Research Article

Neuro Audit: An Interpretable Framework For EEG Privacy Risk Assessment and Empirical Neural Identity-Inference Analysis

Nithyasree KC[1], Abhinayaa S[2], Deepika B[3], Lakshmipriya V[4], Vishnu M[5], Jaganathan M[6], Priyadarshan S[7]

[1–7] Department of IOT, Cybersecurity with Blockchain Technology, Manakula Vinayagar Institute of Technology, Puducherry, India.

Email: Nithyasree KC nithyaasreekcr@gmail.com

Abstract

Electroencephalography (EEG) is more and more being recorded outside of the clinical setting but neural recordings can provide information beyond the purpose for which they are recorded and there are few tools to audit them. NeuroAudit, a web-based platform for EEG privacy and neurosecurity auditing, preprocesses the uploaded recordings using MNE-Python, extracts 19 interpretable features and generates a rule-based heuristic score across four dimensions (Identity, Emotion, Stress and Workload). It then generates rule-based privacy and security recommendations and generates an automated PDF audit report, while treating the raw EEG as a temporary input that can be deleted after processing. A separate research part examines empirical identity-inference vulnerability on the public PhysioNet auditory EEG-biometric dataset (20 subjects, 239 recordings, 5,764 five-second epochs) using Dummy, Logistic Regression and Random Forest models under a recording-disjoint five-fold GroupKFold protocol. The Random Forest reached 64.35% balanced accuracy and 63.82% macro-F1 against a 5% chance level, and 87.03% accuracy after recording-level majority voting. Excluding the heuristic-related feature group reduced the balanced accuracy to 56.49%. However, the heuristic Identity Risk score does not show a clear correlation with recording-level inference success (point-biserial correlation 0.1016, p = 0.1172; bootstrap 95% CI -0.0107 to 0.2098). NeuroAudit is therefore presented as an interpretable auditing baseline and not a trustworthy indicator of privacy vulnerability. We therefore find that calibration of the risk score should be the next direction for research.

Keywords EEG privacy · Neural data · Neurosecurity · Identity inference · Privacy risk assessment · Brain–computer interface.

1 Introduction

Electroencephalography (EEG) measures electrical activity generated by neural populations and is now used well beyond clinical and laboratory settings. Because low-cost wearable headsets and brain-computer interface (BCI) platforms are available for cognitive-state monitoring, education, gaming, wellness and human-computer interaction (Tajdini et al. 2020; Genser et al. 2024), neural recordings are being collected, stored and shared by organisations that don't operate neuroscience laboratories.

EEG is unusual among sensor modalities because a single recording can be used for many inferential applications. Spectral and temporal characteristics have been linked to attention, workload, affective state, and stable individual traits (Chikhi et al. 2022; Zhou et al. 2022; Sabu et al. 2022; Tran et al. 2026), and individual differences in EEG have been used for biometric identification (Maiorana 2020; Wilaiprasitporn et al. 2020; Gómez-Tapia et al. 2022). Recent consumer BCI devices have also shown that untrusted applications can infer private information from neural signals (Landau et al. 2020; Bernal et al. 2022). A recording collected for one purpose might therefore support inferences that the data subject did not anticipate, which has led to calls for neural-data protection and neurorights (Ienca 2021; Genser et al. 2024; UNESCO 2025).

Tooling has failed to keep up with this concern. Current EEG software is concerned with signal analysis, while privacy assessment, security controls and empirical inference testing are normally discussed separately (if at all). Not only does there not seem to be any workflow for a data custodian who wants to know whether a recording is privacy-sensitive and why, but no workflow for the information manager to know if a recording is a crime or not. In such a workflow, two properties are desirable: the assessment should be interpretable so that an auditor can trace a score to the measurable signal characteristics of a given score and that the assessment is evaluated with empirical evidence of what a learning-based adversary can actually infer. These two properties are not mutually exclusive and, as this paper demonstrates, should not be conflated.

NeuroAudit is a web-based platform for EEG privacy and neurosecurity auditing that integrates MNE-Python preprocessing, a 19-feature interpretable representation, a rule-based heuristic risk model for Identity, Emotion, Stress and Workload, privacy and security recommendations based on rule-based privacy and security recommendations, automated PDF audit reporting, and data handling design with raw EEG as a short-term input to be analyzed and not retained in data handling. In addition, a research component in NeuroAudit is a separate research project to evaluate the empirical identity-inference vulnerability on a public EEG-biometric dataset and whether this heuristic score is equivalent to this empirical weakness. The study is organized around three main

components: (i)the production heuristic risk assessment, (ii) an empirical identity-inference attack experiment, and (iii) the relationship between the two. We address the following questions in the paper:

1.RQ1. Can an EEG processing platform offer an interpretable, multidimensional privacy-risk assessment with practical guidance while minimizing retention of raw neural data?

2. RQ2. How well do the EEG characteristics used by the platform allow for subject identity inference under a recording-disjoint evaluation protocol?

3. RQ3. Do the features used by the heuristic assessment contribute to empirical identity inference?

4. RQ4. Is the heuristic Identity Risk score associated with empirical success in identity-inference?

The main contributions of this work are:

1. An integrated EEG privacy-auditing platform that combines signal processing, privacy-risk assessment, recommendations and automated reporting.

2. An interpretable heuristic baseline for multidimensional EEG privacy-risk assessment.

3. A recording-disjoint machine-learning identity-inference evaluation using 19 EEG-derived features.

4. A feature-ablation analysis of the role of heuristic-related features to empirical identity inference.

5. An empirical analysis of the relationship between heuristic Identity Risk and actual inference success, with a negative result that limits the claims that can be made for the heuristic.

6. A privacy-aware architecture in which raw EEG is handled as temporary sensitive input. The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the NeuroAudit platform. Section 4 describes the empirical identity-inference experiment and Section 5 presents the results. Section 6 describes what the results do and do not show, Section 7 presents limitations and future work, and Section 8 concludes.

2. Related work

2.1 EEG as a source of sensitive information.

EEG rhythms have been linked to cognitive workload, attention and affective processing. Theta power tends to increase and alpha power decrease with cognitive workload (Chikhi et al. 2022; Zhou et al. 2022), and frontal alpha asymmetry has been extensively studied as a correlate of affective style and motivation (Davidson 2004; Sabu et al. 2022). Engagement indices based on band power ratios have been used to adjust automation to the operator state (Pope et al. 1995) and differential entropy and band power features are used in EEG-based emotion recognition (Duan et al. 2013; Patel et al. 2021), a field that has been recently recognized as moving towards deep learning models (Wang et al. 2023). The same characteristics that make EEG useful for these purposes can cause privacy issues when recordings are reused for unintended inference.

2.2 EEG biometrics and inference attacks.

Person authentication and identification from EEG has been shown many times and more recently by deep neural networks (Maiorana 2020; Wilaiprasitporn et al. 2020; Gómez-Tapia et al. 2022). The leading contributors are the spatial configuration, spectral content and temporal dynamics, and some spectral properties such as the α peak frequency show considerable individual stability (Tran et al. 2026). On the attack side, surveys of BCI security catalogue attacks that recover private information from consumer BCI signals (Bernal et al. 2022), and Landau et al. (2020) have predicted personality and cognitive measures from resting-state EEG. However, the results obtained in controlled laboratory settings should not be interpreted as any real-world identification capability: the recording hardware, montage, session effects, population diversity and preprocessing all have an impact on the results (Maiorana 2020; Tran et al. 2026). NeuroAudit thus considers empirical subject discrimination as evidence to be measured under a set of defined protocol and not a probability of identity disclosure.

2.3 Privacy and security of neural data.

Solutions to these risks include technical responses to attacks on BCIs (Bernal et al. 2022), rights-based frameworks for mental privacy (Ienca 2021; UNESCO 2025) and benchmarking of company privacy practices with respect to data protection standards (Genser et al. 2024). In practice, these measures translate into access control, encryption in transit and at rest, data minimisation, limited retention, pseudonymisation and controlled sharing. The technical work on privacy-preserving BCIs has focused on federated learning, differential privacy and secure aggregation of EEG data (Paul and Bajwa 2025) and has been thoroughly reviewed by Xia et al. (2023). These approaches are mostly concerned with how to handle neural data safely in the event of a threat. They don't really help a custodian determine which recordings or features need such protection.

2.4 Research gap.

As such, existing work tends to separate EEG signal processing, privacy analysis, security controls and inference vulnerability analysis. NeuroAudit addresses this gap in an integrated auditing workflow where interpretable features provide an explicit risk model, provide actionable guidance and produce a report, and an independent machine

learning experiment measures what an adversary can actually infer. The contribution is presented as an auditing workflow and an honest empirical check on it. It is not presented as a solution to neural privacy in general. neural privacy in general.

3 The NeuroAudit platform

3.1 System architecture

NeuroAudit is a web application with a React, Vite and TypeScript frontend on Vercel and a Python Flask backend on Render. The frontend communicates with the backend via /api route and the backend is sending data to the backend. The backend handles EEG upload, validation, preprocessing with MNE-Python (Gramfort et al. 2013) with NumPy and SciPy, feature extraction, heuristic risk scoring, recommendation generation and PDF report generation with ReportLab. The structured audit results are stored in SQLite. Our machine learning research component uses scikit-learn (Pedregosa et al. 2011) and is separate from the production scoring path.

Fig. 1 summarises the architecture.

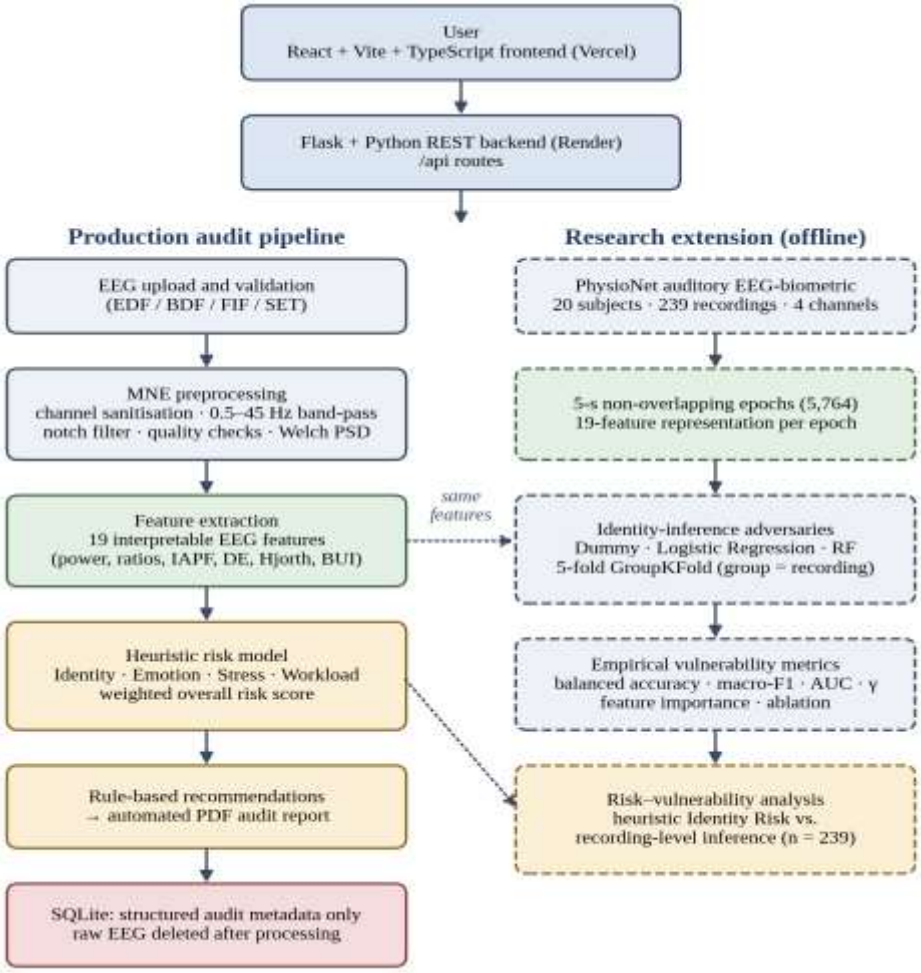


Fig. 1 Architecture of NeuroAudit. The production audit pipeline (left) processes an uploaded recording into a heuristic risk assessment, recommendations and a PDF report, and stores only structured audit metadata. The offline research extension (right) evaluates empirical identity inference on the same 19-feature representation and relates it to the heuristic Identity Risk score. Dashed elements are not part of the production deployment.

3.2 EEG upload, validation and temporary handling

The production pipeline accepts EDF, BDF, FIF and SET recordings. Before processing, the system checks the file extension, rejects empty or invalid files, verifies that the recording can be loaded by MNE, and enforces minimum channel and sample requirements. Uploaded recordings are processed as temporary files, and cleanup is enforced through error-handling logic so that temporary files are removed whether processing succeeds or fails. The production pipeline accepts EDF, BDF, FIF and SET recordings. Before processing, the system checks the file extension, rejects empty or invalid files, verifies that the recording can be loaded by MNE, and enforces minimum channel and

sample requirements. Uploaded recordings are processed as temporary files and cleanup is enforced through error handling logic so that temporary files are removed if processing succeeds or fails.

3.3 EEG preprocessing

Channel names are normalized and sanitised and generic channel names are mapped to standard EEG naming conventions where appropriate. Signals are band-pass filtered between 0.5 and 45 Hz, which retains the main EEG frequency components while decreasing the very low frequency drift and high frequency noise. A 50 Hz or 60 Hz notch filter is applied when it is valid in the Nyquist frequency of the recording. Quality checks are performed before feature extraction. The power spectral density is estimated using Welch's method (Welch 1967) with windows of about two seconds.

3.4 Feature representation

NeuroAudit currently extracts 19 EEG-derived features (Table 1). They cover relative band power, frontal alpha asymmetry, band-power ratios, an engagement index, individual alpha peak frequency, band-wise differential entropy, Hjorth parameters and a composite Brain Usage Index. The representation is intended to capture several aspects of EEG variability instead of relying on a single biomarker.

Table 1 The 19 EEG-derived features used by NeuroAudit

No.	Feature	Internal name	Group
1	Delta relative band power	bp_delta_rel	Relative band power
2	Theta relative band power	bp_theta_rel	Relative band power
3	Alpha relative band power	bp_alpha_rel	Relative band power
4	Beta relative band power	bp_beta_rel	Relative band power
5	Gamma relative band power	bp_gamma_rel	Relative band power
6	Frontal alpha asymmetry	faa	Asymmetry
7	Theta/beta ratio	theta_beta_ratio	Band-power ratio
8	Theta/alpha ratio	theta_alpha_ratio	Band-power ratio
9	Engagement index	engagement_index	Band-power ratio
10	Individual alpha peak frequency	iapf_hz	Spectral peak
11	Delta differential entropy	de_delta	Differential entropy
12	Theta differential entropy	de_theta	Differential entropy
13	Alpha differential entropy	de_alpha	Differential entropy
14	Beta differential entropy	de_beta	Differential entropy
15	Gamma differential entropy	de_gamma	Differential entropy
16	Hjorth activity	hjorth_activity	Hjorth parameters
17	Hjorth mobility	hjorth_mobility	Hjorth parameters
18	Hjorth complexity	hjorth_complexity	Hjorth parameters
19	Brain Usage Index (BUI)	bui	Composite indicator

3.5 Feature definitions

Frontal alpha asymmetry. FAA is the logarithmic difference between right and left frontal alpha power (Davidson 2004):

$$FAA = \ln(\alpha_R) - \ln(\alpha_L) \quad (1)$$

where α_R and α_L denote alpha-band power at the selected right and left frontal sites. FAA requires a suitable left/right frontal electrode pair. The research dataset used in Section 4 does not provide one, which has consequences for the experiment (Section 4.3).

Differential entropy. For a band-limited signal with variance σ^2, differential entropy under a Gaussian assumption is (Duan et al. 2013):

$$DE = \tfrac{1}{2} \ln(2\pi e\sigma^2) \quad (2)$$

and is computed for the delta, theta, alpha, beta and gamma bands.

Hjorth parameters. Activity, mobility and complexity (Hjorth 1970) describe amplitude and frequency-related variation of the signal x(t): activity is the variance of x, mobility is the square root of the ratio of the variance of the first derivative to the variance of x, and complexity is the mobility of the first derivative divided by the mobility of x.

Individual alpha peak frequency. IAPF is the frequency of maximum spectral power within approximately 7.5–12.5 Hz, a feature that is comparatively stable within individuals (Tran et al. 2026). Its distance from a reference frequency is normalised for the identity component:

$$IAPF_{dist} = |IAPF - 10.0| / 2.5 \quad (3)$$

Brain Usage Index. BUI is a composite feature that contributes to the identity component of the heuristic model. It is treated as a candidate feature-level indicator. It is not a validated biometric identity probability, and its inclusion is not evidence that the heuristic score predicts identification.

3.6 Heuristic privacy-risk model

The production risk model is an interpretable, rule-based heuristic baseline. No model training is involved. It produces scores for four dimensions, and the identity dimension combines BUI, channel availability, recording duration and IAPF distance:

$$f_{ch} = \min(1, N_{ch}/16), \quad f_{dur} = \min(1, T_{sec}/30) \quad (4)$$

$$Identity_{base} = 40 \cdot BUI + 30 \cdot f_{ch} + 15 \cdot f_{dur} + 15 \cdot IAPF_{dist} \quad (5)$$

with Identity = clip(round($Identity_{base}$), 20, 96), where N_{ch} is the number of channels and T_{sec} is the recording duration in seconds. The remaining dimensions are:

$$Emotion_{base} = 45 \cdot (|FAA|/2) + 40 \cdot P_{\beta\gamma} + 15 \cdot f_{ch} \quad (6)$$

$$Stress_{base} = 45 \cdot E_{\beta} + 30 \cdot S_{\alpha} + 25 \cdot C \quad (7)$$

$$Workload_{base} = 35 \cdot E_{\theta} + 35 \cdot G + 30 \cdot R_{TBR} \quad (8)$$

where $P_{\beta\gamma}$ is beta–gamma relative power, E_{β} a beta-elevation factor, S_{α} an alpha-suppression factor, C a complexity factor, E_{θ} a theta-elevation factor, G an engagement factor and R_{TBR} a theta/beta-ratio factor. The Emotion, Stress and Workload scores are bounded to 18–94, 15–95 and 15–95 respectively. The overall score combines the four dimensions:

$$Overall = 0.30 \cdot Identity + 0.25 \cdot Emotion + 0.25 \cdot Stress + 0.20 \cdot Workload \quad (9)$$

and is rounded and constrained to the range 10–98. Scores below 40 are labelled LOW, scores from 40 to 69 MEDIUM, and scores of 70 or above HIGH. These labels support auditing and prioritisation. They are not probabilities of privacy loss, and the heuristic model is not treated as ground-truth risk.

3.7 Worked example

Consider a recording with BUI = 0.8124, N_{ch} = 8, T_{sec} = 60 s and IAPF = 9.50 Hz. Then f_{ch} = 0.5, f_{dur} = 1.0 and $IAPF_{dist}$ = 0.2, so $Identity_{base}$ = 32.50 + 15.00 + 15.00 + 3.00 = 65.50, giving an Identity score of approximately 66. The example shows how an auditor can trace a score back to measurable inputs.

3.8 Recommendation engine

The recommendation engine is rule-based and recommendations are generated according to the risk dimensions detected. Typical recommendations are: minimising identifiable EEG metadata, restricting access to raw recordings, using pseudonymous identifiers, encrypting data in transit and at rest, minimising retention, using access-control and audit mechanisms, deleting raw EEG once processing is completed, and sharing only the channels or features needed for the intended purpose, scientifically appropriate. These recommendations are security and privacy guidance. Their quantitative mitigation effectiveness has not been experimentally tested in this study and no specific risk reduction is claimed for any recommendation.

3.9 Automated PDF audit report

For each processed recording, the platform generates a PDF audit report with ReportLab that summarises the extracted features, the dimension-level and overall scores, the risk category and the associated recommendations, so that the assessment can be documented and reviewed outside the application.

3.10 Privacy-aware data handling

NeuroAudit treats raw EEG as sensitive temporary input, and so is to reduce the need to retain neural data. Uploaded recordings are processed briefly and the temporary files are removed after processing, and cleanup is enforced in error handling (finally). Raw EEG is not stored permanently in the SQLite database, and preview traces are not stored. The database maintains structured audit information, and the identity of the ongoing recording is normalised instead of keeping the original uploaded filename as a persistent identifier. Research data is not part of the production deployment, and production uploads are not in version control. These controls were verified in the repository audit described in Section. 5.1.

4 Empirical identity-inference experiment

4.1 Purpose and scope

The research portion is separate from the production heuristic model. It asks if the interpretable features used by NeuroAudit can be exploited by learning-based adversaries to infer subject identity and whether the heuristic Identity Risk score relates to that empirical vulnerability. Because the dataset chosen for the experiment provides subject identity labels, it does not include emotion, stress or workload inference and the results are presented as an identity-inference vulnerability analysis.

4.2 Dataset

The auditory evoked potential EEG-Biometric Dataset v1.0.0 from PhysioNet (Abo Alzahab et al. 2021; Goldberger et al. 2000) was put together to support the development of EEG-based biometric systems and holds two-minute recordings from 20 volunteers. The conditions of the recording are eyes-open and eyes-closed rest and auditory stimulation via in-ear and bone-conduction transducers. The experiment contains 239 recordings from 20 subjects with four EEG channels (T7, F8, Cz and P4) with the nominal sampling frequency of 200 Hz. Table 2 summarises the configuration. The dataset is stored locally and is not deposited to the project repository that only contains research code and result artefacts.

4.3 Epoching and feature extraction

The recordings are split into non-overlapping 5-second epochs, leading to 5,764 epochs and the 19 features of Table 1 are extracted for each epoch. Since the available channels do not have a left/right frontal pair, FAA cannot be computed as it is intended and remains constant at zero throughout the research experiment. This is a limitation of the dataset montage and not of the FAA definition, and the FAA-related results below should be read accordingly.

Table 2 Configuration of the identity-inference experiment

Item	Setting
Dataset	PhysioNet Auditory Evoked Potential EEG-Biometric Dataset v1.0.0
Subjects / recordings	20 subjects / 239 recordings
Channels	4 (T7, F8, Cz, P4)
Nominal sampling frequency	200 Hz
Epoching	5-second non-overlapping epochs (5,764 epochs)
Features	19 EEG-derived features per epoch
Models	Dummy classifier, Logistic Regression, Random Forest
Random Forest settings	n_estimators = 100, max_depth = 10, random_state = 42
Cross-validation	5-fold GroupKFold, grouped by recording (run_id)
Nominal chance level	1/20 = 5%

4.4 Models

Three model categories are evaluated. A Dummy classifier provides a near-chance baseline. Logistic Regression serves as an interpretable linear model. Random Forest (Breiman 2001) serves as the main nonlinear identity-inference adversary, with the settings in Table 2. Any feature scaling for the linear model is performed within the cross-validation pipeline so that no information from test folds is used.

4.5 Recording-disjoint evaluation

Performance is estimated with five-fold GroupKFold cross-validation using the recording identifier (run_id) as the grouping variable, so epochs from the same recording never appear in both training and test folds. The protocol is therefore recording-disjoint. Splitting by recording avoids the leakage that arises when segments of the same recording occur in both training and test sets (Brookshire et al. 2024). Every subject is still represented in the training data through other recordings, so the task is closed-set identification of known subjects. The experiment is not an unseen-subject generalisation test, not a session-disjoint test and not a clinical validation. It measures whether an adversary holding other recordings of a subject can identify that subject from EEG-derived features.

4.6 Metrics and normalised inference advantage

Top-1 accuracy, balanced accuracy, macro-F1 and multi-class ROC AUC are reported. To express improvement over chance, a normalised inference advantage γ is also computed:

$$\gamma = (A_{emp} - A_{chance}) / (1 - A_{chance}) \quad (10)$$

where A_{emp} is the empirical balanced accuracy and $A_{chance} = 1/K$ for K classes. With K = 20, $A_{chance} = 0.05$. γ is a custom descriptive metric of this study and not an established machine-learning standard, and it is not the probability that a real-world adversary will succeed.

4.7 Feature importance and ablation

Importance of features in Random Forest is used to determine if a feature is more relevant to identity inference. Feature ablation then retrains the Random Forest with selected features removed: all 19 features without FAA, without BUI and without the heuristic feature group. The heuristic feature group consists of the 4 features which enter into the Identity heuristic or its neighbouring dimensions in the production model (bp_alpha_rel, bp_beta_rel, bui and faa).

5 Results

5.1 Software verification

Before deployment, a repository audit verified that the frontend production build works, that all 49 of 49 backend tests pass and that the research machine learning pipeline checks pass. The audit also verified that the raw-EEG retention controls were in place, that the research datasets and uploads were not collected in version control, that no dataset or secret files were staged, and that Vercel routing and Render Flask settings were correct. These checks establish software correctness and privacy-control implementation. This study does not report a usability or performance evaluation of our deployed service.

5.2 Identity-inference performance

Table 3 shows the epoch-level results for the recording-disjoint protocol. The Dummy classifier performs at chance (balanced accuracy 5.14% and ROC AUC 0.5007), as expected for 20 balanced classes. Logistic Regression and Random Forest both perform far above chance, and the Random Forest is marginally the best at balanced accuracy 64.35%, macro-F1 63.82% and ROC AUC 0.9507 (Fig. 2a). The normalised inference advantage is γ = 0.6247. The 64.35% figure is the balanced accuracy of an empirical machine-learning identity-inference attack. It is not the accuracy of NeuroAudit's heuristic risk model.

Table 3 Epoch-level identity-inference results (20 subjects, recording-disjoint 5-fold GroupKFold)

Model	Top-1 accuracy (%)	Balanced accuracy (%)	Macro-F1 (%)	ROC AUC
Dummy	5.12	5.14	5.13	0.5007
Logistic Regression	63.93	63.78	63.19	0.9390
Random Forest	**64.40**	**64.35**	**63.82**	**0.9507**

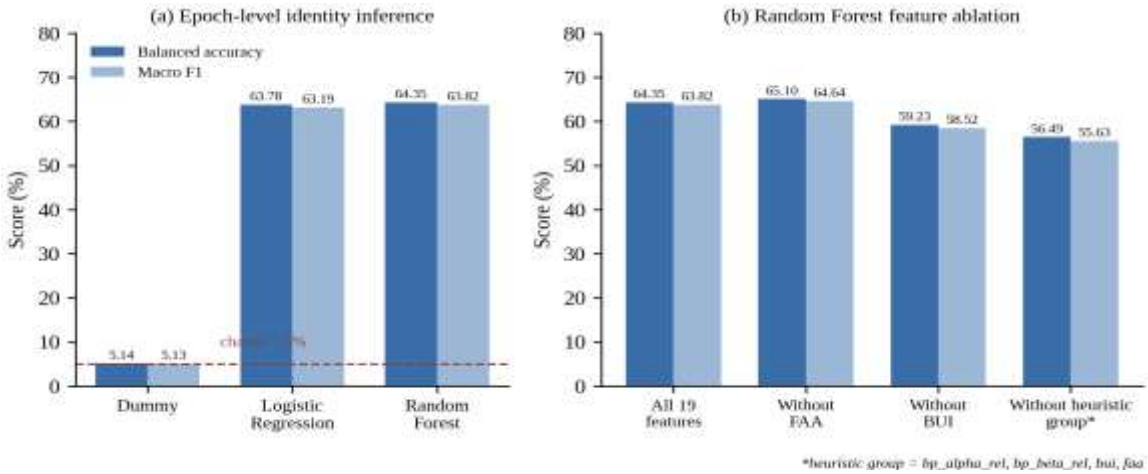


Fig. 2 Identity-inference results. (a) Epoch-level balanced accuracy and macro-F1 of the three models against the 5% chance level. (b) Random Forest feature ablation. The heuristic feature group comprises bp_alpha_rel, bp_beta_rel, bui and faa.

5.3 Feature importance

Table 4 lists the ten features with the highest Random Forest importance. High-frequency relative power, the Brain Usage Index, beta power and the Hjorth parameters rank highest, and the alpha peak frequency and the engagement index also appear. Several heuristic-related features (bp_beta_rel, bui, bp_alpha_rel) are among the top six. FAA was constant at zero in this dataset and so carries no information here.

Table 4 Ten highest-ranked features by Random Forest importance

Rank	Feature	Rank	Feature
1	bp_gamma_rel	6	bp_alpha_rel
2	bui	7	iapf_hz
3	bp_beta_rel	8	engagement_index
4	hjorth_mobility	9	theta_alpha_ratio
5	hjorth_complexity	10	bp_delta_rel

5.4 Feature ablation

Table 5 and Fig. 2b show the ablation results. Removing FAA does not affect performance (balanced accuracy 65.10% versus 64.35%) because FAA is constant in this dataset. The small increase is not interpretable as an informational effect and more likely due to random forest stochasticity. Removing BUI reduces balanced accuracy to 59.23%, and removing the whole heuristic feature group reduces it from 64.35% to 56.49%, a decrease of 7.86 percentage points. So several features that are used in the heuristic assessment also have information to include in an empirical identity-inference model. This does not mean that the heuristic score itself predicts privacy vulnerability, which is the question that we will consider next.

Table 5 Random Forest feature-ablation results

Feature set	Balanced accuracy (%)	Macro-F1 (%)
All 19 features	64.35	63.82
Without FAA	65.10	64.64
Without BUI	59.23	58.52
Without heuristic feature group	56.49	55.63

The heuristic feature group comprises bp_alpha_rel, bp_beta_rel, bui and faa.

5.5 Heuristic risk versus empirical vulnerability

The majority voting in the recording level correctly identified 208 of the 239 recordings and failed on 31 (Table 6), a majority vote accuracy of 87.03%. The average heuristic Identity Risk score was 59.34 for correctly inferred recordings

and 57.75 for poorly inferred recordings, which is 1.59 points in the right direction but small. The point-biserial correlation between Identity Risk and success in inferring identities was 0.1016 (p = 0.1172) and the bootstrap 95% confidence interval, −0.0107 to 0.2098, includes zero. This corresponds to about 1% of the common variance. In this study, the current heuristic Identity Risk score did not seem to be consistent with results in identifying identities.

Table 6 Relationship between heuristic Identity Risk and recording-level inference success

Metric	Result
Total recordings	239
Correctly inferred / incorrectly inferred	208 / 31
Majority-vote accuracy	87.03%
Balanced accuracy	87.08%
Macro-F1	86.94%
Mean Identity Risk, correctly inferred	59.34
Mean Identity Risk, incorrectly inferred	57.75
Point-biserial correlation	0.1016
p-value	0.1172
Bootstrap 95% confidence interval	[−0.0107, 0.2098]

5.6 Epoch-level versus recording-level performance

The Random Forest balanced accuracy achieved at the epoch level (64.35%) and the majority-vote accuracy at the recording level (87.03%) do not present any contradiction to each other, as they are calculated based on different units. This happens because multiple 5-second epochs take place in a single recording; thus, the errors that took place at epoch level get averaged out at recording level due to pooling the epoch-level predictions via majority voting resulting in improved performance at the recording level. Hence, it is not reasonable to compare the two mentioned figures, as they measure different quantities.

6 Discussion

6.1 What the results show

First, EEG-based interpretable features contain information that can be leveraged by a Random Forest to find 20 known subjects with about 13 times the chance level from a single 5-second recording using only 4 channels and a split of the recordings. Second, the features NeuroAudit uses in its heuristic assessment are part of that information. Removing the heuristic feature group reduced the balanced accuracy by 7.86%, so the platform looks at signal characteristics that are relevant for identity inference. Third, aggregation over a recording increases the accuracy to 87.03% because custodians tend to collect recordings from entire recordings and not just isolated epochs.

6.2 What the results do not show

The same results do not necessarily indicate that the heuristic score is a reliable measure of privacy vulnerability. The association between Identity Risk and inference success was not strong and not clear (r = 0.1016, p = 0.1172), and the confidence interval includes zero. This comparison also has limited power with only 31 wrongly inferred recordings so we can look at this result only as an absence of evidence of association and not as evidence of independence. The ablation also shows that some heuristic-related features are informative, not that the heuristic weighting accurately captures inference risk. The experiment does not validate the Emotion, Stress, or Workload dimensions, it is not an unseen subject or session-disjoint test, and the recommendations have not been shown to reduce inference success. It is important to keep these questions separate in order to interpret the study: the heuristic answers which describe interpretable characteristics of signals contribute to a risk indicator, whereas the machine-learning experiment gives a sense of how identifiable subjects are (according to a strict protocol).

6.3 The heuristic as an interpretable baseline

The negative risk vulnerability result is informative rather than disappointing. The heuristic weights were set by design and have not been calibrated to real inference outcomes, so it is not surprising that the score does not measure recording success. Publishing the result limits what we can claim: the heuristic serves as a transparent

baseline that provides us with measurable inputs and helps us generate reports; however, we need to do it in order to properly calibrate it to rank recordings based on real inference vulnerability. A natural way to do that is to use the empirical inference outcomes, such as the out-of-fold recording-level results reported here, as calibration parameters for the identity weights.

6.4 Implications for EEG data governance

The feature-importance and ablation results indicate that even a small number of simple spectral, entropy and Hjorth features can be identified. Data custodians should not think of the features they extract as anonymous, which is consistent with recent calls for stronger protection of neural data (Ienca 2021; Genser et al. 2024; UNESCO 2025). This agrees with the data minimisation and retention controls implemented in NeuroAudit, and supports temporarily storing raw EEG, structured metadata and pseudonymous identifiers, sharing only channels or features that are relevant for a purpose, and feature-level sharing and access control. A possible auditing procedure is to validate a recording, extract features, compute the four-dimensional scores, analyse the high-contributing features, apply the recommended controls, and record the audit report. The influence of those controls on inference success is still to be seen.

7 Limitations and future work

There are several limitations to the conclusions of this study.

Single dataset. The experiment is performed with a single PhysioNet dataset with 20 subjects, 4 channels and a specific setup for the acquisition, so the results should not be generalized to other EEG data sets, devices or populations.

Identity-only validation. The dataset only supports subject identity classification. It does not provide empirical validation of Emotion,Stress or Workload model.

FAA not computable. The available channels (T7, F8, Cz, P4) all lack a suitable left/right frontal pair, therefore FAA is constant at zero in the experiment and a conclusion on FAA cannot be drawn from this study.

Recording-disjoint, not subject-disjoint. The evaluation prevents epochs from one recording from appearing in both training and test folds but all subjects are present in training. It is not a strict unseen subject/session-disjoint generalisation experiment.

Uncalibrated heuristic. The heuristic risk score is interpretable but has not been tested as ground truth, and its Identity component did not find a statistically significant correlation with empirical inference success.

Unvalidated recommendations. The recommendation engine is rule-based, and it has not been tested on the quantitative mitigation effect of these recommendations.

Scale and diversity. More datasets, subjects, sessions and acquisition conditions are needed to validate more accurately for broader validation, as there is well documented EEG variability between subjects and between individuals (Tran et al. 2026).

The recordings also cover many experimental conditions which are not considered separately. Future work will take the following steps:

1. evaluate on larger and more diverse EEG datasets and on different independent datasets;
2. evaluate session-disjoint and unseen subject evaluation;
3. evaluate further privacy-inference tasks such as emotion, stress, workload and demographic characteristics with datasets that provide appropriate labels;
4. evaluate the heuristic risk scores against the inference results; • evaluate the privacy-preserving EEG transformations and privacy-preserving machine learning techniques (Xia et al. 2023; Paul and Bajwa 2025) with respect to inference success and data utility;
5. evaluate on different EEG acquisition conditions and consumer hardware.

8 Conclusion

NeuroAudit is an integrated software platform for EEG privacy and neurosecurity auditing. It combines EEG preprocessing with interpretable feature extraction, a heuristic privacy risk assessment, rule-based security recommendations, and automated audit reporting, and it handles raw EEG as temporary sensitive input. A different machine-learning component showed that EEG-derived features can be used for subject identity inference under a recording-disjoint protocol on a 20-subject public dataset, with a Random Forest generating 64.35% balanced accuracy at epoch level and 87.03% accuracy at recording level. We also found that features used for the heuristic assessment can provide valuable information to the inference model, and the balanced accuracy was reduced to 56.49% when they were removed. However, the heuristic Identity Risk score didn’t seem to be statistically correlated with empirical inference success (correlation 0.1016, p=0.1172). The current system should therefore be seen as an

interpretable privacy-audit baseline that needs further empirical validation, and not as a definitive measure of neural-data privacy vulnerability.

Declarations
Funding: No external funding was received for this study.
Competing interests: The authors state that they do not have any competing interests.
Ethics approval: The study is a secondary analysis of the publicly available EEG dataset (Abo Alzahab et al. 2021). We did not recruit any new human participants and found no new data.
Data availability: The EEG data we investigate in this study are publicly available from PhysioNet as the Auditory Evoked Potential EEG-Biometric Dataset, version 1.0.0, with the permission of PhysioNet to use it. The raw dataset is not redistributed in the project repository. The results of experiments, configuration, feature manifest and risk-vulnerability data are available in the project repository.
The NeuroAudit source code, including the machine-learning research pipeline and result artefacts, is available at https://github.com/Deepika-B-17/Neuroaudit- (main branch, commit bd5e1e2) and a deployed instance of the frontend is available at. https://neuroaudit-tyrm.vercel.app/.

References

[1] Abo Alzahab N, Di Iorio A, Apollonio L, Alshalak M, Gravina A, Antognoli L, Baldi M, Scalise L, Alchalabi B (2021) Auditory evoked potential EEG-Biometric dataset (version 1.0.0). PhysioNet. https://doi.org/10.13026/ps31-fc50
[2] Bernal SL, Celdrán AH, Pérez GM, Barros MT, Balasubramaniam S (2022) Security in brain-computer interfaces: state-of-the-art, opportunities, and future challenges. ACM Comput Surv 54(1):11. https://doi.org/10.1145/3427376
[3] Breiman L (2001) Random forests. Mach Learn 45:5–32
[4] Brookshire G, Kasper J, Blauch NM, Wu YC, Glatt R, Merrill DA, Gerrol S, Yoder KJ, Quirk C, Lucero C (2024) Data leakage in deep learning studies of translational EEG. Front Neurosci 18:1373515. https://doi.org/10.3389/fnins.2024.1373515
[5] Chikhi S, Matton N, Blanchet S (2022) EEG power spectral measures of cognitive workload: a meta-analysis. Psychophysiology 59(6):e14009. https://doi.org/10.1111/psyp.14009
[6] Davidson RJ (2004) What does the prefrontal cortex "do" in affect: perspectives on frontal EEG asymmetry research. Biol Psychol 67:219–233
[7] Duan R-N, Zhu J-Y, Lu B-L (2013) Differential entropy feature for EEG-based emotion classification. In: Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering (NER), pp 81–84
[8] Genser J, Damianos S, Yuste R (2024) Safeguarding brain data: assessing the privacy practices of consumer neurotechnology companies. The Neurorights Foundation, report, April 2024
[9] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101:e215–e220
[10] Gómez-Tapia C, Bozic B, Longo L (2022) On the minimal amount of EEG data required for learning distinctive human features for task-dependent biometric applications. Front Neuroinform 16:844667. https://doi.org/10.3389/fninf.2022.844667
[11] Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. Front Neurosci 7:267
[12] Hjorth B (1970) EEG analysis based on time domain properties. Electroencephalogr Clin Neurophysiol 29:306–310
[13] Ienca M (2021) On neurorights. Front Hum Neurosci 15:701258. https://doi.org/10.3389/fnhum.2021.701258
[14] Landau O, Cohen A, Gordon S, Nissim N (2020) Mind your privacy: privacy leakage through BCI applications using machine learning methods. Knowl-Based Syst 198:105932
[15] Maiorana E (2020) Deep learning for EEG-based biometric recognition. Neurocomputing 410:374–386. https://doi.org/10.1016/j.neucom.2020.06.009
[16] Patel P, Raghunandan R, Annavarapu RN (2021) EEG-based human emotion recognition using entropy as a feature extraction measure. Brain Inform 8:20. https://doi.org/10.1186/s40708-021-00141-5
[17] Paul S, Bajwa G (2025) Privacy-preserving EEG data generation: a federated split learning approach using privacy-adaptive autoencoders and secure aggregation with GFlowNet. In: Proceedings of the 22nd International Conference on Security and Cryptography (SECRYPT). SciTePress

[18] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
[19] Pope AT, Bogart EH, Bartolome DS (1995) Biocybernetic system evaluates indices of operator engagement in automated task. Biol Psychol 40:187–195
[20] Sabu P, Stuldreher IV, Kaneko D, Brouwer A-M (2022) A review on the role of affective stimuli in event-related frontal alpha asymmetry. Front Comput Sci 4:869123. https://doi.org/10.3389/fcomp.2022.869123
[21] Tajdini M, Sokolov V, Kuzminykh I, Shiaeles S, Ghita B (2020) Wireless sensors for brain activity—a survey. Electronics 9(12):2092. https://doi.org/10.3390/electronics9122092
[22] Tran X-T, Vo T-N, Vu S-T, Tran T-T, Nguyen M-D, Do T, Lin C-T (2026) Inter- and intra-subject variability in EEG: a systematic survey. arXiv:2602.01019 (preprint)
[23] UNESCO (2025) Recommendation on the ethics of neurotechnology. UNESCO, Paris
[24] Wang X, Ren Y, Luo Z, He W, Hong J, Huang Y (2023) Deep learning-based EEG emotion recognition: current trends and future perspectives. Front Psychol 14:1126994. https://doi.org/10.3389/fpsyg.2023.1126994
[25] Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 15:70–73
[26] Wilaiprasitporn T, Ditthapron A, Matchaparn K, Tongbuasirilai T, Banluesombatkul N, Chuangsuwanich E (2020) Affective EEG-based person identification using the deep learning approach. IEEE Trans Cogn Dev Syst 12(3):486–496. https://doi.org/10.1109/TCDS.2019.2924648
[27] Xia K, Duch W, Sun Y, Xu K, Fang W, Luo H, Zhang Y, Sang D, Xu X, Wang F-Y, Wu D (2023) Privacy-preserving brain–computer interfaces: a systematic review. IEEE Trans Comput Soc Syst 10:2312–2324
[28] Zhou Y, Huang S, Xu Z, Wang P, Wu X, Zhang D (2022) Cognitive workload recognition using EEG signals and machine learning: a review. IEEE Trans Cogn Dev Syst 14(3):799–818. https://doi.org/10.1109/TCDS.2021.3090217