**1. Data Exploration:**

**1.1 Data Visualization**: Plot raw EEG data and identify any obvious patterns, noise, or artifacts. You've already started on this.

**1.2 Data Description**: Understand the basic statistics of the data. This could include:

* Average amplitude of the EEG signal.
* Standard deviation.
* Frequency distribution (using Fourier Transform).

**1.3 Event Marking**: If the data has annotations (like when a blink occurred), visualize them on the EEG plots. This helps in understanding how the EEG signal changes during these events.

**2. Data Preprocessing:**

**2.1 Noise Removal**: EEG data can be contaminated by various types of noise (e.g., line noise at 50/60 Hz, muscle artifacts). A common way to handle this is by filtering the data.

* **Band-pass filter**: Keeps frequencies within a certain range. Useful to focus on the frequency bands commonly associated with EEG signals (like Delta, Theta, Alpha, Beta, and Gamma bands).

**2.2 Artifact Removal**: Apart from filtering, advanced techniques like Independent Component Analysis (ICA) can be used to identify and remove artifacts.

**2.3 Epoching**: Divide the continuous EEG data into epochs (or segments) centered around specific events (like blinks). This gives you many samples of EEG data for each blink event to analyze.

**3. Feature Extraction:**

**3.1 Time-domain Features**: These could include:

* Peak amplitude.
* Time to peak.
* Area under the curve for each epoch.
* Statistical measures like mean, variance, skewness, kurtosis of the signal.

**3.2 Frequency-domain Features**:

* Power spectral density in specific frequency bands (Delta, Theta, Alpha, etc.).
* Peak frequency.

**3.3 Time-frequency Analysis**:

* Compute spectrograms or wavelet transforms to understand how the frequency content of the EEG signal changes with time.

**4. Biomarker Identification:**

**4.1 Statistical Analysis**:

* Compare the extracted features between the two groups (those with the eye disorder and those without).
* Use tests like t-tests or Mann-Whitney U tests to identify features that are significantly different between the two groups.

**4.2 Machine Learning**:

* Use classifiers like SVM, Random Forest, or Neural Networks.
* Train the classifier on a subset of the data and validate on a separate set.
* Evaluate the classifier's performance using metrics like accuracy, precision, recall, and the ROC curve.

**5. Validation:**

**5.1 Cross-Validation**: Perform k-fold cross-validation to ensure the robustness of the classifier.

**5.2 External Validation**: If available, test the identified biomarker on a separate dataset to ensure it generalizes well to new data.

**Deciding on the Filter:**

Given the objective, you're specifically interested in EEG changes related to blinks. The decision on the best filter would ideally be based on:

* **Literature**: Existing research on EEG characteristics during eye blinks.
* **Exploration**: The frequency content of the EEG signal before and after applying different filters.

If the main signal of interest (related to blinks) is within a specific frequency range (e.g., 1-10 Hz), then a band-pass filter that retains these frequencies while removing others might be ideal.

However, as your error message suggests, if the filter length is longer than the signal length, you'll need to make adjustments either by choosing a different filter design or adjusting the signal (e.g., by padding or using a longer segment of data).

Remember, the choice of filter and its parameters should be made based on both the characteristics of the data and the specific objective of the analysis. In your case, you want to emphasize the EEG signal characteristics that differentiate between individuals with and without the eye disorder.