electronics

Article

Efficient Embedded System for Drowsiness Detection Based on EEG Signals: Features Extraction and Hardware Acceleration

Aymen Zayed [1,2,3,*], Emanuel Trabes [1,4], Jimmy Tarrillo [1,5], Khaled Ben Khalifa [2,6] and Carlos Valderrama [1]

1 Service d'électronique et de Microélectronique, University of Mons, 7000 Mons, Belgium; emanuel.trabes@umons.ac.be (E.T.); jtarrillo@utec.edu.pe (J.T.); carlos.valderrama@umons.ac.be (C.V.)
2 Research Laboratory of Technology and Medical Imaging—LTIM—LR12ES06, Faculty of Medicine of Monastir, Monstir 5000, Tunisia; khaled.benkhalifa@issatso.rnu.tn
3 National Engineering School of Sousse, University of Sousse, Sousse 4054, Tunisia
4 Department of Electronics, Universidad Nacional de San Luis, San Luis 5700, Argentina
5 Electrical and Mechatronic Department, Universidad de Ingenieria y Tecnologia, Barranco 15063, Peru
6 Higher Institute of Applied Science and Technology of Sousse, University of Sousse, Sousse 4003, Tunisia
* Correspondence: aymen.zayed@umons.ac.be

Abstract: Drowsiness detection is crucial for ensuring the safety of individuals engaged in high-risk activities. Numerous studies have explored drowsiness detection techniques based on EEG signals, but these have typically been validated on computers, which limits their portability. In this paper, we introduce the design and implementation of a drowsiness detection technique utilizing EEG signals, executed on a Zynq7020 System on Chip (SoC) as part of a Pynq-Z2 module. This approach is more suitable for portable applications. We have implemented the Discrete Wavelet Transform (DWT) and feature extraction functions as intellectual property (IP) cores, while other functions run on the ARM processor of the Zynq7020.

Keywords: EEG signals; drowsiness detection; Zynq SoC; PynqZ2

check for updates

Academic Editor: Jichai Jeong

Received: 29 November 2024
Revised: 12 January 2025
Accepted: 20 January 2025
Published: 21 January 2025

Citation: Zayed, A.; Trabes, E.; Tarrillo, J.; Ben Khalifa, K.; Valderrama, C. Efficient Embedded System for Drowsiness Detection Based on EEG Signals: Features Extraction and Hardware Acceleration. *Electronics* **2025**, *14*, 404. https://doi.org/10.3390/electronics14030404

1. Introduction

Decreased vigilance [1] is a significant issue across various sectors, including office work, industry, healthcare [2], the military, and construction [3]. This phenomenon is a leading cause of accidents in these environments, often occurring after a state known as passive sleepiness or drowsiness [4]. Drowsiness is a complex state to define, as its manifestation varies from person to person, making its detection particularly challenging. This complexity has led to the development of a range of drowsiness detection methods [5], including vehicle-based systems, which primarily rely on analyzing individual driving habits to detect drowsiness. The two most commonly used indicators in this approach are the steering wheel angle (SWA) and the vehicle's lane position. Despite their potential, these methods are among the least frequently employed due to their inherent limitations. A key challenge is the inability to establish a standardized driving model for drowsy drivers. This challenge is compounded by the considerable variability in driving habits between individuals, making it difficult to extract reliable and consistent features that accurately characterize drowsy driving behavior. Motion-based methods for drowsiness detection focus on analyzing facial expressions, a significant aspect of motion detection. While these methods are considered reliable, they face several limitations. A major drawback is their sensitivity to lighting conditions, which can significantly affect detection accuracy. Additionally, the variability in facial expressions associated with drowsiness across individuals complicates the development of robust models. Furthermore, a critical limitation of these methods is their inability to detect drowsiness before the appearance of physical signs, presenting a significant challenge for effective early detection. On the other hand, physiological signal-based approaches, such as those using EEG [6], EOG [7], ECG [8], and EMG [9] signals, are often considered the most reliable for drowsiness detection, as they utilize direct indicators of drowsiness derived from individuals. However, these methods also have notable limitations. A significant drawback is their sensitivity to artifacts, which can compromise the accuracy of signal analysis. Furthermore, these methods typically require a large number of electrodes to improve detection precision, posing practical challenges in terms of usability and portability.

Among these methods, physiological signal-based detection, particularly using EEG, has gained prominence due to its ability to capture subtle changes in brain activity that precede the onset of drowsiness [10]. Numerous studies in the field have demonstrated that an individual's vigilance level is closely linked to the activation state of the central nervous system, which can be effectively monitored through EEG signals [11–13]. What sets EEG-based methods apart is their precision in detecting drowsiness before any physical signs, such as eyelid closure or changes in body posture, become evident. This early detection capability is critical in preventing accidents and ensuring safety in high-risk environments.

EEG signal-based drowsiness detection systems face significant limitations when it comes to adaptability in real-world conditions [14]. These challenges stem from their reliance on a large number of electrodes to ensure accurate detection, as well as the bulky size of the acquisition tools, the computational demands of the detection system, and the extended detection time, all of which contribute to high power consumption. These constraints make it difficult to implement traditional EEG-based systems in portable or wearable applications [15].

In light of these limitations, developing a fully embedded, portable drowsiness detection system based on EEG signals with a reduced number of electrodes is a promising and intelligent solution. Such a system would enhance adaptability, allowing for more seamless integration into daily life, while still maintaining detection accuracy. Additionally, reducing hardware requirements and energy consumption would make the system more efficient and practical for real-time, on-the-go monitoring, opening up new possibilities for its application in various environments, from workplaces to transportation.

The PYNQ-Z2 [16] is a development board based on the Xilinx Zynq-7000 SoC FPGA, which combines a dual-core ARM Cortex-A9 processor with a reconfigurable FPGA fabric. Designed to make FPGA reconfiguration more accessible, it allows users to create embedded systems and sophisticated hardware applications using Python scripts through the PYNQ (Python Productivity for Zynq) environment.This simplifies the learning process, enables faster prototyping, and is particularly suited for artificial intelligence and digital signal processing (DSP) projects [17–19]. This combination offers great flexibility and high performance for applications requiring intensive real-time processing, with the ability to design optimized data pipelines.

For these reasons, implementing a drowsiness detection approach on an embedded platform such as the PYNQ-Z2 presents a promising solution. This strategy enables the development of a portable, compact, and efficient system, well-suited to the constraints of real-world environments. In this context, this study focuses on evaluating performance trade-offs between detection accuracy, power consumption, and resource utilization within a drowsiness detection system based on a co-design approach. Specifically, it compares the fixed-point and floating-point versions of the Discrete Wavelet Transform (DWT) implementation. These implementations, designed using High-Level Synthesis (HLS), allow for a systematic analysis of the impact of these key parameters to optimize the design and integration of an embedded drowsiness detection system.

The main objective of this study is to propose an embedded system for drowsiness detection based on EEG signals, utilizing a reduced number of electrodes to minimize inter-subject variability. This is achieved by leveraging time–frequency features derived from the Discrete Wavelet Transform (DWT) IP, implemented on the ZYNQ-7020 programmable logic. Additionally, the study explores the most adaptable DWT architecture among Float 32, Float 16, Fixed 32, and Fixed 24, focusing on factors such as power consumption, hardware resource utilization, and execution time.

The remainder of this paper is structured as follows: Section 2 presents related works, highlighting recent studies on drowsiness detection. Section 3 outlines the methodology, detailing the key steps involved in EEG-based drowsiness detection, including database selection, preprocessing, and classification. A clear presentation of the PYNQ-Z2 platform and the implementation details of the proposed system are provided in Section 4. Experimental results, including detection accuracy, execution time, and energy consumption, along with a discussion of the findings, are presented in Section 5. Finally, the conclusions and future perspectives of this work are covered in Section 6.

2. Related Works

This section will review research on embedded systems designed for drowsiness detection, with a particular focus on the effectiveness and precision of EEG signals in identifying the early stages of drowsiness. The discussion will highlight the advantages of EEG-based methods for capturing subtle changes in brain activity associated with drowsiness, which often appear before physical signs manifest. Additionally, this section will address the limitations found in existing systems, such as the reliance on a high number of electrodes and the challenges of adapting these systems to real-life conditions. Several studies further concentrate on the development of embedded systems for drowsiness detection.

In [20], Maior et al. propose a real-time system for detecting drowsiness in drivers by using the Eye Aspect Ratio (EAR) as a primary feature. The EAR is a metric used to quantify the degree of eye closure, which can indicate drowsiness or fatigue based on eyelid movement. The system uses a video camera to capture facial images, particularly focusing on the eye region, and calculates the EAR by measuring the distances between specific points around the eyes. A decrease in EAR over time is indicative of eyelid closure, which corresponds to drowsiness onset. The authors employ a machine learning classifier, typically a Support Vector Machine (SVM) or Random Forest, to classify the detected EAR values into two states: awake or drowsy. The system is designed to operate in real time, with a response time of approximately 1 s. The data processing includes an initial preprocessing step to extract the facial landmarks, followed by EAR calculation and classification. The system achieves an accuracy of 94% and a false positive rate of 3% and a false negative rate of 5%. The video resolution used for detection is typically 640×480 pixels, and the system is designed to function on common computing platforms, such as smartphones or embedded devices, with low computational requirements and 0.75 Watts of power consumption.

Another line of research has focused on drowsiness detection using embedded systems that rely on facial expressions [21]. Florez et al. propose an embedded system designed for real-time driver drowsiness detection using a combination of visual analysis techniques focused on both the eyes and the mouth. The system utilizes Convolutional Neural Networks (CNNs) to process facial images and extract features that can indicate signs of drowsiness, specifically eye closure and mouth movement. Key to the system's approach is the use of the Mouth Aspect Ratio (MAR), which quantifies the vertical and horizontal distances between specific points on the mouth to detect yawning—a common sign of drowsiness. The authors combine the MAR with the Eye Aspect Ratio (EAR), which detects eye closure. By analyzing these two facial features together, the system is able to improve the accuracy of drowsiness detection. The CNN model is trained on a large dataset of labeled facial images, with preprocessing steps that include normalizing and resizing the images to a uniform size of 224 $\times$ 224 pixels before feeding them into the network. The authors report that their system achieves an accuracy of 95%, with a false positive rate of 4% and a false negative rate of 6%. The response time of the system is about 1–2 s, which is suitable for real-time applications. The system is embedded on a low-power device, allowing it to run efficiently with low power consumption, approximately 22 Watts, making it ideal for use in automotive environments.

The work mentioned above focuses on developing embedded systems that primarily rely on facial expressions to detect drowsiness. While this approach can be effective in certain scenarios, it is highly sensitive to lighting conditions, which can impact the accuracy of detection. Furthermore, this method typically identifies drowsiness only after physical signs, such as eye closure or head nodding, have already appeared. In contrast, EEG signals offer a significant advantage by detecting early signs of drowsiness before any physical manifestations occur. EEG-based systems can capture subtle changes in brain activity that precede visible indicators of drowsiness, enabling earlier and potentially more effective interventions. This makes EEG signals a valuable component for developing embedded drowsiness detection systems that provide a proactive solution, enhancing detection accuracy and applicability across various real-world environments.

In [22], Abbas et al. provide a detailed comparison of various fatigue detection methods using multi-sensor systems, smartphones, and cloud platforms. EEG-based systems, which monitor brain activity, achieve an accuracy of 85% to 95% depending on the number of electrodes used (ranging from 10 to 32 electrodes) and the quality of signal preprocessing. They also have a response time of 1–2 s, but consume more power, averaging 15% to 30% of a smartphone battery. ECG-based systems, which rely on heart rate variability, have a slightly lower accuracy of 80% to 90% and a response time of 2–3 s, but they are more power-efficient, consuming 0.41 Watts of the smartphone battery life. Eye-tracking systems, which detect changes in blink rate and gaze direction, offer an accuracy of 80% to 90% with a 1 s response time and moderate power consumption of 0.452 Watts. Accelerometer-based methods, which track vehicle dynamics, show lower accuracy (70% to 85%) and have a longer response time of 2–4 s, but are highly power-efficient, consuming only 0.52 Watts of the battery. When multiple sensors are combined in a multi-sensor fusion system, the accuracy improves to 95% or higher, with a response time between 1 and 3 s. However, these systems consume more power, averaging 0.6 Watts of the smartphone battery, due to the integration of various data streams. The challenges across all methods include sensor integration, real-time data transmission, and power consumption, with multi-sensor fusion systems offering the most robust solutions but also facing increased computational complexity. This work allows us to compare various drowsiness detection methods and conclude that EEG-based detection is the most cost-effective approach, achieving an accuracy of 85% to 95%, depending on the number of electrodes used. Additionally, EEG systems demonstrate a fast response time of approximately 2 s and consume only 1.2 Watts of a smartphone battery. These attributes make EEG-based detection not only efficient in terms of accuracy but also practical for mobile and embedded applications, offering a viable solution for real-time drowsiness monitoring with minimal impact on device performance.

In [23], Gangadharan K et al. propose a drowsiness monitoring system based on a portable, wireless EEG device equipped with eight electrodes for easy setup and enhanced mobility. The EEG signals are captured and processed in real time to differentiate between drowsy and alert states. Before feature extraction, the signals are filtered using a band-pass filter [0.5; 100] Hz to remove artifacts. The study leverages various EEG features, including alpha, beta, and theta wave patterns, and employs a support vector machine (SVM) model for classification, achieving a reported accuracy of 78.3%.

In [24], Arif et al. propose a drowsiness detection system that uses EEG signals to analyze driving fatigue. The system utilizes a total of 19 electrodes placed on the scalp, capturing neurophysiological data processed in segments of 1 s. The EEG signals undergo preprocessing with a band-pass filter operating between 0.5 Hz and 40 Hz to eliminate artifacts and enhance the relevant frequency components. Feature extraction focuses on spectral signatures from specific frequency bands (delta, theta, alpha, and beta), followed by classification using a Support Vector Machine (SVM) model. The reported accuracy of the system is 85.6%, demonstrating its effectiveness in detecting drowsiness based on EEG data.

Wang et al. [25] investigate the application of EEG signals to detect mental fatigue among construction workers. The study utilizes 16 electrodes positioned according to the 10–20 international system, capturing EEG data in segments of 5 s. The preprocessing phase involves filtering the signals with a band-pass filter in the range of 0.5 Hz to 50 Hz to remove noise and artifacts. Feature extraction is performed using statistical metrics, including mean, variance, and entropy. The extracted features are then classified using a Convolutional Neural Network (CNN) model, achieving an accuracy of 85.2% in identifying mental fatigue. This research highlights the potential of EEG and deep learning techniques in enhancing the well-being and safety of workers in the construction industry.

In their work [26], Ren et al. propose a driving fatigue detection method using a radial basis function (RBF) in a two-level learning hierarchy. They use an EEG dataset recorded with 32 electrodes placed according to the international 10–20 system, covering various brain regions. The EEG signals are filtered with a Butterworth band-pass filter ranging from 1 to 45 Hz to remove external noise and isolate relevant frequencies. The filtered signals are then segmented into 10 s windows, allowing for more stable feature extraction. Temporal and spectral features are extracted using discrete wavelet transform (DWT). In the RBF learning hierarchy, the first level distinguishes between wakefulness and light fatigue, while the second level refines the classification between moderate and severe fatigue, achieving an overall 92% accuracy. RBF model parameters, such as the Gaussian function and sigma, are adjusted at each level to optimize classification and minimize errors.

Below (Table 1) is a compact classification and comparative table summarizing the key attributes of the reviewed works.

Table 1. Comparison of drowsiness detection systems.

Study	Method	Features	Accuracy	Electrodes	Response	Advantages	Limitations
Maior [20]	EAR	Eye closure (EAR)	94%	N/A	∼1 s	Low computational needs	Sensitive to lighting
Florez [21]	EAR + MAR (CNN)	Eye and mouth metrics	95%	N/A	1–2 s	Combines multiple indicators	Detects physical signs late
Abbas [22]	Multi-sensor fusion	EEG, ECG, eye-tracking, etc.	95%+	10–32 (EEG)	1–3 s	High robustness	High energy consumption
Gangadharan [23]	Portable EEG (SVM)	Alpha, beta, theta waves	78.3%	8	N/A	Easy setup, mobility	Lower accuracy
Arif [24]	EEG (SVM)	Spectral bands (delta–beta)	85.6%	19	∼1 s	Detects early drowsiness	High electrode count
Wang [25]	EEG (CNN)	Statistical metrics	85.2%	16	∼5 s	Deep learning potential	Longer response time
Ren [26]	EEG (RBF Hierarchy)	Temporal and spectral features	92%	32	∼10 s	Differentiates fatigue levels	Complex setup, many electrodes

The studies presented have achieved promising results in drowsiness detection using EEG signals, demonstrating the effectiveness of EEG-based features as reliable indicators of drowsiness. However, most existing systems employ a high number of electrodes, which limits their practical applicability in real-world scenarios. Such configurations are often cumbersome, require specialized setups, and are challenging to adapt to everyday environments where drowsiness detection is needed. These limitations make it difficult for conventional systems to support wearable or embedded applications that could be integrated seamlessly into users' daily lives. Furthermore, the high electrode count in these systems complicates ease of use, potentially reducing the comfort and mobility of users over extended periods. To address these challenges, the development of an embedded drowsiness detection system that maintains high accuracy with a reduced number of electrodes with a low power consumption is essential. A practical solution would involve designing a system capable of reliably detecting drowsiness in varied and dynamic conditions, such as workplaces or transport environments, where EEG setups must be both efficient and unobtrusive.

3. Methodology

Electroencephalographic (EEG) signals are a highly effective tool for drowsiness detection, owing to their ability to provide detailed insights into the activation state of the nervous system, which closely correlates with an individual's level of vigilance. However, EEG signals are also highly sensitive to various artifacts [27], including muscle activity, eye movements, and environmental noise, which can compromise the accuracy of detection if not properly addressed. As a result, it is essential to extract well-suited features from the EEG data to accurately identify drowsiness [28].

To achieve reliable detection, the process involves several key phases, including data acquisition [29], pre-processing [30], and classification [31]. Data acquisition ensures the collection of high-quality signals from multiple EEG channels, while pre-processing is crucial for removing noise and artifacts, ensuring that only the most relevant brainwave patterns are analyzed. Finally, classification involves identifying drowsiness by analyzing the extracted features using advanced machine learning algorithms or neural networks. In this work, this general processing chain is implemented on the PYNQ-Z2 platform, integrating each phase in a hardware–software framework for efficient, drowsiness detection.

Figure 1 illustrates these critical stages that make up a complete drowsiness detection system based on EEG signals, highlighting the flow from signal acquisition through to the final detection of the drowsiness state. This multi-stage approach ensures a robust and precise system for real-time monitoring of drowsiness, which is crucial for applications in safety-critical environments.

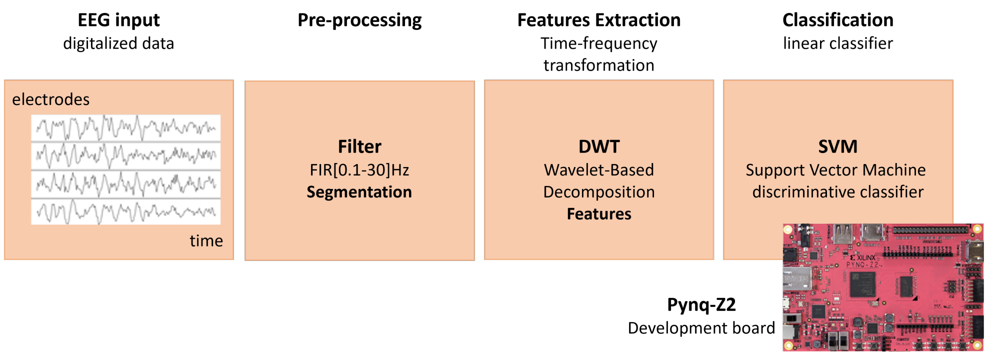

Figure 1. Principal stages for drowsiness detection on PYNQ-Z2.

3.1. EEG Data

In this study, we focus on detecting drowsiness using the ULg Multimodality Drowsiness Database (DROZY) [32], a comprehensive dataset specifically designed for such research. The DROZY dataset consists of two primary data types: video recordings and physiological signals collected from 14 subjects of different genders and ages, all of whom are free from sleep disorders and substance dependencies such as alcohol or drugs. For the purpose of this work, we concentrate solely on the physiological signals.

These physiological signals were recorded using the Embla Titanium polysomnography (PSG) system, a high-precision device that captures a variety of biosignals. The data were stored in European Data Format (EDF) files, which include five EEG channels (Fz, Cz, C3, C4, and Pz), all referenced to A1 according to the international 10–20 electrode placement system. In addition, two electrooculogram (EOG) channels measuring both vertical and horizontal eye movements were recorded, alongside one electrocardiogram (ECG) channel and one electromyogram (EMG) channel. All signals were captured at a sampling rate of 512 Hz, ensuring high temporal resolution for the analysis. Figure 2 illustrates the precise placement of each EEG electrode.

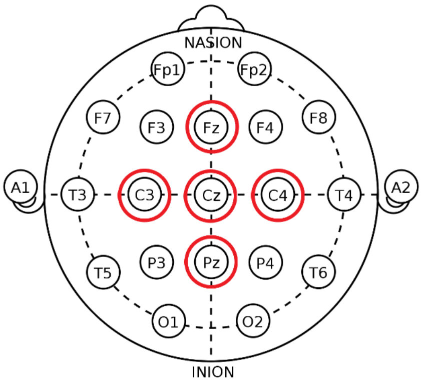

Figure 2. EEG electrodes placement.

The data acquisition protocol involves conducting three psychomotor vigilance tests (PVTs) for 14 subjects, with 10-min recordings of physiological signals. Participants are required to maintain a regular sleep pattern during the week prior to the first PVT. After the initial PVT, they remain awake until the completion of the protocol.

The first test (PVT 1) takes place on the first day at 10:00 AM and lasts one hour. The second test (PVT 2) is conducted at 8:30 PM the same day, with participants prohibited from consuming coffee or energy drinks beforehand. The final test (PVT 32) occurs the following day, between 3:30 and 4:00 AM.

Polysomnography is used to record various physiological signals. In this study, the focus is on five EEG channels: Fz, Cz, C3, C4, and Pz (Figure 3). To label the alertness levels, the Karolinska Sleepiness Scale (KSS) [33] was used, a validated nine-point scale (Table 2) that ranks states of vigilance. For this study, we adopt a binary classification approach to drowsiness detection: levels 0 through 4 on the KSS scale are categorized as the alert state (labeled as ‘0’), while levels 5 through 9 are classified as the drowsy state (labeled as ‘1’).

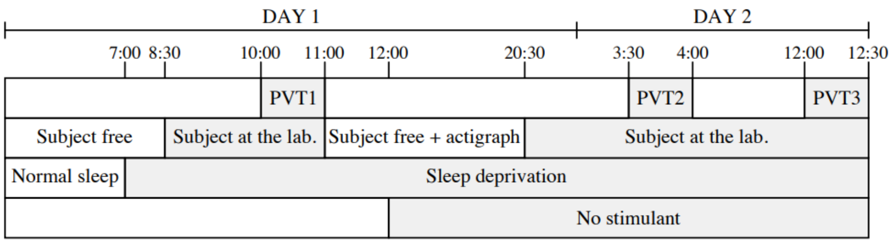

Figure 3. DROZY data acquisition protocol.

Table 2. Karolinska sleepiness scale.

Vigilance State	Scale
Extremely alert	0
Very alert	1
Alert	2
Rather alert	3
Neither alert nor sleepy	4
Some signs of sleepiness	5
Sleepy, but no effort to keep awake	6
Sleepy, but some effort to keep awake	7
Very sleepy, great effort to keep awake, fighting sleep	8
Extremely sleepy, can't keep awake	9

This division allows us to create a robust drowsiness detection system, distinguishing between alert and drowsy states based on physiological signal features. The combination of high-quality EEG, EOG, ECG, and EMG signals and precise KSS labeling provides a rich dataset for developing and testing advanced drowsiness detection algorithms. Based on all the previously mentioned factors, the DROZY database is highly adaptable for testing our embedded system. Another advantage of this database is its use of five EEG derivations, which supports our objective of developing an embedded system with a reduced number of electrodes, aiming to make it adaptable to various environments.

3.2. Preprocessing and Feature Extraction

3.2.1. Preprocessing

EEG signals are an effective tool for detecting vigilance levels due to their ability to reflect brain activity. However, these signals are highly sensitive to artifacts, both physiological, such as muscle, ocular, and cardiac movements, and non-physiological, such as electrode displacement, cable movements, and electrical interference from the acquisition system [27]. Therefore, rigorous preprocessing is essential to improve signal quality and detection accuracy.

To reduce artifacts in EEG signals, three main methods are commonly used: source decomposition [34], blind source separation [35], and filtering [36]. Source decomposition allows for analyzing the EEG signal in subcomponents, isolating artifact sources. A common technique here is Independent Component Analysis (ICA) [37], which separates signals by assuming that sources are statistically independent. Although effective for isolating artifacts related to ocular and muscular movements, ICA requires a substantial amount of data for reliable results.

Blind source separation is also employed to extract brain signals while minimizing artifact interference. However, this method risks eliminating information relevant to drowsiness detection, limiting its adaptability in this field.

Other advanced techniques also exist for denoising physiological signals, such as the Empirical Wavelet Transform (EWT), which has demonstrated significant improvements in eliminating artifacts from signals like ECG and EMG [38]. In contrast, bandpass filtering has proven particularly suitable for drowsiness detection, as demonstrated in several works [39–41]. By restricting the EEG signal to frequencies between [0.1; 30] Hz, this method reduces artifacts while preserving essential information related to vigilance states [42,43]. Therefore, an FIR bandpass filter in this frequency range is used in this study to enhance EEG signal quality while optimizing drowsiness detection.

Following filtering, a crucial step in EEG signal preprocessing is segmentation, which divides the signal into temporal segments, called epochs, facilitating vigilance analysis. Epoch length directly influences drowsiness detection accuracy. In this study, a 10 s epoch length was chosen based on prior research showing its effectiveness in drowsiness analysis [44].

3.2.2. Feature Extraction

Feature extraction is one of the most critical phases in drowsiness detection, as the system's accuracy largely depends on the quality of these features. The most commonly used features in the field of drowsiness detection generally fall into three main categories: temporal features, frequency features, and time–frequency features. According to the literature, the majority of studies on drowsiness detection focus exclusively on frequency features, overlooking the fact that this type of feature does not retain temporal information. For this reason, it seems highly beneficial to extract time–frequency features to ensure more effective and precise detection.

Among the most effective methods for extracting time–frequency features is the Discrete Wavelet Transform (DWT) [45], which allows for detailed, multi-resolution analysis of the EEG signal. Unlike Fourier Transform methods that only focus on frequency information, DWT enables simultaneous localization of both time and frequency information. This makes DWT particularly suitable for drowsiness detection, as it captures transient events within specific frequency ranges that are directly linked to changes in vigilance levels.

DWT works by decomposing the signal into different levels of resolution using a series of wavelet functions. The transformation can be mathematically represented as Equations (1) and (3):

$$A(x[n]) = \sum_k l[k] \cdot x[2n-k] \quad \text{(Low-pass filter)} \tag{1}$$

$$D(x[n]) = \sum_k h[k] \cdot x[2n-k] \quad \text{(High-pass filter)} \tag{2}$$

where $x[n]$ is the input signal $l[k]$ and $h[k]$ are the low-pass and high-pass filter coefficients, respectively. The variable n represents the time index, and k is the index used for the filter coefficients.

Each decomposition level produces approximation A and detail D coefficients [46]. The approximation coefficients represent the low-frequency components of the signal, capturing the broader structure and slower oscillations related to the overall signal trend. Conversely, the detail coefficients capture the high-frequency components, which are linked to rapid signal variations, often reflecting artifacts or sharp transitions.

By iteratively decomposing the EEG signal, DWT provides a hierarchical representation where each level corresponds to a specific frequency band. In the context of drowsiness detection, lower-frequency bands often correlate with deeper vigilance levels, while higher-frequency bands can indicate alertness. This multi-level approach is highly valuable, as it preserves temporal changes in each frequency range, enhancing the effectiveness and accuracy of drowsiness detection systems.

In this work, we will test different wavelets in order to select the most adaptable Discrete Wavelet Transform (DWT) for the drowsiness detection application. The DWTs tested in this work are Haar [47], Daubechies 4 (db4) [48], Symlet [49], Coiflet [50], and Mexican Hat wavelets [51]. Each of these wavelets will be evaluated based on their ability to capture time-frequency features in EEG signals, with an emphasis on achieving a balance between time and frequency resolution, as well as their effectiveness in drowsiness detection. The goal is to identify the wavelet that provides the most accurate and robust performance for embedded systems aimed at drowsiness detection. After extracting wavelet coefficients, feature calculations are conducted to refine and improve the quality of the extracted features. Specifically, for each type of coefficient (approximation or detail), we calculate energy, entropy, mean, and standard deviation as follows (3)–(6):

$$Energy = \sum_{i=1}^{N} |C_i|^2 \tag{3}$$

where C_i represents the wavelet (approximation A or detail D) coefficients and N is the number of coefficients.

$$Entropy = -\sum_{i=1}^{N} p_i \, log(p_i) \tag{4}$$

where p_i is the probability of each coefficient C_i normalized over the total coefficients.

$$Mean = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{5}$$

$$Standard\ deviation = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (C_i - Mean)} \tag{6}$$

3.3. Classification

There are several classification methods based on different machine learning models, each adaptable to specific application areas. In this work, we evaluate the performance of various machine learning classifiers, including Naive Bayes (NB), k-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM), for drowsiness detection tasks. Each classifier is assessed based on its ability to accurately classify drowsy and alert states.

- NB
  The Naive Bayes classifier is a probabilistic machine learning algorithm based on Bayes' theorem, assuming independence among features [52]. Despite this simplification, it performs well in various classification tasks by calculating the probability of a data point belonging to a particular class.
  Bayes' theorem is expressed as follows (7):

  $$P(C|X) = \frac{P(X|C)P(C)}{P(X)}, \tag{7}$$

  where $P(C|X)$ is the posterior probability of class C given the features X, $P(X|C)$ is the likelihood of observing X given class C, $P(C)$ is the prior probability of class C, and $P(X)$ is the overall probability of X.
  In practice, $P(X|C)$ is calculated as the product of the probabilities of individual features, assuming independence (8):

  $$P(X|C) = \prod_{i=1}^{n} P(x_i|C), \tag{8}$$

  where x_i represents the i-th feature of X.
  The Naive Bayes classifier assigns the class with the highest posterior probability, making it computationally efficient and effective for tasks such as text classification and spam filtering.
- k-NN
  The k-NN algorithm is a simple yet effective supervised machine learning method used for classification tasks [53]. It assigns a class to a given data point based on the majority class of its k nearest neighbors in the feature space. The algorithm involves

calculating the distance between the query point and all training samples, identifying the k closest points, and performing majority voting among their classes.
The distance is often computed using the Euclidean distance formula (9):

$$d(x, x_i) = \sum_{j=1}^{n} (x_j - x_{i,j})^2, \tag{9}$$

where x is the query point, x_i is a training point, n is the number of features, x_j is the j-th feature of x, and $x_{i,j}$ is the j-th feature of x_i.
By selecting the majority class among the k nearest neighbors, the k-NN algorithm effectively classifies the query point. Its performance depends on the choice of k and the distance metric used.

- DT
  A DT is a machine learning algorithm that makes decisions by recursively splitting data based on feature values. At each node, the algorithm chooses the best feature to split the data, using criteria like the Gini Impurity [54]. The Gini Impurity is defined as follows (10):

$$Gini(D) = 1 - \sum_{i=1}^{C} p_i^2, \tag{10}$$

  where D is the dataset, C is the number of classes, and p_i is the probability of an instance being classified into class i.
  The feature that minimizes the Gini Impurity is selected for splitting. In drowsiness detection, Decision Trees classify subjects as alert or drowsy based on physiological signal features like EEG. The simplicity and interpretability of DT make them suitable for real-time applications, although they may suffer from overfitting in complex datasets.
- RF
  RF is an ensemble learning method that enhances classification accuracy and robustness by combining multiple decision trees. Each tree in the forest is trained on a random subset of the data, and the final prediction is made by aggregating the outputs of all trees, typically using majority voting for classification tasks [55]. This approach reduces overfitting and improves generalization. In a classification task, the prediction $\hat{y}$ is determined by the majority vote from all trees in the forest (11):

$$\hat{y} = \text{mode}(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T), \tag{11}$$

  where T is the number of trees and $\hat{y}_t$ is the prediction of the t-th tree. In the context of drowsiness detection, Random Forest is well-suited to handle high-dimensional data, such as EEG signals, and provide a robust classification. By leveraging the diversity of individual trees, it effectively manages noisy data and enhances the reliability of the detection system.
- MLP
  A Multilayer Perceptron is a type of Artificial Neural Network widely used for classification tasks. It consists of multiple layers of neurons: an input layer, one or more hidden layers, and an output layer [56]. Each neuron in one layer is fully connected to the neurons in the next layer through weighted connections. MLP is particularly powerful in modeling complex, non-linear relationships in data. During training, the network adjusts the weights of these connections using the backpropagation algorithm, aiming to minimize the error between predicted and actual outputs. The output $\hat{y}$ is determined by the following Equation (12):

$$\hat{y} = f\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{12}$$

where f is the activation function (e.g., sigmoid or ReLU), w_i represents the weights, x_i denotes the input features, b is the bias term, and n is the number of inputs. In drowsiness detection, MLP is effective in modeling the intricate relationships between EEG signal features and drowsiness states. Its strength lies in its ability to learn complex patterns from large datasets, making it a powerful tool for accurate detection in real-time systems.

- SVM
  Support Vector Machines are supervised learning models that aim to find an optimal hyperplane that separates data points into distinct classes. In cases where the data are not linearly separable, the Radial Basis Function (RBF) kernel maps the feature space into a higher-dimensional space, making separation feasible. The RBF kernel is defined by Equation (13):

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right), \tag{13}$$

where x and x' are the feature vectors, and γ is a parameter controlling the kernel's width. This kernel measures the similarity between data points by computing their Euclidean distance. The ability of the RBF kernel to capture non-linear patterns and handle high-dimensional EEG data makes it highly suitable for distinguishing between alert and drowsy states. By maximizing the margin between classes, this method enhances classification performance, improving both generalization and accuracy in drowsiness detection.
SVMs with the RBF kernel have demonstrated effectiveness in classifying EEG features for drowsiness detection, as shown in several studies [57,58].

3.4. Performance Evaluation

To evaluate the performance of the proposed detection method, a binary confusion matrix will be used. This confusion matrix is composed of the following states: a true positive (TP) occurs when drowsiness is correctly predicted while the actual state is drowsiness. A false positive (FP) is recorded when drowsiness is predicted while the actual state is vigilance. Similarly, a true negative (TN) refers to a correct prediction of vigilance when the actual state is also vigilance. Lastly, a false negative (FN) occurs when vigilance is predicted while the actual state is drowsiness. These different states in the confusion matrix will help us calculate the accuracy (14), precision (15), sensitivity (16), F1-score (17), and specificity (18).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

$$F1 - score = \frac{TP}{TP + \frac{FN+FP}{2}} \tag{17}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{18}$$

In this work, the feature set is divided as follows: 70% for training and 30% for testing, for each subject. Subsequently, the overall performance metrics are calculated across subjects. To minimize the risk of overfitting, the k-fold cross-validation method is employed. This approach divides the dataset into multiple subsets, ensuring a more robust evaluation of the models and reducing the risk of overfitting to a single data split.

4. Implementation on PYNQ-Z2

To design an embedded drowsiness detection system that adheres to the principles of portability, compactness, and efficiency, the PYNQ-Z2 platform was selected as the target hardware. Its compact dimensions (87 × 137 mm) make it particularly suitable for developing a portable system, ensuring adaptability to real-world applications such as drowsiness detection. Moreover, the PYNQ-Z2 fulfills the compactness criterion due to its compatibility with optimized co-design methodologies, enabling efficient resource utilization. Additionally, this platform supports low power consumption, real-time processing capabilities, and cost-effectiveness, making it a robust choice for implementing an efficient and practical drowsiness detection system.

4.1. Overview of the PYNQ-Z2 Platform

The PYNQ-Z2 [16] is an open-source development platform based on the ZYNQ-7000 SoC, designed by Xilinx to streamline the development of embedded systems. The ZYNQ-7000 integrates a dual-core ARM Cortex-A9 processor in the PS and a reconfigurable FPGA in the PL, providing a versatile architecture that balances software programmability with high-performance parallel processing. The PS handles high-level program execution, supported by resources such as a floating-point unit and an advanced memory management system, while the PL facilitates user-defined hardware accelerators for computationally intensive tasks, including discrete wavelet transform and feature extraction. This architecture enables the efficient implementation of low-latency, real-time signal processing applications, making it particularly suitable for EEG-based drowsiness detection.

Communication between the PS and PL occurs through AXI (Advanced eXtensible Interface) interconnects [59]:

- AXI4: Used for high-bandwidth data transfers, ideal for applications needing frequent memory access.
- AXI4-Lite: Suited for low-complexity data transfers, particularly for control signals.
- AXI4-Stream: Optimized for continuous data streaming, essential for real-time signal processing, such as EEG signal acquisition and feature extraction.

Figure 4 illustrates the general scheme of the ZYNQ platform.

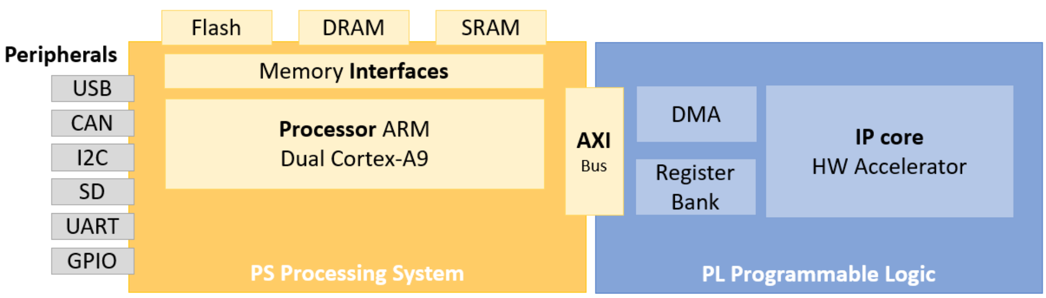

Figure 4. General scheme of the ZYNQ platform.

4.2. Implementation Details

As was mentioned previously, the feature extraction phase is a crucial component in drowsiness detection, as it significantly impacts system accuracy. This phase also demands substantial energy consumption and processing time, and thus it is very important to optimize its implementation.

The time-frequency domain is widely utilized across various applications and is known to enhance robustness in drowsiness detection, particularly by achieving a high accuracy rate. The innovation in this drowsiness detection system lies in the development of an embedded system leveraging time–frequency features. This is accomplished through the implementation of the Discrete Wavelet Transform (DWT) as a hardware IP.

The proposed hardware IP consists of two main components: the first part extracts DWT coefficients, while the second part computes features such as standard deviation (STD), entropy, energy, and mean simultaneously. Additionally, the system evaluates different architectures Float 32, Float 16, Fixed 32, and Fixed 24 to identify the most suitable configuration. This selection is based on key performance metrics, including power consumption, hardware resource utilization, and execution time, while maintaining high detection accuracy.

In this section, we will detail the implementation of our proposed software-hardware system for drowsiness detection. Our approach leverages a hybrid architecture on the Zynq platform, where both software and hardware components work together to optimize performance and resource utilization.

The signal pre-processing tasks, composed of the signal filtering and its segmentation, as well as the classification tasks, are handled in the PS part of the Zynq (ARM-cotex A9). For feature extraction, we implemented both a Discrete Wavelet IP Core (DWC), to compute the DWT using Daubechies 4 (db4) at four decomposition levels, and a Feature Extraction IP Core (FEC) to compute the mean, standard deviation, energy, and entropy features. Both the DWC and the FEC are executed on the PL part of the Zynq.

We implemented both fixed-point and floating-point versions of our system to assess performance trade-offs between precision, resource usage, and power consumption. These implementations were designed using High-Level Synthesis (HLS). We analyze how the different numeric representations affects the speed, area, and power consumption of the resulting design.

4.2.1. Discrete Wavelet IP Core

The Discrete Wavelet IP core (DWC) module is specifically engineered to execute a single level of the wavelet transform. As depicted in Figure 5, the architecture consists of several components, including the input interface for the incoming signal, high-pass and low-pass filter stages for signal decomposition into both D and A coefficients.

We used the AXI4 Stream protocol to efficiently transfer data to the IP core and also to output the resulting D and A coefficients. Using AXI4 Stream allows the IP to be easily connected with various types of embedded processors, enhancing the versatility and usability of the module in different application scenarios.

The DWC stores the incoming data in a 16-word shift register. The central eight words in the register (from the 4th to the 12th word) are fed to the high-pass and low-pass filters. To implement the eight tap convolutions required for both the high-pass and the low-pass filters, we need to compute eight multiplications and to aggregate the resulting values using seven summations. As can be seen in Figure 5, these filters are implemented using four stages. First, the eight required multiplications are performed in parallel. To this end, the shift register, as well as the high and low register, have to be implemented using FF. If we use BRAM, we would be limited to only two reads per clock, which would not permit the eight multiplications to be performed simultaneously. To ensure the use of FF for these three registers, the #ARRAY_PARTITION complete pragma was applied to the shift register and the high and low filter registers. In the second stage, we use four additions to sum the resulting values in pairs. In the third stage, we sum the four values from the previous stage in pairs, to finalize with the last addition in the forth stage.

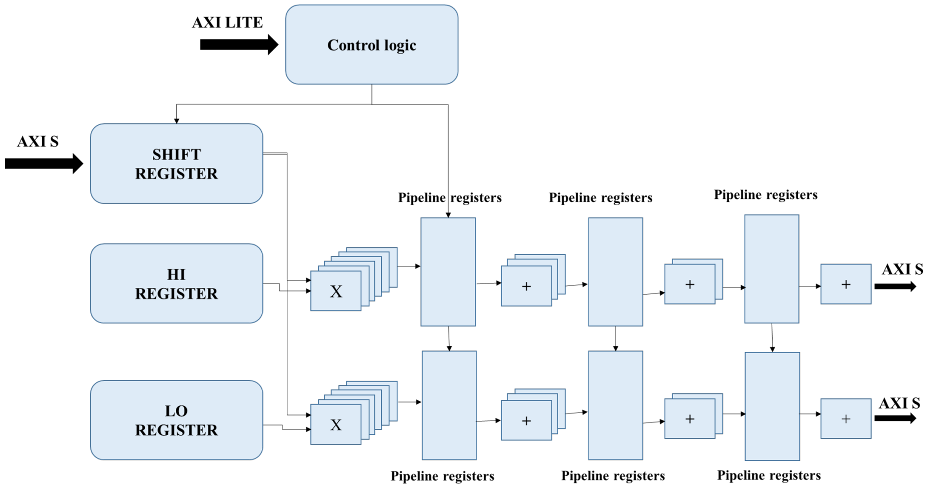

Figure 5. Block diagram for the single stage DWT IP core (DWC). The architecture of the high-pass filter is composed of the blocks at the right of the HI Register. Likewise, the low-pass filter is composed of the blocks at the right of the LO Register. The first 8 multiplications of both the high- and low-pass filters are shown in the replication of the multiplication block. The 3 addition stages are also shown, using a replication of the addition block to represent parallel additions.

The control logic block is responsible for several necessary steps. One of those is the padding of the input signal, which is set to symmetric padding in this implementation. When the first eight words are loaded into the shift register, the control logic writes values into the register replicating contents symmetrically by copying the word at the 0th address into the 15th address, the 1st into the 14th, etc. At the end of the input data stream, when no new data are received from the AXI4 Stream interface, the control logic feeds into the shift register the values from address 0, 2, 4, and so forth, thus achieving the symmetric padding at the end of the data stream.

The downsampling step of the DWT is also handled by the Control logic, by controlling when the output AXI stream signals are written. By writing these signals in one clock cycle and not in the following, this is achieved without any extra utilization of hardware resources.

The described DWC architecture achieves a pipelining initialization interval of one clock cycle, thus maximizing the throughput of the IP.

4.2.2. Features Extraction IP Core

The feature extraction IP core (FEC) was engineered specifically to calculate the following statistical metrics: mean, standard deviation, energy, and entropy. A block diagram of the IP core can be seen in Figure 6.

The input signal is read word by word through an AXI4 stream interface. When all the data coming from the AXI4 stream interface have been read, and the resulting mean, std, energy, and entropy of the signal have been computed, the results are written to a register bank to be read later by the processor through the AXI4 lite interface.

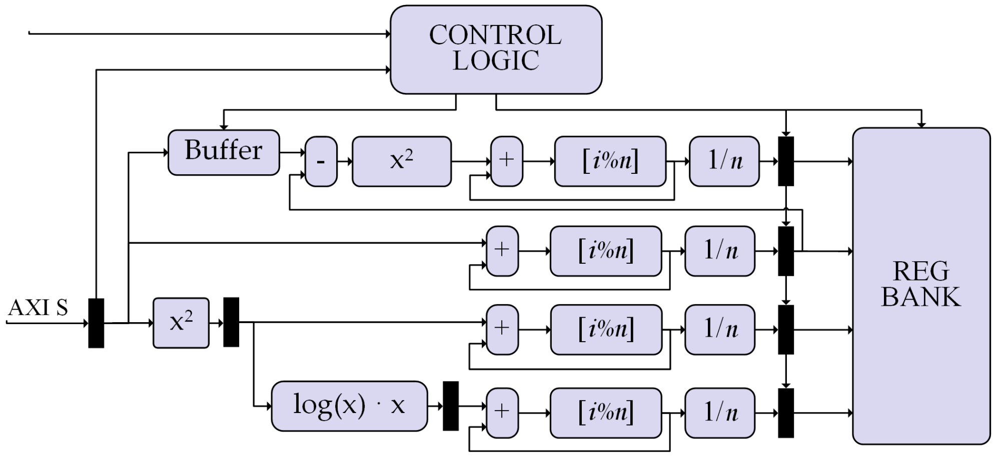

Figure 6. Feature extraction core (FEC). The symbol x is used for the signal present in the current line. The symbol n represents the total length of the data coming from the AXI4 stream interface. The nomenclature $[i\%n]$ is used to describe the ping-pong buffer. Registers are depicted in black.

To compute the mean of the signal, an accumulator is used to add the data as they have been read from the AXI4 stream interface. To avoid data dependencies between the read and write operations of the accumulator, a ping-pong buffering architecture is used. When reading address 0 of the ping-pong buffer, the writing is performed in address 1, and in the next clock cycle, we read from address 1 and write in address 2, repeating this logic until the maximum address of the ping-pong buffer is reached, when we cycle around and write to address 0. In our implementation, a ping-pong buffer with size 16 is needed to achieve a pipeline with an initialization interval of 1 cycle. When all the data have been read from the AXI4 stream interface, we accumulate the values of the ping-pong buffer itself. In a similar fashion as the accumulation described for the DWC, four summation stages are implemented, where we first perform eight summations by pairs of the 16 values in the ping-pong buffer, followed by four summations of the resulting partial additions, finishing with two summations and a last summation to obtain the final result. Afterwards, the resulting value is divided by the length of data in the input stream, and the control logic writes the result into the register bank.

A similar approach was followed to compute both the energy and the entropy of the signal. For the energy, the square of the values coming from the AXI4 stream are first computed, and then the ping-pong accumulation method as described in the previous paragraph is used. For the computation of the entropy, the logarithm of the square of the input value is computed, followed by a multiplication with the square of the value. After that, the same accumulation methodology as before is used. The logarithm is accomplished by the use of the hls_math library provided by Xilinx.

For the computation of the standard deviation, we need first to know the mean of the signal, which is not available as the AXI4 stream data are being read. Because of this, we follow a slightly different approach. First, the totality of the incoming values are stored in a buffer. Then, after all the data have been read from the AXI4 stream and the mean of the signal is available, the control logic starts reading from the aforementioned buffer. The squared difference between the buffer values and the mean is computed, followed by an accumulation approach, as described before.

The architecture described above allows for a pipeline with initialization interval of 1 clock, again maximizing the throughput of the IP core.

4.2.3. Interconnection Architectures for DWT and Feature Extraction

The DWC and FEC cores can be integrated into a single-stage DWT-FE unit (Figure 7, top). In this configuration, the DWC produces two outputs: coefficients A and D. The D coefficients are directly fed into the FEC for immediate processing, significantly reducing storage requirements. To ensure efficient interconnection between the DWC and FEC, a direct feed mechanism is utilized. Simultaneously, the A coefficients are forwarded to subsequent stages for further processing. This pipelined architecture minimizes both storage demands and processing delays by eliminating intermediate memory access for the D coefficients.

A single DWT-FE core can be concatenated to implement a multi-stage configuration, such as a four-stage DWT-FE (Figure 7, middle). However, due to resource constraints, the target architecture cannot support a four-stage DWT-FE implementation for a single channel. To address this limitation, the proposed design employs two single-stage DWT-FE units capable of processing two channels in parallel, with each channel completed in four execution steps. This configuration processes four channels instead of five but satisfies the requirements for processing speed, resource efficiency, and power consumption.

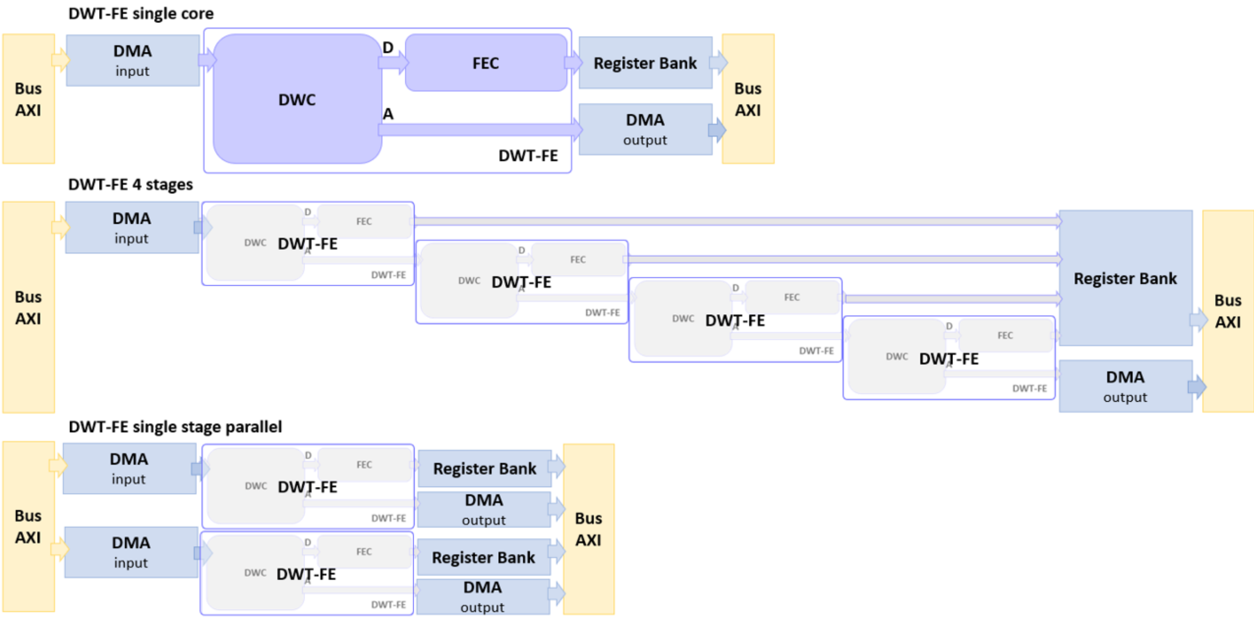

Figure 7. Interconnection architectures for DWT and feature extraction cores: (**top**) single core, (**middle**) 4 stages, and (**bottom**) single-stage parallel.

5. Results and Discussion

5.1. Drowsiness Detection Precision

In this section, we present the results of drowsiness detection using various wavelet transforms and machine learning classifiers to identify the most suitable wavelet transform and classifier for the drowsiness detection application. For this part, accuracy (A) is used as the primary performance metric. In subsequent parts, additional metrics such as precision (P), sensitivity (Sen), F1-score (F1), and specificity (Spe) will be employed to further validate the system's performance.

The results in Table 3 demonstrate that the Support Vector Machine (SVM), when paired with the DWT using the Daubechies 4 (db4) wavelet, is the most efficient classifier for drowsiness detection, achieving an accuracy of 88.37%. The SVM hyperparameters were optimized using the Grid Search method, with the optimal configuration comprising a Radial Basis Function (RBF) kernel, a penalty factor $C = 1$, and a kernel parameter $\gamma = 0.4$. This configuration exhibited superior performance in distinguishing between alert and drowsy states, ensuring high accuracy and robustness across the evaluated performance metrics.

Table 3. Accuracy for different classifiers with different wavelet transforms.

Wavelet	NB A(%)	KNN A(%)	DT A(%)	RF A(%)	MLP A(%)	SVM A(%)
Haar	46.57	48.78	58.9	60.2	61.4	59.71
Symlet	46.81	50.27	60.52	60.72	67.27	67.52
Coiflet	40.22	43.5	47.77	46.89	50.71	51.2
Mexican Hat	47.21	47.7	51.72	50.85	63.79	65.85
db (4)	51.42	66.15	77.01	78.47	81.82	88.37

In this second part, we will analyze the different architectures used for feature extraction. The goal of this analysis is to identify the most effective architecture by evaluating key performance metrics. These results will be compared to those obtained from a computer-based approach.

From the results in Table 4, we observe that the system implemented on the software side of the ZYNQ 7020 (ARM Cortex-A9), as well as the architectures based on a software–hardware co-design (soft-hard), maintains drowsiness detection accuracy comparable to that achieved on a standard computer. However, a slight decrease in accuracy is observed with the fixed-point 24.16 architecture. Despite this, all the tested configurations deliver high overall performance, demonstrating their effectiveness for accurate drowsiness detection using EEG signals.

Table 4. Performance metrics for different architectural approaches for drowsiness detection.

Architecture	A(%)	P(%)	Sen(%)	Spe(%)	F1(%)
PC	88.37	86.43	88.50	88.75	87.67
PYNQ(ARM Cortex A-9)	88.37	86.43	89.60	88.73	87.69
Float 32	88.37	86.43	89.25	88.73	87.69
Float 16	88.37	86.43	89.25	88.73	87.69
Fixed 32 (16)	88.37	86.43	89.25	88.73	87.69
Fixed 24 (16)	88.02	85.71	89.17	88.59	87.28

5.2. Power Consumption

Table 5 presents the power consumption across the various implementations. This comparison highlights the energy efficiency of each architecture, allowing us to identify the most suitable design for minimizing power usage in embedded drowsiness detection applications.

The data in Table 5 clearly show that the proposed architectures significantly reduce power consumption compared to the implementation on a PC. This reduction makes them highly suitable for the energy consumption requirements of embedded systems. Notably, the fixed-point 24.16 architecture stands out as the configuration with the lowest energy consumption.

Table 5. Power consumption for the different implementations.

IP	Total Power (mW)
PC	65,000.0
ARM Cortex A-9	626.5
Float 32	680.0
Float 16	398.0
Fixed 32 (16)	657.0
Fixed 24 (12)	338.0

5.3. Processing Time

In the context of processing the four levels of the discrete wavelet transform (DWT) using extraction functions from the Daubechies 4 (db4) wavelet basis, the CPU implementation using Python recorded an execution time of approximately 2.45 s. In contrast, the integration of the proposed intellectual property (IP) blocks reduced this time to 1.26 s. This significant improvement in processing times observed across different implementations can be attributed to the fact that execution time is primarily influenced by the duration required to transfer data from the processor to the contiguous RAM region, which is essential for direct memory access (DMA) operations. The tradeoff between execution time and accuracy is evident when analyzing the Fixed 24 (12) configuration, which achieves an execution time of 1.26 s, faster than Float 32 and Fixed 32 (16), both of which have an execution time of 1.28 s. This configuration maintains an accuracy of 88.02%, making it suitable for energy-efficient and fast-execution applications. On the other hand, for accuracy-critical applications, configurations such as Float 32 or ARM Cortex A-9, which retain the maximum accuracy of 88.37%, can be employed. However, these configurations come at the cost of increased execution time, with the ARM Cortex A-9 taking 2.45 s and the PC achieving the fastest execution at 0.6 s.

5.4. Hardware Resources

In this section, we present the results obtained using hardware for the HLS implementation. We will analyze the performance of this approach in terms of resource consumption, specifically focusing on the usage of LUTs (Look-Up Tables), LUTRAM, Flip-Flops, BRAM (Block RAM), and DSPs (Digital Signal Processing blocks). Table 6 presents the results.

Table 6. Resource consumption for different IPs.

	LUT	LUTRAM	FF	BRAM
Float 32	21,231	753	26,786	6
Float 16	14,511	588	18,710	14
Fixed 32 (16)	32,152	2084	27,862	43
Fixed 24 (12)	16,898	658	16,174	18

The results show that the IP 24 represents the implementation that consumes the least hardware resources compared to the other architectures.

5.5. Discussion

The results of this study underscore the remarkable effectiveness of the proposed hardware architectures, particularly the fixed-point 24.16 configuration, in addressing the challenges associated with developing embedded drowsiness detection systems. The architectures implemented on the software side of the ZYNQ 7020 platform (ARM Cortex-A9) and the software–hardware co-design achieved an impressive detection accuracy of 88.37%, comparable to that of the traditional PC-based implementation. While a slight decrease in accuracy (88.02%) was observed in the fixed-point 24.16 architecture, this reduction is negligible when weighed against the significant advantages it offers in terms of power consumption efficiency, processing time performance, and resource utilization. Furthermore, the architecture maintained robust performance metrics, with high precision (86.43%), sensitivity (89.6%), and F1-score (87.69%), ensuring reliable and consistent drowsiness detection based on EEG signals across various scenarios. To assess the performance of our system, we employed a second publicly available dataset, SEED-VIG, which is specifically designed for drowsiness detection. This dataset comprises EEG recordings of 21 subjects, each lasting 2 h, and acquired using 12 electrodes: CP1, CPz, CP2, P1, Pz, P2, PO3, POz, PO4, O1, Oz, and O2. The recording protocol involved two driving simulation sessions conducted at distinct times of day: afternoon and evening, with the vigilance levels of the subjects assessed during each session.

Evaluation results using the SEED-VIG dataset revealed an accuracy of 85.69%, precision of 85.62%, sensitivity of 85.58%, F1-score of 85.62%, and specificity of 86.78%. These metrics indicate a high level of performance and demonstrate the effectiveness of the proposed system for drowsiness detection. Despite minor fluctuations in detection precision, which could be attributed to inter-session variability, the overall performance confirms the robustness and reliability of the system.

These findings further validate the generalizability of our approach, showing consistent detection accuracy across different datasets and conditions. This reinforces the applicability of the system for real-time drowsiness monitoring in diverse settings, highlighting its potential for widespread use in practical applications.

In Table 7, we compare our proposed drowsiness detection system with several recent approaches, focusing specifically on EEG-based methods in the intra-subject mode. In [22], the authors utilized a CNN to extract features from 32 EEG channels, achieving an accuracy of 89.65%. This approach was implemented on a smartphone, consuming 2.25 W of power and achieving a response time of 3.7 s. In [23], an SVM classifier was employed to classify EEG signals from eight channels, resulting in an accuracy of 78.3%. This system was implemented on a standard PC with a power consumption of 200 W. Similarly, in [24], SVM was used to classify Power Spectral Density (PSD) features from 19 EEG channels, achieving an accuracy of 85.6% with a response time of 1 s and a power consumption of 200 W. In [25], a CNN was employed to automatically extract features from 16 EEG channels, implemented on a standard PC, achieving 85.2% accuracy with a response time of 5 s and consuming 200 W. Lastly, in [26], the authors used an RBF hierarchy to classify temporal and spectral features from 32 EEG electrodes. This approach, also implemented on a standard PC, achieved an execution time of 10 s with a power consumption of 200 W.

Table 7. Comparative analysis of the proposed method versus recent works in intra-subject mode.

Ref.	Features	Model	Number of Electrodes	Accuracy (%)	Power (Watts)	Time (s)
[22]	EEG	CNN	32	89.65	22	3.7
[23]	EEG PSD	SVM	8	78.3	200	-
[24]	EEG PSD	SVM	19	85.6	200	1
[25]	EEG	CNN	16	85.2	200	5
[26]	EEG temporal and spectral features	RBF Hierarchy	32	92	200	10
Proposed Method	DWT coefficients (time frequency)	SVM	4	88.02	0.338	1.26

Our analysis reveals a significant gap in the development of embedded systems for drowsiness detection based on EEG signals. Compared to these recent approaches, our proposed system achieves an excellent accuracy of 88.02% using only four EEG derivations. This highlights its capability to operate effectively with fewer channels, reducing the complexity of data acquisition and processing. The system also meets the portability criterion, with a compact size of 87 × 137 mm, making it adaptable to various environments. This portability is further enhanced by the reduced number of electrodes, ensuring a simpler and more comfortable setup for the user, thereby increasing its practicality for real-world applications.

Furthermore, it demonstrates significantly reduced response time (1.26 s) and lower power consumption (0.338 W), making it more adaptable to various environments and suitable for real-time applications. These improvements over existing systems not only enhance its practicality for deployment in resource-constrained scenarios but also position it as a superior alternative for embedded EEG-based drowsiness detection systems.

A critical aspect of embedded systems is power efficiency, particularly for applications requiring sustained operation in resource-constrained environments. The proposed architectures demonstrate a dramatic reduction in power consumption compared to the 64 W required by the PC-based implementation and other state-of-the-art solutions, such as the 22 W reported in [22] and the 2.25 W presented in [23]. Notably, the fixed-point 24.16 configuration emerged as the most power-efficient solution, consuming a mere 0.338 W. Balancing high detection accuracy with ultra-low power consumption ensures that the system can operate for extended periods without frequent recharging, a vital feature for portable and wearable applications.

In addition to power efficiency, the proposed system showcases significant improvements in processing time, reducing execution time from the CPU-based Python implementation (2.45 s) to just 1.26 s. This enhancement is vital for applications requiring real-time or near-real-time drowsiness detection to ensure timely intervention in critical situations, such as transportation or industrial operations.

Moreover, the analysis of hardware resource utilization highlights the scalability and practicality of the fixed-point 24.16 architecture. With an efficient utilization of 16,898 LUTs, 658 LUTRAMs, 16,174 Flip-Flops, and 18 BRAMs, this design strikes a delicate balance between resource consumption and performance. The low resource demands not only reduce production costs but also enhance the feasibility of deploying the system in diverse environments where hardware constraints are significant.

In conclusion, the findings of this study firmly establish the proposed hardware architectures, particularly the fixed-point 24.16 configuration, as a robust solution for embedded drowsiness detection systems. By achieving high detection accuracy (88.02%), ultra-low power consumption (0.338 Watts), and efficient hardware resource usage, this architecture is ideally suited for energy-efficient, real-time applications. Its adaptability and scalability make it an ideal candidate for deployment in safety-critical industries such as transportation, healthcare, and industrial automation, where timely and reliable drowsiness detection is essential.

6. Conclusions

Drowsiness is one of the major factors in workplace accidents across various environments (industrial, transportation, healthcare), potentially causing severe and even fatal incidents in some cases. In this work, we propose an embedded drowsiness detection system implemented on the PYNQ-Z2 platform using a hardware–software approach. The proposed system is structured with filtering, segmentation, and classification tasks managed by the ZYNQ-7000 CPU (ARM). The filtering uses an FIR bandpass filter with a range of [0.1; 30] Hz, and segmentation divides data into 10 s segments. For feature extraction, the PL part of the ZYNQ performs time–frequency analysis on EEG signals through the Discrete Wavelet Transform (DWT) (using Daubechies 4) and the feature extraction task. This step extracts coefficients that represent signal energy, entropy, standard deviation, and mean. This feature extraction IP was developed using HLS, with fixed-point and floating-point implementations.

The novelty of this system lies in its ability to combine early drowsiness detection using EEG signals, which can identify drowsiness before physical signs appear, with a practical, embedded implementation. By employing only four electrodes, the system is compact and wearable, addressing real-world usability concerns while maintaining high detection accuracy. Moreover, its physical dimensions of 87 × 137 mm make it portable and adaptable to diverse environments. The inclusion of both time and frequency features extracted via DWT enhances robustness against artifacts and inter-subject variability, ensuring reliable performance across varying conditions. Additionally, the system achieves significant power efficiency (0.338 W) and low resource usage, further emphasizing its suitability for embedded applications.

Our embedded drowsiness detection system achieves an accuracy of 88.02% with the fixed point 24 version, an execution time of 1.26 s, and a power consumption of 0.338 W using only four electrodes.

Future works will focus on implementing this approach on different platforms and processors, such as NVIDIA Jetson Nano, Raspberry Pi, and Intel Movidius Neural Compute Stick, to develop a robust and versatile system. These efforts will include exploring purely hardware-based solutions, such as ASIC implementation, to enhance its applicability for real-time and resource-constrained environments. This will build upon the current implementation, which prioritizes affordability by utilizing a low-cost platform. Additionally, integrating EEG with other types of physiological signals will be explored to further enhance detection capabilities and improve overall accuracy, ensuring reliable performance across various application scenarios.

Author Contributions: Conceptualization, A.Z.; methodology, A.Z., J.T. and E.T.; software, A.Z. and E.T.; validation, K.B.K. and C.V.; formal analysis, A.Z.; investigation, A.Z.; resources, A.Z.; data curation, A.Z.; writing—original draft preparation, A.Z., J.T., E.T. and C.V.; writing—review and editing, A.Z., K.B.K. and E.T.; visualization, K.B.K. and J.T.; supervision, K.B.K. and C.V.; project administration, A.Z., E.T., K.B.K., J.T. and C.V.; funding acquisition, K.B.K. and C.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original data presented in the study are openly available at https://doi.org/10.20944/preprints202405.1615.v1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. van Schie, M.K.; Lammers, G.J.; Fronczek, R.; Middelkoop, H.A.; van Dijk, J.G. Vigilance: Discussion of related concepts and proposal for a definition. *Sleep Med.* **2021**, *83*, 175–181. [CrossRef] [PubMed]
2. Jing, M.J.; Li, H.; Li, C.P.; Wei, X.J.; Lin, W.Q.; Zhu, S.C.; Xu, Y.L.; Li, L.M. The Association between Working Hours with Vigilance and Executive Function of Intensive Care Unit Nurses. *J. Nurs. Manag.* **2023**, *2023*, 3770404. [CrossRef]
3. Wang, H.; Chen, D.; Huang, Y.; Zhang, Y.; Qiao, Y.; Xiao, J.; Xie, N.; Fan, H. Assessment of vigilance level during work: Fitting a hidden Markov model to heart rate variability. *Brain Sci.* **2023**, *13*, 638. [CrossRef] [PubMed]
4. Slater, J.D. A definition of drowsiness: One purpose for sleep? *Med. Hypotheses* **2008**, *71*, 641–644. [CrossRef] [PubMed]
5. Albadawi, Y.; Takruri, M.; Awad, M. A review of recent developments in driver drowsiness detection systems. *Sensors* **2022**, *22*, 2069. [CrossRef]
6. Li, G.; Chung, W.Y. Electroencephalogram-based approaches for driver drowsiness detection and management: A review. *Sensors* **2022**, *22*, 1100. [CrossRef]
7. Das, A.K.; Kumar, P.; Halder, S.; Metia, S. A Statistical Approach for Investigation and Comparison of Fatigue and Drowsiness based on Complexity Parameters of EOGs. *Int. J. Image, Graph. Signal Process.* **2023**, *15*, 39–59.
8. Halomoan, J.; Ramli, K.; Sudiana, D.; Gunawan, T.S.; Salman, M. A New ECG Data Processing Approach to Developing an Accurate Driving Fatigue Detection Framework with Heart Rate Variability Analysis and Ensemble Learning. *Information* **2023**, *14*, 210. [CrossRef]
9. Ma, M.; Luo, X.; Xiahou, S.; Shan, X. A Laguerre–Volterra network model based on ant colony optimization applied to evaluate EMG-force relationship in the muscle fatigue state. *Rev. Sci. Instruments* **2024**, *95*, 065004. [CrossRef]
10. Adams, E.J.; Scott, M.E.; Amarante, M.; Ramírez, C.A.; Rowley, S.J.; Noble, K.G.; Troller-Renfree, S.V. Fostering inclusion in EEG measures of pediatric brain activity. *npj Sci. Learn.* **2024**, *9*, 27. [CrossRef]
11. Latreche, I.; Slatnia, S.; Kazar, O.; Barka, E.; Harous, S. A Review on Deep Learning Techniques for EEG-Based Driver Drowsiness detection systems. *Informatica* **2024**, *48*, 3. [CrossRef]
12. Mohammedi, M.; Mokrani, J.; Mouhoubi, A. An automated and highly efficient driver drowsiness detection and alert system using electroencephalography signals for safe driving. *Multimed. Tools Appl.* **2024**, *83*, 87299–87322. [CrossRef]

13. Minhas, R.; Peker, N.Y.; Hakkoz, M.A.; Arbatli, S.; Celik, Y.; Erdem, C.E.; Semiz, B.; Peker, Y. Association of visual-based signals with electroencephalography patterns in enhancing the drowsiness detection in drivers with obstructive sleep apnea. *Sensors* **2024**, *24*, 2625. [CrossRef] [PubMed]
14. Bajaj, J.S.; Kumar, N.; Kaushal, R.K.; Gururaj, H.; Flammini, F.; Natarajan, R. System and method for driver drowsiness detection using behavioral and sensor-based physiological measures. *Sensors* **2023**, *23*, 1292. [CrossRef] [PubMed]
15. Siddhad, G.; Dey, S.; Roy, P.P. DrowzEE-G-Mamba: Leveraging EEG and State Space Models for Driver Drowsiness Detection. *arXiv* **2024**, arXiv:2408.16145.
16. Zayed, A.; Belhadj, N.; Khalifa, K.B.; Bedoui, M.H. VVC intra prediction decoder: Feature improvement and performance analysis. In Proceedings of the 2022 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS), Cairo, Egypt, 6–9 June 2022; pp. 1–4.
17. Xie, Y.; Majoros, T.; Oniga, S. FPGA-Based Hardware Accelerator on Portable Equipment for EEG Signal Patterns Recognition. *Electronics* **2022**, *11*, 2410. [CrossRef]
18. Chaddad, A.; Wu, Y.; Kateb, R.; Bouridane, A. Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors* **2023**, *23*, 6434. [CrossRef]
19. Cao, H.; Wang, Y.; Banerjee, S.; Cao, Y.; Mou, J. A discrete Chialvo–Rulkov neuron network coupled with a novel memristor model: Design, Dynamical analysis, DSP implementation and its application. *Chaos Solitons Fractals* **2024**, *179*, 114466. [CrossRef]
20. Sathasivam, S.; Saon, S.; Sidek, A.; Som, M.M.; Ameen, H.A. Drowsiness detection system using eye aspect ratio technique. In Proceedings of the 2020 IEEE Student Conference on Research and Development (SCOReD), Batu Pahat, Malaysia, 27–29 September 2020; pp. 448–452.
21. Florez, R.; Palomino-Quispe, F.; Alvarez, A.B.; Coaquira-Castillo, R.J.; Herrera-Levano, J.C. A Real-Time Embedded System for Driver Drowsiness Detection Based on Visual Analysis of the Eyes and Mouth Using Convolutional Neural Network and Mouth Aspect Ratio. *Sensors* **2024**, *24*, 6261. [CrossRef] [PubMed]
22. Abbas, Q.; Alsheddy, A. Driver fatigue detection systems using multi-sensors, smartphone, and cloud-based computing platforms: A comparative analysis. *Sensors* **2020**, *21*, 56. [CrossRef] [PubMed]
23. Gangadharan, S.; Vinod, A. Drowsiness detection using portable wireless EEG. *Comput. Methods Programs Biomed.* **2022**, *214*, 106535. [CrossRef] [PubMed]
24. Arif, S.; Munawar, S.; Ali, H. Driving drowsiness detection using spectral signatures of EEG-based neurophysiology. *Front. Physiol.* **2023**, *14*, 1153268. [CrossRef]
25. Wang, Y.; Huang, Y.; Gu, B.; Cao, S.; Fang, D. Identifying mental fatigue of construction workers using EEG and deep learning. *Autom. Constr.* **2023**, *151*, 104887. [CrossRef]
26. Ren, Z.; Li, R.; Chen, B.; Zhang, H.; Ma, Y.; Wang, C.; Lin, Y.; Zhang, Y. EEG-based driving fatigue detection using a two-level learning hierarchy radial basis function. *Front. Neurorobot.* **2021**, *15*, 618408. [CrossRef]
27. Jiang, X.; Bian, G.B.; Tian, Z. Removal of artifacts from EEG signals: A review. *Sensors* **2019**, *19*, 987. [CrossRef]
28. Saleem, A.A.; Siddiqui, H.U.R.; Raza, M.A.; Rustam, F.; Dudley, S.; Ashraf, I. A systematic review of physiological signals based driver drowsiness detection systems. *Cogn. Neurodyn.* **2023**, *17*, 1229–1259. [CrossRef] [PubMed]
29. Le, V.D.; Carthen, C.D.; Kamaruddin, N.; Tavakkoli, A.; Dascalu, S.M.; Harris, F.C., Jr. Generalized eeg data acquisition and processing system. In Proceedings of the International Conference on Information Technology-New Generations, Las Vegas, NV, USA, 24–26 April 2023; pp. 173–179.
30. Robbins, K.A.; Touryan, J.; Mullen, T.; Kothe, C.; Bigdely-Shamlo, N. How sensitive are EEG results to preprocessing methods: A benchmarking study. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1081–1090. [CrossRef]
31. Hussein, R.M.; Miften, F.S.; George, L.E. Driver drowsiness detection methods using EEG signals: A systematic review. *Comput. Methods Biomech. Biomed. Eng.* **2023**, *26*, 1237–1249. [CrossRef] [PubMed]
32. Massoz, Q.; Langohr, T.; François, C.; Verly, J.G. The ULg multimodality drowsiness database (called DROZY) and examples of use. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–7.
33. Miley, A.Å.; Kecklund, G.; Åkerstedt, T. Comparing two versions of the Karolinska Sleepiness Scale (KSS). *Sleep Biol. Rhythm.* **2016**, *14*, 257–260. [CrossRef] [PubMed]
34. Ille, N.; Nakao, Y.; Yano, S.; Taura, T.; Ebert, A.; Bornfleth, H.; Asagi, S.; Kozawa, K.; Itabashi, I.; Sato, T.; et al. Ongoing EEG artifact correction using blind source separation. *Clin. Neurophysiol.* **2024**, *158*, 149–158. [CrossRef]
35. Yu, J.; Li, C.; Lou, K.; Wei, C.; Liu, Q. Embedding decomposition for artifacts removal in EEG signals. *J. Neural Eng.* **2022**, *19*, 026052. [CrossRef]
36. Sen, D.; Mishra, B.B.; Pattnaik, P.K. A review of the filtering techniques used in EEG signal processing. In Proceedings of the 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–13 April 2023; pp. 270–277.
37. Stone, J.V. Independent component analysis: An introduction. *Trends Cogn. Sci.* **2002**, *6*, 59–64. [CrossRef] [PubMed]

38. Elouaham, S.; Dliou, A.; Jenkal, W.; Louzazni, M.; Zougagh, H.; Dlimi, S. Empirical Wavelet Transform Based ECG Signal Filtering Method. *J. Electr. Comput. Eng.* **2024**, *2024*, 9050909. [CrossRef]
39. Pant, A.; Kumar, A. Hanning FIR window filtering analysis for EEG signals. *Biomed. Anal.* **2024**, *1*, 111–123. [CrossRef]
40. Buzsaki, G. *Rhythms of the Brain*; Oxford University Press: Oxford, UK, 2006.
41. Mahabub, A. Design and implementation of cost-effective simple FIR filter for EEG signal on FPGA. *World Sci. News* **2019**, *125*, 1–17.
42. De Gennaro, L.; Marzano, C.; Fratello, F.; Moroni, F.; Pellicciari, M.C.; Ferlazzo, F.; Costa, S.; Couyoumdjian, A.; Curcio, G.; Sforza, E.; et al. The electroencephalographic fingerprint of sleep is genetically determined: A twin study. *Ann. Neurol.* **2008**, *64*, 455–460. [CrossRef] [PubMed]
43. Saber, M. Removing Powerline Interference from EEG Signal using Optimized FIR Filters. *J. Artif. Intell. Metaheuristics* **2022**, *1*, 8–19. [CrossRef]
44. Stancin, I.; Cifrek, M.; Jovic, A. A review of EEG signal features and their application in driver drowsiness detection systems. *Sensors* **2021**, *21*, 3786. [CrossRef]
45. Zayed, A.; Khalifa, K.B.; Belhadj, N.; Bedoui, M.H.; Sakuyama, C.A.V. Discrete Wavelet Transform Coefficients for Drowsiness Detection from EEG Signals. In Proceedings of the 2023 IEEE International Conference on Design, Test and Technology of Integrated Systems (DTTIS), Gammarth, Tunisia, 1–4 November 2023; pp. 1–6.
46. Zhang, C.; Eskandarian, A. A survey and tutorial of EEG-based brain monitoring for driver state analysis. *arXiv* **2020**, arXiv:2008.11226. [CrossRef]
47. Contoyiannis, Y.; Papadopoulos, P.; Potirakis, S.; Kampitakis, M.; Matiadou, N.; Kosmidis, E. Analysis of Electroencephalography (EEG) Signals Based on the Haar Wavelet Transformation. In *Approximation and Computation in Science and Engineering*; Springer: Cham, Switzerland, 2022; Volume 180, pp. 157–166.
48. Wang, Y.; Li, Z.; Wang, C.; Feng, L.; Zhang, Z. Implementation of discrete wavelet transform. In Proceedings of the 2014 12th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), 2014, Guilin, China, 28–31 October 2014; pp. 1–3.
49. Elshekhidris, I.H.; MohamedAmien, M.B.; Fragoon, A. Wavelet transforms for EEG signal denoising and decomposition. *Int. J. Adv. Signal Image Sci.* **2023**, *9*, 11–28. [CrossRef]
50. Grobbelaar, M.; Phadikar, S.; Ghaderpour, E.; Struck, A.F.; Sinha, N.; Ghosh, R.; Ahmed, M.Z.I. A survey on denoising techniques of electroencephalogram signals using wavelet transform. *Signals* **2022**, *3*, 577–586. [CrossRef]
51. Fouladi, S.; Safaei, A.A.; Mammone, N.; Ghaderi, F.; Ebadi, M.J. Efficient deep neural networks for classification of Alzheimer's disease and mild cognitive impairment from scalp EEG recordings. *Cogn. Comput.* **2022**, *14*, 1247–1268. [CrossRef]
52. Peretz, O.; Koren, M.; Koren, O. Naive Bayes classifier–An ensemble procedure for recall and precision enrichment. *Eng. Appl. Artif. Intell.* **2024**, *136*, 108972. [CrossRef]
53. Khodadadi, N.; Khodadadi, E.; Al-Tashi, Q.; El-Kenawy, E.S.M.; Abualigah, L.; Abdulkadir, S.J.; Alqushaibi, A.; Mirjalili, S. BAOA: Binary arithmetic optimization algorithm with K-nearest neighbor classifier for feature selection. *IEEE Access* **2023**, *11*, 94094–94115. [CrossRef]
54. Charbuty, B.; Abdulazeez, A. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [CrossRef]
55. Palimkar, P.; Shaw, R.N.; Ghosh, A. Machine learning technique to prognosis diabetes disease: Random forest classifier approach. In *Proceedings of the Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021*; Springer: Singapore, 2022; pp. 219–244.
56. Li, H.; Nyirandayisabye, R.; Dong, Q.; Niyirora, R.; Hakuzweyezu, T.; Zardari, I.A.; Nkinahamira, F. Crack damage prediction of asphalt pavement based on tire noise: A comparison of machine learning algorithms. *Constr. Build. Mater.* **2024**, *414*, 134867. [CrossRef]
57. Handayani, L.; Nuraeni, E.; Arsadiando, W.; Pamungkas, A. Fatigue and drowsiness detection using a support vector machine for traffic accident reduction. *J. Intell. Syst. Eng. Appl. Data Sci.* **2024**, *1*, 33–39.
58. Zayed, A.; Belhadj, N.; Ben Khalifa, K.; Bedoui, M.H.; Valderrama, C. Efficient Generalized Electroencephalography-Based Drowsiness Detection Approach with Minimal Electrodes. *Sensors* **2024**, *24*, 4256. [CrossRef]
59. Lázaro, J.; Astarloa, A.; Zuloaga, A.; Araujo, J.Á.; Jiménez, J. AXI Lite Redundant On-Chip Bus Interconnect for High Reliability Systems. *IEEE Trans. Reliab.* **2023**, *73*, 602–607. [CrossRef]