Different Algorithms (*Might*) Uncover Different Patterns:

A Brain-Age Prediction Case Study

FIND ME ON GITHUB

IEEE BIBM 2023

5-8 December 2023 • Istanbul & Turkey

Authors

Tobias Ettling

Computer Science
University of St. Gallen

Dr. Sari Saba-Sadiya

Computer Science
Goethe University Frankfurt

Prof. Gemma Roig

Computer Science
Goethe University Frankfurt

Machine Learning in EEG Research

Feature Extraction

Bagging

Linear Models

Tree-Based-Models

Data Transformation

Gradient Boosting

Interpolation

Unveiling **Insights** through **Model Interpretation**

Do different model architectures find different patterns?

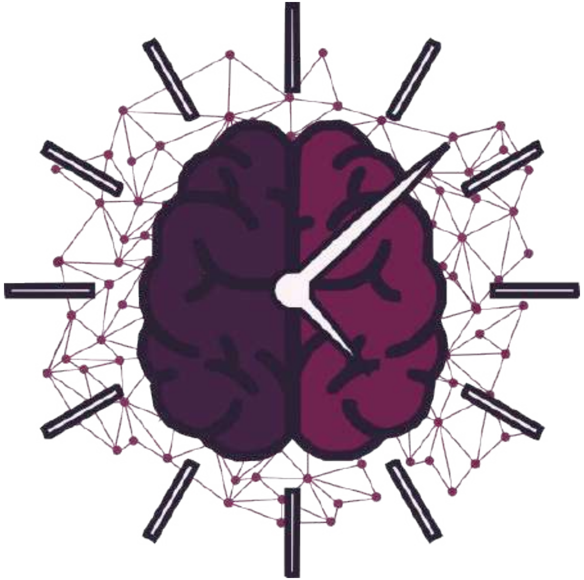
Brain Age Prediction from EEG Signals

Regression Models

ElasticNet
LassoRegression
KernelRidge

Other Models

SVM
MLP



Tree-Based Models

XGBoost
CatBoost
RandomForest

K-Nearest Neighbors

KNN
BaggedKNN

Brain Age Prediction Case Study

Chronological age of the Subject

Brain Age Prediction from EEG Signals

Findings	Age in Years	Study Type	Paper
Central delta features show significant importance to random forest for predicting the age.	5 to 18	longitudinal	[1]
Absolute delta activity decreased with age.	8 to 12	cross-sectional	[5]
Occipital delta magnitude was higher for younger subjects (18 to 50 years) than for older subjects (51 to 85 years).	18 to 85	cross-sectional	[6]
Occipital delta magnitude correlates linearly with the subject age.			
Temporal theta features show significant importance to random forest for predicting the age.	5 to 18	longitudinal	[1]
Theta power decreases with age.	9 to 16	cross-sectional	[7]
	4 to 17	longitudinal	[8]
Increase in alpha activity beginning in posterior regions and ending in anterior regions.	5 to 18	longitudinal	[1]
Peak alpha frequency increases with age.			
Frontal lower and parietal alpha features show importance for predicting the age with random forest.			
Alpha waves show a non linear pattern with ageing.	18 to 85	cross-sectional	[6]
Relative alpha activity increases with age.	8 to 12	cross-sectional	[5]
Beta frequency power shows importance for predicting the age with random forest.	5 to 18	longitudinal	[1]
Spectral flatness of beta band was most important for model predictions.	18 to 58	cross-sectional	[2]
Beta band power positively correlated with age .	8 to 12	cross-sectional	[5]
PSD slope was more negative in pre-teen (under age 13) vs teen subjects (age 13 to 16).	9 to 16	cross-sectional	[7]
Multiscale entropy increased in frontal and central regions with ageing.			

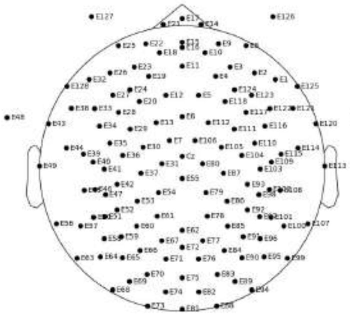
Brain Age Prediction from EEG Signals

Findings	Age in Years	Study Type	Paper
Central delta features show significant importance to random forest for predicting the age.	5 to 18	longitudinal	[1]
Absolute delta activity decreased with age.	8 to 12	cross-sectional	[5]
Occipital delta magnitude was higher for younger subjects (18 to 50 years) than for older subjects (51 to 85 years).	18 to 85	cross-sectional	[6]
Occipital delta magnitude correlates linearly with the subject age.			
Temporal theta features show significant importance to random forest for predicting the age.	5 to 18	longitudinal	[1]
Theta power decreases with age.	9 to 16	cross-sectional	[7]
	4 to 17	longitudinal	[8]
Increase in alpha activity beginning in posterior regions and ending in anterior regions.	5 to 18	longitudinal	[1]
Peak alpha frequency increases with age.			
Frontal lower and parietal alpha features show importance for predicting the age with random forest.			
Alpha waves show a non linear pattern with ageing.	18 to 85	cross-sectional	[6]
Relative alpha activity increases with age.	8 to 12	cross-sectional	[5]
Beta frequency power shows importance for predicting the age with random forest.	5 to 18	longitudinal	[1]
Spectral flatness of beta band was most important for model predictions.	18 to 58	cross-sectional	[2]
Beta band power positively correlated with age .	8 to 12	cross-sectional	[5]
PSD slope was more negative in pre-teen (under age 13) vs teen subjects (age 13 to 16).	9 to 16	cross-sectional	[7]
Multiscale entropy increased in frontal and central regions with ageing.			

Brain Age Prediction Data

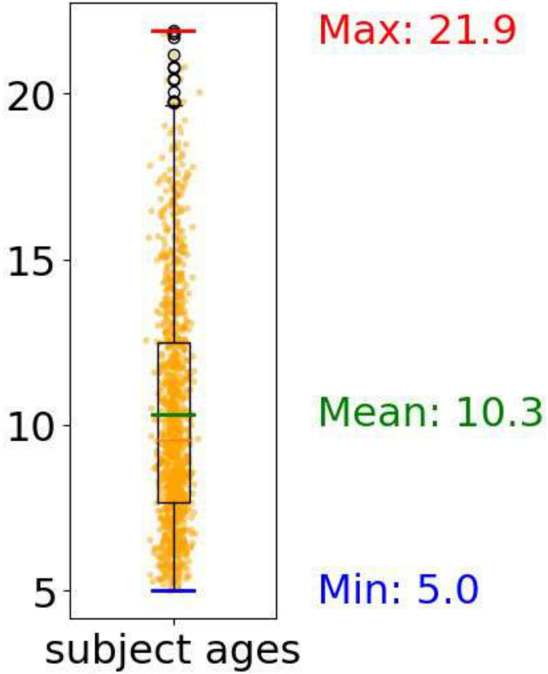
The publically available part of the NeuroTechX brain age prediction dataset.

1200 Subjects **x** 128 Channels **x** 2 States **x** 20/40sec sfreq 500Hz



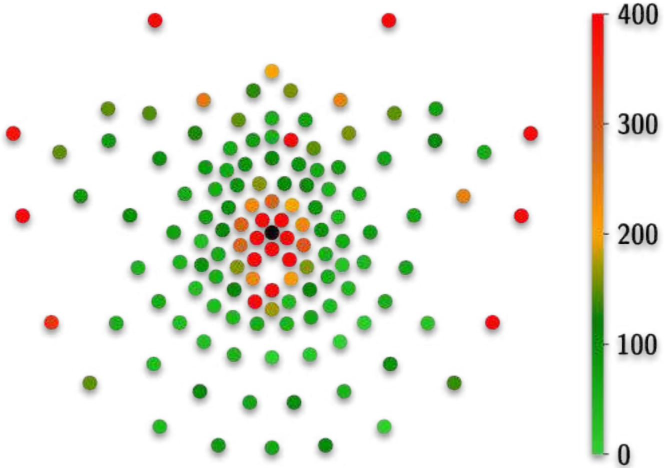
20 sec
eyes open (EO)

40 sec
eyes closed (EC)

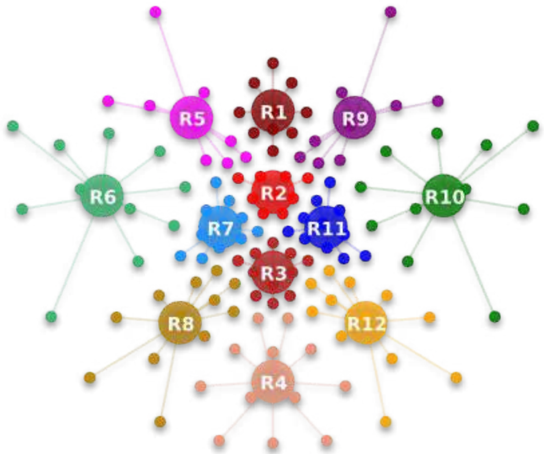


Proposed Pipeline

Channel Rejection Count
Across The Dataset



Regional Channel
Interpolation Map



Training Sets

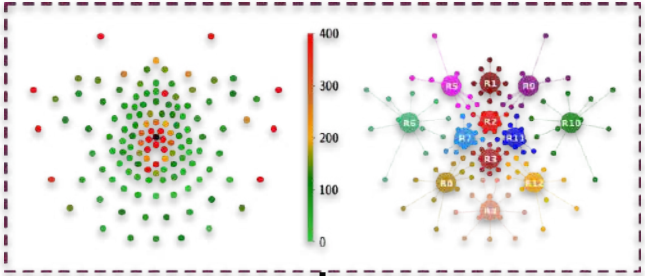
Analyzed Data Sets

Training Set	Channels	State
128-All	128	EC, EO
128-EC	128	EC
128-EO	128	EO
12-All	12	EC, EO
12-EC	12	EC
12-EO	12	EO

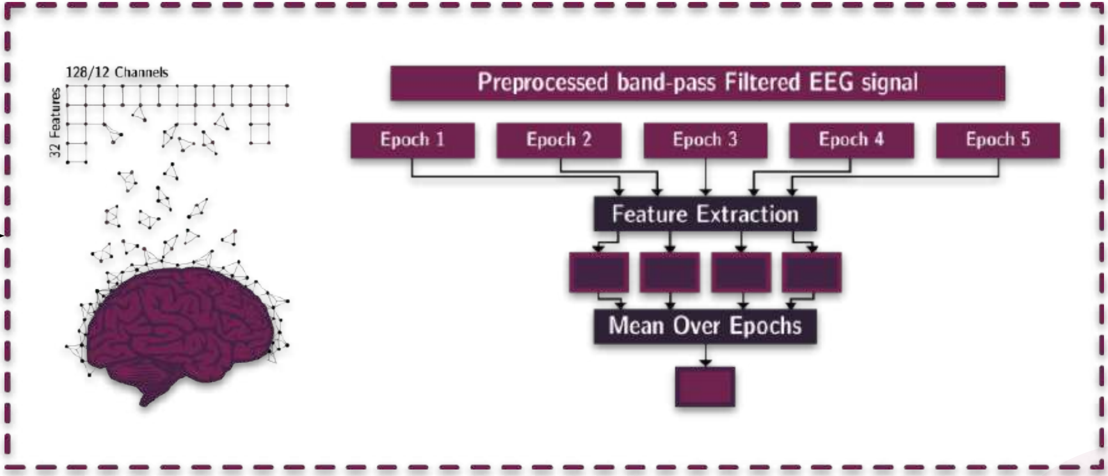
- resting state data:
  - 40 sec eyes closed (EC)
  - 20 sec eyes open (EO)
- number of electrodes:
  - 12 with channel interpolation
  - 128 electrodes without

Proposed Pipeline

Preprocessing / Artifact Rejection / Channel Interpolation



Feature Selection / Feature Extraction

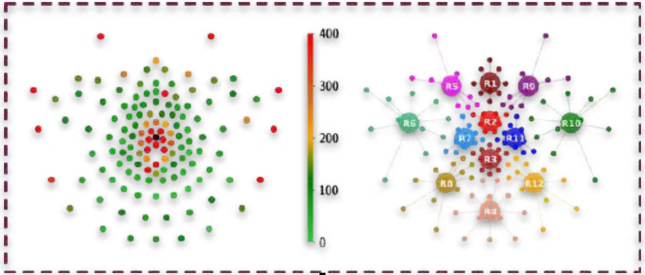


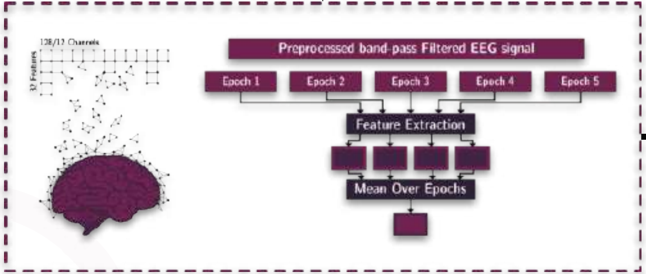
EEG Features

Category	Sub-group	Feature Name	Description
Temporal Features		Mean	Mean of the signal in volt.
Temporal Features		Standard Deviation	The average difference between the signal and its mean in volt.
Temporal Features		Peak-To-Peak Amplitude	Distance between the maximum and minimum measure of the signal.
Temporal Features		Line Length	Physical length of the signal given as the sum of the absolute difference between consecutive time measures of the signal.
Temporal Features		Zero Crossings	How often the signal changes from positive to negative.
Temporal Features		Skewness	Quantifies the asymmetry of the signal.
Temporal Features		Kurtosis	Describes the "tailedness" of a distribution compared to the normal distribution.
Temporal Features		Hjorth Complexity	Measures the complexity or irregularity of the signal
Frequency Features	log-log PSD	Intercept	intercept of the regression line, fitted to the power spectral density (PSD) of the signal with the y-axis.
Frequency Features	log-log PSD	Slope	The slope of the regression line.
Frequency Features	log-log PSD	Mean Square Error	The mean squared error between the regression line and the log-log power spectral density (PSD) of the signal.
Frequency Features	log-log PSD	R2 coefficient	Quantifies how well the regression line fits the log-log power spectral density (PSD) of the signal.
Frequency Features		Band Power $(\delta, \theta, \alpha, \beta, \omega)$	The power of the signal in a frequency band.
Frequency Features		Wavelet Coefficient Energy $(\delta, \theta, \alpha, \beta, \omega)$	The energy of wavelet decomposition coefficients.
Frequency Features		Hjorth Complexity	Measures the complexity or irregularity of the signal's power spectrum.
Statistical Features		Quantile (5%, 25%, 75%, 95%)	Describes the distribution of the signal for percentiles.
Statistical Features		Higuchi Fractal Dimension	Quantifies the fractal complexity or self-similarity of the signal.
Statistical Features		Sample & Approximate Entropy	Quantifies the complexity or regularity of the signal.
Statistical Features		Spectral Entropy	Is the Shannon entropy of the signal's power spectrum.
Statistical Features		SVD Fisher Information	Singular Value Decomposition (SVD) Fisher Information per channel
		Hurst Exponent	Characterizes the long-term memory or self-similarity of the signal.

Proposed Pipeline

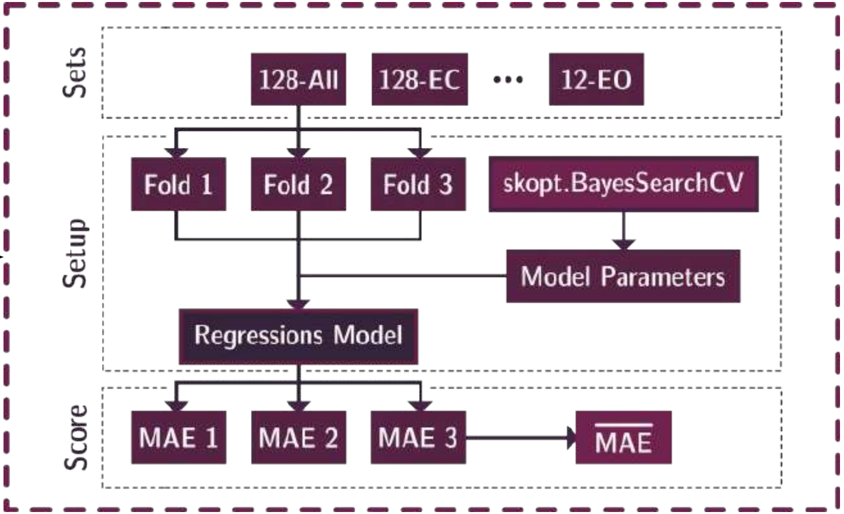
Preprocessing / Artifact Rejection / Channel Interpolation



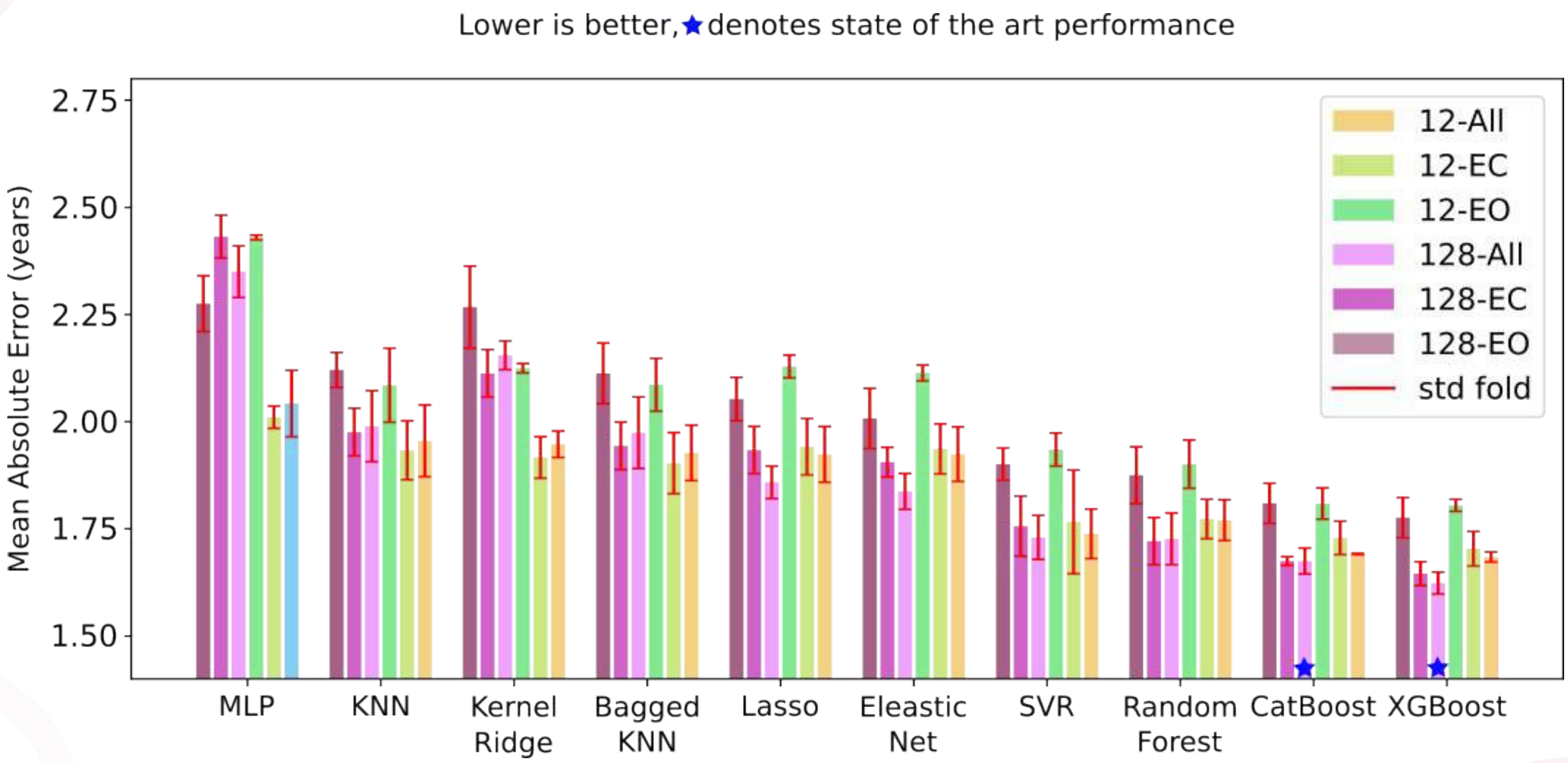


Feature Selection / Feature Extraction

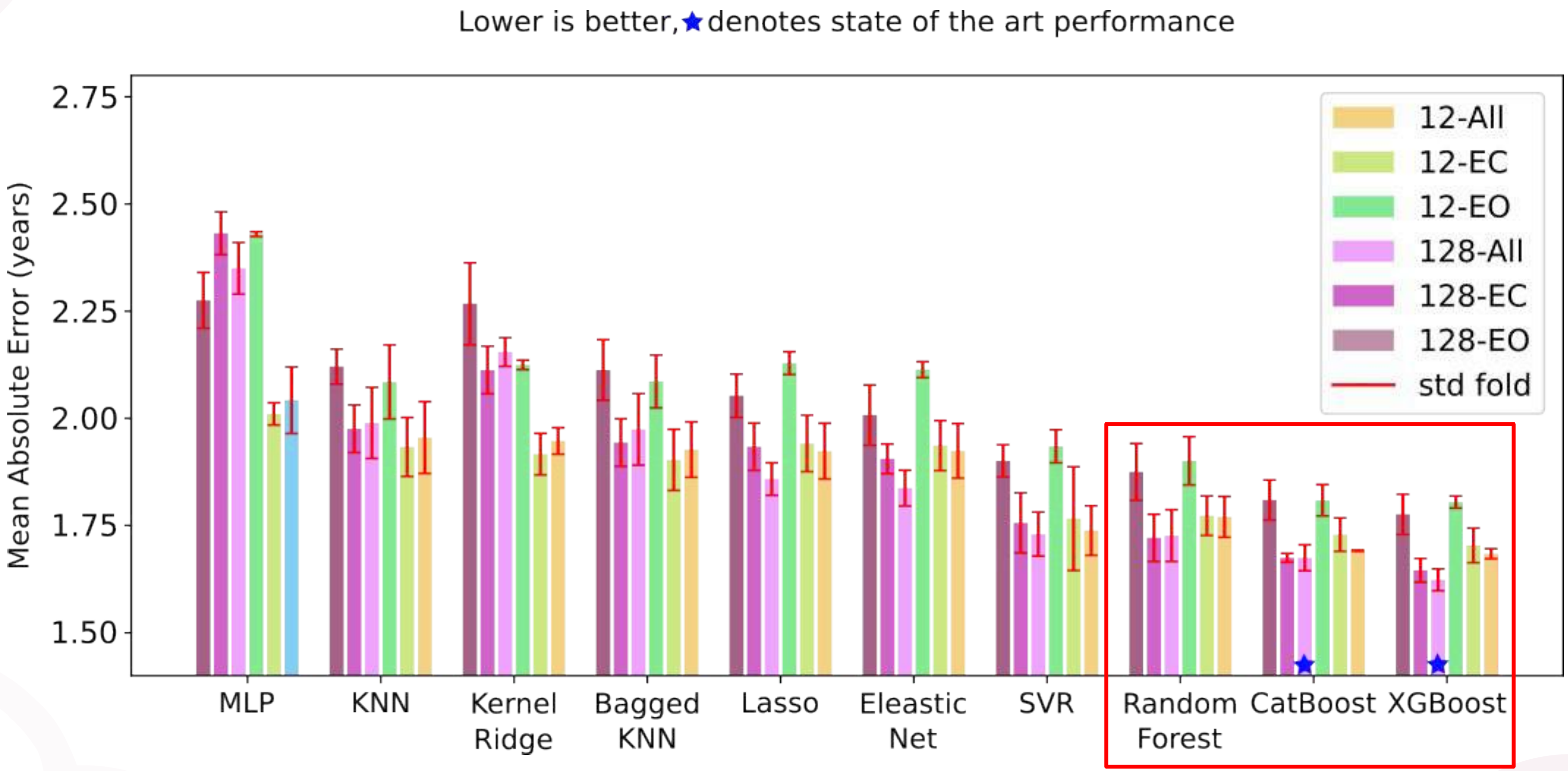
Model Training & Tuning



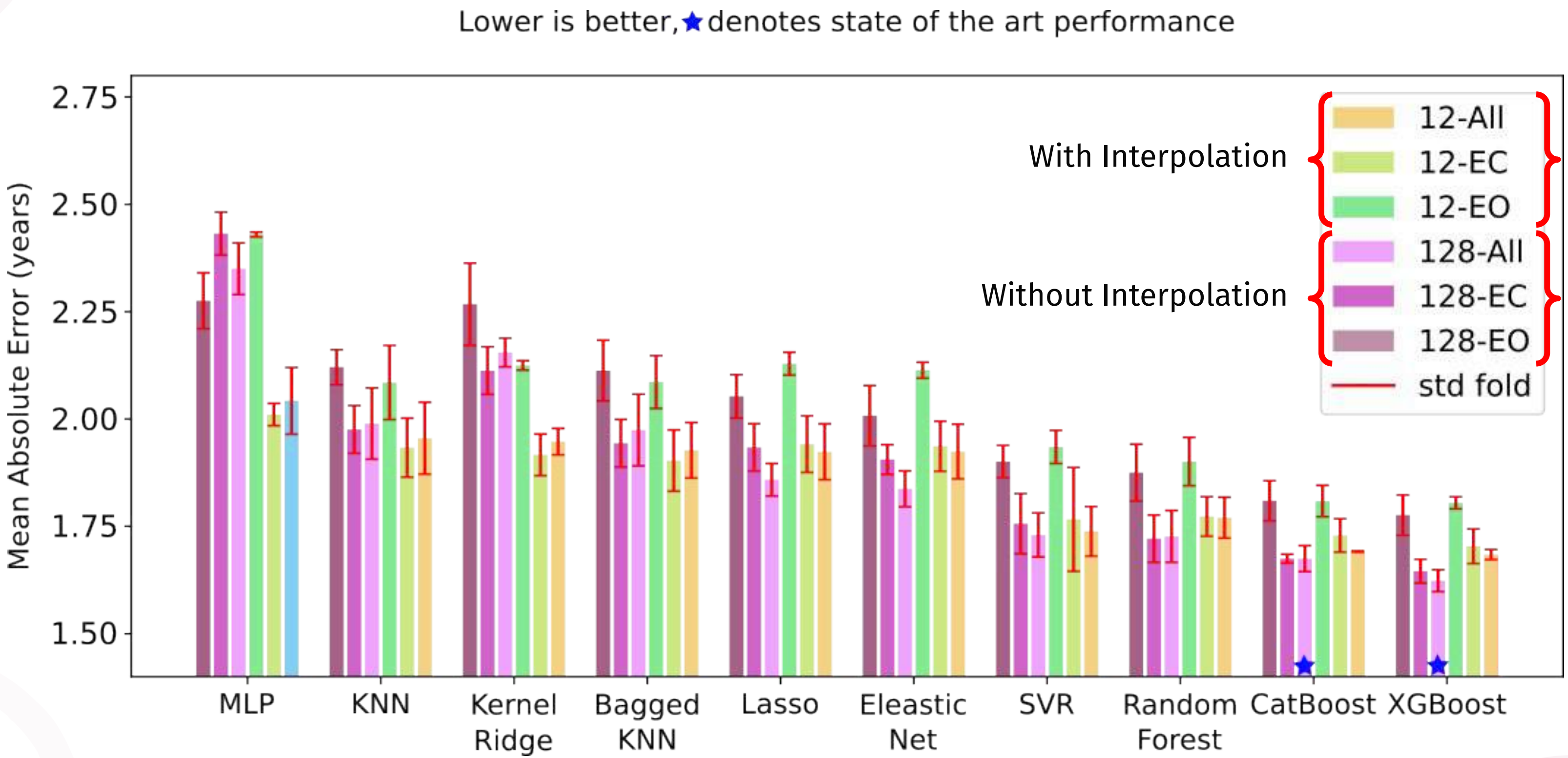
Brain Age Prediction Results



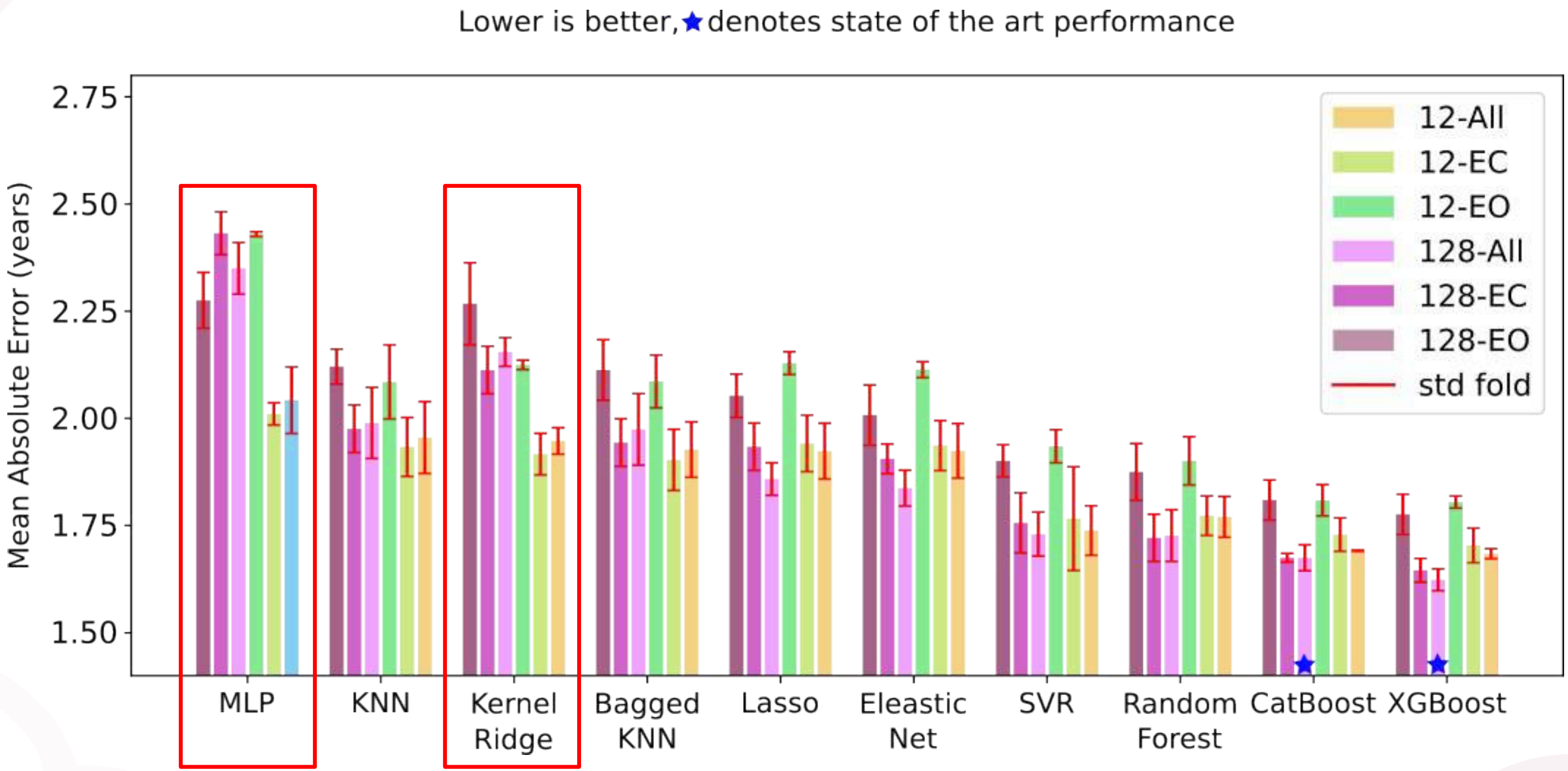
Brain Age Prediction Results



Brain Age Prediction Results

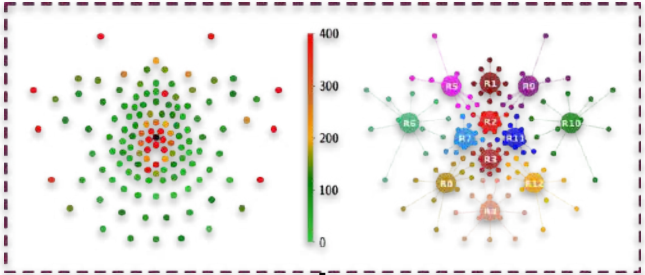


Brain Age Prediction Results

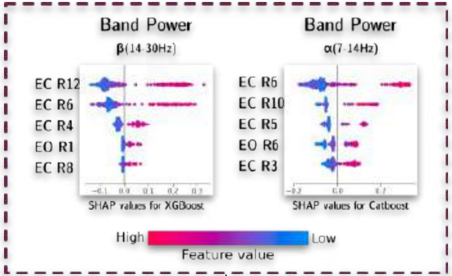


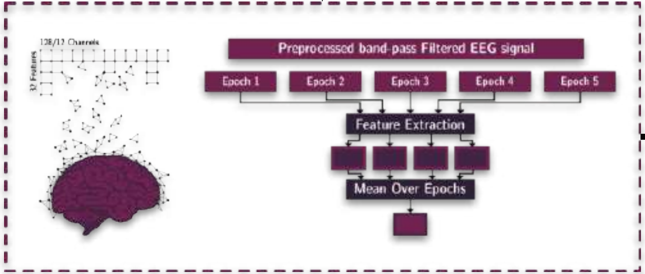
Proposed Pipeline

Preprocessing / Artifact Rejection / Channel Interpolation

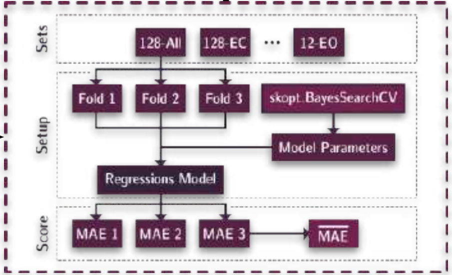


SHAP Value Analysis





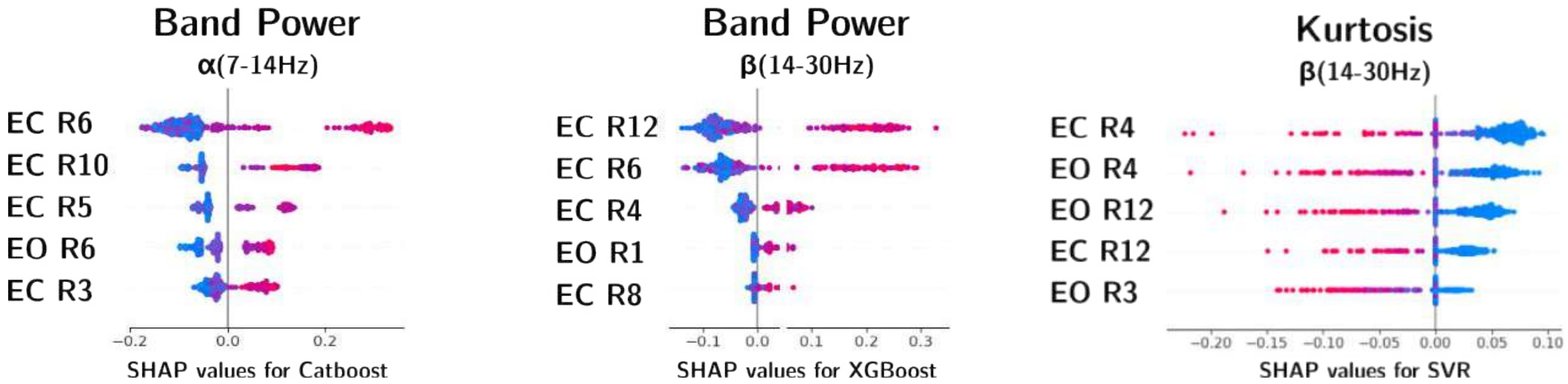
Feature Selection / Feature Extraction



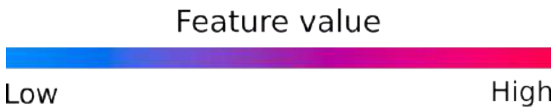
Model Training & Tuning

Can We Replicate Previous Findings?

High frequency activity increase during maturation.
Alpha Band Power [1], [5], [6] and **Beta Band Power [1], [5]**

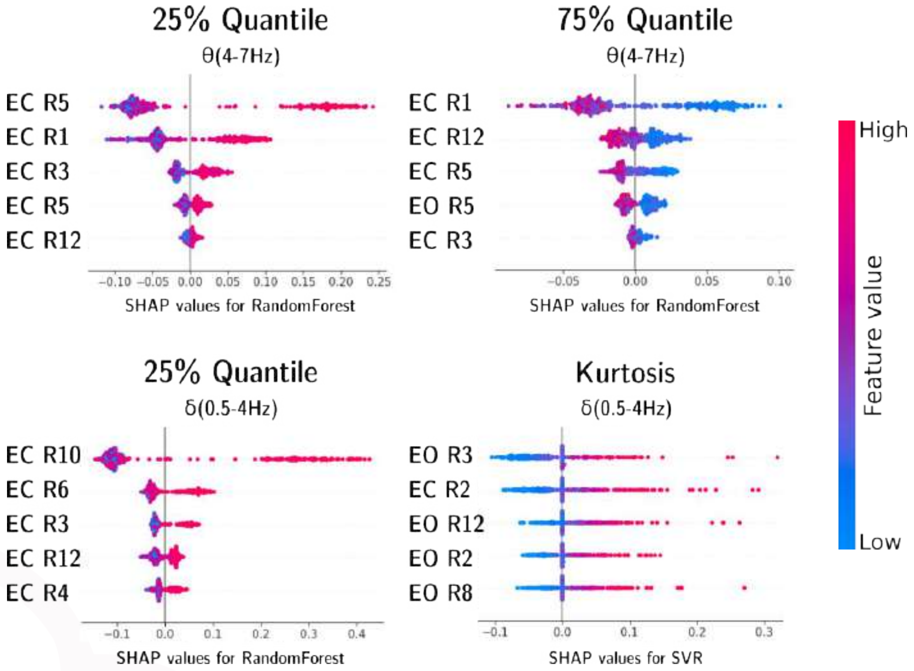


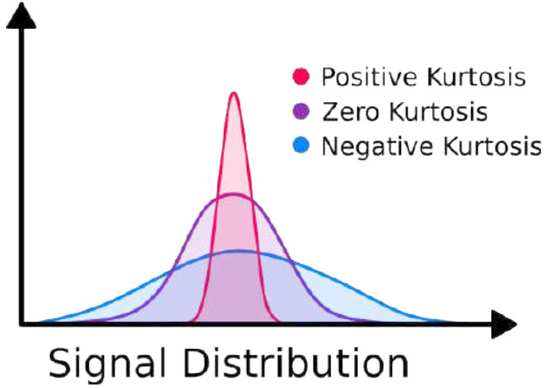
EC: eyes closed
EO: eyes opened
R<n>: regional signal



Can We Replicate Previous Findings?

Low frequency activity decreases during maturation
Delta Band Power [5], [6] and **Theta Band Power [7], [8]**

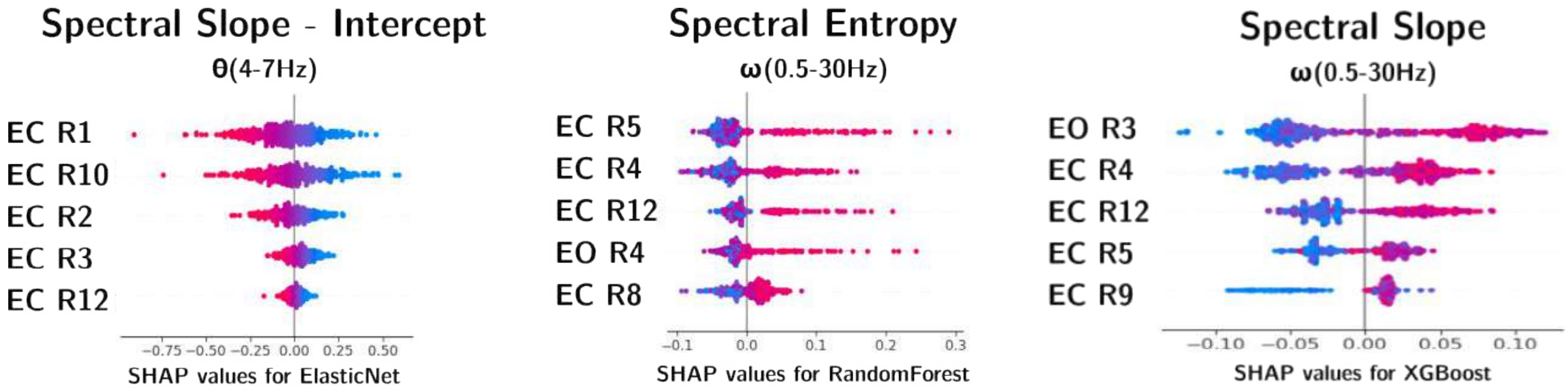




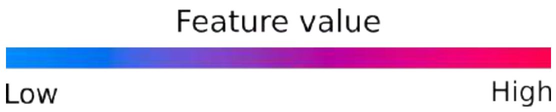
EC: eyes closed
EO: eyes opened
R<n>: regional signal

Can We Replicate Previous Findings?

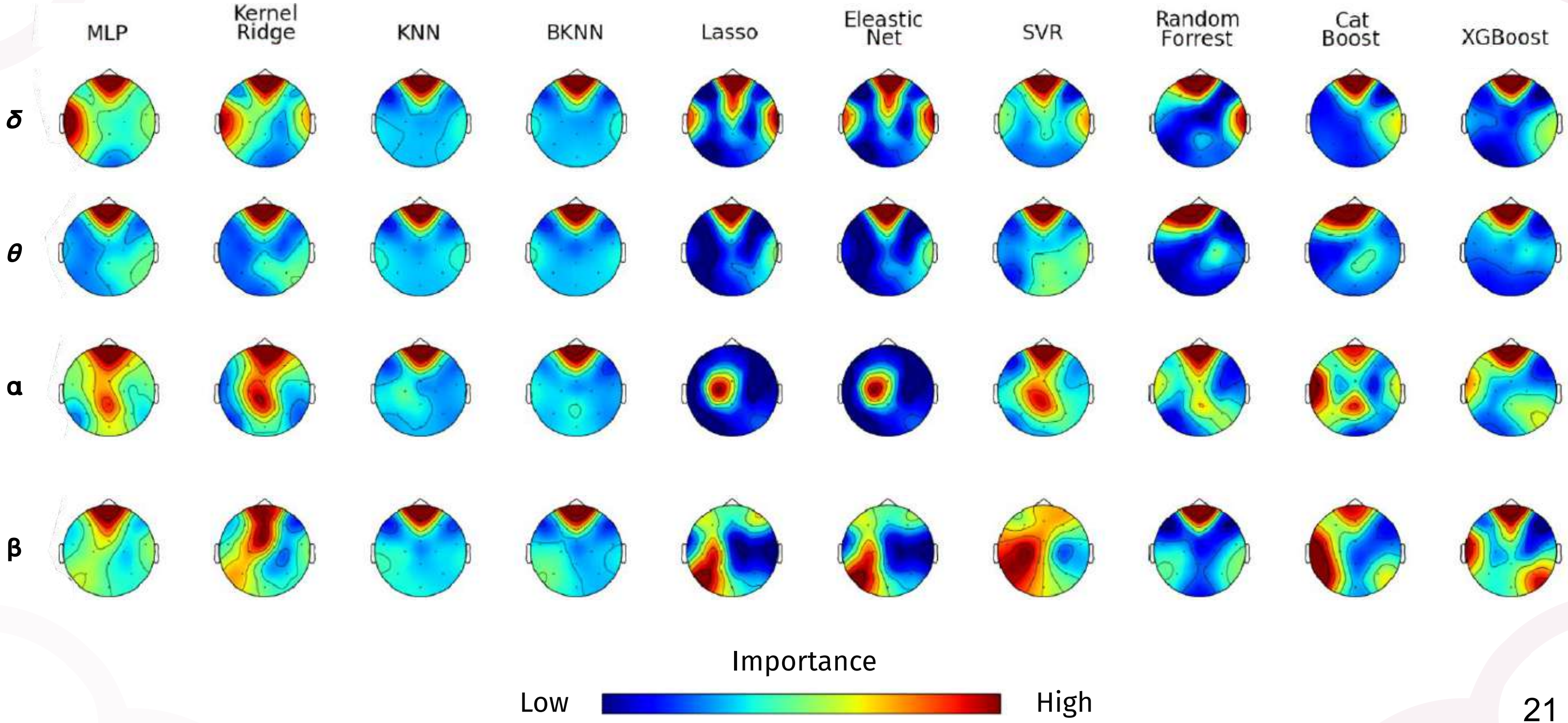
Spectral features show high importance to all of our models.
Beta Spectral Flatness [2] and **Spectral Slope [7]**



EC: eyes closed
EO: eyes opened
R<n>: regional signal

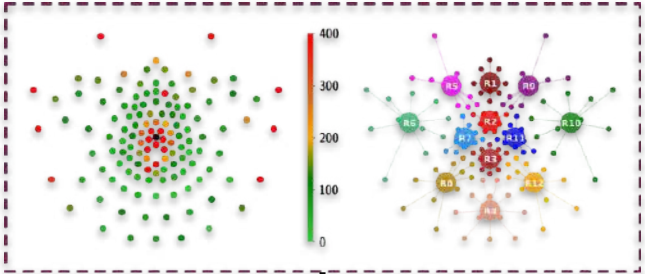


SHAP Values By Region For 12-All

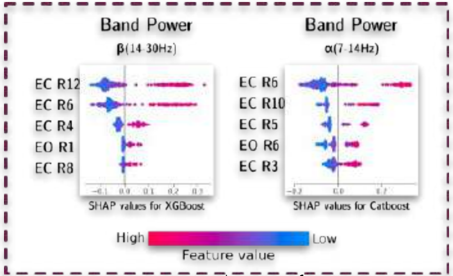


Proposed Pipeline

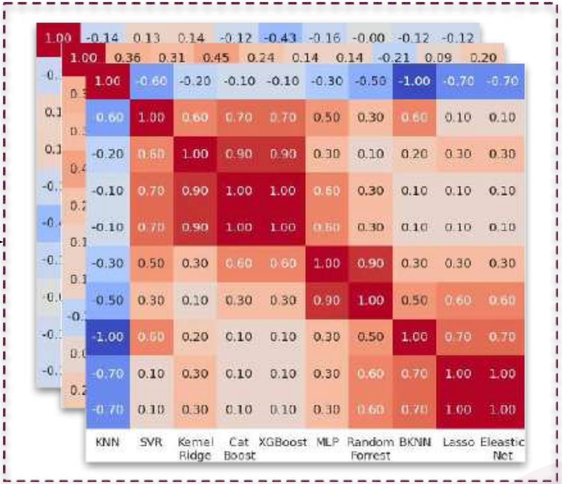
Preprocessing / Artifact Rejection / Channel Interpolation

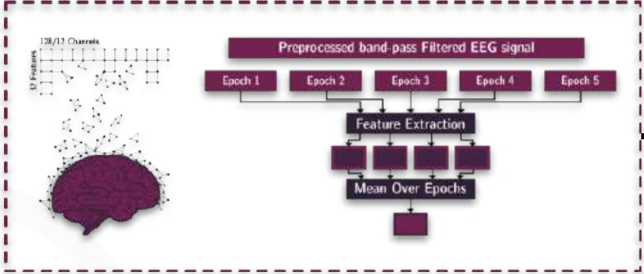


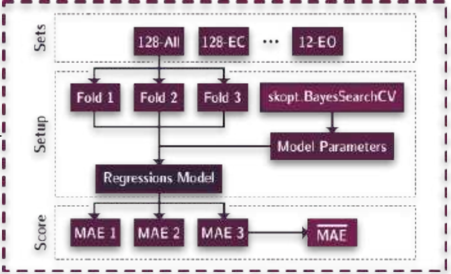
SHAP Value Analysis



Classifier Agreement based on Aggregated SHAP values







Feature Selection / Feature Extraction

Model Training & Tuning

SHAPAgreement

The **SHAPAgreement** measures the agreement in feature importance across different models based on Shapley values.

Group features based on some condition.

Features

$$F = \{f_1, f_2, ..., f_n\}$$

Distinct Feature Groups

$$G_1, G_2, ..., G_m \subset F$$

For each model, compute the group importance.

Group Importance

$$S_{M,G_i} = \sum_{f_j \in G_i} |\phi_M(f_j)|$$

Ranked Group Importance

$$S = (S_{M,G_1}, S_{M,G_2} ..., S_{M,G_m})$$

SHAPAgreement:
Compute the rank order correlation of the group importance.

$$A(M_1, M_2, G) = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$d_i = R(S_{M_1,G_i}) - R(S_{M_2,G_i})$$

SHAPAgreement

The **SHAPAgreement** measures the agreement in feature importance across different models based on Shapley values.

Group features based on some condition.

Features

$$F = \{f_1, f_2, ..., f_n\}$$

Distinct Feature Groups

$$G_1, G_2, ..., G_m \subset F$$

For each model, compute the group importance.

Group Importance

$$S_{M,G_i} = \sum_{f_j \in G_i} |\phi_M(f_j)|$$

Ranked Group Importance

$$S = (S_{M,G_1}, S_{M,G_2} ..., S_{M,G_m})$$

SHAPAgreement:
Compute the rank order correlation of the group importance.

$$A(M_1, M_2, G) = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

$$d_i = R(S_{M_1,G_i}) - R(S_{M_2,G_i})$$

SHAPAgreement

The **SHAPAgreement** measures the agreement in feature importance across different models based on Shapley values.

Group features based on some condition.

Features

$$F = \{f_1, f_2, ..., f_n\}$$

Distinct Feature Groups

$$G_1, G_2, ..., G_m \subset F$$

For each model, compute the group importance.

Group Importance

$$S_{M,G_i} = \sum_{f_j \in G_i} |\phi_M(f_j)|$$

Ranked Group Importance

$$S = (S_{M,G_1}, S_{M,G_2} ..., S_{M,G_m})$$

SHAPAgreement:
Compute the rank order correlation of the group importance.

$$A(M_1, M_2, G) = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

$$d_i = R(S_{M_1,G_i}) - R(S_{M_2,G_i})$$

SHAPAgreement

The **SHAPAgreement** measures the agreement in feature importance across different models based on Shapley values.

Group features based on some condition.

Features

$$F = \{f_1, f_2, ..., f_n\}$$

Distinct Feature Groups

$$G_1, G_2, ..., G_m \subset F$$

For each model, compute the group importance.

Group Importance

$$S_{M,G_i} = \sum_{f_j \in G_i} |\phi_M(f_j)|$$

Ranked Group Importance

$$S = (S_{M,G_1}, S_{M,G_2} ..., S_{M,G_m})$$

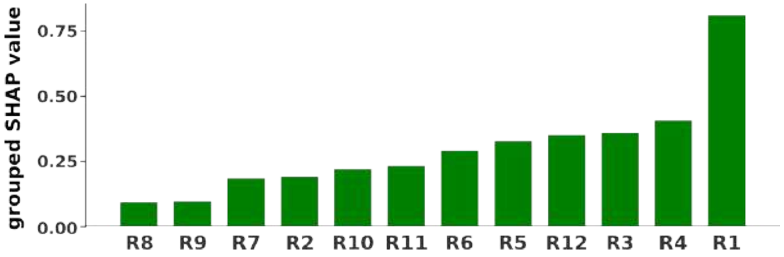
SHAPAgreement:
Compute the rank order correlation of the group importance.

$$A(M_1, M_2, G) = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

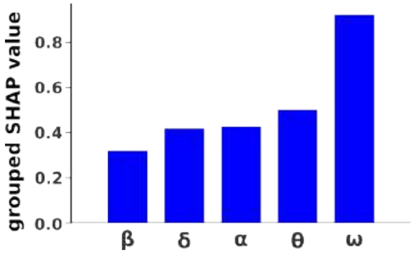
$$d_i = R(S_{M_1,G_i}) - R(S_{M_2,G_i})$$

Rank orders for XGBoost on 12-All

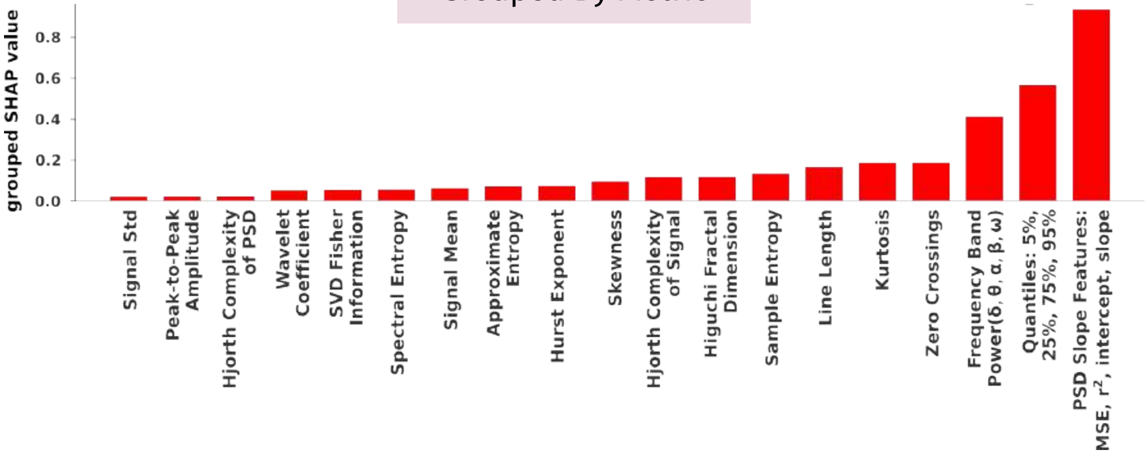
Grouped By Region



Grouped By Frequency

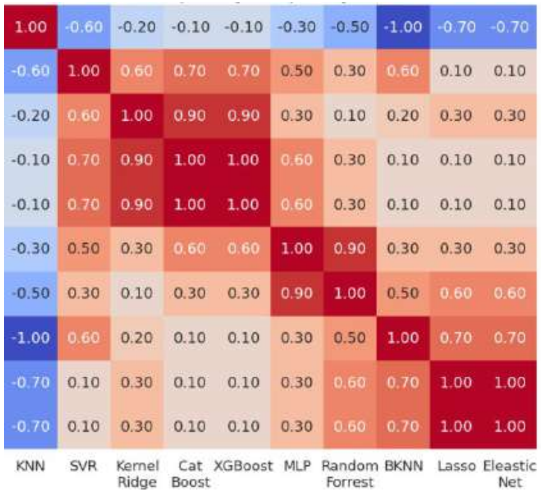


Grouped By Metric

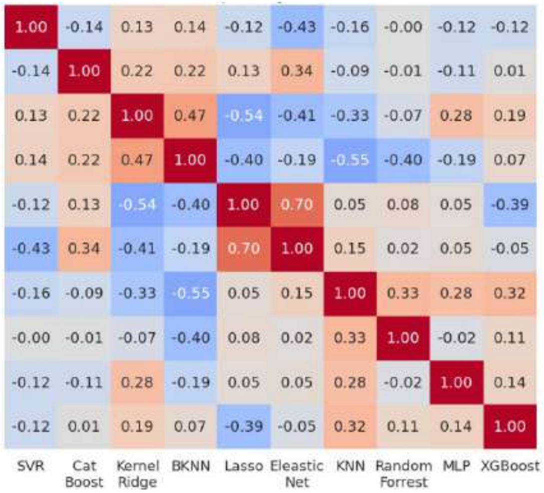


SHAPAgreement

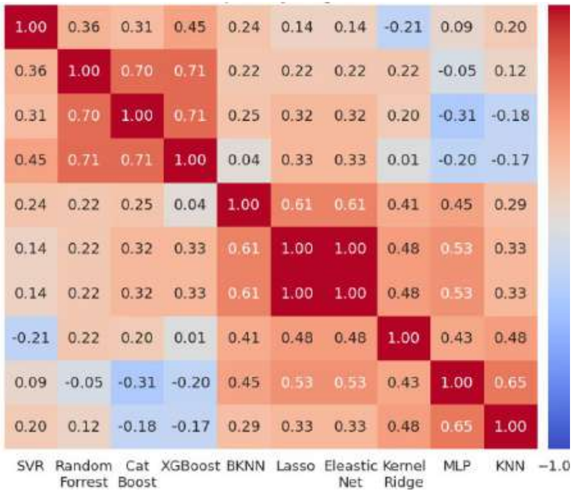
Aggregated By Frequency Band



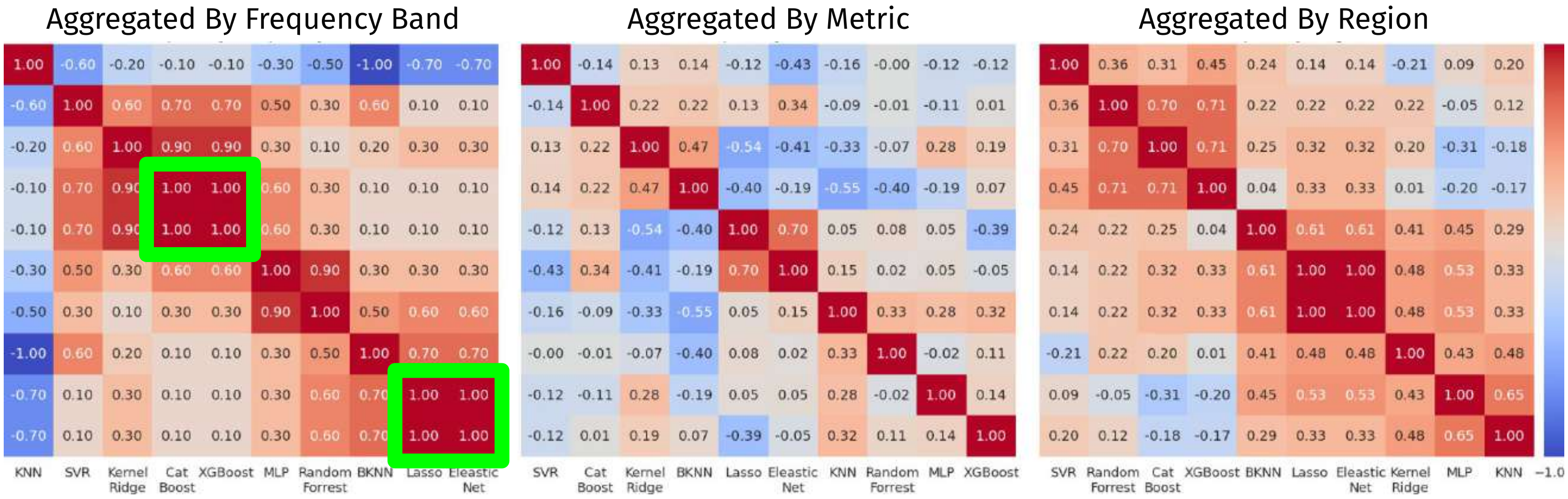
Aggregated By Metric



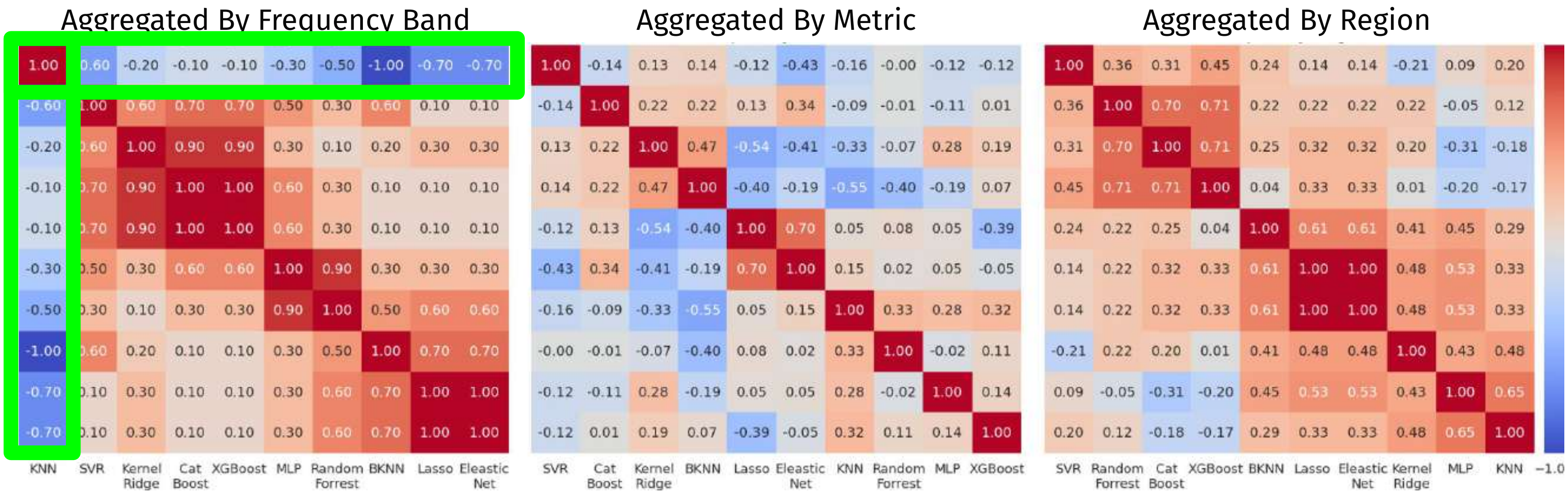
Aggregated By Region



SHAPAgreement

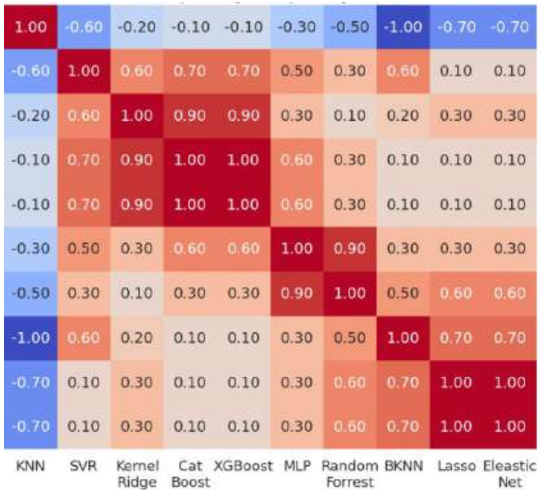


SHAPAgreement

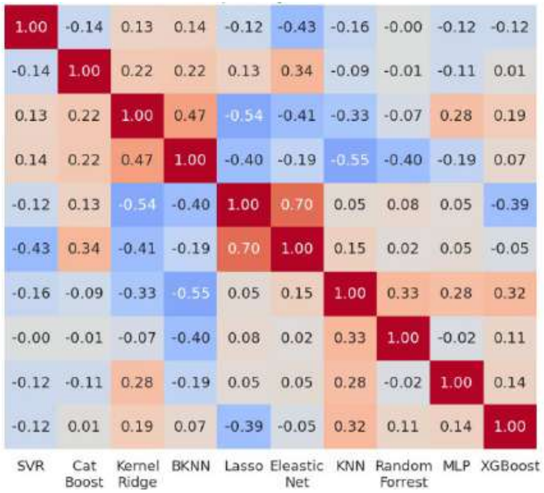


SHAPAgreement

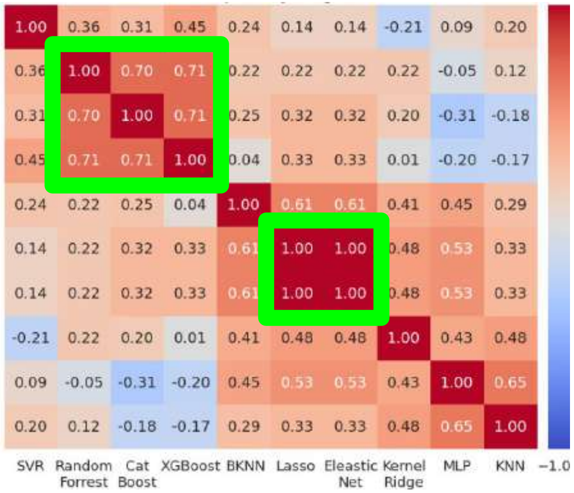
Aggregated By Frequency Band



Aggregated By Metric

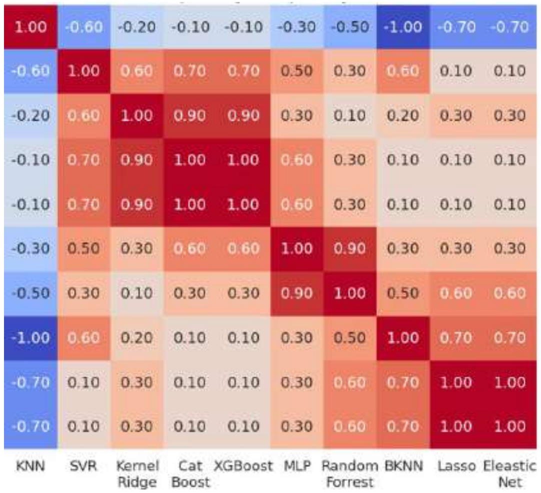


Aggregated By Region



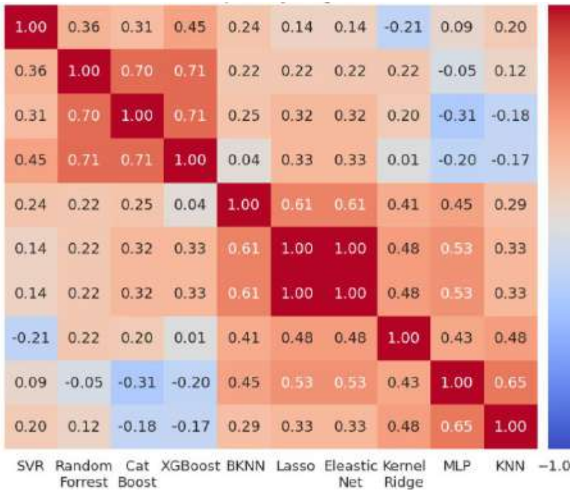
SHAPAgreement

Aggregated By Frequency Band



Aggregated By Metric

Aggregated By Region



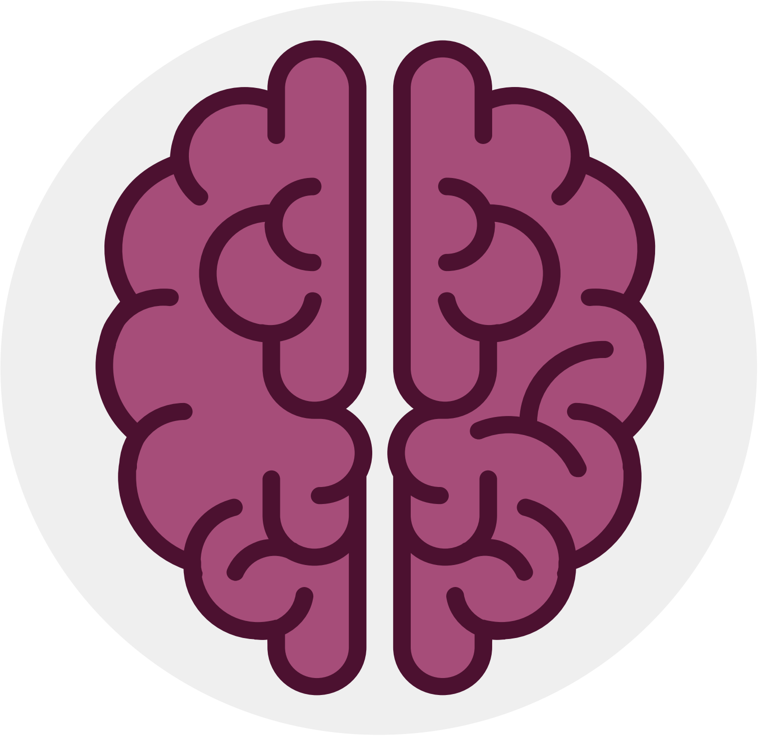
Key Conclusions

Algorithm Choice Matters!

Different 'Algorithms Categories' pick up on different patterns!

SHAP values do not correlate with algorithmic *consistency*!

Thanks For Your Attention!



FIND ME ON GITHUB

IEEE BIBM 2023

5-8 December 2023 • İstanbul & Turkey