Received 20 August 2025, accepted 6 September 2025. Date of publication 00 xxxx 0000, date of current version 00 xxxx 0000.

Digital Object Identifier 10.1109/ACCESS.2025.3608083

Explainable AI-Driven EEG Channel Selection for Accurate and Interpretable Epilepsy Diagnosis

AFIFA JRIDI[1,2,3], KAIS BELWAFI[4], RIDHA DJEMAL[1,2], CARLOS VALDERRAMA SAKUYAMA[3], (Senior Member, IEEE), AND SAMI BACHIR MEJRI[5]

[1]Laboratoire of Advanced Techniques for Medicine and Systems (ATMS), National School of Engineering, University of Sfax, Sfax 3038, Tunisia
[2]Electrical Engineering Department, National School of Engineering, University of Sfax, Sfax 3038, Tunisia
[3]Department of Electronics and Microelectronics (SEMi), University of Mons, 7000 Mons, Belgium
[4]Computer Engineering Department, College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates
[5]Pedagogical Enhancement-CTL, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Corresponding author: Afifa Jridi (afifa.jridi@enis.tn)

ABSTRACT Epilepsy is a complex neurological disorder characterized by recurrent seizures, requiring accurate and timely diagnosis to ensure effective clinical management. This is particularly important in Brain-Computer Interfaces (BCIs), where neurological monitoring plays a key role in assistive and diagnostic applications. In this context, Electroencephalography (EEG) is the primary diagnostic modality; however, traditional manual analysis is labor-intensive, subjective, and prone to inter-expert variability. Automated machine learning techniques offer a promising alternative to manual EEG analysis. The opaque, ''black-box'' nature of many Machine Learning (ML) models often raise concerns about their reliability in clinical settings, which can hinder their practical implementation. To tackle this issue, Explainable Artificial Intelligence (XAI) has emerged as a transformative solution, providing greater transparency and shedding light on the critical features driving model predictions. In this study, we introduce Explainable LIMEbased Selection of EEG Electrodes (XLISEE), a novel approach that not only enhances EEG channel selection but also strengthens interpretability by leveraging XAI methodologies. The suggested approach is evaluated using the dataset provided by Guinea-Bissau Hospital. The most important EEG channels are identified by XLISEE to differentiate between epileptics and healthy subjects, reducing data complexity without sacrificing the precision of the diagnosis. With accuracy rates of 99.76% for both the Support Vector Machine (SVM) and k-Nearest Neighbor (kNN) classifiers, our experimental results show that the suggested method delivers remarkable classification performance. It also attains 100% accuracy, specificity, sensitivity, and F1-score, as well as a Matthews Correlation Coefficient (MCC) of 1.0. Moreover, lower computational demands improve efficiency, making real-time epilepsy detection more practical and feasible for clinical use. Beyond performance gains, integrating XAI enhances clinician confidence by offering interpretable, evidence-based insights, promoting wider adoption of AI-driven diagnostics.

INDEX TERMS Electroencephalography (EEG), epilepsy, adaptive channel selection, explainable artificial intelligence (XAI), local interpretable model-agnostic explanations (LIME), empirical mode decomposition (EMD), independent component analysis (ICA), machine learning.

I. LIST OF ABBREVIATIONS

- **AFN**: False Negative Rate
- **AUC**: Area Under the Curve
- **BCI**: Brain-Computer Interface
- **CBF**: Cerebral Blood Flow
- **CNN**: Convolutional Neural Network
- **C-DRNN**: Convolutional Deep Recurrent Neural Network
- **DTC**: Decision Tree Classifier
- **DWT**: Discrete Wavelet Transform
- **EACS**: Explainable Adaptatif Channel Selection

The associate editor coordinating the review of this manuscript and approving it for publication was Ghulam Muhammad.

- **EEG**: Electroencephalography
- **ELISEE**: Explainable LIME-based Selection of EEG Electrodes
- **EMD**: Empirical Mode Decomposition
- **FE**: Focal Epilepsy
- **fMRI**: Functional Magnetic Resonance Imaging
- **FN**: False Negative
- **FNR**: False Negative Rate
- **FP**: False Positive
- **FPR**: False Positive Rate
- **FFT**: Fast Fourier Transform
- **GE**: Generalized Epilepsy
- **ICA**: Independent Component Analysis
- **IC-RNN**: Inception-based Convolutional Recurrent Neural Network
- **IE**: Idiopathic Epilepsy
- **IMF**: Intrinsic Mode Functions
- **kNN**: k-Nearest Neighbor
- **LDA**: Linear Discriminant Analysis
- **LIME**: Local Interpretable Model-Agnostic Explanations
- **LOOCV**: Leave-One-Out Cross-Validation
- **MCC**: Matthews Correlation Coefficient
- **MEG**: Magnetoencephalography
- **ML**: Machine Learning
- **NEA**: Neuronal Electrical Activity
- **NPV**: Negative Predictive Value
- **PDP**: Partial Dependence Plot
- **PRN**: Pattern Recognition Network
- **PSO**: Particle Swarm Optimization
- **RFC**: Random Forest Classifier
- **ResNet**: Deep Residual Network
- **SHAP**: SHapley Additive exPlanations
- **SNR**: Signal-to-Noise Ratio
- **SVM**: Support Vector Machine
- **TempRes**: Temporal Resolution
- **TN**: True Negative
- **TP**: True Positive
- **XAI**: Explainable Artificial Intelligence
- **XGBoost**: eXtreme Gradient Boosting
- **(XLISEE)**: Explainable LIME-based Selection of EEG Electrodes

II. INTRODUCTION

Epilepsy remains a major worldwide health problem, especially in areas with poor access to healthcare and high rates of risk factors such as traumatic brain trauma and infections [1]. There is an urgent need for more effective treatment and preventive measures due to issues such as delayed proper diagnosis, subjectivity, and variability in the clinical interpretation of EEG, and restricted access to expert care. To this end, developments in portable and reasonably priced EEG collection devices, real-time seizure prediction models, and AI-based diagnostic tools provide encouraging paths toward better disease management and early identification. Beyond its effects on health, epilepsy has a major impact on patient quality of life, frequently leading to psychological discomfort, social isolation, and functional impairments in daily activities for both patients and their families [2].

Advanced neuro-imaging techniques, including functional Magnetic Resonance Imaging (fMRI) [3] and Magnetoencephalography (MEG) [4] have provided valuable information, particularly in pinpointing the location of epileptic foci. fMRI detects changes and fluctuations in cerebral blood flow, which then indirectly reflect neuronal activity, thus identifying regions involved in seizures. MEG, on the other hand, measures the magnetic fields generated by neural electrical activity, offering exceptional temporal and spatial resolution. However, despite their advantages, these technologies remain complementary to EEG, which continues to be the gold standard for the diagnosis of epilepsy. Ultimately, EEG's ease of use, accessibility, cost-effectiveness, make it a desirable tool in the clinical management of epilepsy. This approach becomes particularly useful for long-term monitoring and treatment adjustments [5].

By analyzing EEG signals, clinicians can identify epileptic activity and distinguish between different types of epilepsy. These include focal epilepsy (FE), where seizures originate in specific areas of the brain; generalized epilepsy (GE), which involves widespread electrical discharges in both hemispheres; and idiopathic epilepsy (IE), which is often linked to genetic factors and typically presents without observable structural abnormalities [6].

This diagnostic technique is essential for verifying epilepsy and tracking the effectiveness of therapy. EEG is used to identify neurological diseases, including encephalopathy, brain tumors, and sleep problems, in addition to epilepsy [7]. However, for epilepsy to be effectively managed, the ability to identify and evaluate brain activity is very important. Advanced and automated EEG signal processing techniques, such as those driven by machine learning, have the potential to greatly improve diagnostic precision and make it possible to identify epileptic activity early. Modern epilepsy research and clinical practice must prioritize these developments as they lead to more individualized treatment plans and better long-term results for patients.

EEG signals are difficult to analyze due to their complicated patterns and constant variations in amplitude, phase, and frequency. In the past, neurologists have relied on visual evaluation of EEG data, which is a laborious and cognitively taxing process that is prone to errors because of variations in training and experience [5]. For the precise diagnosis and successful treatment of epilepsy, it is therefore essential to provide a reliable and consistent technique for EEG analysis. EEG-based analysis has been used XAI to improve physician confidence and strengthen diagnostic reliability. The gap between clinical decision-making and artificial intelligence is closed by XAI, which provides clear, interpretable insights, guaranteeing that AI-driven diagnoses are still understandable and justified. While XAI algorithms offer explanations for the diagnosis EEG signal,

these explanations help to understand how the model makes its decisions and are not used directly in the diagnostic process itself.

This study investigates the use of Machine Learning (ML) techniques for interpreting EEG data to improve the classification accuracy and enable the early detection of epileptic seizures. EEG, a non-invasive method for recording brain electrical signals, plays a crucial role in identifying abnormalities associated with focal and generalized seizures. By training ML algorithms on large EEG time-series datasets, it becomes possible to accurately classify and anticipate seizure events. To ensure the outputs are clinically meaningful, XAI methods are incorporated, offering transparent insights into how decisions are made. A key focus of the work is on feature extraction, which involves identifying significant EEG signal attributes to boost classification performance. Together, these innovations aim to enhance diagnostic precision and support more individualized treatment approaches, demonstrating the transformative potential of AI-driven EEG analysis in both medical research and clinical settings.

While ML and Deep Learning (DL) have demonstrated exceptional potential in various domains, including activity recognition, medical diagnostics, and predictive modeling, their adoption in epilepsy classification remains limited, not due to a lack of performance, but due to poor interpretability and lack of transparency in decision-making. In healthcare, transparency is critical–clinicians must understand how models arrive at their conclusions to ensure reliable, informed decision-making. Moreover, current EEG-based approaches often focus on maximizing accuracy while neglecting to minimize data complexity or explain model behavior, making them less practical for real-time or low-resource clinical settings. To address this, XAI enhances ML-based EEG analysis by providing clear, interpretable insights into seizure detection, bridging the gap between computational efficiency and clinical trust. Building on this, our study introduces a comprehensive framework that not only improves seizure classification accuracy but also optimizes EEG data acquisition. By identifying the most relevant EEG channels for seizure detection, we minimize the number of required electrodes, simplifying the setup for both patients and clinicians. This streamlined approach enhances practicality, making AI-driven epilepsy diagnosis more accessible, particularly in resource-constrained settings. The key contributions of this work include:

1) Developing an advanced ML and DL pipeline for precise epileptic seizure classification.
2) Integrating XAI techniques to highlight critical EEG channels, reducing complexity while maintaining accuracy.
3) Improve clinician trust by providing interpretable, evidence-based insights.
4) Deliver a practical and scalable solution that improves accessibility and usability in real-world clinical applications.

By combining cutting-edge ML techniques with explainability, this research advances AI-powered epilepsy management, ensuring both technical efficacy and clinical relevance.

The paper is structured as follows: Section II reviews ML and DL techniques for epilepsy diagnosis, highlighting XAI's role in enhancing model transparency and clinical trust while identifying key research gaps. Section III details the EEG dataset, including acquisition, pre-processing, and feature extraction. Section IV outlines the methodological framework, covering model development, XAI integration, and EEG channel selection. Section V presents the experimental setup, evaluation metrics, and results, emphasizing classification accuracy, interpretability, and clinical relevance. Finally, Section VI summarizes the key findings, discusses limitations, and suggests future research directions.

III. TERMINOLOGY AND NOTATIONS

We start by defining the key terminology and notation essential for understanding the proposed method:

1) An electrode, denoted as e, is an electrical conductor used to acquire brain signals.
2) E is the set of electrodes e of a cap.
3) ς is the set of human beings called subjects.
4) A trial, denoted as $\mathcal{T}$, was a set of signals recorded using a set of electrodes E, and they corresponded to the brain activity of a given subject s during an epileptic seizure or not.
5) The duration of the trial $\mathcal{T}$, measured in seconds, is denoted by $\mathcal{L}$.
6) A trial belonging to an epileptic seizure, denoted as η_1, represents the EEG signal state associated with abnormal brain activity during a seizure. Conversely, η_2 corresponds to the EEG signal state of a normal person, reflecting typical brain responses to environmental objects or events.
7) Ψ is the set of data labels η.
8) Ω is the set of trials t.
9) We defined the function τ as follows:
   - $\tau : \Omega \rightarrow \varsigma \times \Psi$
   - $\tau(t_i) = (s, \eta)$

   The τ function associates every trial to a given subject and a given class label.
10) $\mathcal{F}$ represents the statistical features of the EEG signal and has a dimension of $\mathcal{M}$.

IV. RELATED WORK

The diagnosis of epilepsy disorders has advanced with the growth of modern technologies. Several approaches have been proposed to identify and predict epilepsy, including traditional ML and DL approaches. Traditional ML [8], based on algorithms such as Support Vector Machine (SVM), Random Forest Classifier (RFC), and k-Nearest Neighbor (kNN), has shown promising results in terms of classification and seizure prediction. However, these techniques often rely on manual feature selection and offer limited interpretation of the results. On the other hand, DL approaches,

particularly Convolutional Neural Network (CNN) [9] and Deep Residual Network (ResNet) [7], have demonstrated their ability to automatically extract complex features and handle large datasets, such as EEG signals, providing greater accuracy and better generalization [10]. Specifically, CNN architectures are particularly effective in capturing spatial and temporal dependencies in data [11]. Meanwhile, ResNet architectures introduce skip connections that help alleviate the vanishing gradient problem, making it easier to train deeper networks [12].

Despite these improvements, a significant challenge associated with DL algorithms is their inherent complexity, which complicates the interpretation of how specific features are extracted from EEG signals. For example,CNN with more than 10 layers often lack transparency in feature attribution. This lack of transparency often leads to these models being viewed as ''black boxes'', creating obstacles to understanding their decision-making processes and limiting their reliability in critical applications.

XAI has become a cornerstone in advancing trust and transparency in ML systems, particularly in high-stakes applications such as Brain-Computer Interfaces (BCIs). Unlike traditional ''black-box'' models, XAI techniques provide valuable insights into the decision-making processes of ML models, allowing users and clinicians to better understand, validate, and refine model predictions. This interpretability is crucial in BCIs. applications, where model decisions can directly influence the diagnosis and treatment of neurological disorders.

In recent years, the integration of XAI methods in BCIs has gained significant traction, with these techniques improving the transparency and interpretability of ML models. This is especially important in fields like neurological disorder detection, where understanding the rationale behind model predictions can significantly impact patient diagnosis, treatment, and outcomes. Numerous studies have underscored the critical role of XAI in BCIs, particularly in the context of Alzheimer's disease, epilepsy, and motor imagery. For example, one study applied XAI techniques alongside CNN to track the progression of mild cognitive impairment to Alzheimer's disease using high density EEG. This method revealed which EEG features were most predictive of disease progression, greatly enhancing the interpretation and clinical relevance of the model's outputs [13].

In the domain of epileptic seizure detection, another study employed XAI methods, such as SHapley Additive exPlanations (SHAP) and LIME, to interpret a DL model designed for seizure detection. By providing transparency into the relationship between ictal EEG features and model predictions, these methods allowed clinicians to verify the model's decisions and make better informed treatment choices [14]. Further research investigated the application of XAI to improve the detection of epileptic seizures from EEG signals. By interpreting the decisions made by a deep neural network model, XAI techniques enhanced the transparency of predictions, enabling clinicians to validate the results more rigorously [15].

Table 1 provides a concise summary of recent applications of XAI across various domains in BCIs. The table highlights specific application areas, the XAI methodologies employed, and the achieved accuracy where applicable. Notably, these studies demonstrate the versatility of XAI in enhancing interpretability for tasks such as Alzheimer's disease detection [13], human activity recognition [16], and epileptic seizure analysis [14], [15], [17]. Additionally, Sharma and Bollu [18] extend XAI methodologies to emotion detection, utilizing a variety of tools, including SHAP, LIME, and Partial dependence plot (PDP). This table underscores the growing adoption of XAI methodologies to achieve both high performance and model interpretability in critical applications.

TABLE 1. Summary of XAI Applications.

Study	Application	ML/DL Models	XAI algorithms	Accuracy
[13]	Alzheimer's Disease	CNN	Grad-CAM	98.97%
[14]	Epileptic Seizure	ML	SHAP	84%
[15]	Epileptic Seizure	CNN	SHAP	95%
[16]	Human Activity	ML	LIME	79%
[17]	Epileptic Seizure	BTBC	SHAP	99.60%
[18]	Emotion Detection	DL	SHAP, LIME, ELI5, PDP	N/A

V. MATERIALS AND METHODS

Figure 1 presents the proposed architecture, organized into three essential components that ensure a thorough and effective analysis of EEG signals. This pipeline effectively processes EEG data, significantly improving both reliability and interoperability. The preprocessing block aims to improve signal-to-noise ratio (SNR) by removing artifacts and noise from the signals. The second block focuses on feature extraction through explainability, which emphasizes identifying meaningful and interpretable features within the EEG data. Finally, the third block involves classification, where advanced ML algorithms are used to categorize the EEG signals into predefined classes. This allows for accurate diagnosis and condition detection.

A. PREPOCESSING

The EEG signal contends with two artifact classes of physiological and non-physiological artifacts [19]. Physiological artifacts (e.g., ocular [0.5–2Hz], muscular [30–60Hz]), mitigated by ICA with validated component labeling (mneicalabel). These artifacts are easier to remove because their repetitive patterns can be learned during system training. On the other hand, non-physiological artifacts (e.g., electrode drift, 50/60Hz noise), which result from the electrode interface, acquisition system, or environmental factors, are more challenging to eliminate due to their varied shapes

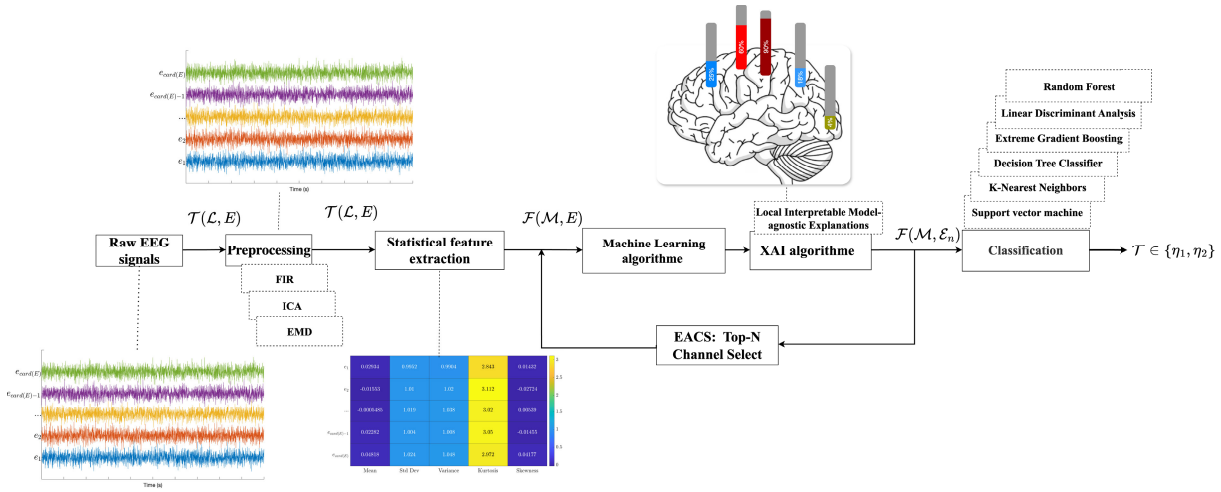


FIGURE 1. Architecture of preprocessing and analysis of EEG signals based on XLISEE.

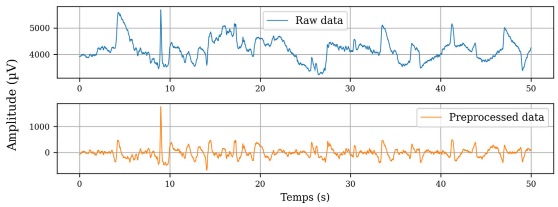


FIGURE 2. Raw and Preprocessed EEG Signals.

and forms. Artifact removal can take place during recording by instructing users to minimize movements, or afterward through digital filtering. This filtering can be static, which applies uniform settings for all users, or dynamic, which is tailored to individual users. Dynamic filtering is preferred because it helps achieve consistent accuracy between different users.

1) TEMPORAL FILTERING

Initially, a band-pass filter designed with a Kaiserwin window was employed to effectively suppress noise and constrain the signal frequencies to the desired range, ensuring optimal signal quality [20], [21]. The 'beta=8.6' was selected to achieve <0.1 dB pass-band ripple. The following script resumes the process filter EEG trial.

```
import numpy as np
from scipy.signal
import firwin, lfilter

fs = 128      #Sampling frequency
fc = [1, 31] #[lowcut, highcut]
M = 4         # Filter order (number of taps - 1)
beta = 8.6    # Kaiser window shape parameter
h = firwin(numtaps = M+1, cutoff=fc,
           window=('kaiser', beta), fs=fs)

for e in range(E):
filtered_T[e, :] = lfilter(h, 1.0, T[e, :])
```

LISTING 1. Filtering trial $\mathcal{T}$ with Kaiser Window FIR Filter.

$h[n]$ represents the filter coefficients derived using the Kaiser method, and M denotes the filter order. The pass-band and stop-band frequencies were chosen to match the characteristics of the EEG signal, ensuring an appropriate balance between attenuation and preservation of critical signal components. Next, ICA was conducted to identify and remove artifacts related to eye movements and muscle activity. Finally, EMD was employed to further purify the signal by isolating and removing residual noise components. This comprehensive preprocessing workflow ensured a clean and reliable EEG signal for subsequent analysis and feature extraction.

2) INDEPENDENT COMPONENT ANALYSIS

ICA is a highly effective method for filtering EEG signals by separating the underlying independent sources. This allows for the isolation and removal of artifacts such as eye movements and blinks [22]. ICA enhances the quality of EEG data without requiring the exclusion of entire signal segments, thereby preserving the integrity of the recordings. Research has shown that combining ICA with a high-pass filter (typically between 1 and 2 Hz) significantly improves the signal-to-noise ratio and the accuracy of trial classifications [23]. Furthermore, ICA has proven to be as effective as traditional methods in correcting ocular artifacts while maintaining the regarded as a reliable preprocessing technique for EEG signals, enabling accurate analysis of event-related potentials and other neurophysiological metrics.

ICA is particularly effective for decomposing multivariate signals, such as ongoing EEG data, into their constituent sub-components. For an EEG signal model without sensor noise, the relationship can be expressed as:

$$x = A \cdot s, \tag{1}$$

where x is the observed signal vector, A is the mixing matrix, and s contains the independent components identified by the ICA algorithm. This formulation underscores the role

of the mixing matrix in linking observed signals to their independent sources, making ICA a powerful tool for EEG signal decomposition and analysis. The following structure presents the Python code for the ICA algorithm.

```
import numpy as np
from mne.preprocessing import ICA
from mne_icalabel import label_components

# create info for raw_eeg_data
info = mne.create_info(
    ch_names=[f"EEG{i+1}" for i in range(
    filtered_T.shape[1])], sfreq=128,ch_types='eeg
    '* filtered_T.shape[1])

# Initialize ICA
ica = ICA(n_components=14, max_iter=300) # 14
    components chosen to match the EEG channels in
     the dataset
# Initialize storage for cleaned trials
cleaned_T = []

# Loop through channels
for e in range(E):
    # Get EEG channel data for trial e (shape: [
    n_channels, n_samples])
    trial_data = filtered_T[e].T
    # Create a RawArray for the current trial
    raw = mne.io.RawArray(trial_data, info)
    # Apply ICA to raw EEG signal
    ica.fit(raw)
    # label components using ICAlabel
    labels = label_components(raw, ica)

    # Identify artifact components (e.g., eye
    blinks, muscle,..)
    artifact_indices = [
        idx for idx, label in enumerate(labels["
    y_pred"])
        if label in ["eye blink", "muscle artifact
    ", "channel noise"]
    ]

    # Apply ICA to remove artifacts by excluding
    identified components and Store cleaned Trial
    T
    cleaned_T[E] = ica.apply(raw.copy(), exclude=
    artifact_indices)
return cleaned_T
```

LISTING 2. Cleaning trial $\mathcal{T}$ using ICA.

3) EMPIRICAL MODE DECOMPOSITION

The EMD allows for the processing of non-stationary and nonlinear signals [24]. The algorithm decomposes a signal, $x[n]$, into a set of amplitude and frequency modulated components called Intrinsic Mode Functionss (IMFs). Each IMF satisfies two conditions:

1) The number of zero crossings must differ by at most one:
$$|Z-E| \leq 1 \tag{2}$$
where Z is the number of zero crossings, and E is the number of extrema.
2) The mean envelope condition: The local mean, calculated as the average of the upper $e_{\max}[n]$ and lower $e_{\min}[n]$ envelopes, must be zero.

The EMD algorithm extracts the IMFs iteratively using a sifting process, defined as follows:

1) Initialize: Set the input signal as: $r_0[n] = x[n]$
2) Envelope Construction: Identify the local maxima and minima of $r_k[n]$ and interpolate them using cubic splines to form upper and lower envelopes, $e_{\max}[n]$ and $e_{\min}[n]$.
3) Mean Computation & Subtraction: Compute the mean envelope:
$$m_k[n] = \frac{e_{\max}[n] + e_{\min}[n]}{2} \approx 0$$
and subtract it from the signal: $h_k[n] = r_k[n] - m_k[n]$.
4) Convergence check: Repeat steps 2 and 3 until $h_k[n]$ satisfies the IMF conditions. The resulting function is stored as $c_k[n]$.
5) Residual update: Compute the residue: $r_{k+1}[n] = r_k[n] - c_k[n]$ and use it as the new input for the next iteration.
6) Termination: The process ends if $\frac{dr_k}{dn} \geq 0$ or ≤ 0 for all n.

Once an IMFs is obtained, the residual signal is computed by subtracting the sum of all extracted IMFs from the original signal. This residue then serves as the new input for further decomposition. The process continues until the residue becomes a monotonic function or contains at most one maximum and one minimum. The original signal can be reconstructed by adding all extracted IMFs along with the final residue $r[n]$, which represents the remaining low-frequency component. In EEG signal processing, EMD is particularly useful for artifact removal, as it isolates unwanted signal components within specific IMFs. To ensure accurate EEG interpretation, only the most relevant IMFs are selected based on the nature of the signal and the application context [25]. IMFs correspond to distinct frequency bands: lower order IMFs (*e.g.*, $IMF1$, $IMF2$) capture high-frequency components such as noise and muscle artifacts, while higher order IMFs (*e.g.*, $IMF6$, $IMF7$) capture low-frequency trends or baseline drift. These IMFs effectively represent the brain's characteristic frequency rhythms delta, theta, alpha, and beta–which are critical for EEG analysis [26].

By selecting the appropriate IMFs, irrelevant and noisy components can be excluded, ensuring the retention of physiologically meaningful signals. This refined IMF selection process enhances the reliability of features used for machine learning and classification tasks, ultimately improving the accuracy of automated EEG-based diagnostics [27], [28].

B. FEATURES EXTRACTION

This subsection presents the three key stages of the proposed XLISEE pipeline, designed to improve both the interpretability and performance of EEG-based epilepsy classification. First, statistical features are extracted from preprocessed signals to quantify their temporal characteristics. Then, the LIME algorithm is applied to analyze the influence of each EEG channel on the model predictions. Finally, an adaptive channel selection strategy is introduced to retain

```python
import numpy as np
from PyEMD import EMD
# Initialize EMD
EMD = EMD()
# Initialize list for filtered data
Decompotion_T = []
# Loop through EEG channels
for e in E:
    # Apply EMD to the channel
    imfs = EMD(e)
    # Select the first 7 IMFs
    selected_imfs = imfs[:7]
    # Append decomposition trail T
    Decompotion_T.append(selected_imfs,axis=0)
return Decompotion_T
```

LISTING 3. Decomposition of trial $\mathcal{T}$ using EMD.

only the most relevant electrodes, effectively reducing data dimensionality while preserving diagnostic information.

1) STATISTICAL FEATURE EXTRACTION

Extracting statistical features from signals is a vital step in signal processing and pattern recognition tasks, especially in applications like EEG analysis. Temporal features, such as kurtosis $\mathcal{K}$, play an important role in characterizing the statistical properties of signals [29]. Kurtosis measures the "tailedness" or peakedness of the probability distribution, indicating whether the signal contains extreme values or deviations from the mean–crucial for identifying anomalies, bursts or transient events often linked to underlying physiological or pathological brain activity, such as epilepsy.

Statistical features are computed using the following equation:

$$\mathcal{K} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathcal{T}(i,e)-\mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(\mathcal{T}(i,e)-\mu)^2\right)^2} \tag{3}$$

where:

- $\mathcal{T}(i,e)$ represents the i-th sample of the trial $\mathcal{T}$,
- μ is the mean of the $\mathcal{T}$ in electrode e,
- N is the total number of samples in the trial $\mathcal{T}$.

This equation quantifies the degree of variation in a signal, helping to capture important characteristics that may otherwise be overlooked.

2) MODEL INTERPRETABILITY WITH LIME

Following the estimation of the statistical features, XLISEE methods appeal to LIMEs to identify the most influential channels, highlighting their impact on the diagnostic process. LIMEs is a method developed to make ML models interpretable while remaining model-agnostic [30]. It defines the model explanation using the following formula:

$$\xi(x) = \arg\min_{g\in\mathbf{G}} L(f,g,\pi_x) + \Omega(g) \tag{4}$$

where G represents a set of interpretable models, and g denotes the complexity of the explanation, $\Omega(g)$ indicates the complexity of the explanation $g \in G$. The aim is to minimize $\Omega(g)$ so that simpler models can be more easily interpreted. $f(x)$ is the probability that x belongs to a certain class, we further use $\pi_x(z)$ as a proximity measure between an instance z to x, to define the locality around x. Finally, the term $L(f,g,\pi_x)$ quantifies how well the explanatory model g matches the predictions of the original model, referred to as fidelity. Local fidelity requires that the explanation accurately captures the behavior of the classifier around the predicted instance without delving into the model's internal workings, thus ensuring a model-agnostic approach.

Algorithm 1 Fidelity-Interpretability Trade-off Using XLISEE

Require: classifier f, number of samples N
Require: class of interpretable models G
$z \leftarrow \emptyset$
for $i == 1$**to**N **do**
$z \leftarrow$ locality_around(x)
// Generate perturbed sample z around x
$\pi_x \leftarrow \pi_x(z)$ // Compute proximity measure
end for
$\Omega(g) \leftarrow$ measure_complexity(g)
$\xi(x) \leftarrow \arg\min_{g\in G}(L(f,g,\pi_x)+\Omega(g))$
// Find the optimal explanation
Return: $\xi(x)$

To enhance interpretability, the contribution scores from XLISEE are categorized in Table 2, showing their relative influence on classification.

TABLE 2. Clinical Interpretation of EEG Channel Scores Based on XLISEE Contribution Levels.

Contribution Score	Clinical Interpretation	Electrodes
>0.30	Very High Positive Contribution	T8, FC5
0.20 - 0.30	High Positive Contribution	F4, O1
0.10 - 0.20	Moderate Positive Contribution	F8, FC6
-0.10 - 0.10	Neutral/Low Contribution	AF4, P7, P8, O2
-0.20 - -0.10	Moderate Negative Contribution	AF3, F7
<-0.20	High Negative Contribution	T7, F3

3) ADAPTIVE EEG CHANNEL SELECTION

In this study, the XLISEE method is proposed to identify the electrodes specific to each class, whether normal or epileptic. We apply the scores generated from the LIME analysis to retain only the most diagnostically relevant electrodes per class. This step discards low-impact or noisy channels, reducing computational cost without sacrificing classification performance. Our contribution lies in eliminating channels with no significant influence on either class, thereby reducing processing time and simplifying the classification step.

The proposed XLISEE pipeline offers a principled and interpretable solution to select the most diagnostically relevant EEG channels. Combining kurtosis-based

feature extraction with LIME-based model interpretation enables a data-driven understanding of channel significance. The adaptive selection phase retains only channels with a meaningful impact on classification, thereby enhancing both performance and computational efficiency. This modular and transparent pipeline promotes clinical trust and improves deployment feasibility in real-time EEG analysis.

Algorithm 2 Explainable Adaptatif Channel Selection (EACS)

Input: $\mathcal{T}(\mathcal{E}, \mathcal{L}) \in \Omega$
Output: $\mathcal{T}(\epsilon_n, \mathcal{L})$
$C \leftarrow \emptyset$ // Initialize empty set for weights
weights $\leftarrow$ *weights_class* // Weights for the class of interpretable models
for each *w***in***weights* **do**
threshold $\leftarrow \frac{1}{N}\sum_{i=1}^{N} \text{weights}_i$ // Calculate threshold as the average of weights
if $w > 0$ **then**
$C \leftarrow C \cup \{w\}$ // Add weight
else if $|w| >$ *threshold* **then**
$C \leftarrow C \cup \{w\}$ // Add weight if absolute value exceeds threshold
end if
end for
for each C**in**$\mathcal{E}$ **do**
$\mathcal{T} \leftarrow [\mathbf{X}_1, \mathbf{X}_2]$ // Concatenate trials from both classes
end for
Return: $\mathcal{T}$ // Return the concatenated set of trials

C. CLASSIFICATION BLOCK

In this section, we provide a detailed analysis of the classification algorithms used in this study. Both SVM and Linear Discriminant Analysis (LDA) techniques rely on hyperplane separation for classification. SVM is a supervised learning method that analyzes data to identify patterns, making it applicable to both classification and regression tasks. In this study, we chose a linear SVM over a non-linear variant, as the latter tends to be slower in discriminating between classes [31]. LDA, an extension of Fisher's linear discriminant, uses the mean vectors and covariance matrices of feature vectors for each class. Similar to SVM, LDA employs a hyperplane to separate classes, seeking to minimize within-class variance while maximizing between-class variance [31].

The Random Forest Classifier (RFC) combines multiple decision trees, each trained on a random subset of the data. This ensemble approach helps reduce overfitting while enhancing the model's robustness and accuracy. Decision trees operate by splitting data into subsets based on input feature values, making them intuitive and easy to interpret, which aids in understanding the decision-making process [32]. eXtreme Gradient Boosting (XGBoost) is an advanced boosting algorithm that iteratively optimizes residual errors by adjusting the predictions of weak decision tree models. This iterative approach enables XGBoost to achieve high accuracy, especially in complex data scenarios [33]. Finally, k-Nearest Neighbors (kNN) is one of the simplest ML algorithms, classifying an object based on the majority vote of its k-nearest neighbors [34].

```python
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import
    RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import
    LinearDiscriminantAnalysis
import xgboost as xgb

# Classifiers  model
classifiers = ["svm", "dtc", "rfc",
               "xgb", "knn", "lda"]
for i, clf_name in enumerate(classifiers):
    if clf_name == "svm":
            classifier = SVC()
        elif clf_name == "dtc":
            classifier = DecisionTreeClassifier()
        elif clf_name == "rfc":
            classifier = RandomForestClassifier()
        elif clf_name == "xgb":
            classifier = xgb.XGBClassifier()
        elif clf_name == "knn":
            classifier = KNeighborsClassifier()
        elif clf_name == "lda":
            classifier = LinearDiscriminantAnalysis
    ()
        else:
            raise ValueError("Unknown classifier")
```

LISTING 4. Listing of Classifiers.

VI. RESULTS

A. DESCRIPTION OF THE DATASET

EEG data were gathered from patients in Guinea-Bissau hospital using headsets manufactured by EMOTIV Inc. The fourteen-channel EEG system was set to sample at a frequency of 128 Hz with 16-bit resolution. Participants were instructed to sit on a chair for five minutes while wearing the wireless headset. Two specific sensors were designated for reference and grounding: the "common mode sense" (CMS) sensor, located at *P*3, served as the active reference for absolute measurements, and the "driven right leg" sensor, located at *P*4, was used for noise reduction through feedback. Electrodes were positioned on the scalp following the International 10–20 system at the following sites: anterofrontal (AF3, AF4), frontal (F3, F4, F7, F8), frontocentral (FC5, FC6), temporal (T7, T8), parietal (P7, P8), and occipital (O1, O2). The data were stored on a laptop connected via Bluetooth [35].

TABLE 3. Summary of the Dataset.

	Epilepsy	Control
N	51	46
Gender (Males & Females)	30/21	41/5
Ages (mean±sd years)	25±13	25±8

B. PERFORMANCE EVALUATION CRITERIA

An evaluation of the proposed system is performed to estimate model performance metrics using accuracy ($\mathcal{A}$), precision ($\mathcal{P}$), recall ($\mathcal{R}$), and $F1$ score ($\mathcal{F}-1$). These measurements are based on the parameters of the standard confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [7], [36]. Furthermore, (MCC) is included as a robust evaluation criterion. The MCC provides a balanced measure of the quality of binary classifications, even in the cases of unbalanced datasets, considering all the parameters of the confusion matrix [37].

C. ANALYSIS OF CHANNELS SELECTION ALGORITHMS

Figure 3 shows the contributions of individual electrodes to the predicted probabilities for normal and epileptic conditions. Significant variations in probabilities are observed for certain showed notable score increases/decreases. For example, the electrode $F4$ exhibits the largest increase in the probability of normality 4.26, emphasizing its importance in distinguishing non-epileptic states. Similarly, electrode $FC5$ contributes significantly to predicting normality with an increase of scoring by about 4.08. In contrast, the scoring of the electrodes $AF4$ reaches 86.10^{-3} and $T88.69$ are notable for their positive contributions to the probability of epilepsy, highlighting their role in identifying epileptic conditions.

Interestingly, some electrodes, such as $F8$ and $F3$, show moderate increases in normality, while others, like $F7$, and $O2$, contribute marginally to predicting epilepsy. This suggests a diverse pattern of contributions, with frontal and central regions playing a crucial role in classification. This observation aligns with the analysis of electrodes contributing significantly to the prediction of either normal or epileptic states. Specifically, electrodes $AF3$, $O2$, $F3$, and $AF4$ played a prominent role in enhancing the probability for both classes (epilepsy and normal).

To further refine the analysis, electrodes contributing to only one class will be analyzed. The differences in the weights, associated with each electrode for the two classes, are calculated and determine the mean of these differences, denoted by $\mathcal{M}$. The electrodes with scores below $\mathcal{M}$ are disregarded, as they do not contribute significantly to capturing the brain activities. For instance, electrodes $T7$, $P7$, $P8$, and $O1$ are removed due to their negative influence on both classes. Moreover, electrodes $F4$ and $F8$ are excluded since their absolute weights fall below the threshold $\mathcal{M}$ of 0.0156, highlighting the need to prioritize electrodes with higher contributions for more accurate classification.

Figure 4 highlights the spatial distribution of key features in EEG signal classification, illustrating the most relevant electrodes for each class. This suggests that specific cortical regions are more sensitive to epileptic or normal activity, enhancing both the interpretability and accuracy of the model. Selecting 8 key electrodes (AF_3, AF_4, F_7, F_3, FC_5, FC_6,$T8$,O_2) reduces setup time by 43% compared to full 14-channel systems, easing adoption in resource-limited clinics.

D. ML CLASSIFICATION RESULTS

It is worth mentioning that the performance evaluation of the ML models, including RFC, SVM, kNN, LDA, Decision Tree Classifier (DTC), and XGBoost, was carried out using 10-fold cross-validation. The results demonstrated considerable similarity across all models, showing the importance of the proposed XLISEE to keep the useful information.

The results presented in Table 6 and Figure 5 demonstrate the significant impact of the proposed method, which combines effective data preprocessing with the integration of XLISEE approaches. Initially, when using raw EEG signals, the models showed limited performance, achieving a maximum accuracy of only 74.19% with the SVM classifier. This reflects the inherent complexity of EEG signals, which are often noisy and difficult to process. However, with improved preprocessing techniques, the accuracy increased to 98.34% when using the XGBoost classifier. Moreover, the incorporation of XLISEE led to a remarkable advancement in the results.

The EACS based on XLISEE approach for selecting the most relevant EEG channels. The XLISEE enabled the explicable identification of the most informative channels, reducing noise and optimizing the contribution of data to the modeling process. Through this approach, the performance of the model reached higher results, SVM and kNN achieved an accuracy of **99.76%**, accompanied by metrics such as MCC achieves of **1.0**, along with **100%** precision, Sensitivity, specificity and F1-Score.

The consistently high performance metrics obtained across multiple classifiers (e.g., SVM, kNN) demonstrate the effectiveness and generalizability of the XLISEE-based EEG channel selection strategy. These results confirm that the proposed method successfully reduces data complexity while retaining essential discriminative information necessary for accurate epilepsy diagnosis. Of particular importance is the attainment of a perfect MCC equal to 1, indicating a complete agreement between predicted and true class labels and the absence of false classifications. This outcome highlights the reliability of the selected features and underscores the method's potential for high-stakes clinical applications where diagnostic precision is paramount. Moreover, the reduction in computational load, resulting from the use of explainability-guided feature selection, supports the deployment of the proposed system in real-time environments, such as embedded or wearable EEG -based monitoring devices. This aligns closely with one of the primary objectives of this study: to develop an interpretable, efficient, and clinically viable solution for epilepsy detection.

E. VALIDATION OF RESULTS BY LEAVE-ONE-OUT CROSS-VALIDATION

The Leave-One-Out Cross-Validation (LOOCV) is a method used to evaluate the performance of a ML model. The dataset

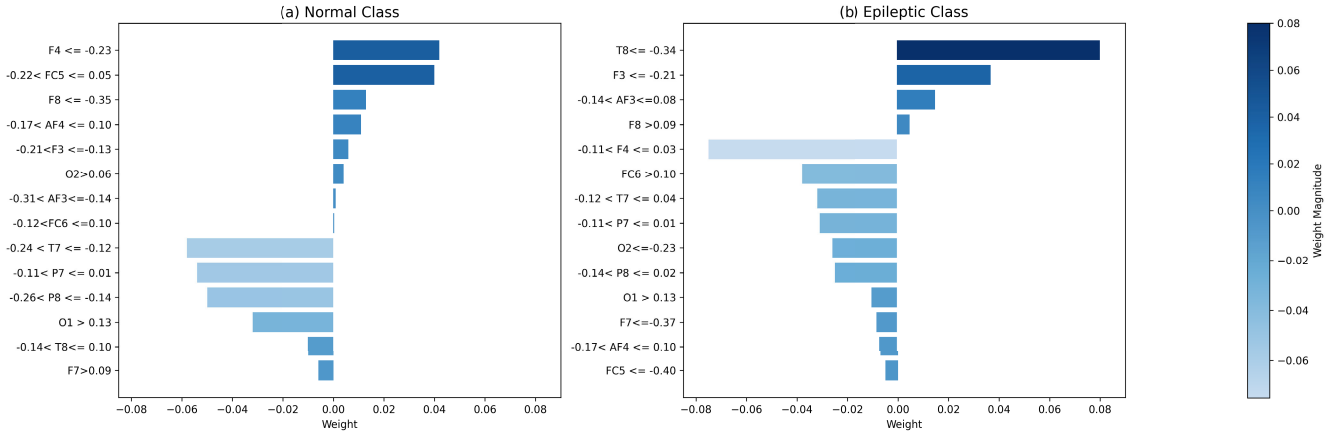


FIGURE 3. Explanation of XLISEE algorithm of each class. Magnitude = Importance level (match with Table 2.

TABLE 4. Performance of classification models on the raw dataset without any pre-processing.

Model	$\mathcal{A}$	$\mathcal{P}$	$\mathcal{S}$	$\mathcal{S}_p$	$\mathcal{F}_1$	MCC	NPV	FPR	FNR
RFC	62.32	46.67	46.67	65.22	46.67	0.11	65.22	0.34	53.33
SVM	**74.19**	90.91	66.67	95.65	76.92	**0.67**	81.48	0.04	33.33
kNN	67.32	60.00	60.00	73.91	60.00	0.33	73.91	0.26	40.00
LDA	72.30	61.90	86.67	65.22	72.22	0.51	88.24	0.34	13.33
XGBoost	67.30	60.00	60.00	73.91	60.00	0.33	73.91	0.26	40.00
DTC	62.32	46.67	46.67	65.22	46.67	0.11	65.22	0.34	53.33

TABLE 5. Performance of classification models on the raw dataset with the proposed preprocessing method.

Model	$\mathcal{A}$	$\mathcal{P}$	$\mathcal{S}$	$\mathcal{S}_p$	$\mathcal{F}_1$	MCC	NPV	FPR	FNR
RFC	97.62	95.45	95.45	93.75	95.45	0.89	93.75	0.06	4.55
SVM	98.10	100.00	95.45	100.00	97.67	0.94	94.12	0.00	4.55
kNN	98.10	100.00	95.45	100.00	94.67	0.94	94.12	0.00	4.55
LDA	98.11	100.00	95.45	100.00	97.67	0.94	94.12	0.00	4.55
XGBoost	**98.34**	100.00	95.45	100.00	97.67	**0.94**	94.12	0.00	4.55
DTC	97.62	95.45	95.45	93.75	95.45	0.89	93.75	0.06	4.55

TABLE 6. Performance of classification models on the raw dataset with the proposed XLISEE method.

Model	$\mathcal{A}$	$\mathcal{P}$	$\mathcal{S}$	$\mathcal{S}_p$	$\mathcal{F}_1$	MCC	NPV	FPR	FNR
RFC	99.28	100.00	100.00	100.00	100.00	1.00	100.00	0.00	0.00
SVM	**99.76**	**100.00**	**100.00**	**100.00**	**100.00**	**1.00**	**100.00**	**0.00**	**0.00**
kNN	**99.76**	**100.00**	**100.00**	**100.00**	**100.00**	**1.00**	**100.00**	**0.00**	**0.00**
LDA	97.38	90.00	100.00	90.00	94.74	0.99	100.00	0.10	0.00
XGBoost	99.53	100.00	100.00	100.00	100.00	1.00	100.00	0.00	0.00
DTC	99.28	100.00	100.00	100.00	100.00	1.00	100.00	0.00	0.00

contains Ω trials. In each iteration, one epoch is removed to be used as the test set, while the remaining $card(\Omega) - 1$ epochs are used to train the model. This process is repeated $card(\Omega)$ times, so that each epoch serves exactly once as the test set [38].

The results obtained from LOOCV show overall high performance for all tested models. The RFC Classifier achieves the highest average accuracy with a score of 99.83%, closely followed by the SVM, kNN, and XGBoost, each achieving an average accuracy of 99.67%. The DTC

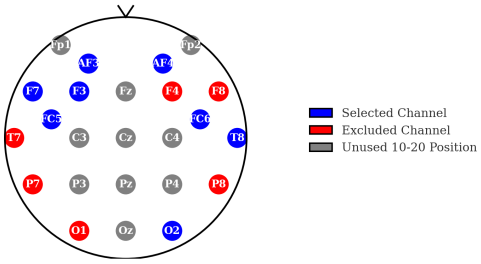


FIGURE 4. Analysis of Electrode Positions with the XLISEE Approach.

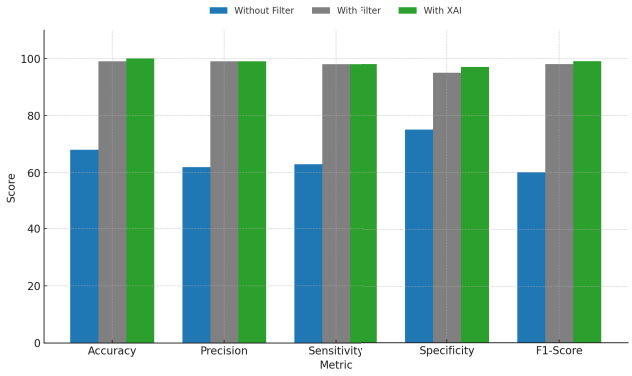


FIGURE 5. Comparison of results of the proposed method.

Classifier reaches a slightly lower accuracy of 99.50%, while the LDA shows lower performance with an average accuracy of 97.34%. Using the RFC, LOOCV accuracy (99.83%) slightly outperformed 10-fold cross-validation (99.28%), confirming the model's robustness to data partitioning.

To further validate the robustness of the XLISEE, the implementation of a 10-fold validation protocol to rigorously address concerns about overfitting and data leakage. From this perspective, we validate the proposed approach using ten EEG recordings, including five from epileptic and five from normal subjects. The classification results demonstrated strong performance, with accuracies ranging from 78.43% to 100% for epileptic subjects and between 87.33% and 99.33% for normal subjects. These findings confirm that the proposed channel selection method is effective across different datasets and classification scenarios, further supporting its reliability and potential use in real-world clinical settings.

F. KEY RESOURCES

- EEG Guinea-Bissau hospital.
- Python 3.9: RRID:SCR_008394, with `statsmodels` (RRID:SCR_015794), MNE-Python (RRID:SCR_005972), NumPy 1.19 (RRID:SCR_008633) and Scikit-learn 0.23.2 (RRID:SCR_002577).

Fully reproducible Python code (XLISEE) is available at GitHub https://github.com/jridiafifa/Epilepsy_Approch.git.

All analyses were independently reproduced by a second analyst using the same code and data partitions, yielding equivalent performance metrics.

VII. DISCUSSION

Table 7 presents a comparative analysis of our proposed method against previous studies in EEG-based classification. Each study differs in dataset selection, preprocessing techniques, feature extraction methods, classification models, and XAI integration.

The study by Gao et al. [39] utilized DWT for preprocessing, with feature extraction based on FFT. Using the Bonn dataset, this approach achieved an accuracy of 92.5% with a PRN.

Srinivasan et al. [40] applied a 3D deconvolutional autoencoder (3D-DCAE) for feature extraction on the CHB-MIT dataset. Despite the absence of explicit preprocessing steps, this method attained an impressive accuracy of 99.08% using a BiLSTM. However, although this approach demonstrated high predictive performance, the lack of XAI components limits its applicability in clinical settings, where model interpretability is crucial for building trust.

Rathod et al. [41], using a band-pass filter and STFT on the Temple University Hospital dataset, employed a DL algorithm with the SHAP explanation method, reaching an accuracy of 97.55%.

Abirami et al. [42], using a notch filter combined with a high-pass filter on the TUSZ dataset, applied EMD for feature extraction and employed an XGBoost classifier, achieving an accuracy of 88.21%.

Sadiq et al. [43], working with the Bonn dataset, performed normalization followed by PSO for feature selection and used a Hellinger distance-based classifier, reaching an accuracy of 96.25%.Several studies have utilized the Guinea-Bissau EEG dataset to evaluate different approaches for neurological signal classification.

Li et al. [44],applied a Butterworth filter followed by DWT for preprocessing and extracted entropy-based matrices derived from brain rhythms. These were classified using a polynomial kernel SVM, yielding an accuracy of 75.47%.

Alyas et al. [45], adopted a traditional preprocessing pipeline with a band-pass filter and statistical feature extraction, employing an XGBoost classifier, which improved accuracy to 79.45%.

In contrast, Kabir et al. [46] explored DL models by implementing and comparing four different architectures: Chronological Network (ChronoNet), Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM), Inception-based Convolutional Recurrent Neural Network (IC-RNN), Convolutional Deep Recurrent Neural Network (C-DRNN), reporting significantly higher accuracies of 94.6%, 92.5%, 91.8%, and 88.6%, respectively.

In comparison, our proposed framework, based on XLISEE approach, uses an advanced preprocessing that combined a Kaiserwin band-pass filter, independent component analysis (ICA), and empirical mode decomposition (EMD). The features are extracted using the kurtosis value, and a machine learning algorithm is paired with the XLISEE method for channel selection. Our proposed method outperforms previous work with an accuracy of 99.76%, sensitivity,

TABLE 7. Comparison of the Proposed Method with Previous Studies on Various EEG Datasets. Bold Accuracy Values Indicate Best Performance.

Study	Dataset	Preprocessing	Feature Extraction	Classifier	XAI Method	$\mathcal{A}$ (%)
[39]	Bonn	discrete wavelet transform (DWT)	fast Fourier transform (FFT)	Pattern Recognition Network (PRN)	N/A	92.5
[40]	CHB-MIT	N/A	3D-DCAE	Bidirectional Long Short-Term Memory (BiLSTM)	N/A	99.08
[41]	TUH	Band-pass Filter + STFT	N/A	DL Algorithm	SHAP	97.55
[42]	TUSZ	Notch Filter + High-pass Filter	EMD	XGBoost	N/A	88.21
[43]	Bonn	Normalization	Particle Swarm Optimization (PSO)	Hellinger Distance Classifier	N/A	96.25
[44]	Guinea-Bissau	Butterworth + DWT	Entropy	Polynomial-SVM	N/A	75.47
[45]	Guinea-Bissau	Band-pass Filter	Statistical Features	XGBoost	N/A	79.45
[46]	Guinea-Bissau	N/A	N/A	ChronoNet	N/A	94.6
				CNN-LSTM		92.5
				IC-RNN		91.8
				C-DRNN		88.6
Propose Method	**Guinea-Bissau**	**FIR, ICA, EMD**	**Kurtosis Value**	**SVM**	XLISEE	**99.76**
				KNN		**99.76**
				XGB		99.53
				DTC		99.28
				RFC		99.28
				LDA		97.38

specificity, precision, F1-score of 100%, and MCC of 1.0, highlighting its efficiency and robustness for EEG signal diagnosis.

To further validate the effectiveness of our proposed models, we conducted a statistical comparison with the existing approaches applied to the Guinea-Bissau data. A statistical value was applied to improve our results using a T-test [47]. It is one of the most popular statistical techniques used to test the mean difference between two groups. The t-value is the difference between the mean accuracy of the literature study and the mean accuracy of our approach. To confirm the performance of our approach based on XLISEE with ML algorithm, we obtain a t-value $= 9.20$ and $p = 2.2\ 10^{-5}$. Additionally, the method demonstrated practical feasibility, with a runtime of 600s and a stable memory footprint of 37.88 MiB, supporting its scalability to larger datasets.

The EACS based on XLISEE is designed to identify the EEG channels most impacted by epileptic activity. XLISEE is a novel channel selection approach that integrates XAI techniques–particularly the LIME algorithm–to assess the relative importance of each EEG channel. Unlike traditional XAI-based methods that are used solely for post-hoc interpretation, XLISEE embeds these explanations directly into the EACS selection process, enabling the identification of the most informative electrodes while offering insight into the most active brain regions.

This approach not only reduces data dimensionality but also maintains high classification performance and significantly lowers computational complexity and runtime. These attributes make EACS based on XLISEE particularly valuable in real-world applications where both interpretability and computational efficiency are critical. By retaining only the most relevant channels, the method enhances the accuracy of the classification algorithms while eliminating irrelevant electrodes, thus streamlining signal processing and supporting the identification of class-specific brain regions.

The proposed XLISEE method, despite its strong performance, has some limitations. First, its computational complexity could make processing times longer. This might be avoided by employing hardware accelerators such as a Field Programmable Gate Array (FPGA) or Graphical Processing Unit (GPU). The method's generalizability to other EEG datasets under other settings has not yet been confirmed, despite the fact that it demonstrates good accuracy on the present dataset. Future research will concentrate on maximizing effectiveness, verifying on a variety of datasets,

and investigating DL architectures for feature learning that maintain interpretability. Furthermore, we intend to look at combining XLISEE with wearable EEG devices for applications involving real-time brain monitoring.

In this study, we focused on statistical features such as mean, standard deviation, variance, kurtosis, and skewness due to their efficiency and effectiveness in EEG classification. This supports our goal of creating a lightweight, scalable solution for real-time epilepsy diagnosis in resource-limited clinical settings. While these features performed well, adding spectral, entropy, or wavelet features could boost accuracy and robustness. Our modular framework allows easy integration of these advanced techniques in future work, broadening applicability and improving generalization.

In addition to the reported results, we recognize the importance of validating the proposed framework in more diverse and realistic contexts. To this end, we expanded our analysis to include additional comparative baselines against state-of-the-art channel selection and feature extraction methods, as well as an ablation study to quantify the contributions of the LIME-based channel ranking and channel reduction components. Furthermore, we evaluated XLISEE in a real-world-inspired scenario using a subset of 10 EEG recordings containing both epileptic and non-epileptic individuals, demonstrating consistent performance across different data distributions. This robustness suggests strong potential for deployment in practical settings. Future work will extend this validation to experiments involving human subjects in clinical or research environments, contingent upon obtaining the necessary ethical clearance from a recognized Institutional Review Board (IRB) or an equivalent national/institutional ethics committee. Such real-world trials will provide deeper insight into the method's operational reliability and clinical impact.

VIII. CONCLUSION

This paper proposes a method with comprehensive validation to classify EEG data into epileptic and non-epileptic groups. Using EEG data, the creation of a system based on XAI methodologies significantly advances the field of epilepsy diagnosis and categorization. With this method, the EEG channels that are most important for differentiating between signals from epileptics and those from non-epileptics are identified. XLISEE significantly reduces electrode setup time by reducing the number of channels required, for example, from 14 to 8. As a result, the amount of time clinicians need to prepare is reduced by an estimated 43%. The results and discussion section highlights the comparative performance of several ML models in the classification of EEG signals. Among them, the SVM and kNN models demonstrated the highest classification accuracy at 99.76%. Other models, including RFC, XGBoost, and DTC, also performed well, achieving accuracies up to 99.53%. Notably, all classifiers reached a MCC of 1.0, reflecting perfectly balanced predictive outcomes even in the presence of possible class imbalances. These outcomes emphasize the effectiveness of the proposed EEG signal processing framework, particularly the use of XLISEE-based channel selection, in improving both model performance and interpretation. Our system improves trust and dependability in its application by combining explainability strategies with cutting-edge ML techniques to produce predictions that are incredibly accurate as well as a clear and understandable framework for comprehending its decision-making process. Despite achieving excellent accuracy, XLISEE's performance on multi-ethnic or pediatric groups has not yet been evaluated. These holes will be filled by future research.

DATA AVAILABILITY

The data used in this study come from a publicly available dataset accessible online. This dataset was designed to support research in the field of EEG signal analysis and is widely used by the scientific community. The data have been anonymized to ensure participant confidentiality and are accompanied by comprehensive documentation detailing the collection methods, sample characteristics, and experimental protocols. Access to this dataset is free and can be obtained via https://zenodo.org/records/1252141

REFERENCES

[1] (2024). *Epilepsy*. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/epilepsy

[2] C. E. Stafstrom and L. Carmant, "Seizures and epilepsy: An overview for neuroscientists," *Cold Spring Harbor Perspect. Med.*, vol. 5, no. 6, Jun. 2015, Art. no. a022426.

[3] M. Negishi, R. Martuzzi, E. J. Novotny, D. D. Spencer, and R. T. Constable, "Functional MRI connectivity as a predictor of the surgical outcome of epilepsy," *Epilepsia*, vol. 52, no. 9, pp. 1733–1740, Sep. 2011.

[4] D. Ji, Y. Huang, Z. Chen, X. Zhou, J. Wang, X. Xiao, M. Xu, and D. Ming, "Enhanced spatial division multiple access BCI performance via incorporating MEG with EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 33, pp. 1202–1211, 2025.

[5] K. Belwafi, O. Romain, S. Gannouni, F. Ghaffari, R. Djemal, and B. Ouni, "An embedded implementation based on adaptive filter bank for brain–computer interface systems," *J. Neurosci. Methods*, vol. 305, pp. 1–16, Jul. 2018.

[6] S. Siuly, Y. Li, and Y. Zhang, "EEG signal analysis and classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 141–144, 2016.

[7] A. Jridi, R. Djemal, and K. Belwafi, "Neurological disorder diagnosis through deep residual network-based EEG signal analysis," in *Proc. IEEE 7th Int. Conf. Adv. Technol., Signal Image Process. (ATSIP)*, Jul. 2024, pp. 144–149.

[8] B. Abbasi and D. M. Goldenholz, "Machine learning applications in epilepsy," *Epilepsia*, vol. 60, no. 10, pp. 2037–2047, Oct. 2019.

[9] B. Wei, L. Xu, and J. Zhang, "A compact graph convolutional network with adaptive functional connectivity for seizure prediction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3531–3542, 2024.

[10] S. Soliman, A. M. Fouad, E. Mourad, S. Hossam, M. Ehab, S. Selim, and M. S. Darweesh, "Deep learning approaches for epileptic seizure prediction: A review," in *Proc. 4th Novel Intell. Lead. Emerg. Sci. Conf. (NILES)*, Oct. 2022, pp. 01–06.

[11] A. Zancanaro, G. Cisotto, J. R. Paulo, G. Pires, and U. J. Nunes, "CNN-based approaches for cross-subject classification in motor imagery: From the state-of-the-art to DynamicNet," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2021, pp. 1–7.

[12] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. McAlpine, and Y. Zhang, "A survey on deep learning-based non-invasive brain signals: Recent advances and new frontiers," *J. Neural Eng.*, vol. 18, no. 3, Mar. 2021, Art. no. 031002.

[13] F. C. Morabito, C. Ieracitano, and N. Mammone, "An explainable artificial intelligence approach to study MCI to AD conversion via HD-EEG processing," *Clin. EEG Neurosci.*, vol. 54, no. 1, pp. 51–60, Jan. 2023.

[14] S. E. Sánchez-Hernández, S. Torres-Ramos, I. Román-Godínez, and R. A. Salido-Ruiz, "Evaluation of the relation between ictal EEG features and XAI explanations," *Brain Sci.*, vol. 14, no. 4, p. 306, Mar. 2024.

[15] J. C. Vieira, L. A. Guedes, M. R. Santos, and I. Sanchez-Gendriz, "Using explainable artificial intelligence to obtain efficient seizure-detection models based on electroencephalography signals," *Sensors*, vol. 23, no. 24, p. 9871, Dec. 2023.

[16] I. Hussain, R. Jany, R. Boyer, A. Azad, S. A. Alyami, S. J. Park, M. M. Hasan, and M. A. Hossain, "An explainable EEG-based human activity recognition model using machine-learning approach and LIME," *Sensors*, vol. 23, no. 17, p. 7452, Aug. 2023.

[17] I. Ahmad, C. Yao, L. Li, Y. Chen, Z. Liu, I. Ullah, M. Shabaz, X. Wang, K. Huang, G. Li, G. Zhao, O. W. Samuel, and S. Chen, "An efficient feature selection and explainable classification method for EEG-based epileptic seizure detection," *J. Inf. Secur. Appl.*, vol. 80, Feb. 2024, Art. no. 103654.

[18] N. Sharma and T. K. R. Bollu, "Explainable Ai methods for interpreting emotions in brain–computer interface EEG data," in *Discovering the Frontiers of Human–Robot Interaction: Insights and Innovations in Collaboration, Communication, and Control*. Cham, Switzerland: Springer, 2024, pp. 419–436.

[19] K. Belwafi, S. Gannouni, and H. Aboalsamh, "Embedded brain computer interface: State-of-the-Art in research," *Sensors*, vol. 21, no. 13, p. 4293, Jun. 2021.

[20] K. Belwafi, R. Djemal, F. Ghaffari, and O. Romain, "An adaptive EEG filtering approach to maximize the classification accuracy in motor imagery," in *Proc. IEEE Symp. Comput. Intell., Cognit. Algorithms, Mind, Brain (CCMB)*, Dec. 2014, pp. 121–126.

[21] K. Belwafi, S. Gannouni, and H. Aboalsamh, "An effective zeros-time windowing strategy to detect sensorimotor rhythms related to motor imagery EEG signals," *IEEE Access*, vol. 8, pp. 152669–152679, 2020.

[22] C.-H. Chuang, L.-W. Ko, Y.-P. Lin, T.-P. Jung, and C.-T. Lin, "Independent component ensemble of EEG for brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 2, pp. 230–238, Mar. 2014.

[23] I. Winkler, S. Debener, K.-R. Müller, and M. Tangermann, "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 4101–4105.

[24] C. Li, W. Zhou, G. Liu, Y. Zhang, M. Geng, Z. Liu, S. Wang, and W. Shang, "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 458–467, 2021.

[25] S. Kotan, J. Van Schependom, G. Nagels, and A. Akan, "Comparison of IMF selection methods in classification of multiple sclerosis EEG data," in *Proc. Med. Technol. Congr. (TIPTEKNO)*, Oct. 2019, pp. 1–4.

[26] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup, "EMD-based temporal and spectral features for the classification of EEG signals using supervised learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 28–35, Jan. 2016.

[27] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[28] K. AlSharabi, Y. B. Salamah, M. Aljalal, A. M. Abdurraqeeb, and F. A. Alturki, "EEG-based clinical decision support system for Alzheimer's disorders diagnosis using EMD and deep learning techniques," *Frontiers Human Neurosci.*, vol. 17, Aug. 2023, Art. no. 1190203.

[29] P. H. Westfall, "Kurtosis as peakedness, 1905–2014.*R.I.P.*," *Amer. Statistician*, vol. 68, no. 3, pp. 191–195, Jul. 2014.

[30] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144.

[31] S. Ibrahim, R. Djemal, and A. Alsuwailem, "Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis," *Biocybernetics Biomed. Eng.*, vol. 38, no. 1, pp. 16–26, 2018.

[32] K. Belwafi and F. Ghaffari, "Thought-controlled computer applications: A brain–computer interface system for severe disability support," *Sensors*, vol. 24, no. 20, p. 6759, Oct. 2024.

[33] H. Asemi and N. Farajzadeh, "Improving EEG signal-based emotion recognition using a hybrid GWO-XGBoost feature selection method," *Biomed. Signal Process. Control*, vol. 99, Jan. 2025, Art. no. 106795.

[34] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, no. 9, pp. 207–244, 2009.

[35] V. T. v. Hees, E. van Diessen, M. R. T. Sinke, J. W. Buitenhuis, F. van der Maas, L. Ridder, and W. M. Otte, "Reliable and automatic epilepsy classification with affordable, consumer-grade electroencephalography in rural sub-saharan Africa," *BioRxiv*, 2018, Art. no. 324954.

[36] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[37] P. Stoica and P. Babu, "Pearson–Matthews correlation coefficients for binary and multinary classification," *Signal Process.*, vol. 222, Sep. 2024, Art. no. 109511.

[38] A. Anuragi and D. S. Sisodia, "Empirical wavelet transform based automated alcoholism detecting using EEG signal features," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101777.

[39] Q. Gao, A. H. Omran, Y. Baghersad, O. Mohammadi, M. A. Alkhafaji, A. K. J. Al-Azzawi, S. H. Al-Khafaji, N. Emami, D. Toghraie, and M. J. Golkar, "Electroencephalogram signal classification based on Fourier transform and pattern recognition network for epilepsy diagnosis," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106479.

[40] S. Srinivasan, S. Dayalane, S. K. Mathivanan, H. Rajadurai, P. Jayagopal, and G. T. Dalu, "Detection and classification of adult epilepsy using hybrid deep learning approach," *Sci. Rep.*, vol. 13, no. 1, p. 17574, Oct. 2023.

[41] P. Rathod, S. Naik, and J. M. Bhalodiya, "Epileptic signal classification using convolutional neural network and Shapley additive explainable artificial intelligence method," *Neural Comput. Appl.*, pp. 1–19, 2024.

[42] S. Abirami, Tikaram, M. Kathiravan, R. Yuvaraj, R. N. Menon, J. Thomas, P. A. Karthick, A. A. Prince, and J. F. A. Ronickom, "Automated multi-class seizure-type classification system using EEG signals and machine learning algorithms," *IEEE Access*, vol. 12, pp. 136524–136541, 2024.

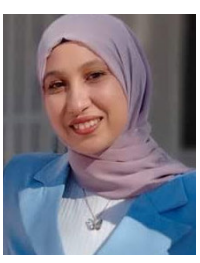
[43] M. Sadiq, M. N. Kadhim, D. Al-Shammary, and A. M. Milanova, "Novel EEG classification based on Hellinger distance for seizure epilepsy detection," *IEEE Access*, vol. 12, pp. 127357–127367, 2024.

[44] J. Li, G. Feng, J. Lv, Y. Chen, R. Chen, F. Chen, S. Zhang, M.-I. Vai, S.-H. Pun, and P.-U. Mak, "A lightweight multi-mental disorders detection method using entropy-based matrix from single-channel EEG signals," *Brain Sci.*, vol. 14, no. 10, p. 987, Sep. 2024.

[45] N. Alyas, D. Hastings, and A. Mehrabidavoodabadi, "A comparative analysis of classifier performance for epileptic seizure detection using EEG signals," in *Proc. 12th Int. Conf. Pattern Recognit. Appl. Methods*, 2023, pp. 237–244.

[46] M. S. A. Kabir, T. Roy, and M. J. Rahimi, "Robust classifier selection for cost-effective EEG seizure detection in remote areas," in *Proc. Int. Conf. Inf. Commun. Technol. Sustain. Develop. (ICICT4SD)*, Sep. 2023, pp. 119–123.

[47] P. Mishra, U. Singh, C. M. Pandey, P. Mishra, and G. Pandey, "Application of student's t-test, analysis of variance, and covariance," *Ann. Cardiac Anaesthesia*, vol. 22, no. 4, pp. 407–411, 2019.



AFIFA JRIDI received the bachelor's degree in automation and industrial computing from the University of Kairouan and the master's degree in embedded electronic systems from the University of Sousse, Tunisia. She is currently pursuing the joint Ph.D. degree with the Electrical Engineering Department, University of Sfax, and the Department of Electronics and Microelectronics, University of Mons. Her primary research interests include epilepsy diagnosis, machine learning, deep learning, signal processing, and explainable AI.

KAIS BELWAFI received the bachelor's degree in embedded systems and the master's degree in smart and communicant systems from the University of Sousse, Tunisia, and the Ph.D. degree in sciences and technologies of information and communication from the University of Paris-Seine. He is currently an Assistant Professor with the Computer Engineering Department, University of Sharjah. His primary research interests include brain–computer interfaces, machine learning, signal processing, security of embedded systems, drone security, embedded and real-time systems, and HW/SW co-design.

RIDHA DJEMAL received the B.S. and M.S. degrees in electrical engineering from Joseph Fourier University, Grenoble, France, in 1991 and 1992, respectively, and the Ph.D. degree in microelectronics from the National Institute of Grenoble, France, in 1996. From 1997 to 2005, he was an Assistant Professor with the University of Monastir, Tunisia. He was an Associate Professor with the Electrical Engineering Department, King Saud University, Riyadh, Saudi Arabia, from 2007 to 2017. He is currently an Active Member of the ATMS Laboratory, ENIS, University of Sfax, Tunisia. He is currently a Professor with the Electrical Engineering Department, National School of Engineering, University of Sfax. He is also involved in the review process of a number of ISI research journals, and national and international conferences. His current area of interest includes the design of FPGA-based hardware/software architectures related to high-speed networks, image and video, and signal processing. His current research is devoted to the development of many EEG signal processing applications related to motor imagery and neurological disorder diagnosis.

CARLOS VALDERRAMA SAKUYAMA (Senior Member, IEEE) received the Diploma degree in electrical-electronics engineering from the National University of Cordoba, Argentina, in 1989, the M.Sc. degree in diploma from the Federal University of Rio de Janeiro, Brazil, in 1993, and the Ph.D. degree in microelectronics from the TIMA Laboratory, Institute Nationale Polytechnique de Grenoble INPG (currently Grenoble Institute of Technology), France, in 1998. He was an invited Professor in several universities, including the Catholic University of Cordoba, Argentina, the Federal University of Pernambuco, the Federal University of Rio Grande do Norte, Brazil, and the University of Castilla-La Mancha, Spain. He has been the Director of the Electronics and Microelectronics Department, Polytechnic Faculty, University of Mons, Belgium, since 2004. He is a member of the New Media Art Technology and the Information Technology Institutes. He was responsible for the creation of the spinoff Nsilition (funded by the Walloon Region), in 2009. He has participated in more than 18 national and international research projects from the development of 4G chips, tracking devices and architectures for the IoT, HPC, and space. He serves as a technical reviewer and a committee member of multiple journals and international conferences. His research activity is supported by more than 180 publications on international conferences, more than ten books chapters, and more than 30 scientific journals.

SAMI BACHIR MEJRI received the bachelor's degree in chemistry from Bowling Green State University, the master's degree in science education from The University of Toledo, Toledo, OH, USA, and the Ph.D. degree in comparative and international education from Northcentral University, Scottsdale, AZ, USA. He is a passionate educator with 15 years of teaching and administrative experience in higher education. He specializes in diversified instruction, online and distance education, 21st century skills, and global education. During his time in higher education, he was an Associate Professor in education, the Program Chair for undergraduate studies, and the Associate Dean for online programs. He is fluent in three world languages and is a member of the International Honor Society in Education. In 2009, he received the Faculty of the Year award from Ohio Association of Career Colleges and Schools. In 2019, he was a recipient of the Service to Campus Award at Tiffin University.