Statistics

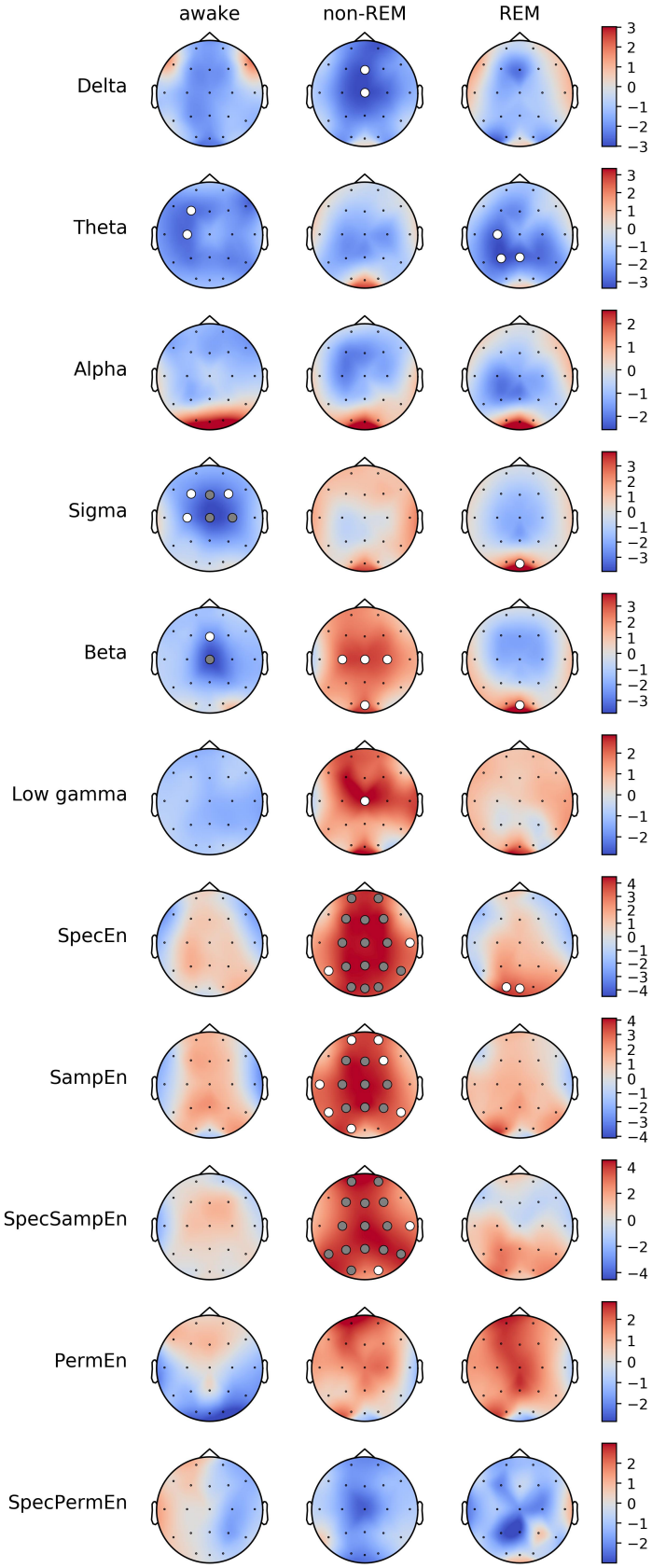

Figure 1: Topographical plots showing the spatial distribution of T-values from exhaustive permutations corrected with tmax statistics of caffeine minus placebo condition across computed features and sleep stages. White dots represent statistical significance at p<0.05 and grey dots show significance at p<0.01.

Single-feature machine learning

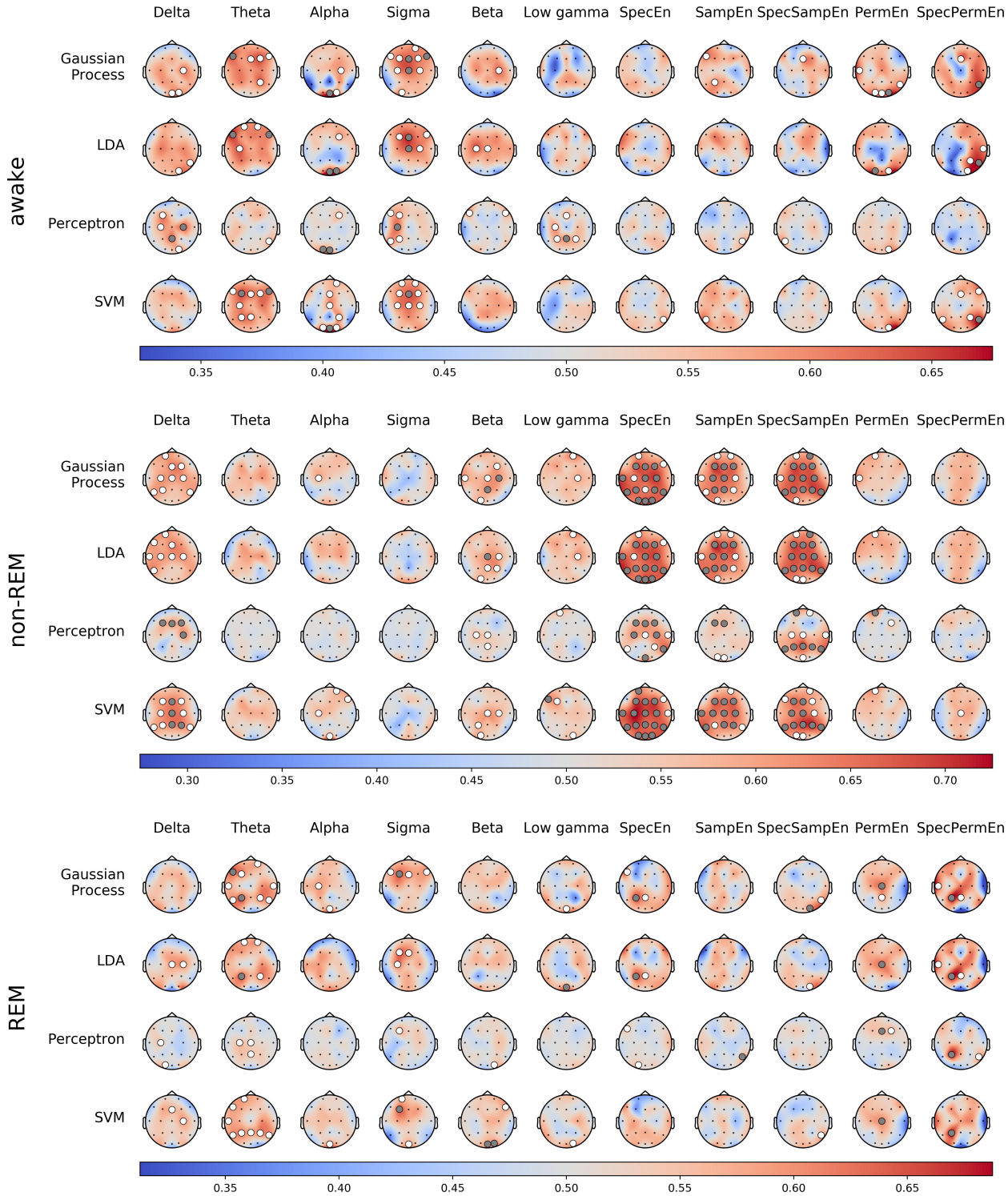

Figure 2: Topographical plots showing the spatial distribution of decoding accuracy of different classifiers on single-feature classification between caffeine and placebo conditions, classifiers were trained separately for each sleep stage. Statistical significance was evaluated using permutation tests with permuted labels, white dots represent significance at $p<0.05$ and grey dots show significance at $p<0.01$.

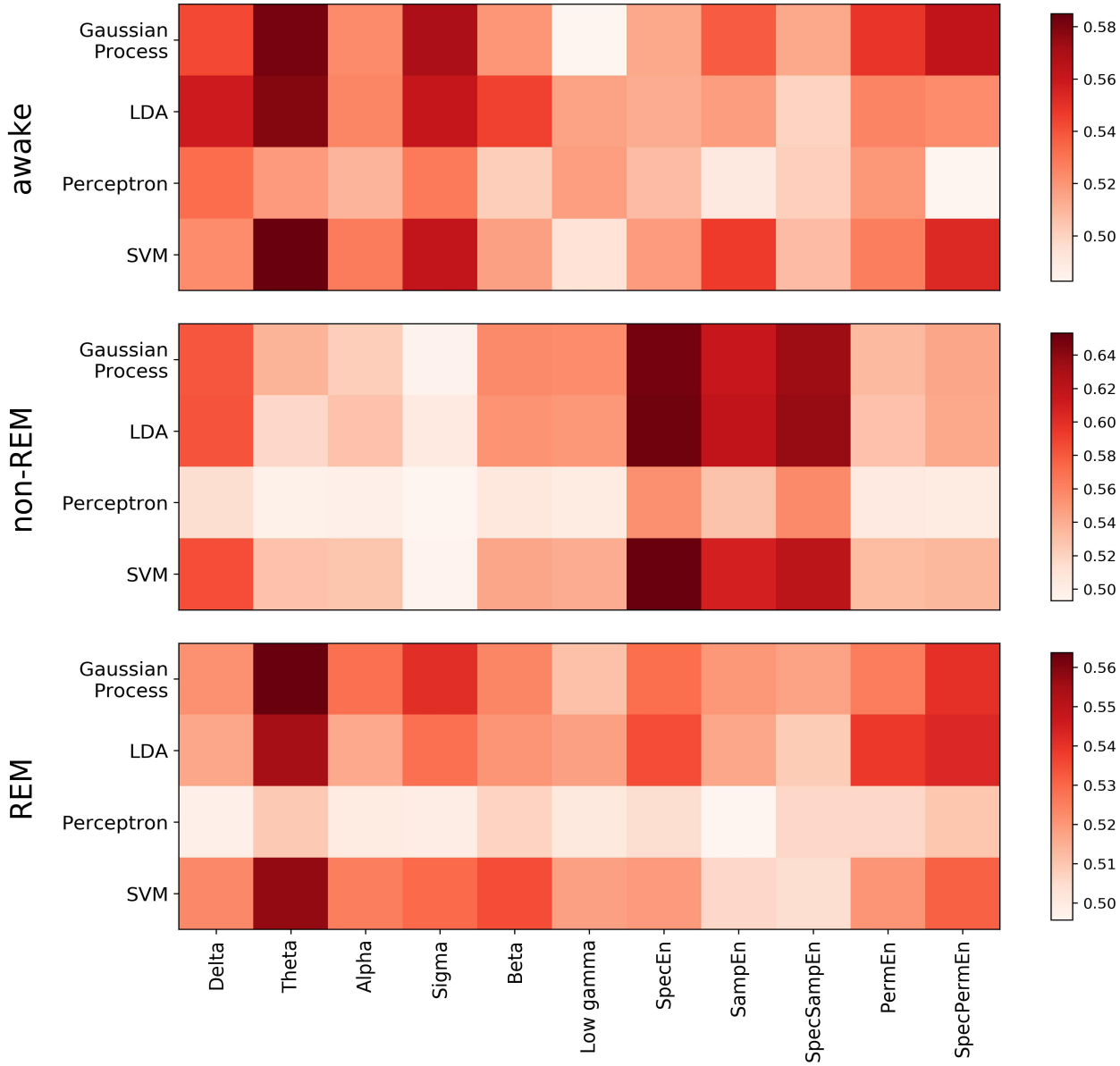

Figure 3: Single-feature, single electrode decoding accuracy between caffeine and placebo condition averaged over electrodes. Classifiers were trained separately for each sleep stage. Each row per sleep stage corresponds to a different classification algorithm while columns show different computed features.

Multi-electrode multi-feature machine learning (random forest)

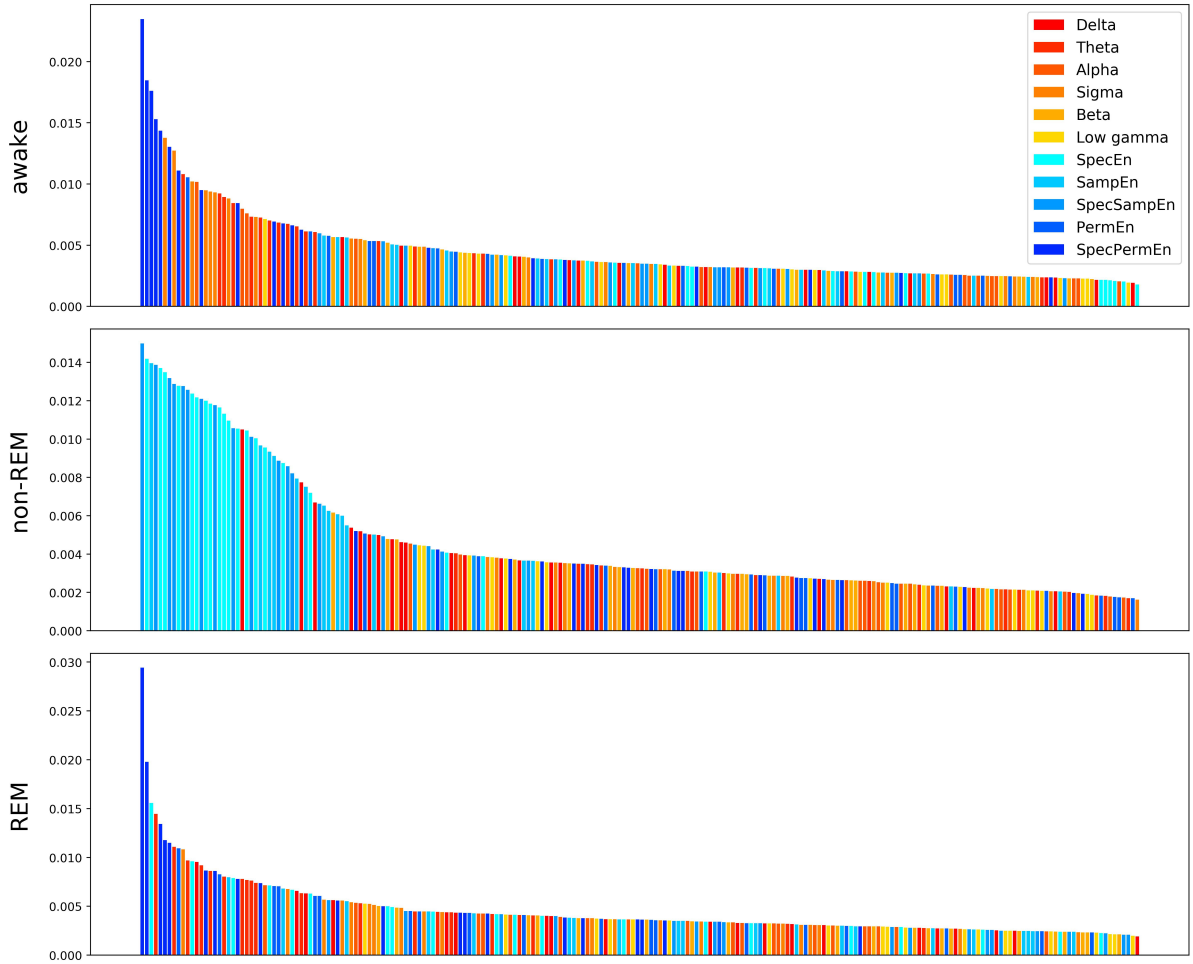

Figure 4: Averaged feature importance of random forests trained on all features and electrodes combined. The height of each bar represents the importance of one feature in one electrode. Classifiers were trained separately for each sleep stage. Colours indicate different computed features, warm colours (red to yellow) show power bands of different frequencies and cold colours (light blue to dark blue) correspond to different entropy measures. The heights of all bars per sleep stage add up to about one (not exactly one since feature importance is averaged over multiple training iterations).

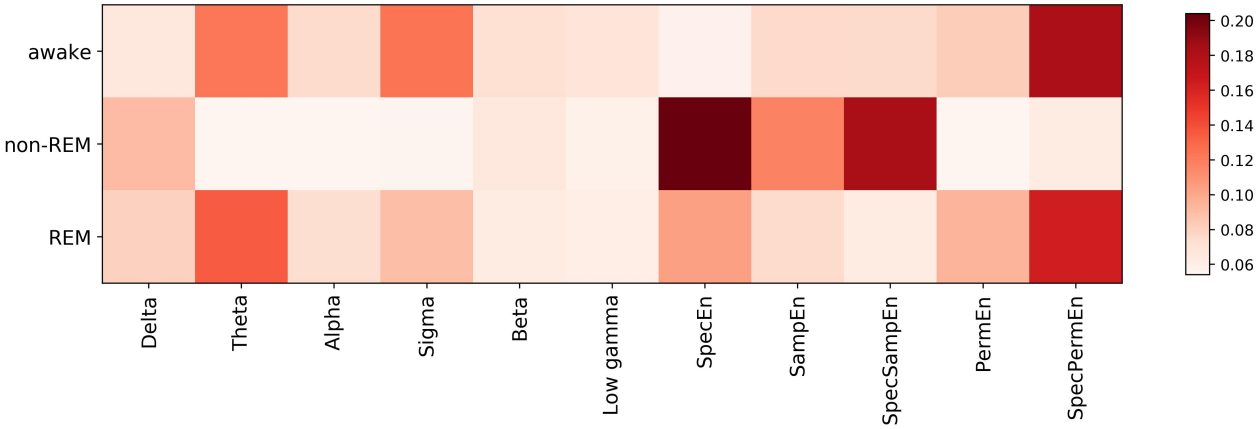

Figure 5: Mean feature importance from random forests (all features and electrodes combined) summed up over electrodes. Random forests were trained separately for each sleep stage (rows). The columns show different computed features (6 power bands and 5 entropy measures). Higher values indicate more importance of a feature, rows add up to about one (not exactly one since feature importance is averaged over multiple training iterations).

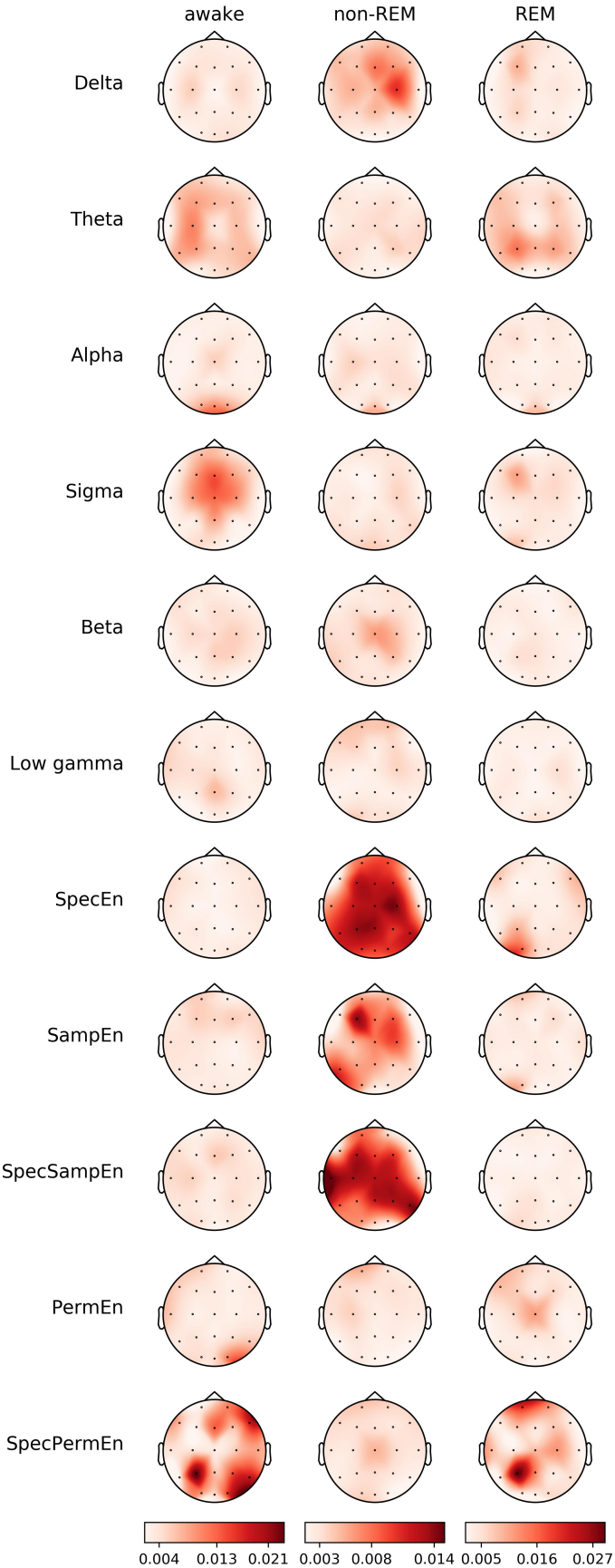

Figure 6: Topographical plots showing spatial averaged feature importance from random forests trained on all computed features (rows) and electrodes (6 power bands, 5 entropy measures, 20 electrodes each). Random forests were trained separately for each sleep stage (columns). Higher values indicate higher importance of a feature in a specific electrode, feature importance adds up to about one inside the columns (not exactly one since feature importance is averaged over multiple training iterations).