Perfect — here is a **single, unified Methods section** that integrates everything we developed across all your scripts: data preparation, correlation analysis, predictive modeling, statistical inference, and model interpretation with SHAP. It is structured, polished, and suitable for submission to a neuroscience journal.

---

# Methods

## Participants and Data Structure

Simultaneous EEG and fMRI data were obtained from $N = 23$ participants, each contributing up to two recording sessions and four experimental blocks per session. For each block, preprocessed EEG recordings and fMRI-derived time series were available. The fMRI signals corresponded to four large-scale brain networks defined using a multimodal surface-based hierarchical Bayesian model (MSHBM): the dorsal attention network (DAN), the default mode network (DMN), and two dorsal attention subnetworks (DNa and DNb).

---

## EEG Time–Frequency Decomposition

EEG data were transformed into time–frequency representations (TFRs) centered on annotated stimulus events. Oscillatory power was estimated with complex Morlet wavelets for frequencies 1–40 Hz in 1-Hz steps, with the number of cycles proportional to frequency ($n_{\text{cycles}} = f/2$), balancing temporal and spectral resolution. Power was computed as the squared magnitude of the convolution between the EEG signal and the wavelet:

$$W_c(f,t) = \left|(x_c * \psi_f)(t)\right|^2,$$

where $x_c(t)$ is the signal at channel $c$ and $\psi_f$ the wavelet at frequency $f$.

For each event, a 20-s pre-stimulus window was divided into $n_{\text{bins}} = 20$ non-overlapping 1-s bins. Within each bin, oscillatory power was averaged, yielding per-event features:

$$X_{\text{event}} \in \mathbb{R}^{C \times F \times B},$$

with $C$ channels, $F = 40$ frequencies, and $B = 20$ time bins. Concatenating across events produced subject-level tensors:

$$X \in \mathbb{R}^{T \times C \times F \times B},$$

where $T$ is the number of events. EEG and fMRI signals were truncated to the minimum common length within each block to ensure temporal alignment.

---

## Preprocessing of EEG Features

To stabilize variance, power values were log-transformed,

$$\tilde{X} = \log_{10}(X + \varepsilon), \quad \varepsilon = 10^{-10},$$

and then standardized to zero mean and unit variance within each subject, separately for every channel–frequency pair:

$$Z = \frac{\tilde{X} - \mu}{\sigma}.$$

For correlation analyses, standardized features were averaged across channels, yielding trial-wise matrices of EEG power as a function of frequency and pre-event time bin.

---

## Subject-wise EEG–fMRI Correlation Maps

For each subject and network, Pearson correlation coefficients were computed between trial-to-trial EEG power and the fMRI time series:

$$r(f, b) = \mathrm{corr}\big(Z(:, f, b),\, Y\big),$$

where $f$ indexes frequency and $b$ time bin. This produced a two-dimensional correlation map ($F \times B$) per subject and network, capturing associations between pre-event EEG activity and network-specific BOLD fluctuations.

At the group level, we tested whether correlations differed from zero using two-sided one-sample $t$-tests at each frequency–time point. $p$-values were corrected for multiple comparisons with the Benjamini–Hochberg false discovery rate (FDR, $\alpha = 0.01$). Group mean correlation maps were visualized as heatmaps, with significant bins marked after FDR correction.

---

## Feature Construction for Prediction

For predictive modeling, EEG features were rebinned both in time and frequency. The 20 one-second bins were aggregated into 10 consecutive 2-s bins covering the 20-s pre-event interval:

$$[0-2],\ [2-4],\ \ldots,\ [18-20]\ \text{s before the event}.$$

Frequencies were collapsed into canonical bands by averaging within defined ranges: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), gamma (30–40 Hz). For band $k$:

$$X_{t,c,k,b} = \frac{1}{|F_k|} \sum_{f \in F_k} Z_{t,c,f,b}.$$

The resulting feature tensor had dimensions:

$$X_{\text{binned}} \in \mathbb{R}^{T \times C \times K \times B},$$

with $K = 5$ frequency bands and $B = 10$ time bins. These features served as inputs to the predictive models.

---

## Predictive Modeling with Gradient-Boosted Trees

We used **extreme gradient boosting regression (XGBoost)** to predict fMRI network activity from EEG features. Predictions are formed by summing over an ensemble of regression trees:

$$\hat{y}_i = \sum_{m=1}^{M} f_m(x_i), \quad f_m \in \mathcal{F},$$

where $x_i$ is the feature vector for sample $i$, $f_m$ a regression tree, and $M$ the number of boosting rounds.

Hyperparameters were fixed at: 300 trees, maximum depth 3, learning rate 0.05, subsample 0.8, and colsample_bytree 0.8.

Performance was evaluated with **leave-one-subject-out (LOSO) cross-validation**. In each fold, one subject was held out for testing, while models were trained on all others. Features were standardized within each fold based on the training data.

Model accuracy was quantified with the coefficient of determination,

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}.$$

Separate models were trained for each target network (DAN, DMN, DNa, DNb), yielding subject-level and overall $R^2$ values.

---

## Statistical Inference on Predictive Performance

For each network, the LOSO procedure produced one $R^2$ value per subject. These subject-level values were tested against chance ($R^2 = 0$):

- **Parametric test:** one-sided one-sample $t$-test, $H_0 : \mathbb{E}[R^2] = 0$ vs $H_1 : \mathbb{E}[R^2] > 0$.
- **Nonparametric tests (robustness):** Wilcoxon signed-rank test and a binomial sign test on the number of subjects with positive $R^2$.

Effect sizes were estimated from the Wilcoxon $z$-score as $d_z = z/\sqrt{n}$. To control for multiple comparisons across the four networks, Wilcoxon $p$-values were corrected with the Benjamini–Hochberg FDR procedure.

We also report descriptive statistics including mean, median, standard deviation, and 95% confidence intervals of $R^2$, computed via the $t$-distribution:

$$\mathrm{CI}_{95\%} : \bar{R}^2 \pm t_{0.975,\, n-1} \frac{s}{\sqrt{n}},$$

where $\bar{R}^2$ is the mean across subjects, $s$ the standard deviation, and $n$ the number of subjects.

Results were visualized using boxplots with individual subject values overlaid, together with mean ± 95% CI markers.

---

## Model Interpretation with SHAP Values

To interpret the trained models, we used **SHAP (SHapley Additive exPlanations)**, a game-theoretic approach that assigns each feature a contribution to the model's prediction. For a given trial $i$ with features $x_i$, the model prediction can be decomposed as:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^{p} \phi_{ij},$$

where $\phi_0$ is the average model output, $\phi_{ij}$ is the SHAP value for feature $j$, and $p$ the number of features. The magnitude $|\phi_{ij}|$ reflects the importance of feature $j$.

Models were trained on **all available data** (rather than LOSO folds) to maximize statistical power for interpretability. SHAP values were reshaped back into their original tensor form (channels $\times$ frequency bands $\times$ time bins) and summarized at multiple levels:

- **Band–time maps:** averaged across channels.
- **Channel–time maps:** averaged across frequency bands.
- **Subject-level summaries:** averaged within subjects, yielding subject $\times$ band $\times$ time arrays.

Both signed SHAP values (direction of effect) and absolute SHAP values (magnitude of importance) were computed, with the latter serving as the primary measure of feature relevance.

---

## Permutation Testing of SHAP Importance

To assess statistical significance of SHAP attributions, we implemented a permutation test. For each of $R = 100$ iterations, the fMRI target time series was permuted **within subjects** (preserving within-subject structure but destroying cross-modal correspondence). A new XGBoost model was refit to the permuted data, SHAP values recomputed, and subject-level band–time importance maps recalculated.

Permutation-based $p$-values were then estimated as:

$$p(f,b) = \frac{1 + \sum_{r=1}^{R} \mathbf{1}\left[\mathrm{SHAP}^{(r)}_{\mathrm{null}}(f,b) \geq \mathrm{SHAP}_{\mathrm{obs}}(f,b)\right]}{R+1},$$

where $\mathrm{SHAP}_{\mathrm{obs}}(f,b)$ is the observed mean absolute SHAP value for band $f$ and time bin $b$, and $\mathrm{SHAP}^{(r)}_{\mathrm{null}}(f,b)$ are permutation-null values.

This yielded empirical significance maps of band–time feature importance, identifying EEG components that most strongly contributed to predicting fMRI network activity.

---

✅ This is now a **single, continuous Methods section** covering everything from raw data to SHAP permutation tests.

Do you want me to also draft a **high-level schematic figure caption** (like a pipeline overview figure: EEG → TFR → band/time binning → XGBoost → LOSO $R^2$ → SHAP + permutation tests) to visually summarize the workflow?

❞ Ask ChatGPT