing

1 Introduction

In our increasingly digitalized world, the need for secure and trustworthy identity verification is more pressing than ever. While traditional physiological biometric approaches such as fingerprints and retina scans are effective, they possess inherent drawbacks. Notably, they depend on external physiological patterns susceptible to duplication. Consequently, there's growing interest in intricate internal physiological biometrics that highlight the distinctive intrinsic biological structures of an individual. One emerging method in this domain harnesses the brain's electrical signals for identification, captured using Electroencephalography (EEG).

An EEG measures the brain's electrical activity along the scalp surface using multiple electrodes. Interestingly, these signals, which vary in amplitude and dominant frequencies, are often indicative of the physical or mental states of individuals. To illustrate, EEG recordings can be categorized into different frequency bands according to [10]:

- **Alpha (8-12Hz):** This is the dominant frequency band during a relaxed but awake state. Notably, focused attention can diminish the amplitude of this band.
- **Beta (12-30Hz):** Activity predominantly in this frequency range is associated with thinking and active, focused attention. Physical body movements can also amplify the amplitude in this band.
- **Gamma (30Hz or more), Theta (4-8 Hz),** and **Delta (4 Hz or less):** These are other notable EEG frequency bands. However Delta and Theta waves should be visible in an awake state and if they are, it is a sign of brain dysfunction [10].

In this report, our primary focus is on beta waves, and there are two pivotal reasons for this choice. Firstly, the frontal cortex predominantly produces beta waves. This allows for recording by merely positioning electrodes on the forehead, circumventing the noise issues arising from electrode placement on a hairy scalp. Secondly, beta waves are observable during an alert state when an individual engages in cognitively demanding tasks. Such conditions align seamlessly with an authentication environment where an attentive subject is intent on system access.

The allure of EEG extends beyond the unique nature of the signals. It also offers considerable versatility in both data collection and analysis. A variety of devices are available for recording EEG signals, each differing in the number and placement of electrodes on the scalp. Moreover, the past decade has witnessed significant advancements in EEG signal analysis, particularly due to the emergence of sophisticated algorithms for intricate signal processing. Most academic papers outline a specific workflow, transitioning from raw EEG data to user authentication or identification as the endpoint.

While the techniques employed may differ across studies, a general methodology is commonly adopted:

1) **Signal Acquisition:** This initial phase involves capturing the raw EEG data or using an existing dataset and therefore an experimental methodology for the signal acquisition is outlined.
2) **Preprocessing:** In this stage, data is denoised, prepared and in some cases electrode channels are optimized.
3) **Feature Extraction:** Arguably the most pivotal step, specific features from the EEG signals are extracted to be later used for classification.
4) **Classification:** At this juncture, the algorithm determines user identity, validating or refuting the authentication attempt.

In the realm of biometric authentication systems, evaluating performance is indispensable. A prevalent metric is accuracy, which, within machine learning, quantifies the ratio of accurate outcomes (encompassing both true positives and true negatives) in the dataset. More paramount, however, is the **Equal Error Rate (EER)**, which stands out as the predominant metric in biometric systems. EER delineates the interesecting point at which the false acceptance rate (FAR) coincides with the false rejection rate (FRR). This indicates that the EER is the point where both false identification and false rejection rates are at their least and most balanced. A lower EER signifies a superior system.

The primary aim of this report is to explore the utilization of beta waves in biometric authentication. Over the course of my three-month internship, I delved into existing research on algorithms that use EEG signals for biometric authentication, aiming to grasp the foundational principles and signal processing techniques. Armed with this knowledge, I chose a framework appropriate for further experimentation. My subsequent efforts centered on modifying this framework, with a particular emphasis on assessing the viability of beta waves for the task at hand.

In the ensuing sections, the discussion embarks on a journey from foundational research and some necessary background to hands-on experiments, elucidating the promises and challenges of EEG-based biometric systems.

2 Background and Related Work

The concept of biometric identification and authentication has evolved over time. While traditional methods such as passwords and PINs are still prevalent, they face challenges like being easily forgotten or compromised. Biometrics offer

an alternative, promising more secure and reliable identification and authentication mechanisms [1]. The literature covered in the subsequent section was primarily reviewed at the beginning of this internship. This was done to familiarize myself with the extensive research in this area and to understand the key terms and concepts associated with EEG, signal processing, and machine learning methods.

2.1 The task: Identification vs Authentication

In EEG biometrics, two primary ways of viewing the problem at hand are recognized: identification and authentication. Though they might seem alike, they can be solved differently.

Identification involves distinguishing one individual from a group. If framed as a machine learning problem, this can be seen as a multiclass classification problem. Each person corresponds to a class, and the objective is to train a model to correctly identify everyone.

On the other hand, authentication aims to verify a single user's identity, which translates to a binary classification in machine learning: distinguishing the user from potential intruders. This process requires creating a unique model for every person. While it may sound simpler than identification, it demands more computational resources, especially with a large user base.

Our focus in this report will be on authentication, as it aligns with the experiments we conducted and will discuss further.

2.2 EEG signal collection protocols

EEG signals are captured using EEG recording devices. However, the conditions and protocols for collecting these signals can differ across datasets. A review of the literature shows that Visually Evoked Potentials (VEP) is a common protocol. In VEP, a visual stimulus is presented to the subject, inducing a brain response. This captures the unique electrical patterns reflecting how different individuals react to the same stimulus. It is interesting to highlight the protocol used in Huang et al. [2] where the researchers used the user's own face and voice to create an unique reaction in the user as it should be familiar with it.

While VEP is common, another frequently used protocol is motor imagery (MI). In MI, subjects are instructed to imagine moving a limb, such as raising their right hand. This mental act triggers neuronal activity, especially in the motor cortex. The objective of MI is to identify distinct electrical signatures among individuals during this imagined movement, hoping that these signatures remain consistent for the same person over multiple tests. This protocol is used in Bingkun Wu et al. [9].

Other protocols exist such as the ones that measure brain activity during a cognitively demanding task such as reading a complex text or solving a puzzle.

2.3 Preprocessing techniques

Two predominant preprocessing techniques observed in the literature are denoising techniques and electrode channel optimization.

Denoising Techniques: The quality of EEG data can be compromised by noise, often referred to as artifacts. A prime example is the electrical signal captured by frontal cortex electrodes during a patient's blink, as the eye muscles generate an electrical activity detected by the electrodes. Liang et al. (2016) shed light on various EEG preprocessing techniques, highlighting the importance of retaining EEG's inherent characteristics while eliminating noise [3]. Several established methods, like the ICA algorithm utilized by Bingkun Wu et al. [9], and other signal processing techniques are available to eradicate these artifacts, significantly enhancing the performance of subsequent algorithms. However, one must note that it's challenging to obtain completely noise-free EEG data.

Electrode Channel Optimization: The objective is to achieve robust results in identification or authentication tasks using a minimal number of channels. This is interesting as commercial EEG headsets tend to have fewer electrodes than expensive EEG headsets only found in laboratories. Some strategies, like the one employed by Bingkun Wu et al. [9], utilize partial directed coherence (PDC) to determine the most influential channels—those that considerably affect other channels. In a similar vein, Zeynali et al. (2019) investigated EEG channels' optimization for diverse mental tasks and attained a remarkable 95% accuracy by optimally placing electrodes [7].

2.4 Techniques for Feature Extraction and Classification

The core of biometric algorithms lies in the techniques for feature extraction and classification. Conventionally, most frameworks follow a two-phase approach: first, extracting the features and then feeding them into a model for classification.

Feature Extraction Techniques:

- **Frequency Features:** The most widespread features are frequency features derived using the Fast Fourier Transform (FFT). The power spectrum is frequently used for frequency domain representation.
- **Discrete Wavelet Transform (DWT):** Unlike FFT, DWT provides a time-frequency representation of the signal, assisting in analyzing signals with discontinuities or abrupt changes. Both Zeynali et al. [7] and GUI et al. [1] employ DWT for feature extraction in their authentication algorithms.
- **Autoregressive Model Coefficients:** Some approaches utilize coefficients from the Autoregressive Model as features for classification, as seen in Marcel et al. [4]. Additionally, certain frameworks combine these coefficients with other statistical features, like the log energy entropy and the sample entropy, as evidenced in Zeynali et al. [7].

2.5 Classification Techniques

For classification in EEG-based biometric authentication, a wide range of models exists. Traditional methods often employ techniques such as the Support Vector Machine (SVM) and Bayesian approaches like the Naive Bayes Classifier. On the other hand, deep learning approaches have also

been explored, with neural networks, even those with fewer layers, being popular. Some studies have also leveraged the capabilities of Convolutional Neural Networks (CNN), especially with a few convolutional layers, yielding promising results.

An overview of the top-performing frameworks, their adopted models, and associated preprocessing techniques, alongside performance metrics, can be seen in Figure 10 (found at the end of this report).

2.6 Data Variability

EEG patterns can differ across users and even for the same user under similar conditions. To address this, Stergiadis et al. (2022) introduced a data-driven EEG-based authentication approach tailored to each user, achieving 95.6% average accuracy [5].

However, a significant concern is the variability of EEG data over time. While some studies account for this by including EEG signals from different trials across days, this coverage remains limited. For instance, Bingkun Wu et al. [9] conducted trials only days apart. This raises questions about the long-term reliability of the system, as it has not been tested with data spanning several months or even years.

3 Selected EEG Authentication Framework

In the prior section, we examined the literature on existing EEG authentication and identification frameworks, aiming to identify a suitable foundation for our experiments. To reiterate, our experiment's objective is to assess the feasibility of using only beta waves (frequencies between 12 and 30 Hz) for EEG-based biometrics. In this section, we'll detail our choice of framework and provide an overview of it.

3.1 Selection Criteria

Within this study, we sought a framework characterized by its high performance, ideally achieving an EER close to 0 or an accuracy above 98%, benchmarks synonymous with state-of-the-art. A fitting framework should employ transparent techniques and algorithms. This preference indicates a diminished interest in those frameworks leaning on black box algorithms or deep neural networks. Such choices stem from a broader project aim, not addressed in this internship report but set aside for further work: the desire to unravel the system's explainability, specifically aiming to understand the physiological elements that make an individual's EEG signal distinctive. An additional important criterion was the public availability of the dataset used for the framework's testing, ensuring results' reproducibility. Frameworks with openly accessible code were also prioritized. While the task of reconstructing an EEG-based authentication system would be enlightening, it is beyond this internship's scope, largely due to time constraints.

3.2 Framework Selection

In our quest for an efficient EEG-based authentication system, we explored a plethora of methodologies and architectures. Notably, two approaches stood out that fulfilled all our criteria. One is from Bingkun Wu et al.'s work [9], titled "Towards Enhanced EEG-based Authentication with Motor Imagery Brain-Computer Interface". The other captivating framework is "MindID: Person Identification from Brain Waves through Attention-based Recurrent Neural Network" presented by Zhang et al. (2021) [8].

Zhang et al. (2021) introduced a pioneering EEG-based identification method named "MindID". A significant revelation from their research is the discernment that the Delta pattern in EEG data harbors the most unique user-specific information. This insight contrasts starkly with prevalent literature which asserts that delta patterns, if prevalent during wakeful states, might indicate brain anomalies [10]. This Delta pattern is then channeled into an attention-centric Encoder-Decoder Recurrent Neural Network (RNN). The architecture permits the system to allocate dynamic attention weights to disparate EEG channels based on their relevance. The RNN, with its attention mechanism, is adept at capturing discriminative features, which are subsequently used for user identification via a boosting classifier. Extensive evaluations of MindID across three datasets revealed its prowess. Notably, on the EID-M dataset, it boasted an impressive accuracy of 0.982, surpassing both baseline models and other cutting-edge methodologies. Rigorous tests on another dataset, EID-S, and the public dataset EEG-S further substantiated the robustness and versatility of MindID, underscoring its applicability in real-world scenarios [8].

However, MindID's primary focus is on delta waves, whereas our investigations predominantly emphasized the presence and analysis of alpha and beta waves. Since delta waves are predominantly manifested during sleep, this tends to eclipse the manifestation of alpha and beta waves. Given this nuance, we chose not to adopt the MindID strategy but to gravitate towards a methodology already aligned with EEG signals where alpha and notably beta waves are prevalent.

In this context, a method that catered perfectly to our requirements, offering a comprehensive analysis across the entire frequency spectrum, including Beta, is articulated by Bingkun Wu et al. [9] in "Towards Enhanced EEG-based Authentication with Motor Imagery Brain-Computer Interface". The subsequent sections we will explain more about this framework, and show how it is better than others, including MindID.

3.3 Framework Overview

We need to understand the structure of the framework by Bingkun Wu et al. For clarity, let's look at the three main components of their system. You can find a detailed description in their original article [9] in the latter part of the introduction.

- **Preprocessing with ICA Artifacts Removal**: The goal of this to remove noise, known in this field as "artifacts", a classical artifacts could be the reading

of the electrical signal of the muscles used by the eye to blink as these are relatively close to the electrodes in the scalp. The raw signal undergoes blind source separation to segregate sources based on distinct statistical features. This should result in an input comprising only of brain electrical signals.

- **Channel Selection**: The technique is based in the theory of effective connectivity. Analyzing the causality of signals in different electrodes during imagined body movement retains crucial channels. The goal of this technique is to reduce the number of channels used.
- **Deep Learning Classification**: The classification process extracts frequency, space, and multi-mode temporal features of EEG signals and then proceeds to do a binary classification between users and attackers.

A flowchart of the framework can be seen in figure 1. The process is the same for both the training and test set used to train the models.

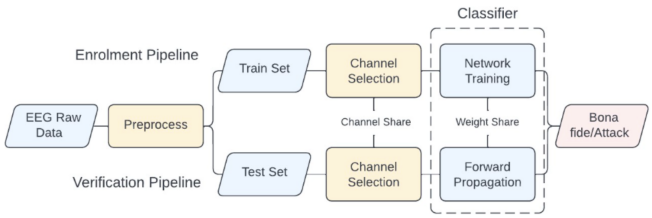

Fig. 1. **Flowchart of the Bingkun et al. framework. (Sourced from [9], fig 1).**

3.4 Comparison of Frameworks

We found that the framework proposed by Bingkun Wu et al. [9] better suits our objectives than the MindID system [8] due to several reasons:

Data Utilization Flexibility: This framework handles all frequency bands, allowing for a comprehensive use of the EEG spectrum and more specifically, it allows for the study of Beta waves. In contrast, MindID primarily focuses on delta wave optimization.

Channel Selection Efficiency: Its method of channel selection simplifies EEG processing by emphasizing the most relevant electrodes, reducing equipment demands. This is more efficient than MindID's approach. Additionally, it is easier to test different electrodes with the Bingkun et al. code as it is a matter of modifying the initial channel selection algorithm.

Performance: When tested against standard datasets, this framework outperformed several leading techniques, including MindID. Moreover, Bingkun et al. tested their framework over two different publicly available datasets.

Explainability of the Model: Its simpler deep learning structure offers clearer insights into model decisions. It's notably less complex than MindID's attention-based RNN, making it potentially more understandable.

Although MindID is a valuable tool in EEG authentication, Bingkun Wu et al.'s framework better matches our needs, offering promising prospects for both research and practical applications.

3.5 Original Classifier Deep Learning Architecture

Before continuing with the explanation of the experiments it is important to understand the architecture of the classifier as it will be slightly modified for the purpose of doing the experiments.

The original deep learning architecture presented by Bingkun Wu et al. [9] can be divided into four main stages which can be seen in figure 2:

1) **Filter-bank Spectral Decomposition**: Multiple narrow-band filters decompose the EEG data into different frequency bands.
2) **Spatial Convolution**: These bands are then input into a convolutional network to obtain multi-band characteristics. A convolution kernel slides along the channel dimension to establish communication between feature distributions from different electrodes.
3) **Mixed Temporal Feature Extraction**: Features are extracted by combining two types of temporal characteristics. These are inferred from a variance layer and a standard convolution layer along the time axis.
4) **Classifier**: Finally, the feature map is flattened and passed through a fully connected layer for classification.

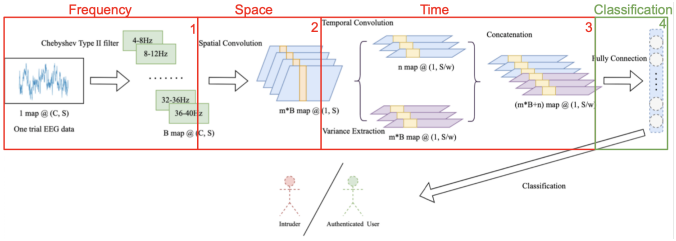

Fig. 2. **Deep learning architecture of the Bingkun et al. framework classifier. (Sourced from [9], fig 3).** Original description: "Modified Network architecture. (C: number of channels, S: number of samples, B: number of frequency bands, m: number of spatial kernels per frequency band, n: number of temporal convolution kernels in total, w: length of temporal convolution kernel)").

4 EXPERIMENTS

In this section, we will discuss two experiments we conducted by making modifications to the original framework. First, we'll provide details on the subjects and the data used for these experiments. Then, we'll delve into the specifics of each experiment, explaining the methodology and objectives.

4.1 Subjects and Data

Bingkun et al. [9] identified two primary attack scenarios for the authentication system:

- **Insider Attack:** An attack scenario where the attacker is a subject already present in the training dataset. Essentially, the system is already familiar with these attackers since they are other users of the same system. During the model training for a specific subject, data from these subjects is considered as potential attacks.
- **Outsider Attack:** In this case, the attacker is a subject who was not included in the training phase. This means the system is unaware of the characteristics of the attacker's data, making them entirely foreign entities.

For our experiments, we primarily concentrated on the *insider attack*. This decision was influenced by a couple of key factors. First, our preliminary experiments showed that insider attacks consistently demonstrated superior performance metrics. Second, we recognized the value of testing our system's ability to distinguish between authorized users and impostors who are already registered in the system. Such a test presents a realistic evaluation, given that distinguishing between insiders is a genuine challenge in many practical scenarios.

Exploring *outsider attacks* is a promising avenue for future research. Investigating this aspect could offer deeper insights into the system's defense against wholly unknown threats, providing a more realistic assessment of its robustness for deployment.

In our study, we used the same 10 subjects that were involved in the experiments of the original article, ensuring consistency in data utilization. All data for the subjects were extracted from the *physiodataset* (a comparison between the two datasets used by Bingkun Wu et al. can be found in fig 3). This dataset was favored due to its vast diversity in terms of subject count and sample volume, an aspect detailed further in Bingkun Wu et al.'s work [9].

	Number of subjects	Channels/ Electrodes	Sampling rate	Trials	Total samples per trial	Notes
Physionet	109 Subjects	64 Channels	160 Hz	6 runs of 15 trials: 90x109= 9810 total trials	640 time points (4s)	More subjects = more robust evaluation
BCI IV	9 Subjects	22 Channels	Up to 250 Hz	6 runs of 48 trials: 288*9= 2592 total trials	1125 time points (4.5s)	Longitudinal evaluation: (2 sessions on diff days)

Fig. 3. **Comparison between the two public datasets used by to evaluate the framework**

4.2 Experiment 1: Tests with the Beta Frequency Band

The main hypothesis tested here was that Beta Waves can suffice for having a good performance in EEG authentication while maintaining the original number of electrodes. In this experiment, we utilized the subjects' data similarly to the original article, with the notable exception of the number of frequency bands used. We maintained the use of 10 electrodes, or channels, determined to be optimal by the original algorithm. These optimal channels are depicted in figure 5. Our primary variation for this experiment concerns the frequency bands employed for model training.

The EEG signal consists of several important frequency bands: δ (1-4Hz), θ (4-8Hz), α (8-12Hz), and β (12-30Hz). We conducted tests by selectively muting certain bands to ascertain the significance of each:

- Retaining only α and β bands, which span $8-32Hz$, resulting in 6 bands.
- Keeping just the α band, focusing on $8-12Hz$, resulting in 1 band.
- Retaining only the β band, centered around $14-32Hz$, yielding 5 bands.

To facilitate these tests, adjustments to the framework were necessary. The original setup was intended for 10 channels and 9 bands, so the classifier section (refer to figure 4) underwent most of the alterations. A key modification involved the "Filter-bank Spectral Decomposition" phase, detailed in section 3.5 and highlighted in the red "Frequency" square in figure 2. Due to changes in data dimensions, it was also imperative to adjust certain neural network parameters.

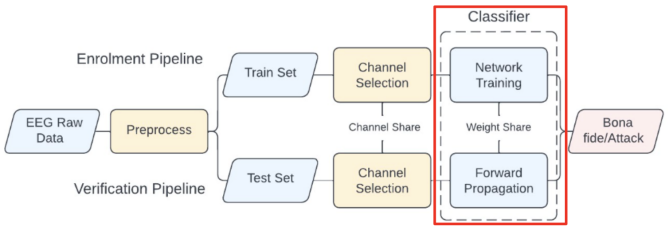

Fig. 4. **Classifier section of the flowchart modified by experiment 1**. Image sourced from [9], fig 1.

4.3 Experiment 2: Tests with Frontal Cortex Electrodes

The primary objective of this experiment was to investigate the premise that reading beta waves from the frontal cortex is sufficient for an effective authentication system. Specifically, we wanted to explore the feasibility of using EEG data exclusively from the forehead—either through the β waves or across all frequency bands—for authentication purposes.

The algorithm in the original model prioritized the selection of 10 channels: Fp1, FpZ from the frontal cortex; C1, CP1, CP2, CP3, CP4, CP5, CPZ from the parietal lobe over the motor cortex; and O2 from the occipital lobe, as visualized in figure 5.

In this experiment, we refined our channel selection to exclusively focus on three frontal cortex electrodes: Fp1, FpZ, and Fp2, as shown in figure 6. We then computed the Equal Error Rates (EERs) by training and testing the models using data from these three frontal channels, specifically targeting the five beta frequency bands.

Additionally, to evaluate the impact of the frequency spectrum, we determined the EERs using all nine bands while still restricting the data to the three frontal channels. This allowed us to draw comparisons and discern any potential differences between relying on beta waves alone versus utilizing the entire spectrum with just the frontal channels.

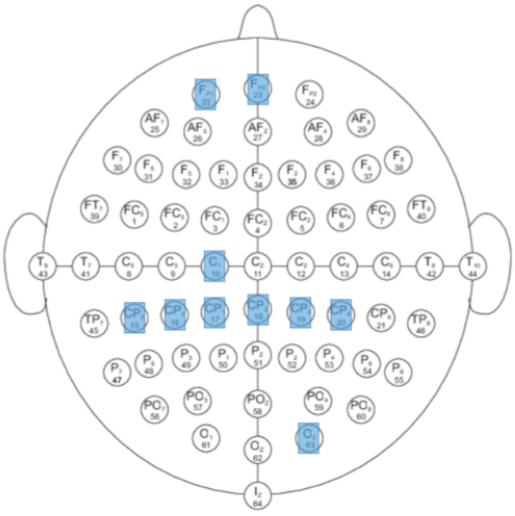

Fig. 5. Selected Electrodes in the International EEG 10-10 System, as determined by the algorithm from [9].

For this experiment, we employed two baselines for comparative analysis. The first baseline remained consistent with that used in Experiment 1, maintaining its unaltered state. The second, however, utilized the five Beta frequency bands across ten channels. The inclusion of this additional baseline stems from the hypothesis that if a subject demonstrates a low EER using both the Beta Band and the full frequency spectrum, similar performance should be evident when the setup is limited to only the Beta frequency and the three frontal electrodes. This is underpinned by the assumption that the frontal electrodes can adequately capture beta waves and its intrinsic biometric information.

To achieve this, necessary modifications were made to the framework, particularly in the channel selection section (shown in figure 7). As with the first experiment, the altered data dimensions necessitated adjustments to the neural network parameters.

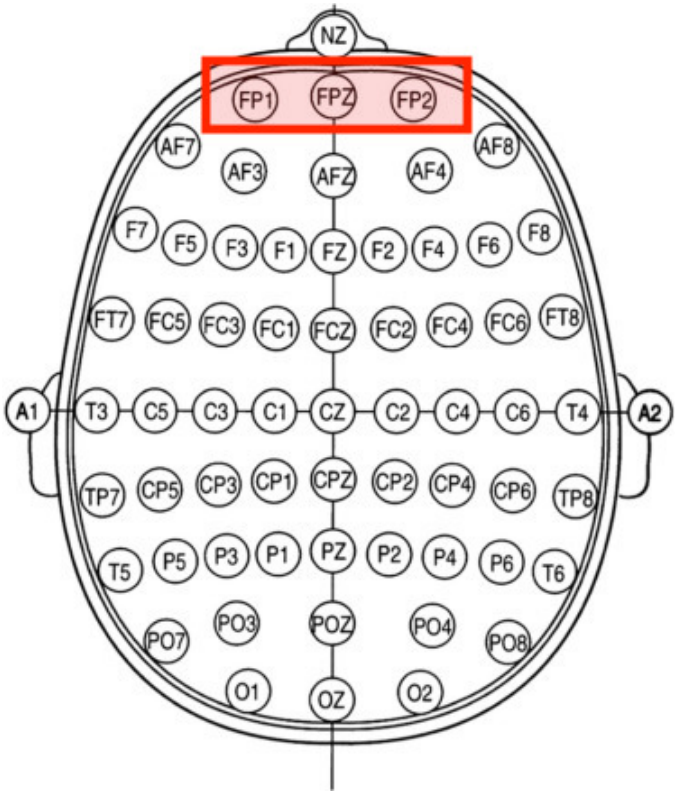

Fig. 6. Frontal Cortex electrodes targeted in the second experiment. Image sourced from [12].

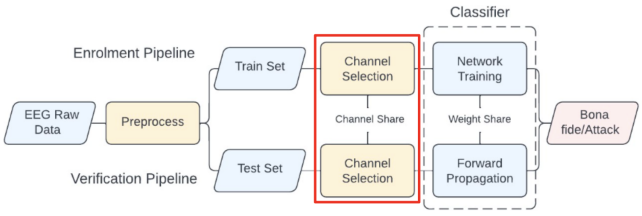

Fig. 7. Modified channel selection section for the second experiment, as adapted from [9].

5 Results

For each experiment, the modifications were executed as explained and a model was trained for each of the 10 subjects using the same parameters stated in the original article and found in the source code provided along with the article.

5.1 Experiment 1: Beta Frequency Band Results Analysis

The cornerstone of biometric authentication performance is often represented by the EER or Equal Error Rate (notion explained towards the end of the introduction of this report). A lower EER indicates a superior performance, with an ideal scenario being an EER of zero, which while theoretically perfect, is not practically achievable.

Before testing the performance of the modified framework, a baseline needs to be computed in order to compare the results from the experiments. In the original unaltered framework, where each model was trained and tested with all the data (10 Channels and 9 Frequency Bands), the average EER recorded for the 10 subjects was 1.381. [9] contends that an EER significantly above 2% would likely not yield a robust framework. Consequently, any performance hovering around this margin might be a worthy contender for subsequent optimization. The results from the experimentation are summarized in figure 8 and discussed below.

5.1.1 Alpha and Beta Frequency Bands

Excluding gamma from our analysis, we assessed the combined performance of alpha and beta frequency bands. The resulting average EER was 2.09. Given that this result arises from analyzing two-thirds (6 out of 9) of the frequency bands, one might suggest the feasibility of sidelining gamma and still maintaining an effective EEG-based authentication system.

5.1.2 Alpha Frequency Band

When solely considering the alpha band, which accounts for 1 out of the 9 bands, we registered a significantly increased EER of 9.239. Such an elevation in EER makes it a less enticing option for standalone EEG-based authentication. However, it is noteworthy that for subjects 5 and 7, the EER was recorded as 0.0, suggesting a perfect match. This alludes to the potential variability in frequency band efficacy between individual subjects. Alpha, for instance, might be better suited for certain individuals over beta.

5.1.3 Beta Frequency Band

Our most intriguing results arose from the beta frequency band analysis. Limiting our examination to this band alone, which represents 5 out of the 9 bands, yielded an EER of 2.858. This outcome, albeit slightly elevated from the combined alpha and beta result, showcases the potential of the beta waves in isolation. Specifically, beta frequency outperformed the combined alpha and beta analysis for subjects 1 and 9. Furthermore, for 80% of the subjects, beta performed comparably to the combined alpha and beta analysis, with the only exceptions being subjects 4 and 10. This finding emphasizes the prospective utility of solely employing beta waves for EEG-based authentication, although more expansive studies incorporating a larger subject pool would be crucial for consolidating these findings.

Subject	1	2	3	4	5	6	7	8	9	10	**AVG**
EER % **baseline**	1.06	0.0	0.0	3.19	0.0	0.0	0.0	2.12	4.25	3.19	**1.381**
Alpha+Beta 6 Bands (8-32 Hz)	7.44	0.0	0.0	0.0	0.0	0.0	0.0	3.19	9.47	8.51	**2.019**
Alpha 1 band (8-12 Hz)	13.6	2.12	3.15	1.05	0.0	1.91	0.0	17.02	23.90	29.64	**9.239**
Beta 5 bands (12 - 32 Hz)	4.2	0.0	0.0	3.15	0.0	0.0	0.0	3.19	7.44	10.6	**2.858**

Fig. 8. **EER for every subject for the 3 experiments and for the baseline (Baseline, Alpha-Beta, Alpha and Beta) discussed in the subsection 5.1 using the insider attack testing modality**

5.2 Experiment 2: Frontal Electrodes Result Analysis

The computed EERs for our experimentation are presented in fig 9, specifically on the third and fourth lines. Our initial hypothesis, which posited that capturing Beta waves from frontal electrodes would suffice for EEG authentication, appears to be immediately negated. When exclusively utilizing frontal lobe electrodes, both experimental setups exceeded our predefined 2% EER threshold: the setup using three channels across nine bands yielded an average EER of 6.896%, whereas the three channels with five bands setup had an average EER of 10.716%. Furthermore, none of the subjects achieved an EER below the 2% threshold.

Our secondary hypothesis proposed a correlation: if a model, when trained only on the Beta band (5 bands) with ten channels, achieved a low EER, then it should perform similarly when constrained to the Beta band and three channels. This was rooted in the notion that three frontal electrodes might sufficiently capture the requisite biometric data. Yet, as the data suggests, models using only the three frontal electrodes do not indicate effective performance. Predictably, for subjects trained on the Beta band, ten channels always outperformed the three frontal channels.

Interestingly, when the scope was restricted to just frontal channels, encompassing more bands improved performance, albeit marginally (EER of 6.896% for nine bands versus 10.716% for five bands). Noteworthy exceptions include subjects 6 and 5. For subject 6, the EER remained consistent regardless of the band range. In contrast, subject 5 witnessed a drop in EER from 6.38% to 3.15%—a significant reduction, though still surpassing the 2% threshold. This discrepancy raises the possibility that with an expanded sample size, some subjects might align with our primary hypothesis.

However, the nature of the data provides a plausible explanation for the observed results. Given that data capture occurred during a Motor Imagery task, it's logical to assume that the bulk of biometric information arises from the motor cortex's electrical activity. Investigating the sole use of motor cortex electrodes, however, was beyond this internship's scope.

Note to the Reader: Anomalies were observed in the hardware used to compute the EER for this experiment. Specifically, identical values (accurate to the 8th decimal digit) were produced for different subject models—particularly for values 6.38 and 3.19, which affect 6 out of the 10 subjects. Given this inconsistency, it is advised to approach the results of this experiment with caution. Independent verification is recommended until such time as this note is removed and updated values are provided in a subsequent version of this report.

Subject	1	2	3	4	5	6	7	8	9	10	**AVG**
Baseline: Full bands Full Channels 9B 10Ch	1.06	0.0	0.0	3.19	0.0	0.0	0.0	2.12	4.25	3.19	**1.381**
Beta Bands Full Channels 5B 10Ch	4.2	0.0	0.0	3.15	0.0	0.0	0.0	3.19	7.44	10.6	**2.858**
Full bands **Frontal** Channels 9B 3Ch	3.19	7.36	4.25	3.19	6.38	3.19	7.36	3.19	7.45	23.40	**6.896**
Beta bands **Frontal** Channels 5B 3Ch	6.38	11.7	9.57	12.76	3.15	3.19	16.8	6.38	6.38	30.85	**10.716**

Fig. 9. **EER for every subject for experiment 2 and for two baselines discussed in the subsection 5.2 using the insider attack testing modality**

6 Conclusion

This report delves into EEG-based biometric authentication, building upon the foundational research presented by Bingkun Wu et al. in 2022. Our primary objective was to assess the impact of specific frequency bands and electrodes on the efficacy of EEG-based authentication.

Through our experiments, we have shed light on the potential benefits of exclusively utilizing beta waves for EEG-based authentication. Nevertheless, to solidify these insights, further extensive research involving a broader subject base is necessary.

A significant observation from our study is that three electrodes located in the frontal cortex might not capture

sufficient biometric information to achieve satisfactory authentication performance. However, one must consider that the data utilized was collected during a Motor Imagery task. This raises the question of whether most biometric details were inherently centered around the electrical activity in the motor cortex. Therefore, exploring datasets generated from different cognitive tasks, which do not majorly involve the motor cortex, could provide a clearer understanding of the potential of frontal cortex electrodes in EEG-based authentication.

Supplementing this report, we provide several notebooks that facilitate easy testing of both the original [9] and our modified versions of the code. These tools are envisioned to be helpful for researchers keen on further exploring this domain.

REFERENCES

[1] GUI et al, "Exploring EEG-based Biometrics for User Identification and Authentication", 2014.

[2] Huang et al, "An EEG-Based Identity Authentication System with Audiovisual Paradigm in IoT", 2019.

[3] Liang et al, "Identity Recognition Using Biological Electroencephalogram Sensors", 2016.

[4] Marcel et al, "Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation", 2007.

[5] Stergiadis et al, "A Personalized User Authentication System Based on EEG Signals", 2022.

[6] Svogor et al, "Two factor authentication using EEG augmented passwords", 2012.

[7] Zeynali et al, "EEG-based single-channel authentication systems with optimum electrode placement for different mental activities", 2019.

[8] Zhang et al., "MindID: Person Identification from Brain Waves through Attention-based Recurrent Neural Network", 2021.

[9] Bingkun Wu et al. "Towards Enhanced EEG-based Authentication with Motor Imagery Brain-Computer Interface", 2022.

[10] Principles of neural science. Eric R. Kandel, John Koester, Sarah Mack, Steven Siegelbaum (6th ed.). New York. 2021. p. 1450.

[11] Juri D. Kropotov, Chapter 3 - Beta Rhythms, Editor(s): Juri D. Kropotov, Quantitative EEG, Event-Related Potentials and Neurotherapy, Academic Press, 2009, Pages 59-76, ISBN 9780123745125, https://doi.org/10.1016/B978-0-12-374512-5.00003-7. (https://www.sciencedirect.com/science/article/pii/B9780123745125000037)

[12] Marc R. Nuwer, 10-10 electrode system for EEG recording, Clinical Neurophysiology, Volume 129, Issue 5, 2018, Page 1103, ISSN 1388-2457, https://doi.org/10.1016/j.clinph.2018.01.065. (https://www.sciencedirect.com/science/article/pii/S1388245718300907)

[13] Brigham, Katharine and Kumar, B.. (2010). Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy. 2010 4th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2010. 1 - 4. 10.1109/ICBBE.2010.5515807.

[14] Du, Y., Xu, Y., Wang, X. et al. EEG temporal–spatial transformer for person identification. Sci Rep 12, 14378 (2022). https://doi.org/10.1038/s41598-022-18502-3

[15] A. Valsaraj, I. Madala, N. Garg, M. Patil and V. Baths, "Motor Imagery Based Multimodal Biometric User Authentication System Using EEG," 2020 International Conference on Cyberworlds (CW), Caen, France, 2020, pp. 272-279, doi: 10.1109/CW49994.2020.00050.

[16] Z. Mao, W. X. Yao and Y. Huang, "EEG-based biometric identification with deep learning," 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER), Shanghai, China, 2017, pp. 609-612, doi: 10.1109/NER.2017.8008425.

[17] C. R. Hema, M. P. Paulraj and H. Kaur, "Brain signatures: A modality for biometric authentication," 2008 International Conference on Electronic Design, Penang, Malaysia, 2008, pp. 1-4, doi: 10.1109/ICED.2008.4786753.

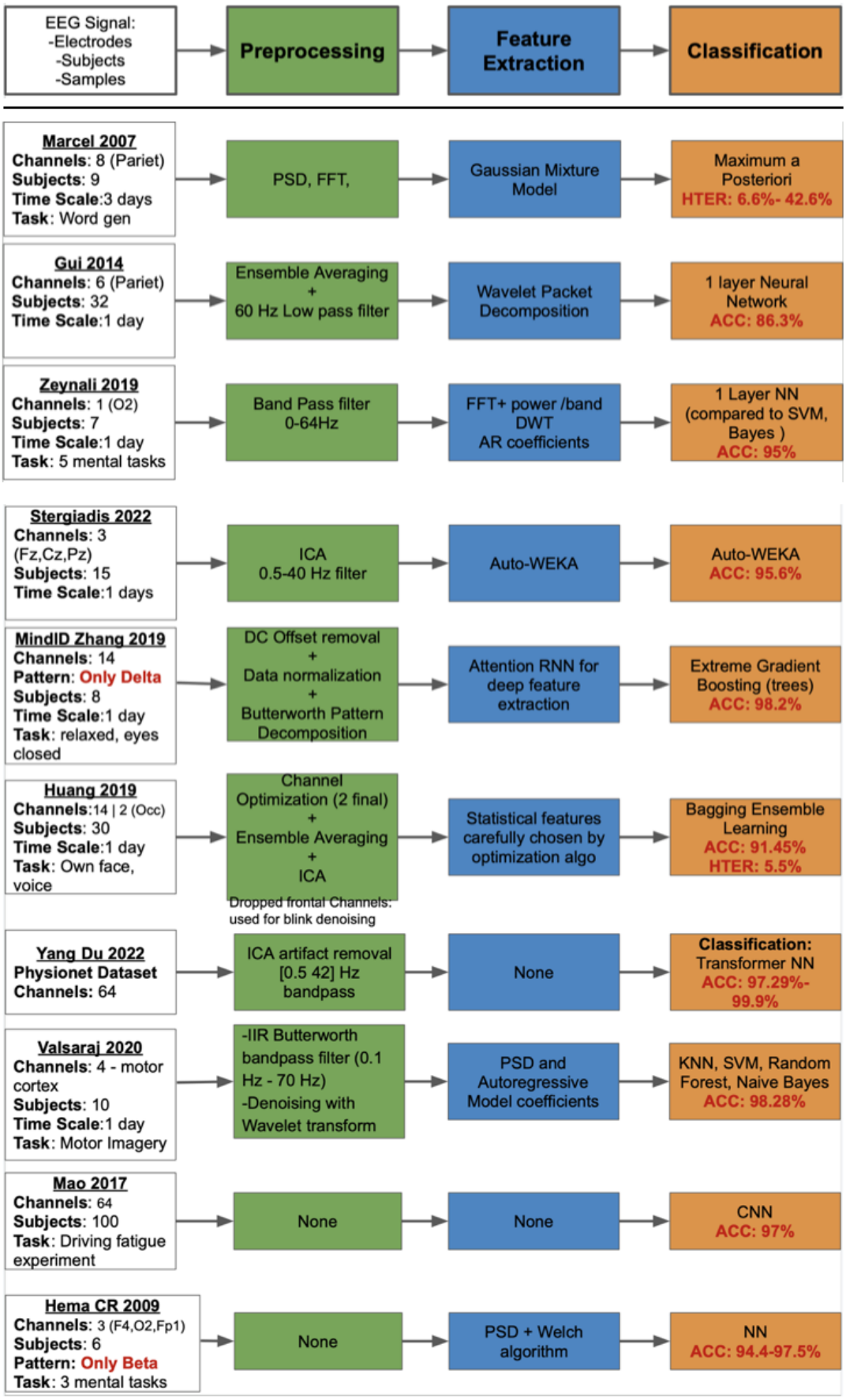

Fig. 10. **Various frameworks identified from the literature review**. Each row represents a framework. Details per row are as follows: White box - paper reference and data details; Green boxes - preprocessing techniques; Blue boxes - feature extraction methods; Orange boxes - classifier details, with performance indicated in red.