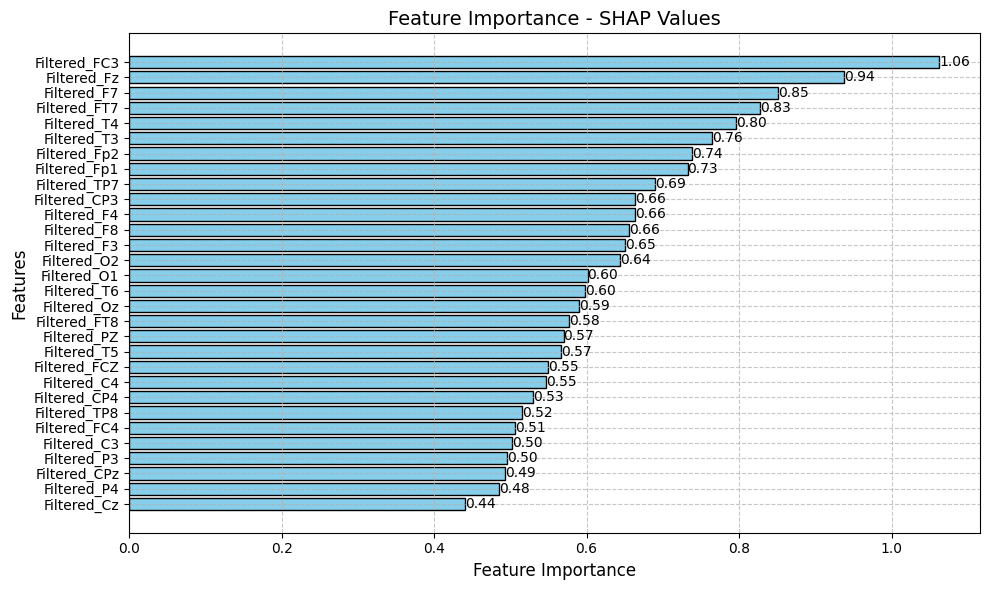
Feature Selection - Shapley Approximation (shap)

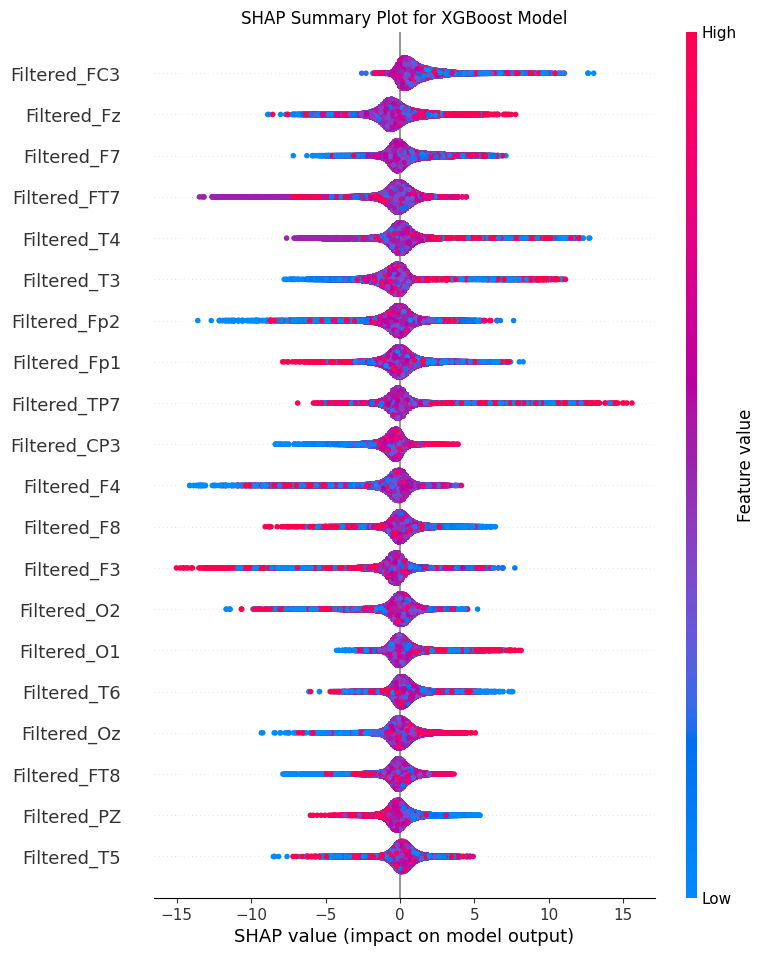
The dataset is split into training and testing sets, with 80% for training and 20% for testing. XGBoost classifier is trained on the training data. GPU acceleration is utilized for faster computation, and the number of trees is set to 5000. SHAP is employed for interpreting the XGBoost model predictions. SHAP values are calculated for the test dataset to understand the contribution of each feature to the model's output. Feature importance is determined based on the absolute mean SHAP values across all instances in the test dataset.

**Feature Importance Plot:** A horizontal bar plot is created to illustrate the importance of each feature, with annotations showing the numerical values of importance.



**Figure:** Feature Importance SHAP Values

**SHAP Summary Plot:** A summary plot is generated to visualize the distribution of SHAP values for each feature.



**Figure:** Summary Plot for Feature Impact Distribution.

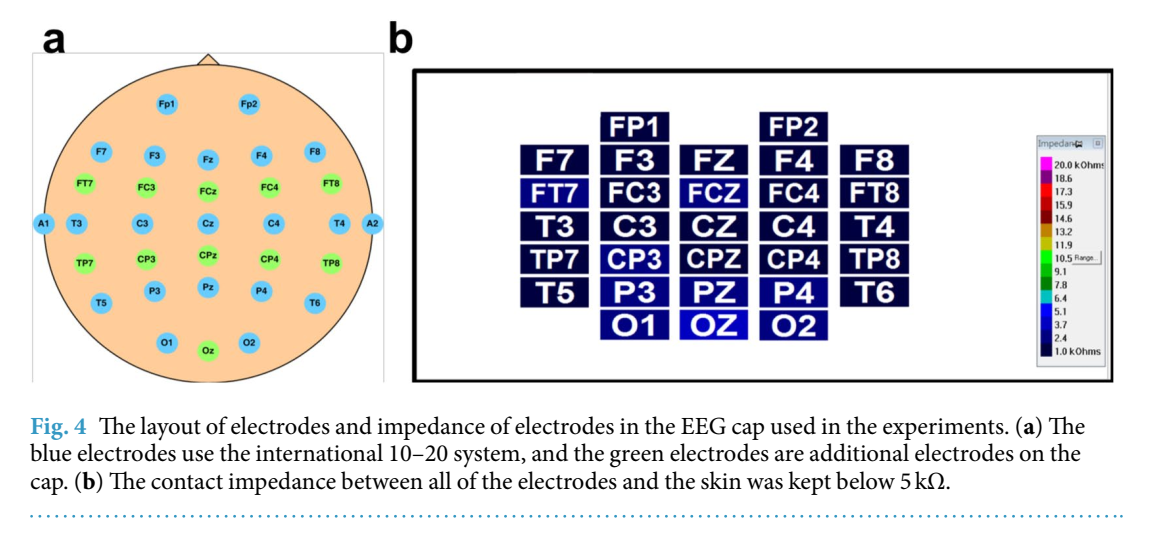


Image Source: <https://www.nature.com/articles/s41597-019-0027-4>