Report

MSc. Data Science & AI

Class: Biomedical Signal Processing

Spectral analysis and filtering

Professor: Vicente Zarzoso

Student: Aglind Reka

Contents

1. Abdominal ECG during pregnancy

1.1 Plot the 8 leads. Remark that the 4th abdominal lead is particularly noisy

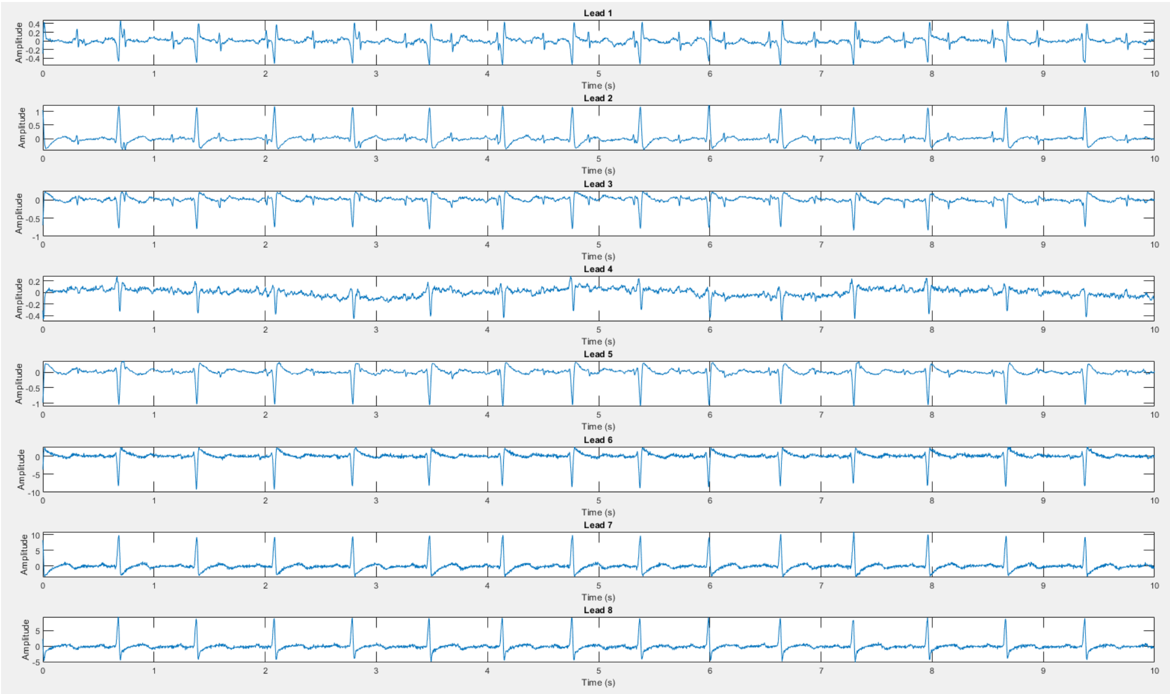

Figure 1: Ploting of the 8 leads

As depicted in Figure 1, the waveform in the fourth lead exhibits a distinct pattern, deviating from the uniformity observed in the other seven leads. This variance prompts suspicion of potential noise interference.

Upon examining the 8-lead ECG recording, it's evident that the fourth abdominal lead displays distinct irregularities compared to the other leads.

The noise in the fourth lead likely comprises irregular fluctuations or distortions, potentially stemming from factors like muscle artifacts or electrode movement. This noise can obscure important ECG features and affect diagnostic accuracy.

1.2 Using the spectral estimation method of your choice, compute and plot the power spectral density (PSD) of the noisy lead.

In this exercise, I first load the ECG data. Select the noisy lead, which is the 4th abdominal lead in this case. Compute the power spectral density (PSD) of the noisy lead using the pwelch function. Plot the PSD on a logarithmic scale to better visualize the frequency components of the signal.

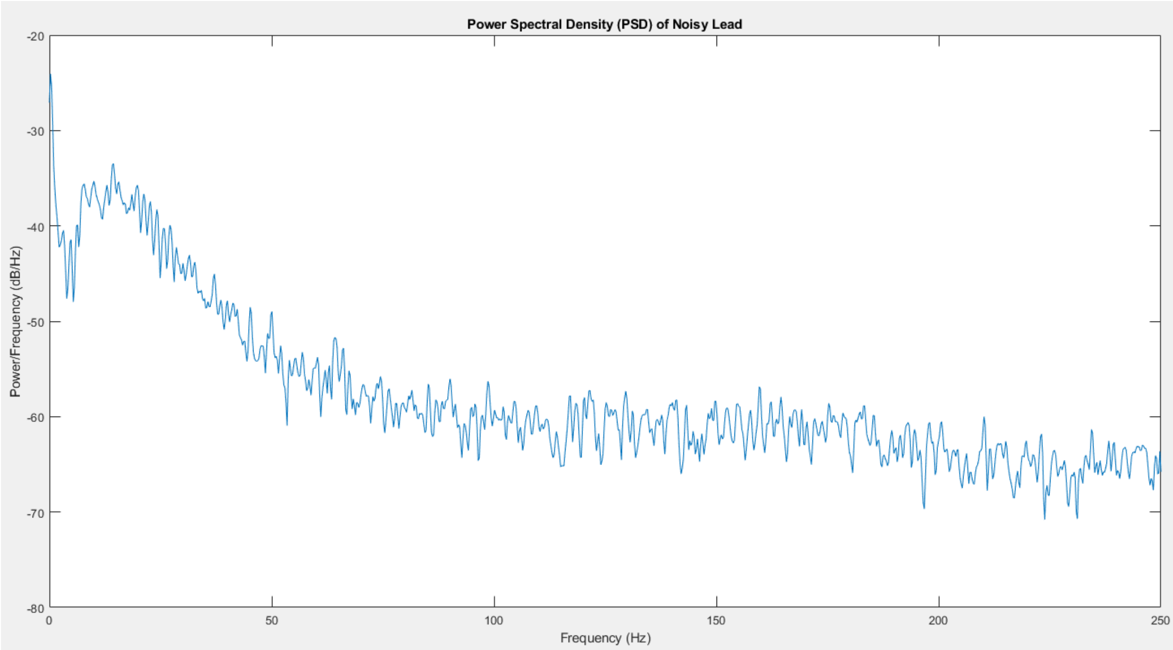

Figure 2: Power Spectral Density (PSD) of the noisy lead.

The pwelch function estimates the Power Spectral Density (PSD) of a signal by computing the Discrete Fourier Transform (DFT) of windowed segments of the signal and then averaging them to provide a smoothed representation of the signal's frequency content.

1.3 By observing the estimated PSD, identify the frequency intervals of noise and interference contaminating the signal of interest.

Based on the frequency range provided in the graph (0-250 Hz), frequency intervals of noise and interference contaminating the signal of interest can be identified as follows:

Noise Frequency Intervals: In the frequency range from 0 to 10 Hz, where the PSD decreases, it's likely that the signal contains relatively low power levels and is dominated by noise. This interval may correspond to baseline wander or low-frequency noise inherent in the recording environment.

Interference Frequency Intervals: Between 11 and 20 Hz, where the PSD increases, there could be interference from sources such as power line noise or muscle artifacts. The sharp increase in power in this frequency range suggests the presence of unwanted interference.

Beyond 20 Hz: Frequencies beyond 20 Hz exhibit a decrease in PSD, indicating diminishing power at higher frequencies. While some interference may still be present, the decreasing trend suggests that the signal of interest becomes less contaminated as frequency increases beyond this point.

1.4 Design appropriate frequency filters to reduce noise and interference.

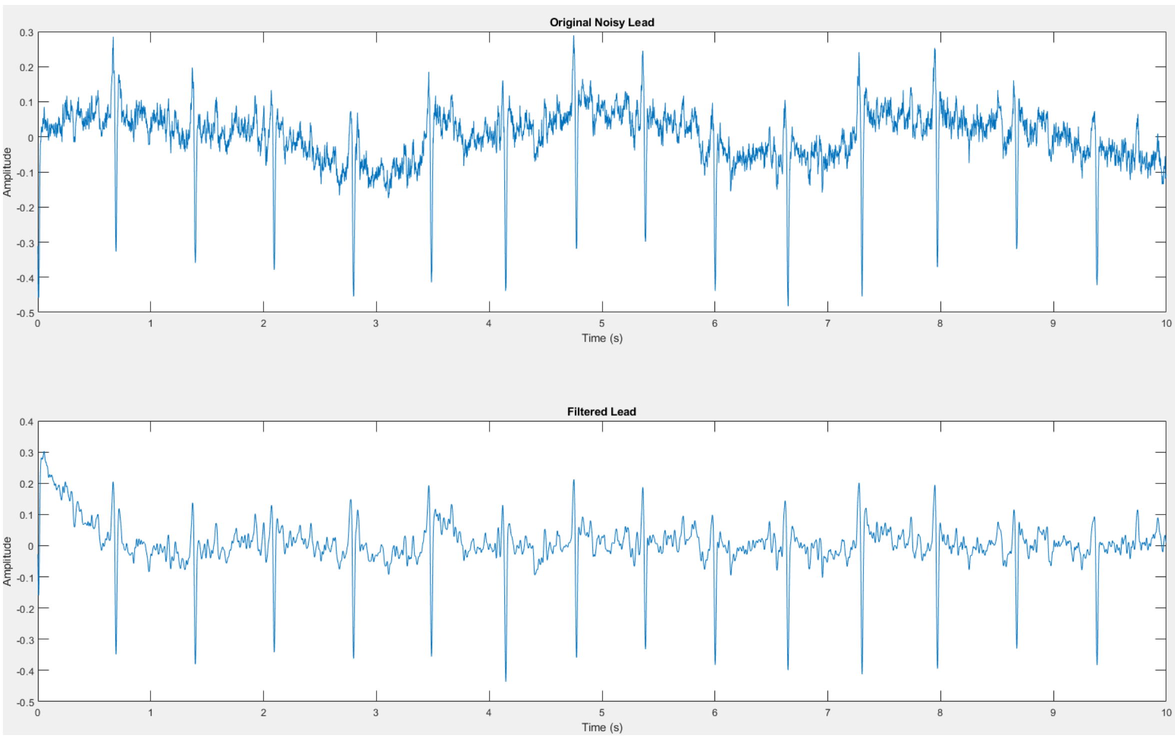

Figure 3: Noise Reduce

As illustrated in Figure 3, the filtered signal exhibits a noticeable improvement compared to the original noisy signal. In the noisy signal, the waveform appears irregular, with fluctuations and inconsistencies in its pattern. However, after applying the filter, the signal becomes uniform and free from noise. The clean signal displays regular and consistent waveforms, devoid of the fluctuations observed in the noisy signal. This stark difference highlights the effectiveness of the filtering process in removing noise and enhancing the clarity of the signal.

1.5 Apply the filters to the observed ECG and confirm their effectiveness by analyzing the output signal in both time and frequency domains.

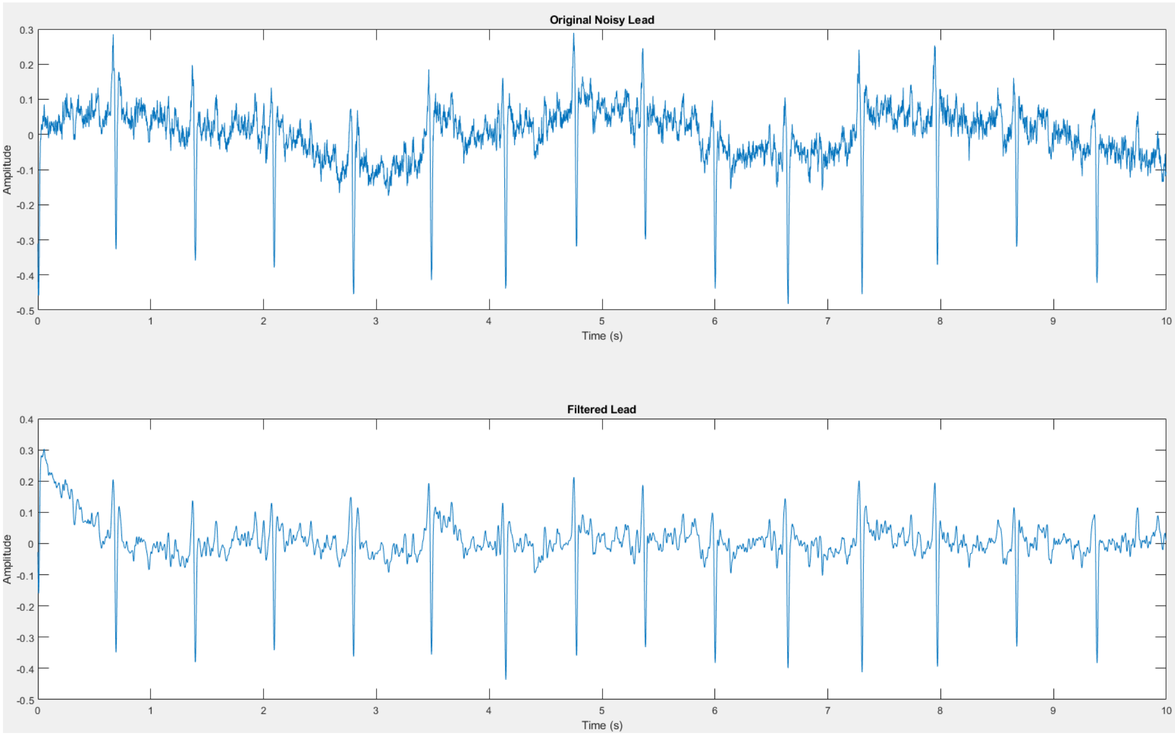

Figure 4: Noise Reduce

As depicted in Figures 4 and 5, the noise reduction is clearly evident in both the time and frequency domains. In Figure 5, the filtered Power Spectral Density (PSD) plot displays a notably smoother waveform with reduced fluctuations, indicating successful noise reduction. This improvement underscores the efficacy of the filtering process in enhancing signal quality.

Furthermore, with the cut-off frequency applied, the noise has been effectively eliminated, as evidenced by the near-horizontal line observed in the waveform. In the frequency domain, the presence of noise is markedly reduced, leading to a cleaner representation of the underlying signal.

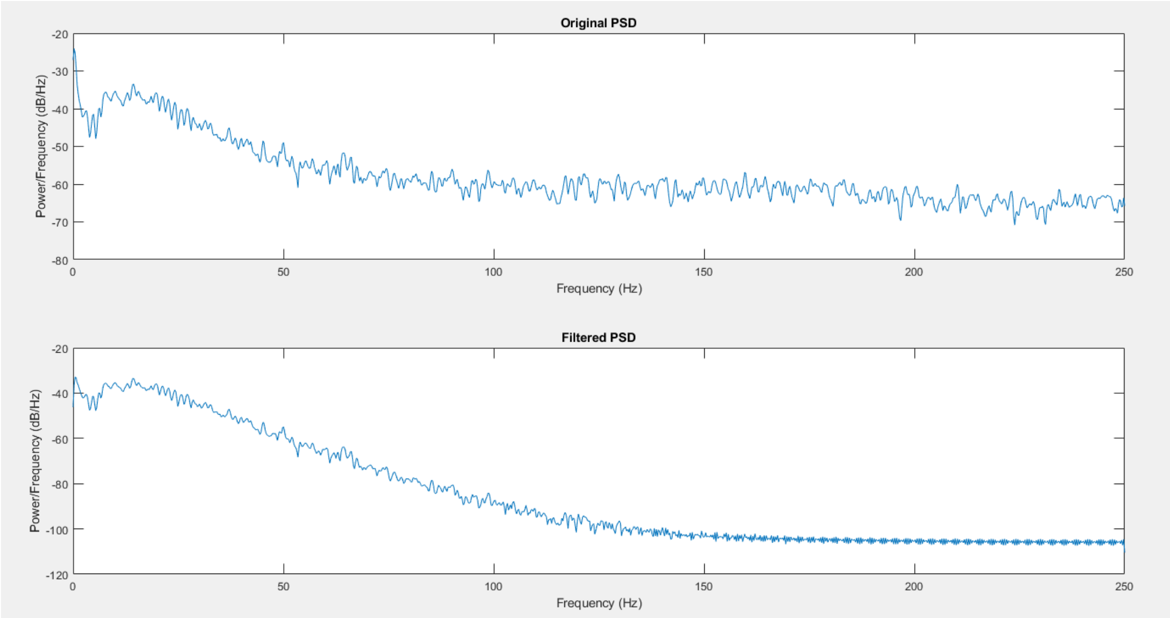

Figure 5: analyzing the output signal in both time and frequency domains

2. Atrial Fibrillation Analysis

2.1 Create MATLAB functions to compute the DF and SC of the signals given as inputs. As a PSD estimator, use the Welch method with the typical window type and size, overlap factor and number of FFT points found in the literature. Find also in the literature the classical frequency band for SC computation.

After reviewing the literature, I came across a paper by Garibaldi et al. (2012) [1], which utilized the Welch method for spectral analysis with specific parameters. In their study, they employed the following parameters: a window size of 4096 samples, an overlap of 2048 samples, and an FFT size of 8192. They focused on analyzing the spectral characteristics within the classical band of 3-9 Hz for computing the spectral concentration (SC). After I created the function **computeDFandSC** for the computation of DF and SC.

2.2 Compute the DF and SC in lead V1 of all ECG records in the provided AF dataset.

The code extracts ECG signals from lead V1 for each patient in matrices Xa and Xva, computes the dominant frequency (DF) and spectral concentration (SC) for these signals, and then displays the results for each patient.

2.3 With the help of box-and-whiskers plots, compare the distributions of the spectral parameters of the recorded data (Xva) versus those of the estimated atrial activity (Xa) over the patient population.

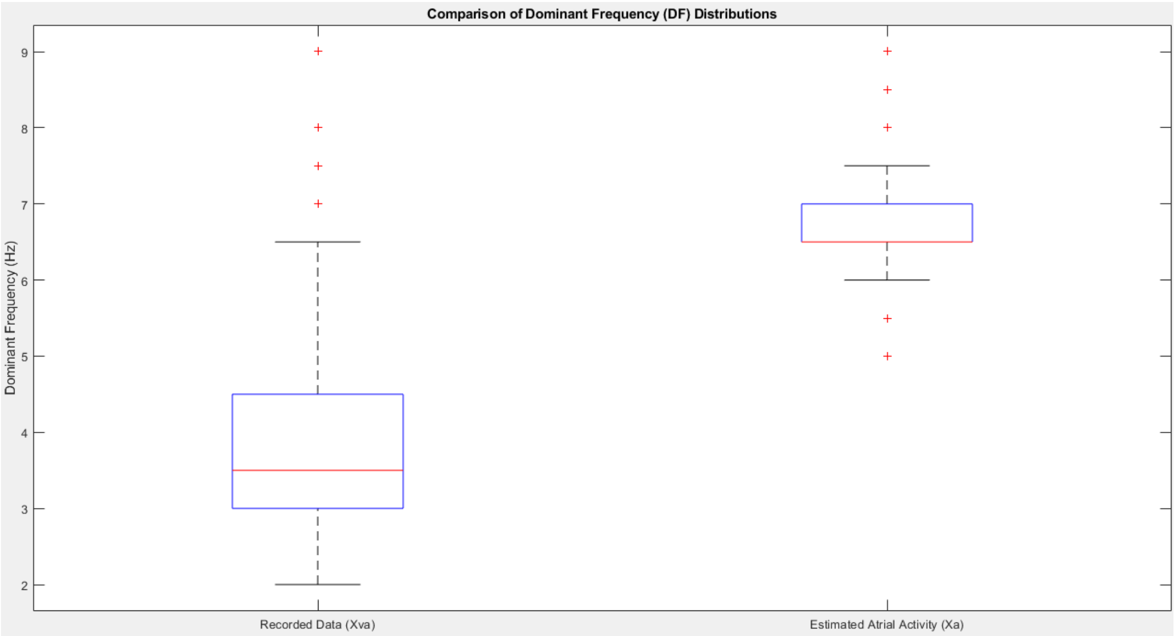

Figure 6: Comparison of Dominant Frequency (Df) Distributions

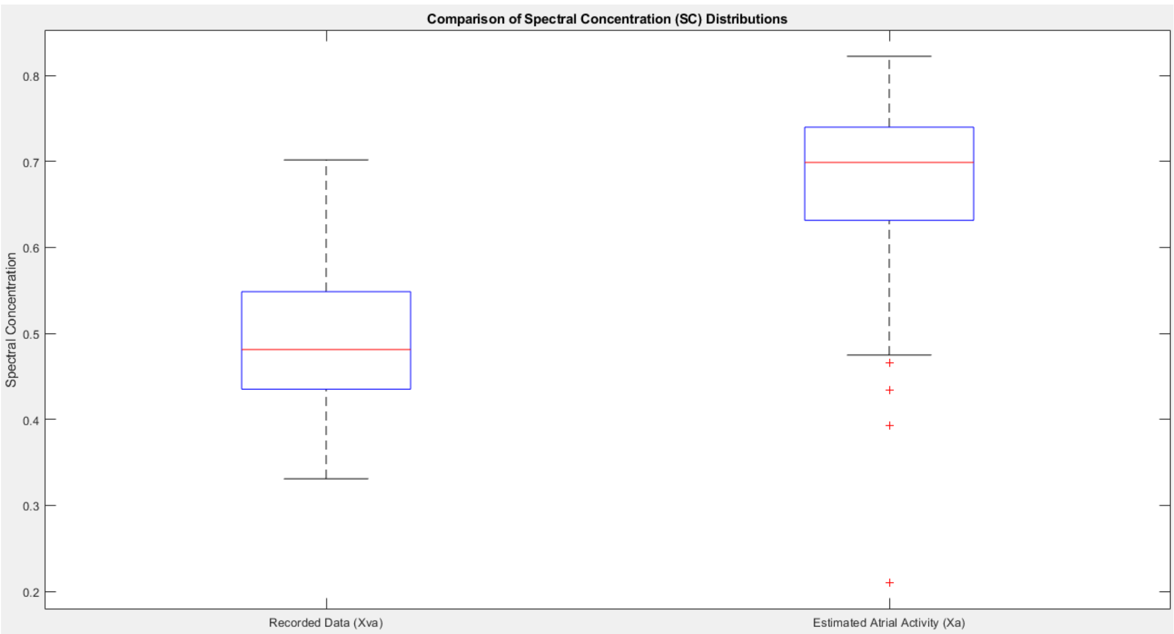

Figure 7: Comparison of Spectral Concentration (SC) Distributions

As depicted in Figure 6, the distributions of Dominant Frequency in the two populations, namely **Xva** and **Xa**, exhibit notable disparities. In the case of **Xva**, approximately 50% of the dataset falls within the range of 3 to 4.5, with a median value of 3.5. The maximum value observed is 6.5, while the minimum is 2. Additionally, the presence of four outliers is noticeable.

In contrast, for **Xa**, roughly 50% of the dataset is concentrated within the range of 6.5 to 7, encompassing the interquartile range between the first and third quartiles, where the median coincides with the first quartile. Notably, outliers are observed both below the minimum and above the maximum values in this case.

As depicted in Figure 7, the distributions of Spectral Concentration in the two populations, namely **Xva** and **Xa**, exhibit notable disparities even in this case. In the case of **Xva**, approximately 50% of the dataset falls within the range of 0.45 to 0.55, with a median value of 0.48. The maximum value observed is 0.7, while the minimum is 0.33. Additionally, I don't see the presence of outliers.

In contrast, for **Xa**, roughly 50% of the dataset is concentrated within the range of 0.63 to 0.74, encompassing the interquartile range between the first and third quartiles, where the median coincides with 0.7. Notably, outliers are observed below the minimum of 0.48.

2.4 Design a classifier to predict AF recurrence using DF and SC as features. Show its performance using 5-fold cross-validation.

Table 1: Average Classification Accuracy

	Xa	**Xva**
Logistic Regression	58%	57.33%
SVM	58.22%	58.40%
Random Forest	48.00%	42.67%

After evaluating logistic regression, SVM, and Random Forest classifiers using 5-fold cross-validation, I found that SVM yields the most favorable results among them.

2.5 Repeat parts 2 and 4 by averaging, for each patient, the spectral parameters over all precordial leads (V1 to V6)

Table 2: Average Classification Accuracy (V1 to V6)

	Xa	**Xva**
Logistic Regression	60%	60.20%
SVM	60%	59.62%
Random Forest	53.33%	40%

After evaluating logistic regression, SVM, and Random Forest classifiers using 5-fold cross-validation, In V1 to V6, I found that Logistic Regression yields the most favorable results among them.

2.6 Conclude on the ability of spectral parameters DF and SC to predict AF recurrence in the given dataset.

Based on the provided table of average classification accuracy for datasets Xa and Xva using different classifiers (Logistic Regression, SVM, and Random Forest), several observations can be made:

- **Consistency of Results:** Across both datasets Xa and Xva, the Logistic Regression and SVM classifiers demonstrate similar average classification accuracies, indicating consistency in their performance.
- **Higher Accuracy with Logistic Regression and SVM:** Both Logistic Regression and SVM achieve higher average classification accuracies compared to Random Forest for both datasets. This suggests that, in this specific context, Logistic Regression and SVM may be more effective in predicting AF recurrence than Random Forest.
- **Decrease in Accuracy for Xva with Random Forest:** There is a notable decrease in average classification accuracy for Xva when using Random Forest compared to Xa. This indicates that the Random Forest classifier may not generalize well to the unseen data in Xva, potentially due to overfitting or other issues.
- **Overall Performance:** While Logistic Regression and SVM exhibit relatively consistent performance across both datasets, Random Forest shows lower accuracy, especially for Xva. This suggests that further investigation or refinement may be necessary when using Random Forest for this particular classification task.

In summary, based on the provided results, Logistic Regression and SVM appear to be the more reliable classifiers for predicting AF recurrence in this dataset, with Logistic Regression achieving slightly higher accuracy overall. However, additional analysis, such as considering other evaluation metrics or exploring feature engineering techniques, may provide further insights into the predictive ability of spectral parameters DF and SC for AF recurrence prediction.

References

[1] Michele Garibaldi et al. “Predicting catheter ablation outcome in persistent atrial fibrillation using atrial dominant frequency and related spectral features”. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2012, pp. 613–616.