# Nonlinear Feature Extraction for EEG Signals Using Entropy and Fractal Complexity Metrics

## Abstract

This study presents a systematic approach for extracting nonlinear features from electroencephalogram (EEG) signals. A set of five nonlinear complexity features—Sample Entropy, Permutation Entropy, Approximate Entropy, Lempel-Ziv Complexity, and Higuchi Fractal Dimension—were computed over sliding windows of EEG data. The preprocessing includes signal downsampling and window segmentation to ensure computational efficiency and temporal resolution. The extracted features are visualized using boxplots, kernel density estimation (KDE), and correlation heatmaps to explore inter-feature relationships and distributions. This work enables robust and interpretable feature extraction for downstream applications like seizure detection, mental state classification, and BCI systems.

## I. INTRODUCTION

EEG-based brain signal analysis has emerged as a vital tool in neuroscience, healthcare, and brain-computer interfaces (BCIs). Traditional linear methods are often insufficient to capture the underlying dynamical properties of EEG signals due to their inherent non-stationarity and complexity. This paper explores nonlinear feature extraction techniques applied to multichannel EEG signals to quantify irregularity, unpredictability, and complexity using entropy and fractal-based measures.

## II. DATA PREPROCESSING

A. Dataset  
The input EEG data is assumed to be stored in a cleaned `parquet` format (`eeg\_dataset\_cleaned.parquet`) with at least five channels of continuous EEG recordings sampled at 256 Hz.

B. Downsampling  
To reduce computational cost, all EEG channels are downsampled using an FIR filter and zero-phase filtering:  
f\_s^{new} = 256 / 16 = 16 Hz  
This allows sufficient temporal resolution for low-frequency components while increasing processing speed.

C. Window Segmentation  
The signal is divided into fixed-duration windows of 5 seconds:  
Window size = 16 Hz \* 5 sec = 80 samples  
Each window is processed independently to extract time-localized features.

## III. NONLINEAR FEATURE EXTRACTION

For each 5-second window of each EEG channel, five nonlinear features are computed. A `safe\_compute` wrapper ensures the robustness of the process by handling computation errors gracefully.

A. Sample Entropy (SampEn)  
Measures the complexity and self-similarity of the time-series. It evaluates the likelihood that patterns remain similar over successive comparisons.  
SampEn(m, r) = -ln(A / B)  
where A and B are counts of matches of length m+1 and m, respectively.

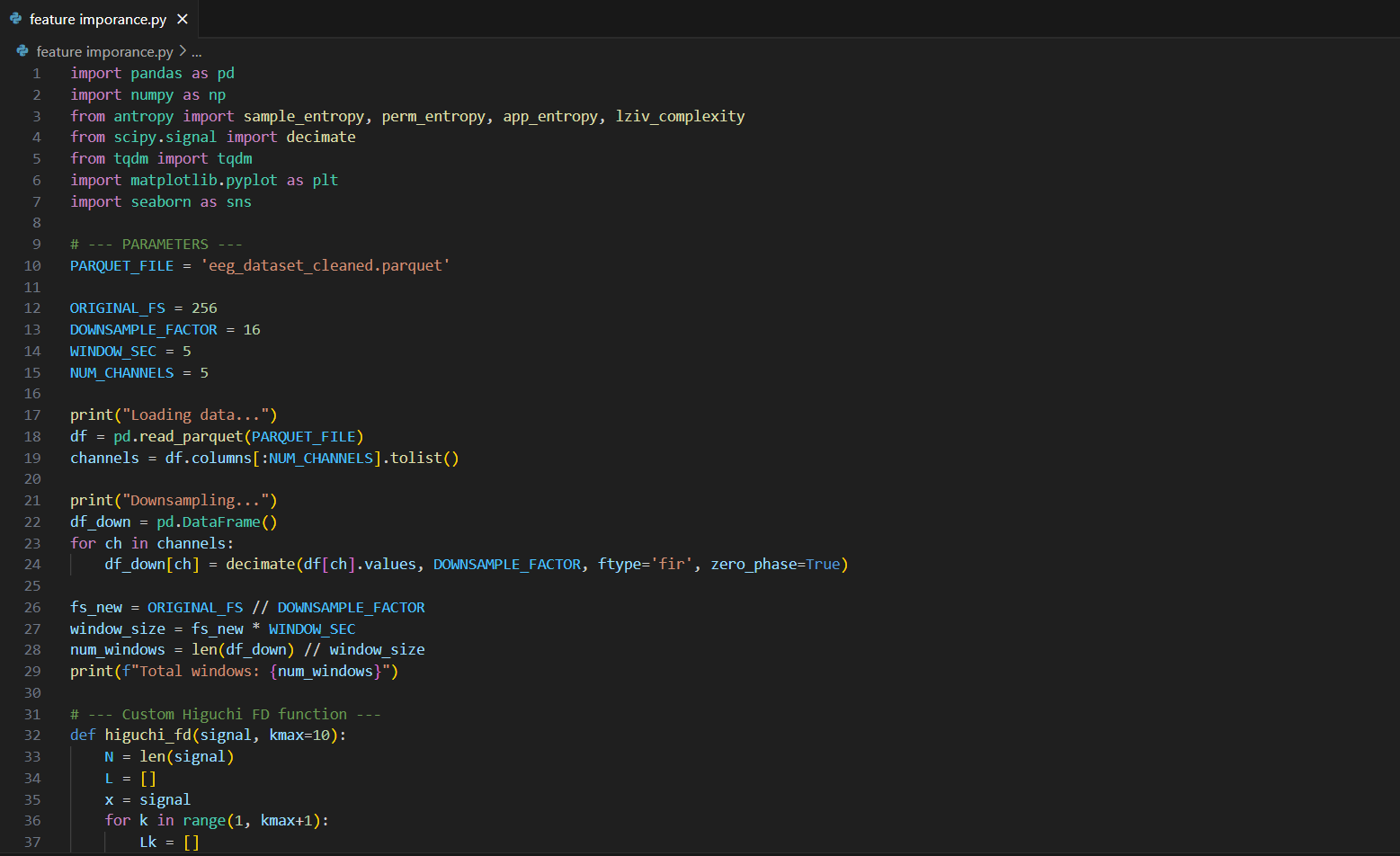
B. Approximate Entropy (ApEn)  
Quantifies the predictability of fluctuations in the signal by evaluating the logarithmic probability that similar patterns remain similar.  
ApEn(m, r, N) = Φ^m(r) - Φ^{m+1}(r)

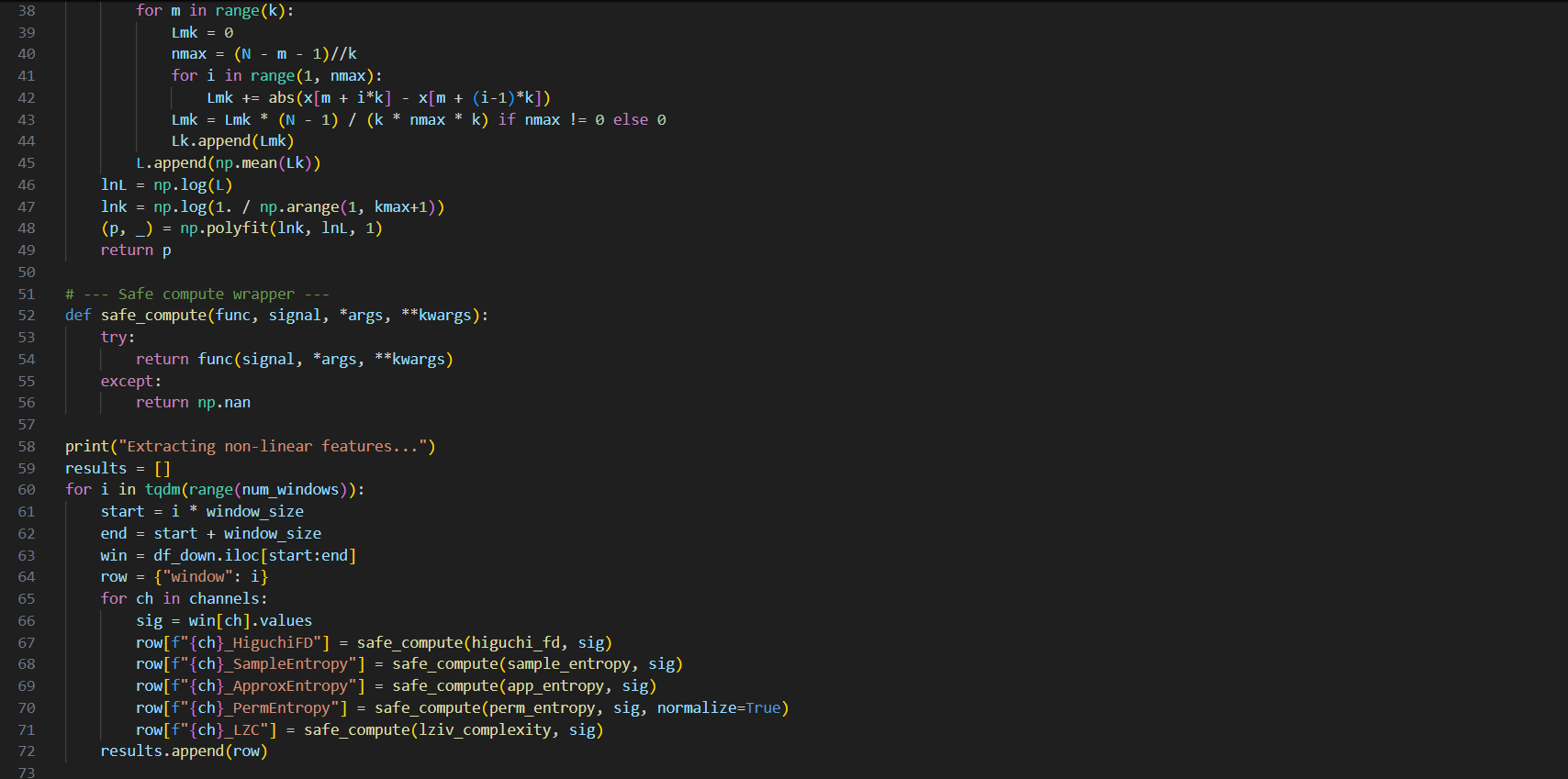
C. Permutation Entropy (PermEn)  
Calculates the entropy based on the order relations between values. It reflects the signal's dynamical complexity and is robust to noise.  
PermEn = -Σ p\_i log(p\_i)  
where p\_i represents the relative frequency of each permutation.

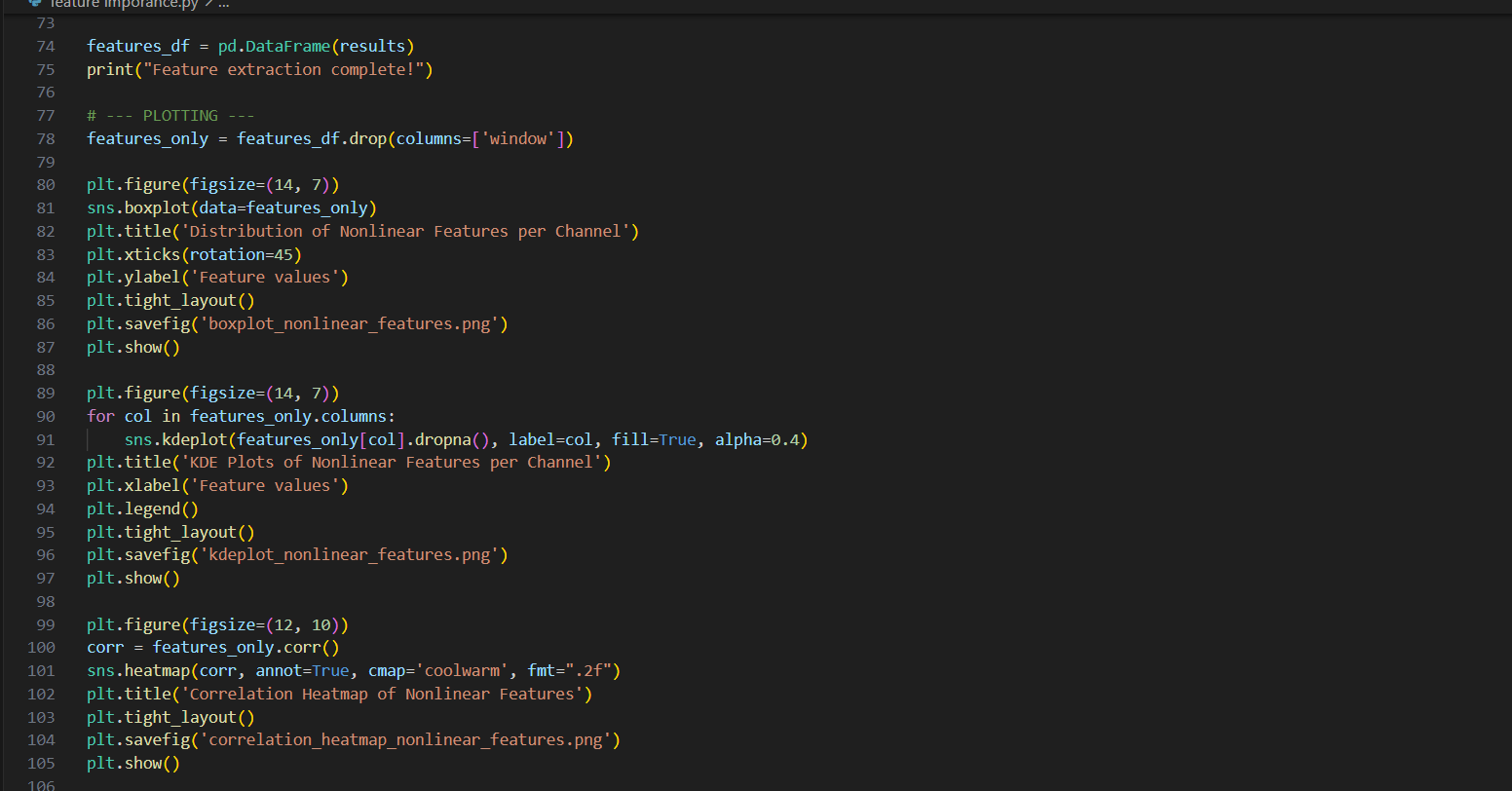
D. Lempel-Ziv Complexity (LZC)  
This metric captures the randomness of the binary string derived from the signal. The signal is thresholded (e.g., using median) and converted into a binary string. LZC quantifies the number of distinct substrings and their recurrence rate.

E. Higuchi Fractal Dimension (HFD)  
Fractal-based approach to estimate the signal’s complexity and self-similarity across multiple temporal scales. For each scale k, a curve length L(k) is calculated, and the slope of the log-log plot:  
FD = ∂ln(L(k)) / ∂ln(1/k) gives the fractal dimension.

**CODE SNIPPETS:**

****

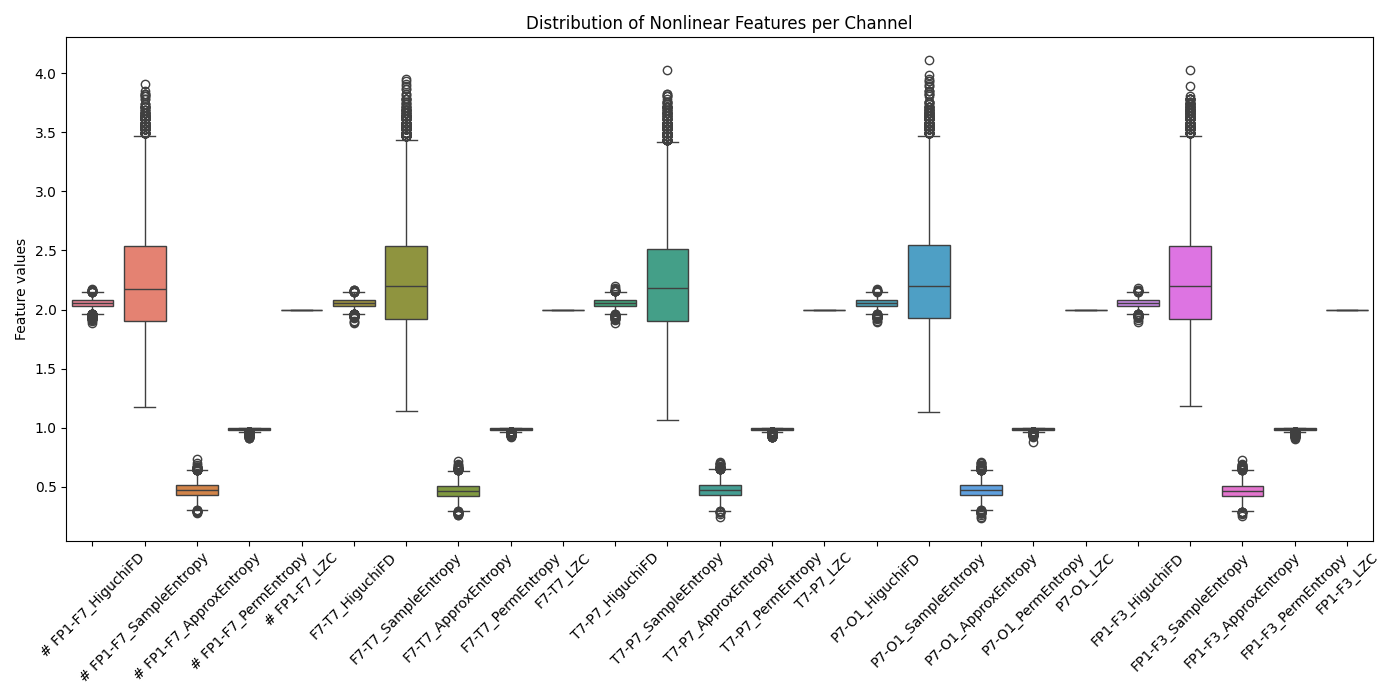
****

****

## IV. RESULTS AND VISUALIZATION

After feature extraction, results are aggregated into a DataFrame with shape:  
(num\_windows, num\_features)

1. **Boxplot Analysis**  
   Boxplots provide a summary of the central tendency and variability for each feature across all windows, revealing outliers and skewness.



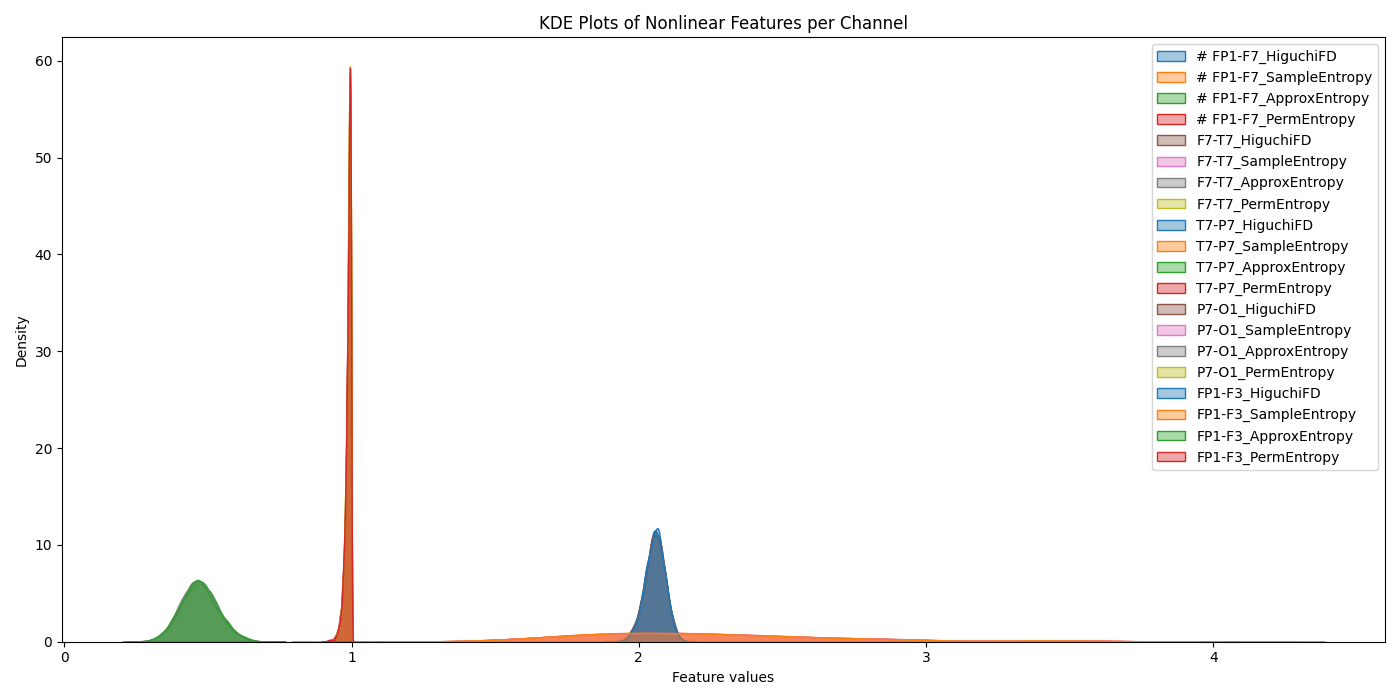
**Analysis:**

The distribution of five nonlinear features—Higuchi Fractal Dimension (HiguchiFD), Sample Entropy, Approximate Entropy, Permutation Entropy, and Zero Crossings (ZC)—was analyzed across five bipolar EEG channels (Fp1-F7, F7-T7, T7-P7, P7-O1, Fp1-F3). The following key insights were observed:

* HiguchiFD exhibited a wide range with numerous outliers across all channels, indicating high signal complexity but also substantial variability.
* Sample Entropy showed low variance and a consistent distribution, suggesting it is a robust and reliable feature for distinguishing signal dynamics.
* Approximate Entropy displayed moderate spread and outliers, highlighting its sensitivity to data length and signal irregularity.
* Permutation Entropy values were tightly clustered near 1.0 for all channels, implying limited discriminative power due to its near-constant behavior.
* Zero Crossings (ZC) were uniformly distributed around a fixed value with negligible variation, indicating minimal contribution to distinguishing signal characteristics.

Overall, Sample Entropy and HiguchiFD appear to be the most informative nonlinear features, whereas Permutation Entropy and ZC may have limited utility in EEG-based seizure classification.

**B.KDE Distribution**  
KDE plots offer a smoothed view of the feature distributions, helping visualize their discriminative power and clustering tendencies.



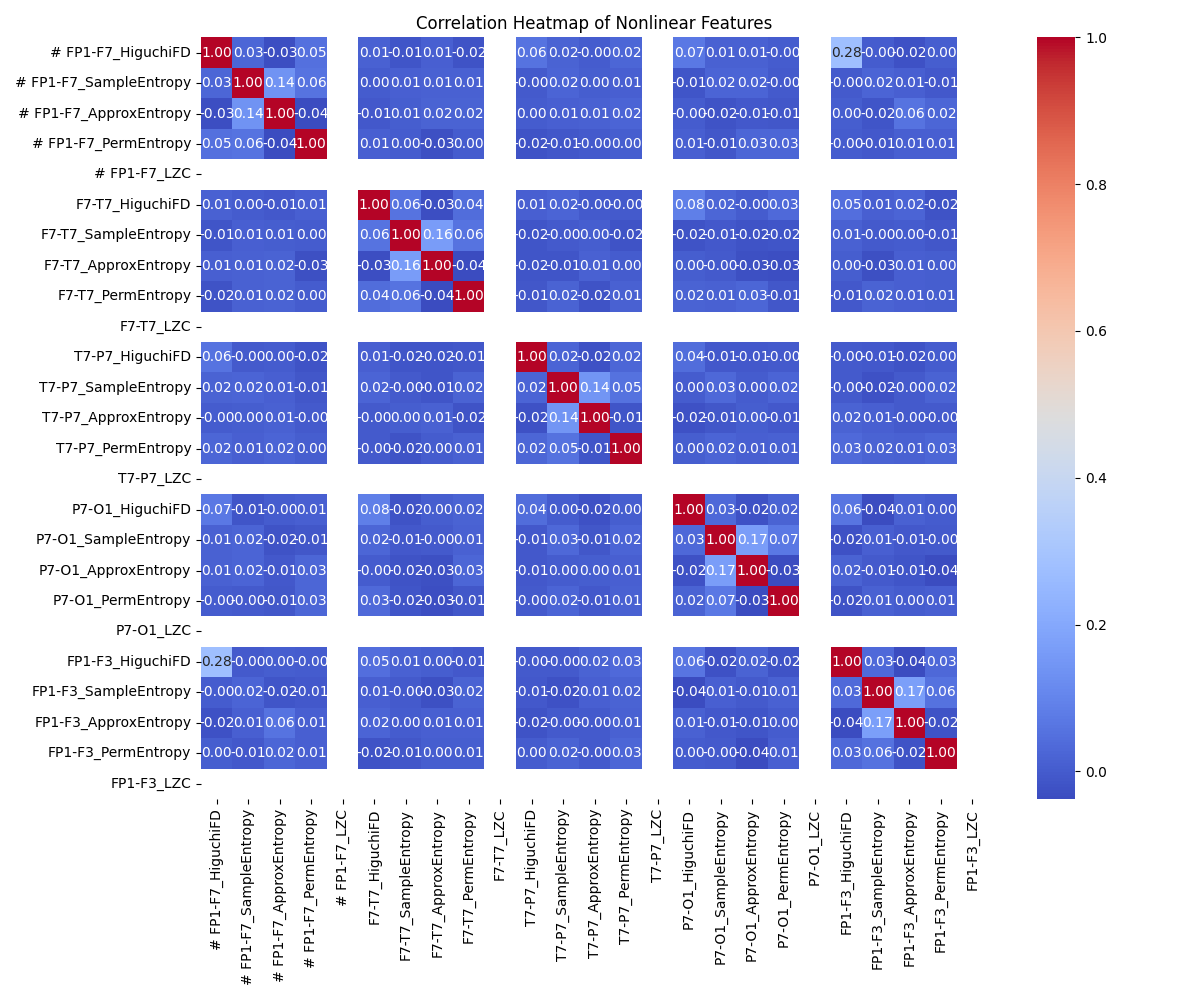
**Analysis:**

The KDE plot presents the probability density distribution of four nonlinear features—Higuchi Fractal Dimension (HiguchiFD), Sample Entropy, Approximate Entropy, and Permutation Entropy—across five EEG channels (Fp1-F7, F7-T7, T7-P7, P7-O1, Fp1-F3). The major observations are as follows:

* Permutation Entropy exhibits a sharply peaked density around 1.0 across all channels, indicating near-constant values. This confirms its low discriminativevariability, which may reduce its utility in feature-based classification tasks.
* Approximate Entropy distributions are narrow and centered around 0.45, showing consistent values with low variance across all channels.
* Sample Entropy shows distributions centered near 2.0, with slightly broader spread than Approximate Entropy. It maintains a moderate level of variability, indicating good potential for class separation.
* HiguchiFD presents wider, more skewed distributions centered around 2.5–3.0, highlighting higher signal complexity and inter-channel variability.

In summary, the KDE plot supports the boxplot findings: HiguchiFD andSample Entropy provide more diverse and informative distributions, while Permutation Entropy is overly concentrated, limiting its effectiveness for discriminative analysis.

1. **Correlation Heatmap**  
   A Pearson correlation matrix illustrates linear relationships among the features, highlighting redundancy or complementary information between metrics.



**Analysis:**

The correlation heatmap displays pairwise Pearson correlation coefficients among nonlinear features—Higuchi Fractal Dimension (HiguchiFD), Sample Entropy, Approximate Entropy, Permutation Entropy, and Lempel-Ziv Complexity (LZC)—computed across five EEG channels. Key observations are as follows:

* Low Cross-Feature Correlation: Most inter-feature correlation coefficients lie within the range of -0.1 to 0.1, indicating minimal redundancy among nonlinear features. This supports their combined use for improving feature diversity in classification models.
* Channel Independence: Features extracted from different channels show negligible inter-channel correlation (e.g., Fp1-F7 features vs. T7-P7 features), reinforcing the spatial independence of nonlinear EEG characteristics.
* Mild Intra-Feature Correlation: Within the same channel, weak correlations are observed, such as between HiguchiFD and Approximate Entropy for Fp1-F7 (≈ -0.03) and between Sample Entropy and Permutation Entropy for Fp1-F3 (≈ 0.17). These values suggest marginal shared structure but still indicate distinct signal dynamics.
* Diagonal Dominance: All diagonal values equal 1.0, validating the identity correlation of features with themselves.

In conclusion, the heatmap confirms that the selected nonlinear features are mutually complementary, channel-independent, and offer minimal overlap, thus enhancing the feature robustness for EEG-based seizure classification or cognitive state analysis.

## V. DISCUSSION

The five selected nonlinear metrics capture diverse aspects of EEG signal complexity:  
- Entropy measures (SampEn, ApEn, PermEn) quantify unpredictability and dynamic order.  
- LZC evaluates randomness through symbolic dynamics.  
- HFD measures fractal scaling behavior.

Their combination yields a rich feature space suitable for classification tasks in neuroscience, including seizure detection, sleep staging, and mental workload assessment.

## VI. CONCLUSION

This work demonstrates a robust and interpretable pipeline for nonlinear EEG feature extraction using entropy and fractal-based techniques. The features extracted are visualized and analyzed to understand their statistical behavior and interrelationships. This method is suitable for real-time and offline applications in clinical and cognitive neuroscience.