Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Developmental Differences in Gamma-Band Brain Connectivity During Propofol and Sevoflurane Anesthesia in Children

Junpeng Zhang
Sichuan University
Rui Zhao
zhaorui.scu@gmail.com

Sichuan University
Lifang Yang
Children's Hospital Affiliated to Xi'an Jiao Tong University
Liming Cheng
Kunming Children's Hospital
Jingping Wang
Harvard Medical School

Research Article

Keywords: Propofol, Sevoflurane, Electroencephalogram, Source space, gamma band, Machine learning, Age-dependent effects

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11430305/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Developmental Differences in Gamma-Band Brain Connectivity During Propofol and Sevoflurane Anesthesia in Children

Junpeng Zhang, PhD[1], Rui Zhao, BS[1], Lifang Yang, MD[2,*], Liming Cheng, MD[3], Jingping Wang, MD, PhD[4].

[1]College of Electrical Engineering, Sichuan University, Chengdu, China, [2]Department of Anesthesiology, Children's Hospital Affiliated to Xi'an Jiao Tong University, Xi'an, Shaanxi, China, [3]Department of Anesthesiology, Kunming Children's Hospital, Kunming, Yunnan, China, [4]Department of Anesthesia, Critical Care and Pain Medicine Massachusetts General Hospital, Harvard Medical School, Boston, USA. [*]indicates corresponding author.

Corresponding author

Lifang Yang, MD

Department of Anesthesiology, Children's Hospital Affiliated to Xi'an Jiao Tong University, Xi'an, Shaanxi, China

Email: yanglf@fmmu.edu.cn

Abstract

Background: Propofol and sevoflurane are commonly used pediatric anesthetics with distinct molecular targets; however, their differential modulation of developing brain networks across age groups remains poorly understood. Unlike scalp low-frequency EEG, source-space gamma-band functional connectivity reflects cortical processing and may offer a sensitive window into drug-specific effects. This study aimed to identify distinct gamma-band connectivity patterns distinguishing propofol from sevoflurane anesthesia using explainable machine learning.

Methods: We analyzed 19-lead EEG data from 196 children undergoing general anesthesia, divided into an internal derivation cohort (n = 150) and an external validation cohort (n = 46). Patients were stratified into three age groups (≤ 42 months(G1), 43--83 months(G2), and ≥ 84 months(G3)). Cortical activity was reconstructed via source imaging using eLORETA. Machine learning models were developed to classify anesthetic types based on gamma-band connectivity features. Model performance was evaluated using internal and external validation, and Shapley Additive Explanations (SHAP) were applied to identify key anatomical regions driving network reorganization.

Results: Model predictive performance exhibited a distinct age gradient. The internal validation Area Under the Curve (AUC) significantly improved from G1 (e.g., MLP AUC = 0.703), G2(MLP AUC = 0.794), to G3 (MLP AUC = 0.922), with external validation confirming robust generalization in older children (e.g., SVM AUC = 0.773). SHAP analysis based on 150 optimal features revealed age-dependent network shifts: distinguishing anesthetics relied on primary sensory-limbic interactions in G1, shifted to default mode network and orbitofrontal connectivity in G2, and depended on long-range parahippocampal and occipito-frontal networks in G3.

Conclusions: Source-space *gamma*-band connectivity combined with explainable machine learning effectively distinguishes propofol from sevoflurane anesthesia in children. These findings reveal that distinct, drug-specific modulation patterns mature alongside the developing brain, highlighting a pronounced age gradient that lays the groundwork for age-adapted precision

anesthesia monitoring.

Key words: Propofol; Sevoflurane; Electroencephalogram; Source space; *gamma* band; Machine learning; Age-dependent effects

Clinical implications

a. What is already known about the topic

Propofol and sevoflurane produce distinct EEG patterns, and anesthetic EEG characteristics vary with age in children. However, age-related differences in source-space gamma-band connectivity between these agents remain poorly characterized.

b. What new information this study adds

Source-space gamma-band connectivity distinguished propofol from sevoflurane more accurately in older children, with different connections contributing to classification across age groups. These findings support further investigation of age-adapted, agent-specific EEG monitoring in pediatric anesthesia.

Abbreviations

EEG, electroencephalogram
ssEEG, source spatial electroencephalogram
ML, machine learning
eLORETA, exact low-resolution brain electromagnetic tomography
LightGBM, lightweight gradient boosting machine
CatBoost, categorical boosting
XGBoost, extreme gradient boosting
GBM, gradient boosting machine
IFS, incremental feature selection
TCI, target controlled infusion

Introduction

Propofol and sevoflurane are the most common intravenous and inhaled anesthetics used in induction of general anesthesia[1,2]. Studies have shown that they regulate the state of consciousness through different neural circuits and molecular targets[3,4]. In particular, due to the immature neural circuits in children's brain development, the specific effects on brain functional networks of different anesthetic drugs may be more significant[5,6]. Electroencephalogram (EEG), as a non-invasive, portable and high temporal resolution measurement technology of brain activity, has been widely used in monitoring the depth of anesthesia and studying the mechanism. However, EEG analysis suffers from fundamental methodological constraints, such as volume conduction, low spatial resolution, signal aliasing, and reference electrode bias. These issues can obscure the true nature of neural dynamics in deep cortex and inter-areal functional connections[7,8,9].

In order to cope with the above limitations, source space EEG (ssEEG) analysis can construct brain functional networks with more physiological specificity and anti-interference ability by reverse projecting scalp potentials to individual cortical anatomical structures, combined with phase coherence analysis and graph theory methods[10,11]. Furthermore, ssEEG can quantify the dynamic changes of brain topological properties such as global efficiency and clustering coefficient caused by anesthetics, providing a more reliable network neuroscience basis for revealing the mechanism

of action of different anesthetics[12].

However, although source space analysis can provide accurate anatomical localization, it produces high-dimensional, nonlinear and complex features. Traditional statistical methods are often difficult to accurately capture drug-specific biomarkers from the massive information of the whole frequency band and the whole brain region. In recent years, machine learning (ML) algorithms have shown excellent classification and feature recognition capabilities in processing complex biomedical signals, and have been successfully applied to decoding the depth of anesthesia and predicting individual sensitivity to anesthetics[13,14,15]. Inspired by this, this study aims to combine source spatial coherence analysis and ML algorithms to screen the key frequency bands and core network features to distinguish the anesthesia induction mechanism of propofol and sevoflurane, and to provide a new theoretical basis for the precise monitoring of clinical anesthesia depth and the study of drug mechanism.

Methods

Study design and population

This retrospective, multicenter study analyzed EEG data from 196 children who underwent general anesthesia between January 2025 and October 2025. The study protocol was approved by the Ethics Committee of Xi'an Children's Hospital and Kunming Children's Hospital (approval number: 2019OF05-02), and the requirement for informed consent was waived due to the retrospective nature of the study involving analysis of routinely collected clinical data. All procedures were conducted in accordance with the Declaration of Helsinki. The dataset was partitioned into an internal derivation cohort (n = 150) sourced entirely from Xi'an Children's Hospital, and an independent external validation cohort (n = 46) from Kunming Children's Hospital to evaluate multicenter generalization.

To account for neurodevelopmental differences, the internal derivation cohort (75 receiving propofol and 75 receiving sevoflurane) was evenly stratified into three age groups, with 50 patients per group (25 propofol and 25 sevoflurane): Group 1 (≤ 42 months), Group 2 (43--83 months), and Group 3 (≥ 84 months). A strict subject-independent, stratified 5-fold cross-validation strategy was employed for model training and internal evaluation. To ensure completely balanced representation, the data partitioning was strictly stratified by both age group and anesthetic agent. Specifically, each of the five folds comprised 30 patients, containing exactly 10 patients from each of the three age groups (5 propofol and 5 sevoflurane per age group). In each iteration, four folds (120 subjects) were utilized for model training, and the remaining fold (30 subjects) served as the internal validation set. The final internal performance of the model was calculated as the average across all five validation iterations. Finally, the independent external validation cohort (n = 46) was tested to verify the model's robustness. The distribution of this external cohort across the predefined age groups was as follows: Group 1 (4 propofol, 14 sevoflurane), Group 2 (9 propofol, 7 sevoflurane), and Group 3 (8 propofol, 4 sevoflurane). Notably, the external cohort showed imbalanced distribution in Group 1 and Group 3, which may affect the generalizability assessment and should be interpreted with caution.

The inclusion criteria were as follows: (1) Age from 6 months to 13 years; (2) American Society of Anesthesiologists (ASA) physical status grade I-II; (3) Elective surgery under general anesthesia; (4) The quality of EEG data during anesthesia maintenance was good.

Exclusion criteria included: (1) Prior general anesthesia; (2) Patients with nervous system diseases, including epilepsy, cerebral palsy, intellectual disability; (3) Developmental delay; (4) Craniofacial deformity; (5) Genetic syndrome; (6) Long-term use of drugs affecting the central nervous system; (7) The use of propofol and sevoflurane was contraindicated; (8) There were too many artifacts in EEG data to obtain enough valid data segments.

Anesthesia details

Propofol Group. Electrocardiogram, non-invasive blood pressure, pulse oxygen saturation were monitored after entering the operating room. Propofol was administered via target-controlled infusion (TCI) for intravenous induction, starting at a target plasma concentration of 0.5 μg/mL. The concentration was then gradually increased according to the child's level of consciousness until the child became unresponsive to verbal commands, corresponding to a Ramsay Sedation Scale score of 5, with regular spontaneous breathing and stable hemodynamics. After a 5-minute observation period under this state, sufentanil 0.3 μg/kg and vecuronium 0.1 mg/kg were administered. Five minutes later, video laryngoscope-guided orotracheal intubation was performed and mechanical ventilation was initiated to maintain the end-tidal carbon dioxide partial pressure between 35-45 mmHg. Intraoperative anesthesia maintenance was adjusted according to patient status index (PSI), 95% spectral margin frequency, density spectral array, electromyography, and suppression ratio obtained by Sedline depth of anesthesia monitor (Masimo Inc, Irvine, California)[16]. The target PSI range was maintained at 25-50, and the 95% spectral edge frequency was kept between 8-12 Hz to ensure comparable anesthetic depth between groups. Furthermore, the target plasma concentration of TCI was set at 2-4 μg/mL and remifentanil 0.2-0.3 μg/(kg·min) was continuously pumped. Propofol infusion was stopped 5 min before the end of surgery, and remifentanil infusion was stopped at the end of surgery.

Sevoflurane Group. With monitors connected upon arrival in the operating room, induction was initiated using a prefilled-circuit technique. Sevoflurane was administered at an initial concentration of 8% in 100% oxygen with a fresh gas flow of 4 L/min. Once consciousness was lost, the inspired sevoflurane concentration was progressively decreased until the child achieved a Ramsay Sedation Scale score of 5 while maintaining regular spontaneous breathing and hemodynamic stability. After observing this state for 5 minutes, sufentanil 0.3 μg/kg and vecuronium 0.1 mg/kg were administered. Video laryngoscope-guided orotracheal intubation followed five minutes later, and mechanical ventilation parameters were set identical to those for the propofol group. Anesthesia was maintained by adjusting the concentration of sevoflurane inhalation according to SedLine (target PSI: 25-50; 95% spectral edge frequency: 8-12 Hz) , and remifentanil 0.2-0.3 μg/(kg·min) was continuously pumped. Sevoflurane inhalation was stopped 5 min before the end of surgery, and remifentanil infusion was stopped at the end of surgery.

Data collection for both groups was initiated once a stable anesthetic maintenance phase was achieved, characterized by unresponsiveness (Ramsay Sedation Scale score of 5), regular spontaneous breathing, hemodynamic stability (heart rate and blood pressure within 20% of baseline values), and consistent processed EEG indices (PSI 25-50) after induction, and was carried out in a controlled setting to minimize artifact.

EEG acquisition and preprocessing

The standard 19-lead EEG acquisition system was used, and electrodes were placed according to the international 10-20 standard[17]. The reference electrode was set to the bilateral mastoid, the

ground electrode was placed at the frontal midline (FPz), and the impedance between the electrode and the scalp was controlled below 10 kΩ to ensure the signal quality. The sampling frequency was set at 500 Hz with a sampling precision of 16 bits, and all data were collected during a stable period of anesthesia maintenance to ensure signal stationarity and comparability between groups. Meanwhile, the main indicators of anesthesia depth monitor, electrocardiogram and pulse oxygen saturation were recorded simultaneously, which provided an objective basis for maintaining a stable anesthesia state and high-quality EEG signal acquisition.

The EEG preprocessing procedure was as follows: (1) A 0.5–70 Hz finite impulse response bandpass filter was applied to remove baseline drift and high-frequency noise; (2) a 50 Hz notch filter was applied to eliminate power line interference; (3) highly correlated artifacts, including electrooculogram, electromyogram, and electrocardiogram signals, were identified and removed using independent component analysis (ICA); (4) the ICA-corrected data were manually inspected to further exclude any residual artifacts; (5) the data were re-referenced to the global average reference to minimize systematic bias associated with reference electrode selection. Following preprocessing, data quality was assessed by examining the power spectral density to ensure that the results met the standardized requirements for subsequent analysis.

Source space reconstruction and coherence calculation

Source space localization and signal reconstruction. The EEG signal was reversely mapped to the cortical source space using the exact low-resolution brain electromagnetic tomography (eLORETA) for source space reconstruction. Based on the standard MNI152 brain template[18], a forward head model with boundary elements was constructed, including the three-layer structure of scalp, skull and brain, to accurately simulate the characteristics of head volume conductor. Subsequently, the anatomical structure of the source space was restricted to the cortical gray matter, and the cerebral cortex was divided into 68 anatomically clear brain regions according to Desikan-Killiany (D-K) atlas[19], covering 34 in each of the left and right hemispheres. For each brain region, the first principal component of the current density of all source points in the region was extracted by principal component analysis, which was used as a time-series signal representing the neural activity of the brain region.

Coherence calculation and feature extraction. After the source signals of 68 brain regions were successfully reconstructed, they were band-pass filtered at 30-90 Hz to extract the neural oscillatory activity in the gamma band. Subsequently, the continuous source signals were segmented into multiple data segments with a window length of 10 seconds and no overlap, and 40 valid data segments without artifacts were screened for each subject for subsequent analysis. The Welch periodogram method was used to estimate the power spectral density and cross-spectral density of the signal[20], and the amplitude squared coherence was calculated to quantify the frequency synchrony and functional connectivity strength between brain regions. The square coherence can be calculated using the following formula:

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{[S_{xx}(f) \times S_{yy}(f)]}$$

Here, Sxy(f) is the cross-spectral density of brain regions x and y at the frequency f, and Sxx(f) and Syy(f) are the self-spectral densities of the two brain regions, respectively. A 68 × 68 whole-brain functional connectivity matrix was generated for each data segment, and its lower triangular part was extracted as the feature vector. Finally, all features were standardized by Z-score to eliminate dimensional effects.

Statistical Analysis

All data analyses were performed in the Python 3.9 environment (Python Software Foundation, Amsterdam, Netherlands). Kolmogorov-Smirnov test was used to assess the normality of continuous variables for clinical data. Continuous variables with normal distribution were expressed as mean ± standard deviation, and Student's *t* test was used for comparison between groups. Continuous variables that did not follow a normal distribution were presented as medians and interquartile ranges, and comparisons between groups were performed using the Mann-Whitney U test. Categorical variables were presented as n (%) and compared using chi-square and Fisher exact tests.

In this study, within the framework of nested cross-validation, the performance of four gradient boosting ensemble models - light gradient boosting machine (LightGBM), categorical boosting (CatBoost), extreme gradient boosting (XGBoost), and gradient boosting machine (GBM) was systematically evaluated. Additionally, support vector machine (SVM), k-nearest neighbors (KNN), multi-layer perceptron (MLP), deep neural network (DNN), and tree-based models including Random Forest (RF), Decision Tree, AdaBoost, and ExtraTrees were evaluated. A two-stage feature selection strategy combining LASSO regularization with incremental feature selection (IFS) was used to determine the optimal feature subset. Specifically, LASSO was first applied within each training fold to perform initial feature shrinkage, with the optimal regularization parameter (λ) determined via internal 5-fold cross-validation. Then, IFS was employed to identify the optimal feature subset size by incrementally adding features ranked by their LASSO coefficients. DeLong non-parametric test was used to pairwise compare the area under the curve (AUC) of different models with Bonferroni correction applied for multiple comparisons (adjusted $\alpha = 0.05$/number of comparisons) , and calibration curve and decision curve analysis were used to evaluate the clinical applicability. Finally, the overall performance of the model was evaluated by AUC, accuracy, sensitivity, specificity and other multi-dimensional indicators. In order to deeply understand the decision-making mechanism of the model, Shapley additive explanations (SHAP) was introduced to analyze the interpretability of the optimal model, and the contribution of each brain functional connectivity feature to classification prediction was quantified and visualized from the global and local levels.

All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant (or adjusted $P < 0.05$ after Bonferroni correction for multiple comparisons).

To prevent overfitting and ensure optimal model performance, hyperparameter tuning was conducted within the training folds using a Grid Search strategy combined with 5-fold cross-validation. For tree-based models (LightGBM, XGBoost, CatBoost, RF), key hyperparameters including the number of estimators, maximum depth, learning rate, and subsample ratio were optimized. The specific search ranges were: n_estimators [50, 100, 200], max_depth [3, 5, 7, 10], learning_rate [0.01, 0.05, 0.1, 0.2], and subsample [0.6, 0.8, 1.0]. For the SVM model, the penalty parameter (C) and kernel specific parameters (e.g., gamma for RBF kernel) were tuned with search ranges: C [0.1, 1, 10, 100], gamma ['scale', 'auto', 0.001, 0.01, 0.1]. To deeply understand the decision-making mechanisms of the models, Shapley Additive exPlanations (SHAP) analysis was applied, to quantify and visualize the contribution of each brain functional connectivity feature to the classification predictions at both the global and individual levels.

Results

Baseline characteristics

Within the internal derivation cohort (n = 150), the overall average age was 5.92 ± 3.72 years, comprising 57.3% male patients, with an average body weight of 21.93 ± 12.16 kg. To account for developmental neurobiology, the cohort was further stratified into three distinct age groups (≤3.5 years, 3.5–6.5 years, and >6.5 years). Overall, there were no significant differences in baseline demographic and clinical characteristics between the propofol and sevoflurane groups (all $P > 0.05$). Detailed demographic data and statistical comparisons, subdivided by these three age groups and anesthetic type, are presented in **Table 1**. Notably, the overall Patient Status Index (PSI) values during the steady-state maintenance of anesthesia were strictly matched between the propofol and sevoflurane groups (46.50 ± 11.29 vs. 45.49 ± 8.87, $P = 0.543$). This strict matching ensures that the subsequent EEG differences observed are driven by drug-specific mechanisms rather than disparate depths of anesthesia.

Feature selections

Initially, the brain connectivity analysis yielded a 68 × 68 symmetric coherence matrix for each subject. By extracting the unique off-diagonal elements (i.e., the strictly lower triangular part, excluding self-connectivity on the diagonal), a high-dimensional space comprising 2,278 initial connectivity features was generated.

To mitigate the risk of overfitting and strictly prevent data leakage, the feature selection process was embedded within the 5-fold cross-validation loop. In each iteration, the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm was applied to the training set (four folds). Notably, the optimal penalty parameter (λ) for LASSO was automatically determined via internal cross-validation to ensure objective feature shrinkage. The predictive performance of these top-ranked features was then incrementally evaluated on the internal validation set (the remaining fold) across 12 machine learning models (SVM, KNN, MLP, DNN, LightGBM, CatBoost, XGBoost, GBM, RF, DecisionTree, AdaBoost, and ExtraTrees).

The mean Area Under the Curve (AUC) across the cross-validation iterations initially increased rapidly, followed by a plateau. When the number of selected features reached 150, the models achieved relatively high and stable predictive performance. For instance, at this threshold, the SVM model reached a mean AUC of approximately 0.67 in Group 1, 0.723 in Group 2, and 0.90 in Group 3, with minimal performance gains observed upon incorporating additional features.

Given that the feature rankings could vary slightly across the five independent training iterations, a frequency-based stability approach was utilized to finalize the feature subset. The 150 features that most consistently appeared among the top rankings across all five folds were ultimately retained as the optimal, most stable brain connectivity features for subsequent classification modeling and interpretation.

Model construction and performance evaluation

Based on the 150 optimal connectivity features selected via the LASSO and IFS framework, 12 machine learning models were constructed to differentiate the EEG patterns induced by propofol and sevoflurane. Model performance exhibited a pronounced age-dependent progression (**Table 2, Figure 1, Figure 2**). Discrimination improved significantly with advancing age; for instance, the internal validation AUC of the MLP model progressively increased from 0.703 in Group 1 (≤ 42 months), to 0.794 in Group 2 (43–83 months), and reached 0.922 in Group 3 (≥ 84 months). This

age gradient was also starkly reflected in the subject-wise accuracy distribution (**Figure S1**). The youngest cohort exhibited high inter-individual variability with a lower overall median accuracy (**Figure S1A**), whereas the oldest cohort demonstrated concentrated, high-stability classification performance across iterations (**Figure S1C**). Pairwise comparisons using the DeLong test confirmed this age-dependent progression, demonstrating that the AUC of the MLP model in Group 3 was significantly superior to its performance in Group 1 (adjusted $P < 0.001$). Similar trends were observed across multiple algorithms including SVM and tree-based ensembles (**Figure 1A, 1D, 1G**).

Furthermore, cross-sectional model comparisons within the oldest cohort (Group 3) revealed that the MLP model possessed the highest discriminative power, significantly outperforming the SVM model (AUC 0.922 vs. 0.904; DeLong test, adjusted $P < 0.05$).Crucially, quantitative calibration analysis and decision curves revealed a distinct "age gradient effect" in model reliability (**Table 3; Figure 2**). Brier scores universally decreased across advancing age cohorts. In the toddler cohort (Group 1), highly overlapping network responses to anesthetics caused most models to output overconfident probabilities (Brier scores ranging 0.25-0.31). Despite its high AUC, SVM exhibited suboptimal calibration here. Instead, bagging-based ensemble trees (e.g., Random Forest) demonstrated superior calibration (RF Eavg=0.0407); their conservative voting mechanism better reflected the intrinsic neurophysiological overlap and clinical uncertainty inherent to immature brains. Conversely, in the oldest cohort (Group 3), calibration errors dropped substantially (Brier scores 0.13-0.16; **Figure 2J**). Here, highly nonlinear models (MLP, CatBoost) emerged as optimal calibrators (MLP Brier=0.1374), translating to higher net clinical benefit in decision curve analysis (**Figure 2K**). This transition provides robust, data-driven evidence that maturing brain networks develop highly specific pharmacological signatures.

Conversely, in the oldest cohort (Group 3), calibration errors dropped substantially (Brier 0.13-0.16). Here, highly nonlinear models (MLP, CatBoost) emerged as optimal calibrators (MLP Brier=0.1374). This transition provides robust data-driven evidence that maturing brain networks develop highly specific pharmacological signatures, enabling advanced models to execute precise probability fitting with minimal error.

SHAP analysis

To unravel the neurobiological mechanisms differentiating the two anesthetics across developmental stages, we employed SHAP analysis on the optimal models. The feature importance rankings demonstrated a profound age-dependent shift in brain network topology (**Figure 2,Figure 3**). In the G1, the most discriminative features were predominantly concentrated in the interactions between primary sensory (auditory/visual) cortices and the limbic system, such as the connectivity between the right temporal pole and the right pars orbitalis. This connectivity showed the highest mean |SHAP| value of 0.082 (SD = 0.015) across all features. To further elucidate these decision-making mechanisms at the individual level, SHAP waterfall plots were generated(**Figure 3**). These plots demonstrate how specific regional connectivities—mapped to their precise anatomical definitions (**Figure 3D**, **H**, **L**)—drive the model's predictions in typical highly-confident cases versus borderline cases across different age groups. This suggests that early differentiation of these agents hinges on their distinct blockade patterns of basal emotional centers.

As children develop (G2), the key discriminative nodes shifted towards the isthmus of the cingulate and the orbitofrontal cortex. The isthmus cingulate-left lateral occipital connectivity exhibited the highest importance (mean |SHAP| = 0.095, SD = 0.018). The prominence of these regions highlights

that the divergent suppression of the Default Mode Network (DMN) and early self-awareness networks becomes the core pharmacological difference. Finally, in the oldest cohort (Group 3), the discriminative network matured into long-range connectivity, completely dominated by the parahippocampal gyrus (memory network), lateral occipital cortex, and frontal regions (high-level cognition). The parahippocampal gyrus-right frontal pole connectivity was the most discriminative feature (mean |SHAP| = 0.113, SD = 0.021). As illustrated by the individual-level SHAP analysis (**Figure 3I-L**), these specific long-range uncouplings consistently provided the strongest positive and negative contributions to the decision margins in mature brains, contrasting sharply with the sensory-limbic dominance seen in the toddler cohort (**Figure 3A-D**). This indicates that in mature pediatric brains, distinguishing propofol from sevoflurane primarily relies on their distinct disruption of brain-wide long-range communication and high-level cognitive integration.

External validation

Independent external validation further corroborated the age-dependent developmental pattern observed in our internal cohorts (**Table 2**). The older cohorts (Groups 2 and 3) demonstrated robust cross-center consistency. For instance, the SVM model achieved an external AUC of 0.773 in Group 3, confirming that mature neural networks yield highly conserved and specific pharmacological signatures that generalize well across different clinical centers. In contrast, the toddler cohort (Group 1) exhibited expected cross-center variability and degraded generalization (e.g., KNN external AUC = 0.439). This performance drop is likely attributable to the inherent heterogeneity of early developmental brain networks, where high inter-individual variability and overlapping anesthetic responses make cross-center generalization inherently challenging, a phenomenon strongly supported by the highly dispersed subject-level accuracy observed in our internal derivation cohort (**Figure S1A**). The imbalanced sample distribution in the external cohort (Group 1: 8 propofol vs 4 sevoflurane) may have further contributed to this reduced generalizability. This performance drop highlights the intrinsic heterogeneity of early developmental brain networks, where high inter-individual variability and overlapping anesthetic responses make cross-center generalization inherently challenging.

Discussion

In this study, we successfully developed explainable machine learning models based on source-space gamma-band functional connectivity to distinguish the distinct brain network topographies induced by propofol and sevoflurane anesthesia in children. The principal findings demonstrate a pronounced "age gradient effect" in both model discriminative performance and underlying network mechanisms. Furthermore, SHAP analysis revealed an age-dependent evolution of drug-specific pharmacological signatures: the differentiation of the two anesthetics relied primarily on primary sensory-limbic interactions in toddlers, shifted to default mode network (DMN) and orbitofrontal connectivity in preschool children, and ultimately depended on long-range parahippocampal and occipito-frontal networks in older children.

Traditional anesthetic EEG studies mostly rely on power spectral analysis or simple coherence analysis at the sensor level[3,4,5]. However, the EEG signal is deeply affected by volume conduction effects, and it is difficult to truly reflect the interaction of specific anatomical regions of the cortex[10]. Meanwhile, children's skulls are thinner and have large electrical conductivity changes, which further aggravates signal aliasing[21,22]. By utilizing eLORETA source localization, we mitigated these issues. Similar to the recent thinking of Tu et al.[13] and Jin et al.[14] in using ML algorithms to decode the depth of anesthesia or individual sensitivity, this study demonstrated the unique

advantages of advanced ML algorithms (such as MLP, CatBoost, and SVM) in mining high-dimensional, nonlinear topological features that simple temporal or spectral features might miss.

A striking finding of this study is the aforementioned "age gradient effect" observed in both predictive discrimination and model calibration. Quantitative calibration revealed that in the toddler cohort, most algorithms produced overconfident yet erroneous predictions, whereas conservative bagging ensemble models like Random Forest provided the best calibration. This computationally mirrors the clinical reality: immature brain networks lack specific, highly differentiated responses to different anesthetics. However, as the brain matures (Group 3), the calibration error dropped substantially. Highly nonlinear models emerged as optimal calibrators, proving that mature neural networks develop highly specific pharmacological signatures, enabling advanced models to execute precise probability fitting with minimal error.

The neurobiological rationale underlying these age-dependent gamma-band signatures warrants deep consideration. Gamma oscillations (30-90 Hz) are fundamentally generated by the rhythmic inhibition of fast-spiking, parvalbumin-positive (PV+) GABAergic interneurons onto pyramidal cells[27]. Propofol primarily enhances GABA_A receptor-mediated inhibition, leading to a profound, synchronized suppression of cortical networks. Sevoflurane, while also acting on GABA_A receptors, additionally exhibits significant multi-target effects, notably blocking NMDA receptors and two-pore-domain potassium channels. During early brain development (Group 1), the GABAergic system undergoes a functional shift from excitatory to inhibitory, and PV+ interneuronal networks are not yet fully mature[28]. This relative immaturity likely explains the high overlap and poor discriminability (higher Brier scores) observed in toddlers, as neither drug can yet interact with a fully established PV+ interneuron network. As the brain matures into late childhood (Group 3), the cortical excitation/inhibition balance stabilizes. At this stage, propofol's pure GABAergic enhancement and sevoflurane's broader receptor antagonism distinctly alter gamma-band synchronization, particularly manifesting in the specific long-range uncoupling (e.g., parahippocampal and occipito-frontal networks) clearly identified by our SHAP analysis.

Importantly, previous studies on the mechanism of anesthesia mostly focused on power changes in low frequency bands such as α and δ, which are directly related to the depth of anesthesia[5,12,23]. In contrast, gamma oscillations are usually associated with local cortical information processing and the implementation of higher cognitive functions such as feature binding[24]. Focusing on this band allowed us to capture the sophisticated trajectory of cognitive development. For instance, our finding of long-range frontal-occipital uncoupling in older children aligns with adult patterns; this conclusion provides a more sophisticated and complex supplement to Liang et al.[11], who demonstrated that propofol mainly inhibits long-range feedforward connectivity in the adult study. On the other hand, regarding the prominent frontal network changes observed in preschool children (Group 2), Kim et al.[4] have reported that sevoflurane may cause stronger α-wave coherence in the prefrontal lobe than desflurane. Our results confirm that sevoflurane uniquely modulates frontal and orbitofrontal microcircuits. This difference may be due to the fact that sevoflurane, as an inhaled anesthetic, has a wide range of effects on N-methyl-D-aspartic acid receptors and potassium channels in addition to gamma-aminobutyric acid Type A receptors, thus leading to a different change in cortical excitation/inhibition balance than propofol[1,15].

The clinical translational value of these results is significant. When bispectral index (BIS) and patient status index (PSI) are applied in children, the differences in EEG characteristics caused by

different drugs are often ignored[25,26]. Our findings suggest fundamental, age-specific differences in how propofol and sevoflurane reorganize brain functional networks, even when achieving similar clinical depths of anesthesia. Furthermore, these distinct network connectivity patterns may be linked to varying trajectories of postoperative neurocognitive recovery or emergence delirium. Future longitudinal studies could explore whether these age-specific gamma-band features can serve as biomarkers to predict postoperative neurological complications in pediatric patients.

Limitations

Despite the encouraging results of this study, several limitations remain. First, the EEG source localization analysis was based on a standardized MNI head model rather than individualized magnetic resonance imaging (MRI) models, which may affect the absolute spatial precision of the source localization. Second, although source space analysis was utilized, its inherent spatial resolution is constrained by the 19-lead EEG spatial sampling density, making it challenging to fully characterize ultra-fine cortical network activity. Third, the data collected covered only the steady-state phase of anesthesia maintenance, failing to capture the dynamic temporal evolution of brain networks during induction and emergence. Finally, while an independent external validation cohort was employed to demonstrate robustness, the sample size of the external dataset (n = 46) was relatively small and exhibited imbalanced distribution across age groups and anesthetic types, which limited the generalizability testing, particularly for the youngest cohort (Group 1) where inter-individual heterogeneity is highest. Additionally, the retrospective design limits our ability to control for all potential confounding factors, and the relatively short data collection period (January-October 2025) may not capture seasonal variations in clinical practice. Future large-scale, prospective multicenter studies with high-density EEG (e.g., 64 or 128 channels) and individualized head models are necessary to further validate the clinical applicability of these age-adapted monitoring models.

Conclusion

In conclusion, this study demonstrates that source-space gamma-band functional connectivity combined with explainable machine learning can effectively distinguish the distinct brain network topographies induced by propofol and sevoflurane in children. Crucially, we identified a pronounced "age gradient effect," wherein both the predictive performance of the models and the underlying pharmacological signatures evolve significantly with brain maturation. The discriminative functional networks shift from primary sensory-limbic interactions in toddlers to default mode and orbitofrontal connectivity in preschool children, ultimately maturing into long-range parahippocampal and occipito-frontal uncoupling in older children. These findings reveal that anesthetics selectively disrupt actively developing neural hubs at specific developmental stages. This highlights the critical need to move beyond "one-size-fits-all" EEG monitoring, laying a robust theoretical foundation for the development of age-adapted, drug-specific precision monitoring systems in pediatric anesthesia.

References

[1] Rigouzzo A, Khoy-Ear L, Laude D, Louvet N, Moutard ML, Sabourdin N, et al. EEG profiles during general anesthesia in children: A comparative study between sevoflurane and propofol. Paediatr Anaesth. 2019; 29(3): 250-257.

[2] Yuan I, Bong CL, Chao JY. Intraoperative pediatric electroencephalography monitoring: an updated review. Korean J Anesthesiol. 2024; 77(3):289-305.

[3] Akeju O, Westover MB, Pavone KJ, Sampson AL, Hartnack KE, Brown EN, et al. Effects of sevoflurane and propofol on frontal electroencephalogram power and coherence. Anesthesiology. 2014; 121(5):990-8.

[4] Kim YS, Kim J, Park S, Kim KN, Ha Y, Yi S, et al. Differential effects of sevoflurane and desflurane on frontal intraoperative electroencephalogram dynamics associated with postoperative delirium. J Clin Anesth. 2024; 93:111368.

[5] Cornelissen L, Kim SE, Purdon PL, Brown EN, Berde CB. Age-dependent electroencephalogram (EEG) patterns during sevoflurane general anesthesia in infants. Elife. 2015; 4: e06513.

[6] Zhang Y, Liang ZH, Wang X, Zhang N, Zhu HT, Wang DX, et al. Age-Dependent Entropic Features During Propofol Anesthesia in Developing Brain. Anesth Analg. 2025.

[7] Ferree T C, Eriksen K J, Tucker D M. Regional head tissue conductivity estimation for improved EEG analysis. IEEE Transactions on Biomedical Engineering. 2002; 47(12): 1584-1592.

[8] Gevins A, Le J, Brickett P, et al. Seeing through the skull: advanced EEGs use MRIs to accurately measure cortical activity from the scalp. Brain Topography. 1991; 4(2): 125-131.

[9] Odabaee M, Tokariev A, Layeghy S, et al. Neonatal EEG at scalp is focal and implies high skull conductivity in realistic neonatal head models. Neuroimage. 2014; 96: 73-80.

[10] Zhao X, Wang Y, Zhang Y, Wang H, Ren J, Yan F, et al. Propofol-Induced Anesthesia Alters Corticocortical Functional Connectivity in the Human Brain: An EEG Source Space Analysis. Neurosci Bull. 202; 37(6):563-568.

[11] Liang Z, Chang Y, Liu X, Cao S, Chen Y, Wang T, et al. Changes in information integration and brain networks during propofol-, dexmedetomidine-, and ketamine-induced unresponsiveness. Br J Anaesth. 2024; 132(3):528-540.

[12] Desowska A, Berde CB, Cornelissen L. Emerging functional connectivity patterns during sevoflurane anaesthesia in the developing human brain. Br J Anaesth. 2023; 130(2):e381-e390.

[13] Tu Z, Zhang Y, Lv X, Wang Y, Zhang T, Wang J, et al. Accurate Machine Learning-based Monitoring of Anesthesia Depth with EEG Recording. Neurosci Bull. 2025; 41(3):449-460.

[14] Jin X, Liang Z, Li F, Li X. Evaluating individual sensitivity to propofol through EEG complexity and information integration: from neural dynamics to precision anesthesia. J Neural Eng. 2025; 22(3).

[15] Kumar R, Skowno J, von Ungern-Sternberg BS, Davidson A, Xu T, Zhang J, et al. Quantitative electroencephalogram and machine learning to predict expired sevoflurane concentration in infants. J Clin Monit Comput. 2025; 39(5):999-1014.

[16] Kamata K, Lipping T, Yli-Hankala A, Jäntti V, Yamauchi M. Spurious electroencephalographic activity due to pulsation artifact in the depth of anesthesia monitor. JA Clin Rep. 2021; 7(1):35.

[17] Herwig U, Satrapi P, Schönfeldt-Lecuona C. Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. Brain Topogr. 2003; 16(2):95-9.

[18] Ota K, Nakazato Y, Oyama G. Developing and validating an elderly brain template: A

comprehensive comparison with MNI152 for age-specific neuroimaging analyses. Neuroimage. 2025; 320:121473.
[19] Jao CW, Lau CI, Lien LM, Tsai YF, Chu KE, Hsiao CY, et al. Using Fractal Dimension Analysis with the Desikan-Killiany Atlas to Assess the Effects of Normal Aging on Subregional Cortex Alterations in Adulthood. Brain Sci. 2021; 11(1):107.
[20] Gallet C, Julien C. The significance threshold for coherence when using the Welch's periodogram method: Effect of overlapping segments. Biomedical Signal Processing and Control. 2011; 6(4): 405-409.
[21] Odabaee M, Tokariev A, Layeghy S, Mesbah M, Colditz PB, Ramon C, et al. Neonatal EEG at scalp is focal and implies high skull conductivity in realistic neonatal head models. Neuroimage. 2014; 96:73-80.
[22] Cornelissen L, Bergin AM, Lobo K, Donado C, Soul JS, Berde CB. Electroencephalographic discontinuity during sevoflurane anesthesia in infants and children. Paediatr Anaesth. 2017; 27(3):251-262.
[23] Purdon PL, Sampson A, Pavone KJ, Brown EN. Clinical Electroencephalography for Anesthesiologists: Part I: Background and Basic Signatures. Anesthesiology. 2015; 123(4):937-60.
[24] Liu C, Han T, Xu Z, Liu J, Zhang M, Du J, et al. Modulating Gamma Oscillations Promotes Brain Connectivity to Improve Cognitive Impairment. Cereb Cortex. 2022; 32(12):2644-2656.
[25] Davidson AJ, Skowno J. Neuromonitoring in paediatric anaesthesia. Current Opinion in Anaesthesiology. 2019; 32(3):370-376.
[26] de Heer IJ, Bouman SJM, Weber F. Electroencephalographic (EEG) density spectral array monitoring in children during sevoflurane anaesthesia: a prospective observational study. Anaesthesia. 2019; 74(1):45-50.
[27] Buzsáki, G., & Wang, X. J. (2012). Mechanisms of gamma oscillations. Annual review of neuroscience, 35, 203-225.
[28] Ben-Ari, Y. (2002). Excitatory actions of GABA during development: the nature of the nurture. Nature reviews neuroscience, 3(9), 728-739.

Table 1. Demographic and Perioperative Characteristics of the Study Population

Characteristic	Internal cohort		External validation cohort	
	Propofol(n=75)	Sevoflurane(n =75)	Propofol(n=21)	Sevoflurane(n=25)
Sex, No. (%)				
Male	35 (46.7)	42 (56.0)	11 (52.3)	13 (52.0)
Female	40 (53.3)	33 (44.0)	10 (47.7)	12 (48.0)
Age, mean (SD), y	4.98 (3.46)	4.37 (3.72)	5.65 (3.61)	5.12 (3.28)
Group 1				
No.	25	25	4	14
Male, No. (%)	13 (52.0)	15 (60.0)	2 (50.0)	8 (57.1)
Age, mean (SD), y	1.65 (0.62)	1.53 (0.81)	1.91 (0.90)	1.95 (0.91)
Weight, mean (SD), kg	11.14 (2.19)	10.73 (2.2)	11.88 (2.30)	12.02 (2.73)
At the start of surgery, mean (SD)				
Heart rate, bpm	92.5 (12.1)	114.06 (18.21)	100.8 (18.9)	120.38 (13.90)
Systolic blood pressure, mmHg	84.3 (9.6)	81.16 (10.03)	89.7 (11.4)	87.47 (14.25)
Diastolic blood pressure, mmHg	38.1 (8.1)	39.98 (6.4)	45.3 (10.7)	42.19 (11.66)
PSI	47.8 (12.4)	51.18 (10.42)	58.3 (11.2)	53.75 (9.79)
SEF-L, Hz	18.2 (3.3)	20.63 (3.18)	21.2 (3.3)	20.96 (2.01)
SEF-R, Hz	18.3 (3.7)	20.76 (2.78)	20.0 (3.9)	21.20 (2.14)
At extubation				
Heart rate, bpm	140.7 (18.6)	143.64 (24.51)	130.3 (22.7)	143.37 (17.97)
Respiratory rate, breaths/min	26.9 (7.8)	22.43 (5.37)	22.7 (6.5)	20.93 (3.94)
Systolic blood pressure, mmHg	97.1 (11.4)	97.56 (14.10)	97.6 (12.1)	98.67 (15.33)
Diastolic blood pressure, mmHg	47.2 (9.4)	53.62 (12.84)	51.6 (12.7)	54.41 (15.14)
PSI	74.3 (11.2)	60.79 (18.79)	74.9 (11.8)	62.11 (17.99)
SEF-L, Hz	14.7 (8.8)	12.98 (7.23)	11.2 (5.6)	15.06 (5.11)
SEF-R, Hz	14.5 (8.2)	12.05 (7.54)	12.3 (5.3)	13.92 (5.67)
Time from propofol discontinuation to extubation, mean (SD), min	35.0 (14.2)	26.89 (8.35)	29.2 (12.1)	25.56 (8.88)
Propofol Ce/Sevoflurane MAC at extubation, mean (SD), μg/mL	0.45 (0.35)	0.036 (0.048)	0.51 (0.26)	0.04 (0.05)
Surgical duration, mean (SD), min	61.2 (58.7)	41.38 (34.50))	43.2 (37.7)	36.66 (29.46)
Anesthesia duration, mean (SD), min	112.8 (62.6)	80.45(36.10)	92.3 (38.9)	83.00 (37.05)
Group 2				
No.	25	25	9	7
Male, No. (%)	14(56.0)	13 (52.0)	5 (55.6)	3 (42.9)
Age, mean (SD), y	5.12 (1.19)	5.67 (0.97)	5.51 (1.10)	5.13 (0.93)
Weight, mean (SD), kg	19.45 (4.55)	19.81 (4.06)	21.38 (5.13)	19.76 (4.35)
At the start of surgery, mean (SD)				
Heart rate, bpm	78.3 (15.4)	99.94 (23.34)	84.9 (12.4)	108.32 (25.91)
Systolic blood pressure, mmHg	90.1 (11.5)	92.44 (8.52)	91.8 (10.4)	90.04 (8.77)

Diastolic blood pressure, mmHg	50.9 (12.2)	48.41 (8.54)	50.5 (12.3)	47.48 (7.46)
PSI	48.1 (8.4)	42.38 (6.65)	53.0 (11.4)	47.33 (12.93)
SEF-L, Hz	17.8 (4.1)	18.40 (3.11)	19.3 (4.2)	19.74 (3.96)
SEF-R, Hz	18.3 (4.1)	18.76 (2.57)	20.5 (5.0)	20.60 (2.86)
At extubation				
Heart rate, bpm	100.2 (19.6)	124.84 (21.33)	107.2 (25.8)	134.64 (16.90)
Respiratory rate, breaths/min	21.0 (5.7)	19.88 (6.30)	18.3 (6.8)	18.56 (4.51)
Systolic blood pressure, mmHg	99.9 (11.6)	103.53 (10.62)	98.7 (11.0)	99.40 (11.69)
Diastolic blood pressure, mmHg	53.9 (10.0)	61.13 (10.94)	56.4 (11.5)	58.64 (14.14)
PSI	69.2 (15.8)	66.44 (16.89)	71.5 (12.2)	66.92 (20.41)
SEF-L, Hz	14.0 (7.3)	11.23 (6.44)	15.2 (9.0)	12.84 (5.29)
SEF-R, Hz	12.9 (7.1)	11.68 (7.21)	13.6 (8.0)	12.38 (5.65)
Time from propofol discontinuation to extubation, mean (SD), min	26.5 (8.2)	22.50 (6.39)	29.1 (12.7)	25.32 (7.81)
Propofol Ce/Sevoflurane MAC at extubation, mean (SD), μg/mL	0.69 (0.20)	0.07 (0.05)	0.76 (0.29)	0.052 (0.051)
Surgical duration, mean (SD), min	30.7 (24.3)	27.81 (16.60)	40.5 (28.4)	29.56 (24.90)
Anesthesia duration, mean (SD), min	76.4 (29.6)	65.03 (22.58)	85.7 (31.7)	73.84 (29.55)
Group 3				
No.	25	25	8	4
Male, No. (%)	16 (64.0)	15 (60.0)	4 (50.0)	2 (50.0)
Age, mean (SD), y	9.39 (1.69)	10.15 (2.17)	9.39 (1.69)	9.85 (3.11)
Weight, mean (SD), kg	30.73 (10.71)	35.32 (12.01)	30.73 (10.71)	34.34 (13.59)
At the start of surgery, mean (SD)				
Heart rate, bpm	69.4 (9.2)	91.13 (17.61)	76.4 (11.6)	99.50 (14.91)
Systolic blood pressure, mmHg	95.7 (16.1)	89.93 (9.73)	99.5 (11.4)	91.93 (9.28)
Diastolic blood pressure, mmHg	49.6 (8.0)	48.88 (8.08)	51.5 (10.0)	48.43 (9.60)
PSI	45.9 (6.8)	43.73 (7.54)	49.2 (15.1)	46.39 (8.14)
SEF-L, Hz	17.2 (3.2)	18.28(2.80)	18.0 (5.8)	19.59 (3.39)
SEF-R, Hz	17.1 (3.0)	18.40 (2.97)	18.4 (5.4)	19.98 (3.20)
At extubation				
Heart rate, bpm	91.2 (23.5)	104.36 (18.66)	99.7 (17.6)	113.43 (24.66)
Respiratory rate, breaths/min	17.9 (4.7)	16.89 (4.47)	18.2 (3.8)	18.85 (5.51)
Systolic blood pressure, mmHg	105.2 (14.8)	110.23 (14.19)	108.7 (12.1)	110.29 (15.40)
Diastolic blood pressure, mmHg	61.3 (14.6)	71.02 (12.46)	62.0 (12.0)	65.82 (10.63)
PSI	67.5 (16.8)	75.71 (11.05)	71.0 (15.4)	73.11 (11.61)
SEF-L, Hz	16.8 (7.4)	16.10 (7.39)	14.2 (7.7)	13.89 (8.41)
SEF-R, Hz	17.1 (7.0)	16.52 (7.38)	16.1 (4.9)	12.88 (8.40)
Time from propofol discontinuation to extubation, mean (SD), min	18.9 (5.5)	22.79 (5.68)	20.5 (10.7)	23.25 (6.70)
Propofol Ce/Sevoflurane MAC at extubation, mean (SD), μg/mL	1.06 (0.28)	0.086 (0.044)	1.00 (0.33)	0.08 (0.07)
Surgical duration, mean (SD), min	40.5 (22.7)	41.11 (28.96)	42.7 (33.5)	56.00 (42.92)
Anesthesia duration, mean (SD), min	78.7 (30.8)	75.40 (30.04)	83.4 (38.0)	94.46 (47.23)

Table 2. Predictive performance of machine learning models for classifying anesthetic states across age cohorts and validation datasets

Note: Values are presented as mean (95% confidence interval) for AUC. Accuracy, sensitivity, and specificity are reported as percentages. Bold values indicate the best performance within each age group and validation type. AUC = area under the curve; SVM = support vector machine; KNN = k-nearest neighbors; MLP = multi-layer perceptron; CatBoost = categorical boosting.

Models	Group	Datasets	AUC(95% CI)	Accuracy	Sensitivity	Specificity
		Training	0.889(0.859-0.901)	83.6%	81.1%	84.9%
	1	Internal validation	0.677(0.651-0.712)	64.3%	60.6%	**67.9%**
		External validation	**0.729(0.666-0.787)**	**64.6%**	73.0%	56.9%
		Training	0.852(0.873-0.835)	82.5%	80.6%	81.8%
SVM	2	Internal validation	0.724(0.704-0.752)	70.8%	**67.7%**	73.8%
		External validation	0.712(0.654-0.767)	60.6%	**75.6%**	95.6%
		Training	0.968(0.944-0.983)	92.3%	89.7%	90.6%
	3	Internal validation	0.904(0.891-0.905)	84.4%	80.5%	88.1%
		External validation	0.773(0.722-0.819)	**70.9%**	64.4%	77.5%
		Training	0.821(0.809-0.848)	79.5%	74.7%	81.2%
	1	Internal validation	**0.704(0.681-0.717)**	63.1%	**61.4%**	64.8%
		External validation	0.535(0.471-0.599)	55.8%	48.0%	**63.1%**
		Training	0.895(0.921-0.875)	86.2%	87.9%	90.8%
CatBoost	2	Internal validation	0.799(0.779-0.819)	70.8%	65.2%	76.4%
		External validation	**0.765(0.710-0.821)**	**67.5%**	35.6%	**95.4%**
		Training	0.938(0.915-0.951)	89.3%	87.1%	91.6%
	3	Internal validation	0.894(0.874-0.904)	**85.3%**	**83.3%**	87.2%
		External validation	0.745(0.690-0.797)	67.2%	62.5%	71.9%
		Training	0.754(0.738-0.799)	72.6%	70.6%	75.7%
	1	Internal validation	0.672(0.643-0.691)	62.5%	57.4%	67.6%
		External validation	0.439(0.376-0.503)	44.5%	48.0%	41.2%
		Training	0.821(0.804-0.856)	78.6%	78.2%	80.5%
KNN	2	Internal validation	0.733(0.708-0.764)	65.0%	60.6%	69.3%
		External validation	0.752(0.700-0.801)	63.4%	30.0%	96.9%
		Training	0.904(0.894-0.916)	89.1%	86.3%	89.0%
	3	Internal validation	0.801(0.783-0.819)	75.1%	73.1%	77.0%
		External validation	0.760(0.710-0.810)	66.9%	**76.9%**	56.9%
		Training	0.833(0.815-0.853)	80.4%	76.9%	79.9%
	1	Internal validation	0.703(0.680-0.723)	**64.4%**	61.0%	67.7%
		External validation	0.664(0.605-0.728)	61.7%	**83.1%**	41.9%
MLP		Training	0.917(0.893-0.928)	88.7%	89.2%	91.7%
	2	Internal validation	**0.794(0.770-0.806)**	**73.0%**	67.0%	**78.9%**
		External validation	0.699(0.639-0.756)	62.5%	35.0%	90.0%
	3	Training	0.979(0.966-0.986)	90.9%	95.6%	94.8%

	Internal validation	**0.922(0.901-0.932)**	84.6%	80.3%	**88.7%**
	External validation	0.691(0.630-0.748)	69.4%	63.1%	**75.6%**

Table 3. Quantitative Calibration Metrics of Machine Learning Models

Note: Brier score ranges from 0 (perfect calibration) to 1 (worst calibration). Emax = maximum calibration error; Eavg = average calibration error. Lower values indicate better calibration. RF = Random Forest; GBM = gradient boosting machine; DNN = deep neural network; MLP = wide multi-layer perceptron.

Age Cohert	Model	Berier Score ↓	Emax ↓	Eavg ↓
Group1	SVM	0.3005	0.3166	0.1914
	CatBoost	0.2615	0.2735	0.1552
	LightGBM	0.2818	0.3509	0.2012
	RF	0.2192	0.0802	0.0407
	XGBoost	0.2841	0.3468	0.2122
	GBM	0.2448	0.2531	0.1395
	KNN	0.2725	0.2822	0.1895
	DecisionTree	0.3495	0.4524	0.2550
	AdaBoost	0.2349	0.1401	0.0960
	ExtraTrees	0.2186	0.0857	0.0503
	DNN	0.3197	0.3320	0.1975
	MLP	0.3133	0.3226	0.2056
Group2	SVM	0.2691	0.2797	0.1473
	CatBoost	0.1994	0.2456	0.1346
	LightGBM	0.2352	0.3343	0.2049
	RF	0.1839	0.1463	0.0539
	XGBoost	0.2331	0.3327	0.1804
	GBM	0.2017	0.2342	0.1304
	KNN	0.2372	0.2542	0.1523
	DecisionTree	0.3392	0.3871	0.1981
	AdaBoost	0.2162	0.2867	0.1747
	ExtraTrees	0.1853	0.1206	0.0560
	DNN	0.2721	0.3242	0.1997
	MLP	0.2612	0.2856	0.1905
Group3	SVM	0.1505	0.3000	0.1206
	CatBoost	0.1422	0.1561	0.0624
	LightGBM	0.1549	0.2863	0.1234
	RF	0.1593	0.1802	0.0959
	XGBoost	0.1557	0.2036	0.1197
	GBM	0.1501	0.1550	0.0739
	KNN	0.1910	0.1410	0.0790
	DecisionTree	0.2618	0.3511	0.1937
	AdaBoost	0.2041	0.3274	0.2316
	ExtraTrees	0.1606	0.1788	0.1084

DNN	0.1490	0.3329	0.1631
MLP	0.1374	0.3253	0.1247

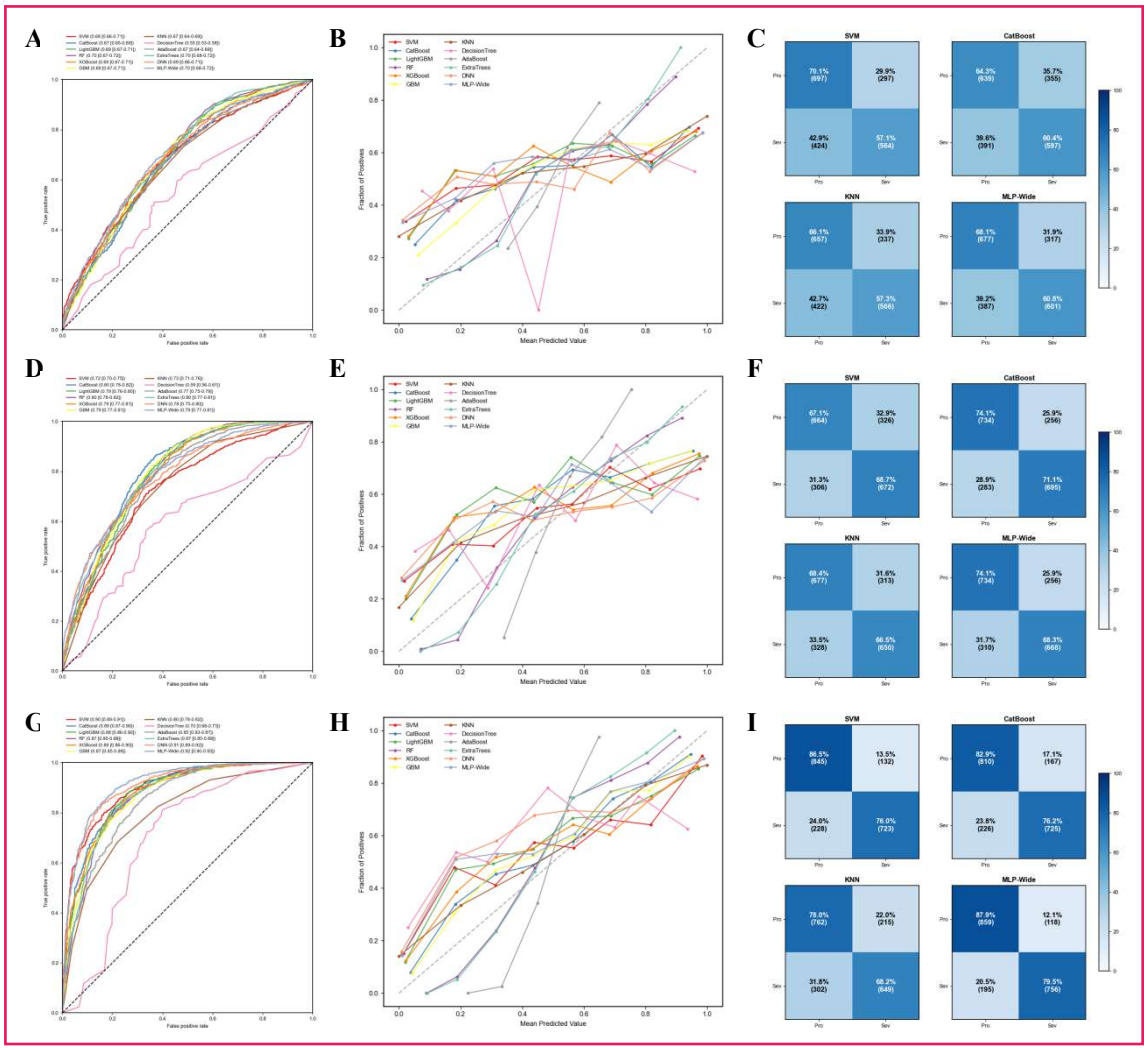


Figure 1. Performance and clinical utility of machine learning models across pediatric age groups. Rows represent Age Group 1 (≤ 3y 6m), Group 2 (> 3y 6m to ≤ 6y 7m), and Group 3 (> 6y 7m). Columns display (from left to right): ROC curves, Calibration plots, and Confusion matrices. (A, D, G) ROC curves for SVM, MLP, CatBoost, and KNN models. AUC values with 95% CIs are provided, indicating discriminative power. (B, E, H) Calibration plots showing predicted vs. observed probabilities; the dashed diagonal represents perfect agreement. (C, F, I) Confusion matrices summarizing classification rates for Propofol (Pro) and Sevoflurane (Sev).

AUC=area under the curve. CatBoost=Categorical Boosting. KNN=k-nearest neighbors. MLP=multi-layer perceptron. ROC=receiver operating characteristic. SVM=support vector machine.

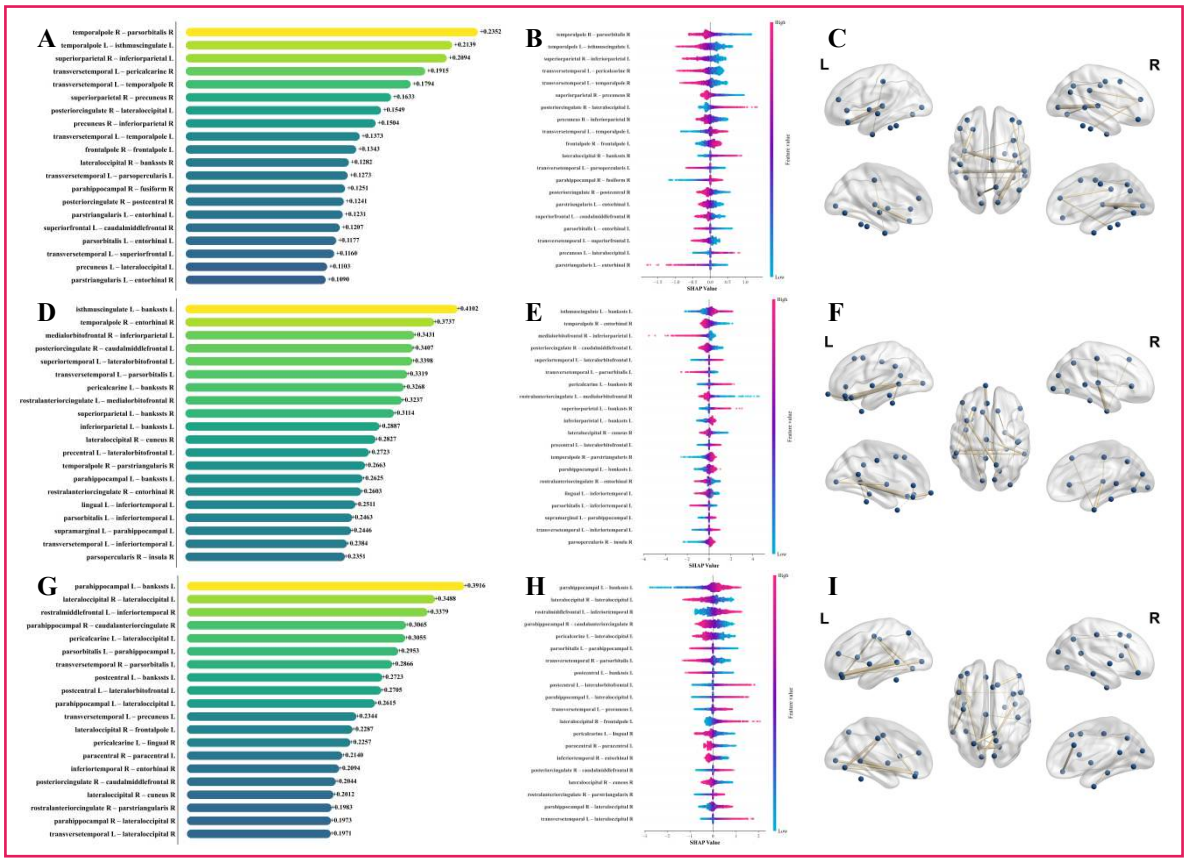


Figure 2. Comprehensive evaluation of model performance across three pediatric age cohorts. Panels are organized by age group: Group 1 (A – D), Group 2 (E – H), and Group 3 (I – L). (A, E, I) ROC curves illustrating discriminative power for SVM, MLP, CatBoost, and KNN models; AUC values with 95% CIs are provided. (B, F, J) Calibration plots comparing predicted versus observed probabilities; the dashed diagonal represents perfect calibration. (C, G, K) Decision curves evaluating clinical net benefit compared to "Treat All" and "Treat None" benchmarks. (D, H, L) Confusion matrices summarizing classification accuracy for Propofol (Pro) and Sevoflurane (Sev). AUC=area under the curve. DCA=decision curve analysis. ROC=receiver operating characteristic.

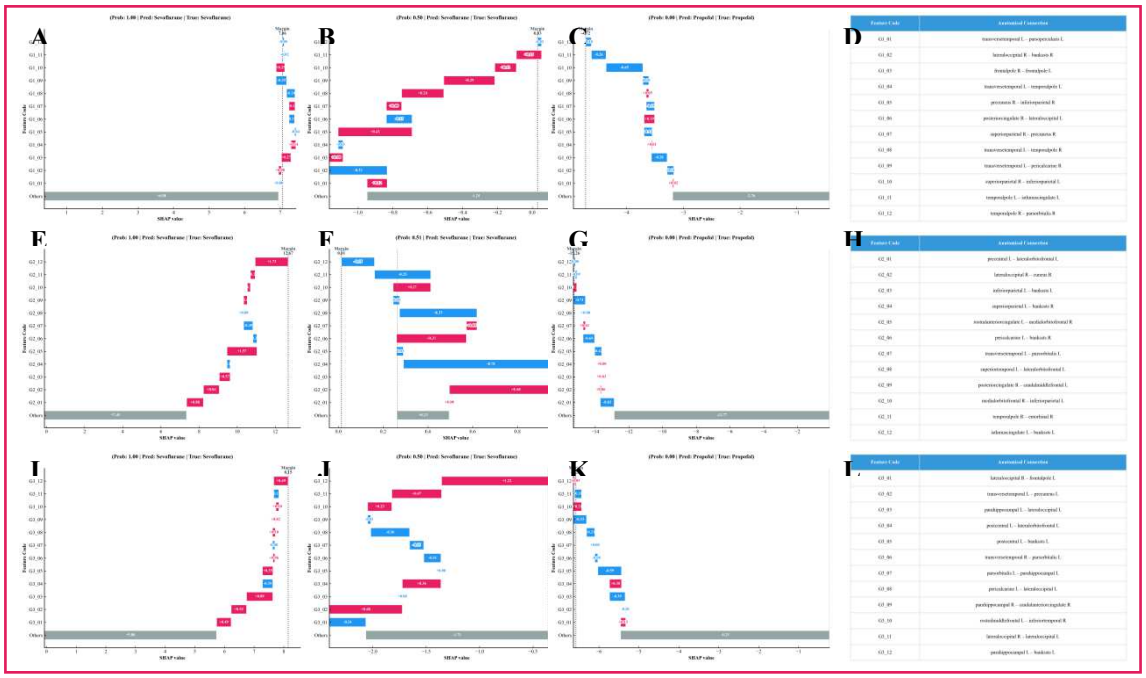


Figure 3. Individual-level SHAP waterfall analysis and feature-to-anatomy mapping.
Panels follow age stratification: Group 1 (A – D), Group 2 (E – H), and Group 3 (I – L).
Waterfall plots (A, E, I) display typical Sevoflurane (Sev) cases correctly classified with high confidence (P≈1.00). Panels (B, F, J) show borderline cases (P≈1.00) with conflicting feature impacts. Panels (C, G, K) illustrate typical Propofol (Pro) cases (P≈0.50). Red and blue bars indicate positive and negative contributions to the SVM decision margin, respectively. (D, H, L) Mapping tables provide anatomical definitions for the specific feature codes (e.g., G1_01) used in each corresponding age group.Pro=Propofol. Sev=Sevoflurane. SHAP=Shapley Additive Explanations.

Supplemental Materials

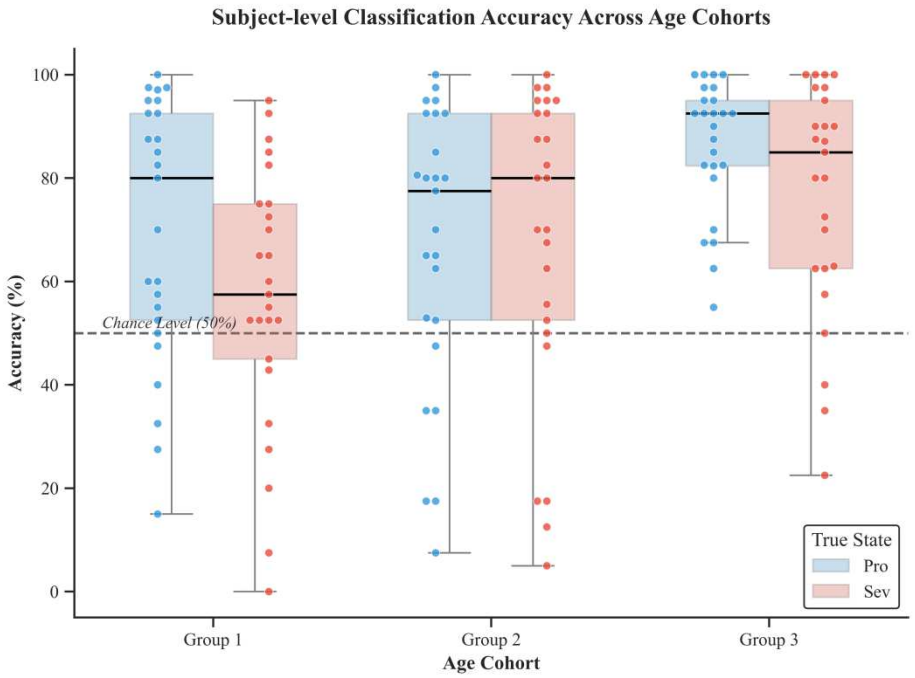


Figure S1. Subject-wise accuracy distribution within the internal inference set across three pediatric age cohorts. Panels correspond to specific age groups: (A) Group 1 (≤3y 6m), (B) Group 2 (>3y 6m, ≤6y 7m), and (C) Group 3 (>6y 7m). The x-axis represents individual subject IDs within each respective cohort. The y-axis indicates the classification accuracy. For each subject, a box plot displays the distribution of accuracy across multiple validation iterations (e.g., k-fold cross-validation runs), showing the median (central line), interquartile range (IQR, box), and potential outliers (points). This panel provides a granular view of the performance, highlighting the high variability and lower overall median accuracy in the youngest cohort (A), contrasted with the increased stability and highest median performance in the oldest cohort (C), reinforcing the developmental 'age gradient' effect.