Journal of Neural Engineering

OPEN ACCESS

RECEIVED
21 May 2024

REVISED
19 August 2024

ACCEPTED FOR PUBLICATION
17 October 2024

PUBLISHED
29 October 2024

PAPER

Neural decoding and feature selection methods for closed-loop control of avoidance behavior

Jinhan Liu[1,2,*], Rebecca Younk[3], Lauren M Drahos[3], Sumedh S Nagrale[3], Shreya Yadav[3], Alik S Widge[3,4] and Mahsa Shoaran[1,2,4]

[1] Institute of Electrical and Micro Engineering, EPFL, Lausanne, Switzerland
[2] Neuro-X Institute, EPFL, Geneva, Switzerland
[3] Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, United States of America
[4] These authors jointly supervised this work.
[*] Author to whom any correspondence should be addressed.

E-mail: jinhan.liu@epfl.ch

Keywords: neural decoder, defensive behavior, machine learning, psychiatric brain-machine interfaces, neuro-marker

Abstract

Objective. Many psychiatric disorders involve excessive avoidant or defensive behavior, such as avoidance in anxiety and trauma disorders or defensive rituals in obsessive-compulsive disorders. Developing algorithms to predict these behaviors from local field potentials (LFPs) could serve as the foundational technology for closed-loop control of such disorders. A significant challenge is identifying the LFP features that encode these defensive behaviors. *Approach.* We analyzed LFP signals from the infralimbic cortex and basolateral amygdala of rats undergoing tone-shock conditioning and extinction, standard for investigating defensive behaviors. We utilized a comprehensive set of neuro-markers across spectral, temporal, and connectivity domains, employing SHapley Additive exPlanations for feature importance evaluation within Light Gradient-Boosting Machine models. Our goal was to decode three commonly studied avoidance/defensive behaviors: freezing, bar-press suppression, and motion (accelerometry), examining the impact of different features on decoding performance. *Main results.* Band power and band power ratio between channels emerged as optimal features across sessions. High-gamma (80–150 Hz) power, power ratios, and inter-regional correlations were more informative than other bands that are more classically linked to defensive behaviors. Focusing on highly informative features enhanced performance. Across 4 recording sessions with 16 subjects, we achieved an average coefficient of determination of 0.5357 and 0.3476, and Pearson correlation coefficients of 0.7579 and 0.6092 for accelerometry jerk and bar press rate, respectively. Utilizing only the most informative features revealed differential encoding between accelerometry and bar press rate, with the former primarily through local spectral power and the latter via inter-regional connectivity. Our methodology demonstrated remarkably low training/inference time and memory usage, requiring <310 ms for training, <0.051 ms for inference, and 16.6 kB of memory, using a single core of AMD Ryzen Threadripper PRO 5995WX CPU. *Significance.* Our results demonstrate the feasibility of accurately decoding defensive behaviors with minimal latency, using LFP features from neural circuits strongly linked to these behaviors. This methodology holds promise for real-time decoding to identify physiological targets in closed-loop psychiatric neuromodulation.

1. Introduction

Fear and anxiety serve as adaptive defensive responses to threats, a phenomenon observed across a variety of species [1]. These responses are evolutionary mechanisms designed to enhance survival by preparing an organism to confront or flee from immediate danger [2, 3]. However, the same reactions, when excessive or misplaced, can significantly disrupt an individual's overall quality of life [4, 5]. Anxiety disorders, characterized by disproportionate and persistent fear and anxiety, are among the most prevalent psychiatric conditions [6, 7]. Fear and anxiety contribute to the manifestation of a wide array of psychiatric disorders, underscoring their critical role in mental health [8, 9].

Fear and anxiety are often studied through the lens of defensive behaviors, a set of responses or patterns elicited in the face of perceived threats [10]. Rodents exhibit various defensive behaviors in response to actual or potential threats [11, 12]. These behaviors are often used to model human illnesses, because anxiety can also be viewed as an excessive response to potential or actual threats [11, 13]. Therefore, exposure to a threatening stimulus evokes defensive responses that resemble emotional states related to fear and anxiety [12]. Recent studies indicate that the defensive patterns observed in normal human subjects show notable similarities to those of laboratory rodents. This parallel supports the hypothesis that rodent defensive behaviors may be reasonable models of similar behaviors in human anxiety disorders [14, 15]. The subjective human experience of fear is not the same as innate defensive behaviors in lower vertebrates, but those defensive behaviors are the closest available model [16]. Furthermore, both human fear/anxiety and rodent defensive behavior load onto the same frontal-amygdala circuits [17–19]. Hence, animal defensive behaviors offer a valuable model for understanding negative-valence processes in humans [20–22]. As a result, the excessive or contextually inappropriate exhibition of these behaviors can serve as a model for certain aspects of human psychiatric disorders [23, 24].

The long-term goal of modeling defensive and anxious behavior is to develop new treatments. Direct electrical stimulation of the brain is a particularly promising approach to that translation. Brain stimulation specifically improves the symptoms of multiple fear/anxiety disorders [25–28]. The approach involves the precise targeting of specific brain areas to modulate dysfunctional neural circuits associated with these conditions, which allows direct targeting of mechanisms discovered through animal models. Within the field of psychiatric brain stimulation, there is a strong drive towards closed-loop therapies [29–32]. The symptoms of psychiatric disorders, and of fear/anxiety disorders in particular, vary over time, and only some of those symptom states require neurostimulation. A closed-loop brain-machine interface (BMI) system that uses real-time neural activity from the subject to guide stimulation could help develop effective, precisely tailored therapies that stimulate only when it will be beneficial [29, 33–36].

There exists a main challenge for developing such closed-loop BMIs: we need a neural decoder that is capable of estimating the disorder symptom or behavior in real-time [37–39]. The development of highly accurate, fast, and memory efficient decoders is essential for optimizing the therapeutic outcomes, ensuring that stimulation protocols are dynamically adjusted to the fluctuating patterns of neural dysregulation associated with psychiatric disorders [40]. Consequently, advancements in BMI technology and decoding algorithms hold the promise of revolutionizing the treatment landscape for patients with psychiatric conditions, offering hope for more personalized and effective interventions.

Decoding plays a pivotal role in neural engineering and the analysis of neural data [41–43]. It leverages activity recorded from the brain to forecast behaviors or symptoms [44–47]. These predictions, derived from decoding, can be utilized to manipulate devices or to enhance our understanding of the brain's involvement in disorders [48–50]. This is achieved by assessing the extent of information that neural activity conveys about a symptom or behavior and examining how this information varies across different brain areas, experimental conditions, and states of disorder [51–53]. Decoding psychiatric states poses unique modeling challenges due to the complex and widespread network of brain regions involved in neural processes linked to neuropsychiatric states and behaviors, particularly in disorders such as chronic pain, addiction, or post-traumatic stress disorder (PTSD) [30, 38, 54–57]. It is also important to emphasize decoding from local field potentials (LFP) as opposed to single-neuron recordings. Single-neuron activities can be highly informative and were the foundation of early successful human motor decoding examples [41, 43, 58]. Single unit signals also underpinned a recent study demonstrating decoding of anxiety/threat-related behaviors [59]. These signals, however, are unstable over long periods of time (i.e. the decades that a clinical BMI might need to function) and require sampling at rates above 10 kHz, dramatically increasing system power requirements. LFPs, in contrast, carry rich behaviorally relevant signals [60–62] and can be stable for years [63]. They may specifically carry defensive information, and therefore they can also be used to decode defensive behaviors. For instance, one study showed that freezing states could be partially classified using 4 Hz LFP power [64].

In essence, neural decoding represents a regression or classification challenge that links neural signals to specific variables [65]. Machine learning (ML) has emerged as a pivotal technique for elucidating the intricate patterns of neural activity, as well as the individual variations in brain function correlated with symptoms and behaviors [66]. Its utility is particularly pronounced when the primary research objective is to enhance predictive accuracy, a goal partly attributed to ML's proven efficacy in addressing nonlinear challenges [67, 68]. Despite recent advances in ML techniques, the decoding of neural activity frequently employs traditional approaches such as linear regression (LR) and support vector machine (SVM) [69–71]. The adoption of contemporary ML tools for neural decoding might yield not only a substantial performance improvement but also the possibility of gaining more profound insights into neural functionality, as shown in recent studies. Encoding-decoding frameworks, based on linear state-space models, have decoded mood and cognitive state fluctuations from multi-site intracranial electrocorticogram (ECoG) or stereo-electroencephalography (sEEG) signals [38, 69]. A multi-layer perceptron (MLP) has been utilized to forecast depressive states in human patients from local field potential (LFP) signals [72]. A decoder leveraging random forest (RF) models has been developed for the prediction of multi-class affective behaviors via intracranial electroencephalography (iEEG) recordings from the human mesolimbic network [73]. A discriminative cross-spectral factor analysis model was utilized for identifying a brain-wide oscillatory pattern for predicting resilient versus susceptible mice to stress [74]. Episodes of mental fatigue and changes in vigilance were precisely decoded from ECoG signals in non-human primates (NHPs), using a gradient boosting classifier [75, 76].

In the aforementioned studies, beyond the decoding model utilized for prediction, the neuro-markers derived from neural data were crucial for decoding efficacy. The majority of previous efforts to detect psychiatric symptoms and behaviors in humans have focused on classical spectral power features [38, 69, 73, 77, 78]. It is not clear that spectral power is the best feature for decoding complex cognitive-emotional phenomena such as fear/defensive behaviors. For instance, spectral power features were outperformed by cross-region connectivity metrics when attempting to decode cognitive task engagement [70, 79]. Spectral (wavelet entropy), temporal (Hjorth parameters), and connectivity features (partial directed coherence and phase locking index) have all been identified as significant markers for detecting mental fatigue [75]. In contrast, shifts in depressive states were more influenced by variations in spectral power features within the subcallosal cingulate than by coherence and phase-amplitude coupling [72]. Consequently, the importance of spectral power vs. other neuro-markers for modeling and decoding defensive behaviors and fear expression requires further investigation.

Here, we studied the decoding of defensive behaviors from the prefrontal cortex (PFC) and amygdala, which together comprise a circuit believed to regulate the expression of threat/defense versus safety behaviors [24, 80–82]. In prior rodent work, the balance between defensive and safety behaviors was associated with theta band (5–8 Hz) LFP synchrony between the infralimbic cortex (IL) and basolateral amygdala (BLA) [60, 64, 83]. Therefore, IL-BLA LFP connectivity and power features are promising targets for the development and testing of decoding algorithms that could be used in closed-loop psychiatric BMIs. At the same time, prior work focused on simple categorical analyses (t-tests between groups) and did not consider the more clinically relevant question of how to decode imminent behavior at the timescale of milliseconds to seconds. Rapid decoding would be crucial for a closed-loop BMI aimed at mitigating anxious or avoidance behavior in humans. It is not clear that the same LFP features that broadly discriminate two groups will be able to predict moment-to-moment behavior. Similarly, past studies that employed decoding methods used them primarily to identify when and where specific information was encoded [59], or to identify behavior at longer timescales [64].

We thus developed a behavioral decoder based on IL-BLA LFP signals from rats undergoing a tone-shock conditioning and extinction protocol [84, 85]. Beyond conventional band power features, we explored and exploited a broad array of neuro-markers derived from the LFPs, across spectral, temporal, and connectivity domains. We considered three defensive behaviors: freezing, bar press suppression (bar press rate), and accelerometry, specifically the jerk (first derivative) calculated from a 3-axis head-mounted accelerometer. Freezing and bar press suppression are canonical defensive behaviors that have been studied for decades [20–22, 86–88]. Accelerometry jerk is a newer metric we have proposed and shown to correlate with, but not fully overlap the two other measures [89]. We developed a decoding framework based on Light Gradient-Boosting Machine (LightGBM), which outperformed other state-of-the-art ML-based decoders in both accuracy and latency. Our approach included a methodology to assess feature importance and a feature selection strategy utilizing the SHapley Additive exPlanations (SHAP), effectively reducing the dimensionality of the feature space. Band power and band power ratio between channels emerged as critical for decoding defensive behaviors, with the high-gamma band being particularly predictive compared to other frequency bands. By prioritizing highly-ranked neuro-markers, we enhanced decoding performance beyond that with solely band power features. Consequently, this study underscores the effectiveness of our proposed ML framework in the precise and rapid decoding of defensive behaviors within a closed-loop psychiatric Brain-Machine Interface (BMI) system.

2. Methods

2.1. Animals and behavior paradigm

We utilized 16 adult Long Evans rats, with weights ranging from 250 to 350 grams. Initially, rats were pair-housed in plastic cages for at least 7 days to facilitate acclimation to the research facility. Subsequent to this acclimation period, the rats underwent daily handling for 5 days to mitigate handling-related stress, after which they were individually housed in plastic cages. To prepare for experimental procedures, food intake was restricted to 10 grams per day until each rat achieved 85%–95% of its initial body weight. Thereafter, the animals were allocated 15–20 grams of food daily to maintain their weight within this specified range throughout the behavioral experiments. During the first three days of food restriction, sucrose pellets were introduced into the home cages to acquaint the rats with the reward, thereby facilitating the learning of bar-pressing behavior.

The behavioral training and experiments were conducted in the Coulbourn conditioning chambers, with dimensions of 30.5 × 24.1 × 21 cm. These chambers were equipped with a grid floor, consisting of rods spaced 1.6 cm apart and with a diameter of 4.8 mm, to facilitate the delivery of foot shocks. An aluminum wall of the chamber incorporated a retractable bar and a food trough for monitoring reward-seeking behaviors, while a speaker mounted on the opposite wall emitted sound stimuli. Additionally, a camera with an attached wide-angle lens was positioned outside the conditioning chamber, above the speaker unit, to record video footage through the chamber's plexiglass top.

Initially, rats were trained to execute bar presses to obtain sucrose pellets. They subsequently underwent electrode implantation and participated in a post-surgical behavioral paradigm. To provoke defensive behaviors, the rats were exposed to a tone-shock conditioning protocol, which comprised three phases: habituation/conditioning, extinction, and extinction recall, as shown in figure 1(a). Electrophysiological and video recordings were systematically carried out during each experimental session. During the habituation phase on day 1, rats encountered 5 trials of the conditioned stimulus (CS: a 30-s, 82 dB tone). This was followed by the conditioning phase, where they experienced 7 instances of the CS paired with the unconditioned stimulus (US: a 0.6 mA, 0.5-s foot shock) immediately after the CS. On day 2, the extinction phase consisted of 20 presentations of the CS alone, without the US, in the same chamber. On day 3, to evaluate extinction memory (recall), the CS was presented 6 times without the US.

Reward-seeking behavior, indicated by bar presses, functioned as a dynamic measure for defensive behavior, with a decrease in pressing activity interpreted as an elevated threat response. Bar press events were captured in the electrophysiology event data using a Data Acquisition System (DAQ) (USB 6343-BNC or PCIe-6353, National Instruments, Woburn, MA, USA). These event data were subsequently processed to isolate bar press incidents and their associated timestamps. Assessment of freezing behavior was conducted through offline video analysis, employing a Logitech HD Pro Webcam C910 equipped with a Neewer Digital High Definition 0.45 × Super Wide-angle Lens. The footage, captured at a rate of 24 frames per second with Debut Professional software, was analyzed using ANY-maze, which utilizes its integrated freezing detection functionality to assign a 'freezing score' to each frame. This score increased with more significant changes in pixels between consecutive frames. Meanwhile, accelerometry data were collected continuously at a 30 kHz sampling rate via the RHD 2132 electrophysiology headstage, which includes a built-in 3-axis accelerometer. These data were logged using the Open Ephys acquisition system, a widely used open-source platform for *in-vivo* electrophysiology research [90]. The synchronization of accelerometry records with video data was accomplished by aligning the 'tone on' events observed in both datasets. There were sparse bar press events for a few rats during the conditioning phase due to the bar press suppression resulting from foot shocks. We chose 10 rats with no less than 5 bar presses in the conditioning phase to ensure the threat responses were well-encoded in the following neural decoding study. All 16 rats were used for the analysis in habituation, extinction, and recall sessions.

2.2. Electrodes and surgery

Each electrode bundle was comprised of 8 nickel-chromium recording microwires, each measuring 12.5 μm in diameter, accompanied by one reference wire of the same diameter as the recording wires, and a platinum-iridium stimulating channel with a diameter of 127 μm [91]. The stimulation channel was used for another set of experiments not reported here, and no stimulation occurred during any of the data reported in this paper. These recording and stimulating components were collectively bundled within a 27-gauge stainless steel cannula, which also served as a pathway for the return current during stimulation. The recording wires were bonded to the individual pins of an Omnetics connector using silver paint, while the stimulating wire was soldered to a mill-max connector, enabling concurrent recording and stimulation at the same site. A ground wire was affixed to the connector, and the entire bundle was safeguarded with epoxy. Prior to surgery, the electrodes were sterilized using Ethylene Oxide (EtO).

The electrode arrays were surgically implanted into the left infralimbic cortex (IL) (+3 mm anterior-posterior (AP), +0.5 mm medial-lateral, and −3.95 mm dorsal-ventral (DV) from the brain surface) and the basolateral amygdala (BLA) (−2.28 mm AP, +5 mm medial-lateral, and −7.5 mm DV from the brain surface). Dental cement was utilized to secure the electrodes and to construct a protective head cap for the animals. The ground wire was securely wrapped around a skull screw prior to the implant fixation. A minimum recovery period of seven days was allowed for the animals before starting physiological experiments.

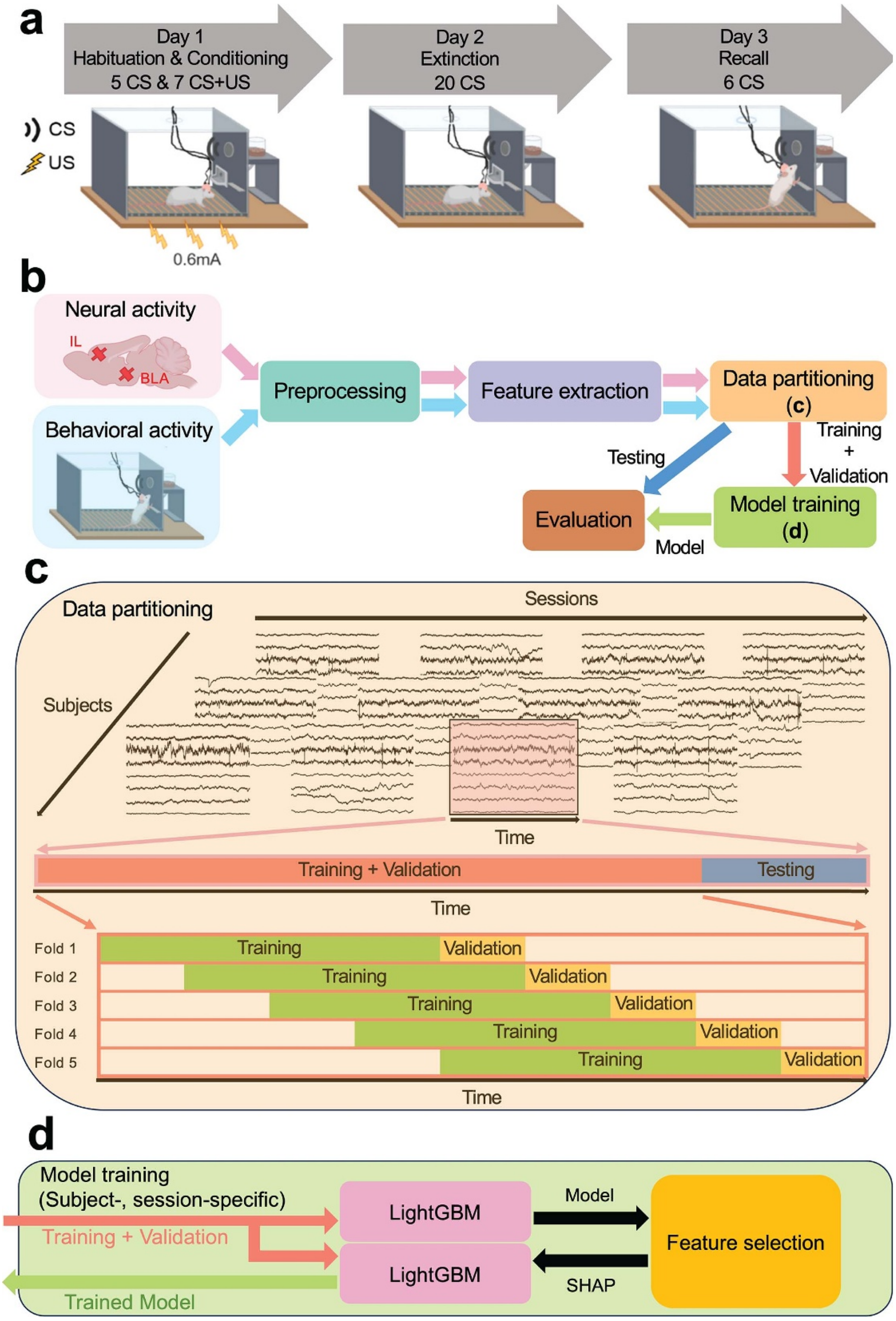

Figure 1. Experimental paradigm and proposed ML framework for decoding defensive behaviors. (a) The three-day tone-shock conditioning protocol. The experiment consisted of three phases: habituation/conditioning, extinction, and extinction recall, with electrophysiological and behavioral data recorded during each phase. (CS: Conditioned stimulus. US: Unconditioned stimulus) (b) The proposed ML framework contained modules including neural data preprocessing, feature extraction in three representation domains, data partitioning into training, validation and testing sets as shown in (c), decoding model training as shown in (d), and model evaluation. (c) Data partitioning process in each recording session for each subject. A held-out testing set was taken at the end of the session, and the beginning of the session was split in a sliding window 5-fold cross-validation paradigm. (d) The procedure of training the decoding model. SHAP values were measured using the first trained LightGBM model with all features, and the second LightGBM was then trained using the high-rank features in the order of their SHAP values and subsequently used for final model evaluation.

2.3. Electrophysiology and data processing

The electrophysiological signals, specifically local field potentials (LFPs), were recorded at a sampling rate of 30 kHz using an Open Ephys acquisition system throughout all experimental sessions. The recording headstage was interfaced with two mill-max male-male connectors, each comprising eight channels, through an adaptor.

Quantification of defensive behavior adhered to the methodology established in a prior study [89]. We calculated and attempted to decode three separate types of defensive behavior: freezing, bar press rate/suppression, and accelerometry rate of change (jerk). Freezing is the most common assay of defensive behavior in rodents, particularly in tone-shock conditioning paradigms as used in this study. It therefore places our results in context in the broader literature. Bar press rate is used less frequently because it requires extensive operant pre-training, but as we showed in [89], bar pressing is only partially correlated with freezing. That is, it captures a different aspect of defensive behavior that may be differentially encoded in IL-BLA activity. The same is true for accelerometry jerk: it correlates partially with the other two metrics, and can be used to compute an extinction/recall analysis, but does not measure precisely the same type of defensive reaction [89]. Jerk can be thought of as capturing the vigor or rapidity of response, and may be able to capture more active forms of defense such as darting [92].

The jerk is defined as the rate of change in total (3-axis) acceleration, calculated as:

$$j(t) = \left| \frac{d\sqrt{V_X^2(t) + V_Y^2(t) + V_Z^2(t)}}{dt} \right| \quad (1)$$

where $j(t)$ is the accelerometry jerk as a function of time t, and $V_X^2(t)$, $V_Y^2(t)$, and $V_Z^2(t)$ are the voltages of the accelerometer in X, Y, and Z axes, respectively. The accelerometry jerk was downsampled from its original sampling rate of 30 k samples/second to 1k samples/second, and was then smoothed with a Gaussian filter using a 200-sample window to remove non-biological noise transients. Bar press events and their corresponding timestamps were extracted from the electrophysiological recordings, with timestamps being resampled to 1k samples/second. The counts of presses was binned into each 1-ms time interval, and then these counts were smoothed using a Gaussian filter with a 1k-sample window. This process transformed the data from a discrete count of events into an approximation of a continuous press rate, hereby referred to as the bar press rate. The resampling process utilized the *downsample*() function in Matlab, while Gaussian smoothing was executed with the *smoothdata*() function in Matlab.

A total of 8 recording channels were obtained from bipolar-referenced LFP signals, with 4 channels from each of the IL and BLA regions. These bipolar-subtracted channels were subsequently band-pass filtered within the range of 1–150 Hz using a 3rd-order zero-phase infinite impulse response (IIR) Butterworth filter. Subsequently, line noise was removed by applying a notch filter at 60 Hz and its harmonics. The LFP data was then demeaned across the time series for each channel. For each subject and recording session, we visually inspected the neural data and excluded time epochs that exhibited clear non-neural artifacts, such as significant sharp voltage transients. For extracting features in various frequency bands, the neural data were processed using 3rd-order zero-phase IIR Butterworth band-pass filters across 7 frequency bands: 1–4 Hz (delta), 4–8 Hz (theta), 8–13 Hz (alpha), 13–30 Hz (beta), 30–50 Hz (low-gamma), 50–80 Hz (gamma), and 80–150 Hz (high-gamma). Phase and amplitude were extracted from the band-pass filtered signal via Hilbert transform. Cross-spectral density was estimated on the neural signals before band-pass filtering using the Multitaper method in the MNE package. The other preprocessing steps were implemented using the SciPy package.

Both behavioral data and the neuro-markers were computed in overlapping 1 s sliding windows with a 0.2 s step size. Behavioral measurements were quantified by averaging the measures of accelerometry jerk, bar press rate, and freezing score within each window.

2.4. Neuro-marker extraction

To investigate the neural representations in various aspects and enhance the accuracy of decoding defensive behaviors in our model, we extracted 17 types of neuro-markers across 3 representation domains as neural features for each window, as detailed in table 1. We chose these features based on existing evidence that, in general, local power and cross-region connectivity between IL and BLA have been linked to defensive behaviors in past research [83, 93]. Additionally, we computed time domain features that are pivotal in identifying patterns of neural activity associated with specific behaviors or pathological states, owing to their simplicity and the direct interpretation of neural dynamics [94–96].

In the spectral domain, band power (BP) was quantified across 7 frequency bands [38, 97, 98]. Relative band power (RBP) refers to the power in a specific frequency band relative to the total signal power [94, 99, 100]. Band power ratio between bands (BPRB) facilitates the pairwise comparison of power levels across different bands within a single channel [99–101].

Regarding temporal features, line length (LL) calculates the absolute differences between successive time points [94, 97, 102]. Hjorth parameters (HP) reflect statistical attributes including variance, mean frequency, and frequency variation [94, 103–105]. Maximum (Max) and minimum (Min) represent the extreme values within the window [94]. Nonlinear energy (NE) gives an estimate of the energy content of the neural signal [106]. Skewness evaluates the asymmetry of the distribution of instances within the window [107]. Approximate entropy (ApEn) and sample entropy (SampEn) assess the existence of patterns within a sequence of instances [108, 109].

In the connectivity domain and cross-region representations, band power ratio between channels (BPRC) enables pairwise power level comparison

Table 1. Neuro-markers extracted in the spectral, temporal, and connectivity domains.

Spectral domain	
Band power[a,b]	$\mathrm{BP}_j = \frac{1}{T}\sum_{t=1}^{T} y_j^2(t)$
Relative band power[a,b]	$\mathrm{RBP}_j = \frac{\mathrm{BP}_j}{\frac{1}{T}\sum_{t=1}^{T} y^2(t)}$
Band power ratio between bands[a,b]	$\mathrm{BPRB}_{jk} = \frac{\mathrm{BP}_j}{\mathrm{BP}_k}$
Temporal domain	
Line length[a]	$\mathrm{LL}= \sum_{t=1}^{T-1} \lvert y(t+1) - y(t)\rvert$
Hjorth parameters[a,c]	$\mathrm{Act}= \sigma^2(y(t))$, $\mathrm{Mob}= \sqrt{\sigma^2(\frac{dy(t)}{dt})/\sigma^2(y(t))}$, $\mathrm{Com} = \mathrm{Mob}(\frac{dy(t)}{dt})/\mathrm{Mob}(y(t))$
Maximum[a]	$\mathrm{Max}= \max_{t=1}^{T} y(t)$
Minimum[a]	$\mathrm{Min}= \min_{t=1}^{T} y(t)$
Nonlinear energy[a]	$\mathrm{NE}= \frac{1}{T-2}\sum_{t=1}^{T-2} y^2(t+1) - y(t)y(t+2)$
Skewness[a,c,d]	$\mathrm{Skewness} = \frac{\sum_{t=1}^{T}(y(t)-\bar{y})^3}{(T-1)\sigma^3}$
Approximate entropy[a,e]	$\mathrm{ApEn}= \frac{1}{T-1}\sum_{t=1}^{T-1}\log C_u^2 - \frac{1}{T-2}\sum_{t=1}^{T-2}\log C_u^3$
Sample entropy[a,e]	$\mathrm{SampEn}= -\log(A^2/B^2)$
Connectivity domain	
Band power ratio between channels[a,b,f]	$\mathrm{BPRC}_{jmn} = \frac{\mathrm{BP}_{jm}}{\mathrm{BP}_{jn}}$
Coherence[a,f,g]	$\mathrm{Coh}_{mn} = \sum_f \frac{\lvert G_{mn}(f)\rvert^2}{G_{mm}(f)G_{nn}(f)}$
Phase amplitude coupling[a,f,h,i]	$\mathrm{PAC}_{mn} = \lvert \frac{1}{T}\sum_{t=1}^{T} a_m(t)e^{i\theta_n(t)}\rvert$
Phase locking value[a,f,i]	$\mathrm{PLV}_{mn} = \lvert \frac{1}{T}\sum_{t=1}^{T} e^{i(\theta_m(t)-\theta_n(t))}\rvert$
Pearson correlation[a,c,d,f]	$\mathrm{Corr}_{mn} = \frac{1}{N\sigma_m\sigma_n}\sum_{t=1}^{T}(y_m(t)-\bar{y}_m)(y_n(t)-\bar{y}_n)$
Band Pearson correlation[a,b,c,d,f]	$\mathrm{BCorr}_{jmn} = \frac{1}{N\sigma_{jm}\sigma_{jn}}\sum_{t=1}^{T}(y_{jm}(t)-\bar{y}_{jm})(y_{jn}(t)-\bar{y}_{jn})$

[a] $y(t)$ is the time-series neural signal within a window of length T, where $t \in \{1, 2, \ldots T\}$.

[b] j, k are the i^{th} and j^{th} bands of the neural representations.

[c] σ is the standard deviation of $y(t)$, and σ^2 is the variance.

[d] $\bar{y}$ is the mean of $y(t)$.

[e] $C_p^r = \frac{1}{(T-r+1)}$[number of q such that $q \leqslant T-r+1$ and $d[l_r(p), l_r(q)] \leqslant 1$],
$A^2 = \frac{1}{(T-3)(T-2)}\sum_{p=1}^{T-2}\sum_{q=1,q\neq p}^{T-2}$[number of times that $d[\lvert l_3(q) - l_3(p)\rvert < 1]$],
$B^2 = \frac{1}{(T-3)(T-2)}\sum_{p=1}^{T-2}\sum_{q=1,q\neq p}^{T-2}$[number of times that $d[\lvert l_2(q) - l_2(p)\rvert < 1]$], where
$l_r(p) = \{y(p), y(p+1), \ldots, y(p+r-1)\}$, and
$d[l_r(p), l_r(q)] = max_{v=1,2,\ldots,r}(\lvert y(p+v-1) - y(q+v-1)\rvert)$.

[f] m, n are the m^{th} and n^{th} channels of the neural representations.

[g] $G_{mn}(f)$ is the cross-spectral density between $y_m(t)$ and $y_n(t)$ at frequency f, and $G_{mm}(f)$ is the auto-spectral density of $y_m(t)$ at frequency f.

[h] $a(t)$ is the amplitude of $y(t)$.

[i] $\theta(t)$ is the phase of $y(t)$.

across channels from two distinct brain regions [99]. Coherence (Coh) quantifies the similarities of neural oscillation between channels [110]. Phase-amplitude coupling (PAC) captures the linkage between the phase of a low-frequency band and the amplitude of a high-frequency band between channels [95, 111, 112]. Phase locking value (PLV) describes the phase relationship consistency between signals from different channels [64, 113]. Pearson correlation (Corr) and band Pearson correlation (BCorr) quantify functional connectivity between channels, across the full band and within individual frequency bands, respectively [79].

Here, BP, RBP, BPRC, Coh, PLV, and BCorr were assessed across the aforementioned 7 frequency bands. PAC analysis was performed between the amplitudes of the low-gamma, gamma, and high-gamma bands and the phases of the theta and alpha bands, with 6 phase-amplitude combinations. BPRB comparisons were made between each pair of the 7 bands, with 21 band-band combinations in total. BP, RBP, BPRB, LL, HP, Max, Min, NE, skewness, ApEn, and SampEn were calculated for each individual channel. BPRC, Coh, PAC, PLV, Corr, and BCorr were derived only from channel pairs between IL and BLA, with 16 channel-channel combinations.

ApEn and SampEn were computed using the MNE-Features package.

2.5. Dataset partitioning and characteristics

After extracting the neuro-markers and behavioral data, we partitioned them into three distinct datasets for subsequent use in training the decoding model, selecting high-rank features, and evaluating the model's performance. For each subject, we divided the data from each recording session into training, validation, and test sets, as depicted in figure 1(c). The models were further trained and evaluated on these datasets in a subject-dependent, session-dependent manner.

A held-out test set, constituting 20% of the entire recording, was designated from the final 20% of each behavior session, while the initial 80% served as the training and validation sets. The separation between the training and validation sets employed a sliding-window 5-fold cross-validation paradigm. The time series data were evenly divided into 9 windows. In the first fold, the initial 4 windows formed the training set, and the 5^{th} window served as the validation set. From the second to fifth folds, we sequentially shifted the training and validation sets by one window forward in time, ensuring that validation sets were different across folds and incorporating validation sets from preceding folds into the training sets of subsequent folds. Therefore, the division ratios for training and validation sets versus test sets, and training sets versus validation sets, were maintained at 80%–20%. This method respected the chronological sequence of the time series data by consistently organizing the datasets in a training-validation-testing order. This organization ensured that the model was always trained on historical data and validated/tested on subsequent data, thereby preventing data leakage across the temporal dimension. The size of training, validation, and test sets is shown in table A1 and appendix A. The test set was utilized for the final evaluation of the model's decoding accuracy, trained using the complete training and validation sets. Feature selection and hyperparameter optimization were conducted based on the model's validation set performance, trained on the training set data.

Our dataset covers both neural and behavioral responses under a structured experimental paradigm. Its richness and uniqueness lie in its detailed temporal resolution and the simultaneous recording of multiple modalities (neural signals and behavioral measures including accelerometry, reward-seeking, and freezing). This integration allows for advanced modeling of the neural correlates of behavior, providing insights into the neural dynamics underlying avoidance behaviors. Additionally, the specific conditioning paradigm, featuring sequential phases of habituation, conditioning, extinction, and recall, enables a nuanced analysis of conditioned/unconditioned behavioral responses. Furthermore, the integration of these diverse neuro-markers across multiple domains not only improves the accuracy of decoding models applied to predict behavioral outcomes, but also enables a deeper understanding of the underlying neurophysiological processes, therefore making it a valuable resource for both exploratory and predictive studies in defensive behaviors.

2.6. Decoding model

A diverse array of machine learning (ML) models has been employed for neuropsychiatric tasks and brain-machine interface applications, including LR [69, 110], SVM [64, 70, 74, 114, 115], RF [116, 117], and artificial neural network (ANN) [118, 119]. Moreover, gradient-boosted decision trees (GBDT) have demonstrated promising performance in previous neurophysiological task studies [97, 98, 104, 114, 120]. In this work, we utilized a GBDT-based model named Light Gradient-Boosting Machine (LightGBM), known for its efficiency in reducing data instances and features through gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) [121]. GOSS retains data instances with gradients above a certain threshold while randomly discarding instances with smaller gradients, thereby maintaining the data's substantial contribution to information gain. EFB efficiently reduces the number of effective features by bundling mutually exclusive features—those not taking non-zero values simultaneously—into a single feature. By leveraging GOSS and EFB, LightGBM achieves superior computational speed and lower memory usage compared to other GBDTs, without compromising the accuracy intrinsic to GBDT models.

In addition to LightGBM, we evaluated a variety of ML models widely applied in neurophysiological research, employing our proposed feature set as outlined in table 1. These models encompass traditional ML approaches such as LR, SVM with both Linear (SVM-Lin) and radial basis function kernels (SVM-RBF) [122], the tree-based RF model [123], and ANN models with diverse architectures, including MLP [68], long short-term memory (LSTM) [124], convolutional neural networks (CNN) [125], and Residual Networks (ResNet) [126]. The details of the training procedure, model architectures, and hyperparameter selection are presented in appendix C. In our preliminary decoding analysis shown in table 2, by leveraging all neural features, we assessed the decoding accuracy for accelerometry jerk and bar press rate across the aforementioned ML models, averaged over subjects in each recording session. Performance evaluation was conducted using both the coefficient of determination (R^2) and the Pearson correlation coefficient (r) metrics. It should be noted that here, R^2 is not the squared Pearson correlation coefficient, and its value lies within the range of $(-\infty, 1]$. A negative

Table 2. Performance of ML models for decoding defensive behaviors averaged across subjects in each recording session. The performance was evaluated using the coefficient of determination (R^2) and the Pearson correlation coefficient (r). The best results are **bolded**.

Behavior	Metric	Session	LR	SVM-Lin	SVM-RBF	RF	MLP	LSTM	CNN	ResNet	LightGBM
Accelerometry jerk	R^2	Habituation	−59.52	−140.8	0.4495	0.4581	0.4668	0.4554	0.4614	0.4545	**0.4677**
		Conditioning	−1099	−3.593	0.3931	**0.6324**	0.6240	0.6301	0.6239	0.6216	0.6310
		Extinction	−62.73	−6.935	0.4815	0.4913	0.4878	0.4646	0.4779	0.4592	**0.4952**
		Recall	−2.040	−0.3346	0.4916	0.5417	0.5214	0.5391	0.5449	0.5390	**0.5515**
	r	Habituation	0.4660	0.5038	0.6895	0.6974	0.6914	0.6827	0.6934	0.6817	**0.6998**
		Conditioning	0.2497	0.5543	0.7052	0.8417	0.8313	0.8357	0.8224	0.8132	**0.8425**
		Extinction	0.5340	0.5443	0.7037	0.7147	0.6951	0.6782	0.6978	0.6747	**0.7187**
		Recall	0.5515	0.5940	0.7320	0.7688	0.7614	0.7712	0.7676	0.7615	**0.7753**
Bar press rate	R^2	Habituation	−466.0	−719.3	−8.451	0.3299	0.3105	0.3201	0.3281	0.3188	**0.3306**
		Conditioning	−638.8	−92.94	−20.55	0.2819	0.2697	0.2610	0.2734	0.2713	**0.2848**
		Extinction	−22.16	−49.67	−13.56	0.3713	0.3576	0.3654	0.3722	0.3578	**0.3798**
		Recall	−2.027	−2.517	−9.068	0.3746	0.3618	0.3597	0.3610	0.3576	**0.3761**
	r	Habituation	0.3626	0.3194	——[a]	0.5995	0.5796	0.5809	0.6064	0.5856	**0.6113**
		Conditioning	0.3202	0.2899	——[a]	**0.5480**	0.5313	0.5298	0.5334	0.5331	0.5435
		Extinction	0.4536	0.3432	——[a]	0.6115	0.6072	0.6105	0.6204	0.6101	**0.6237**
		Recall	0.3540	0.3248	——[a]	0.6249	0.6134	0.6029	0.6212	0.6251	**0.6368**

[a] Not applicable due to the fact that the model produced a single constant value as its output for all inputs, and thus a Pearson correlation coefficient could not be computed.

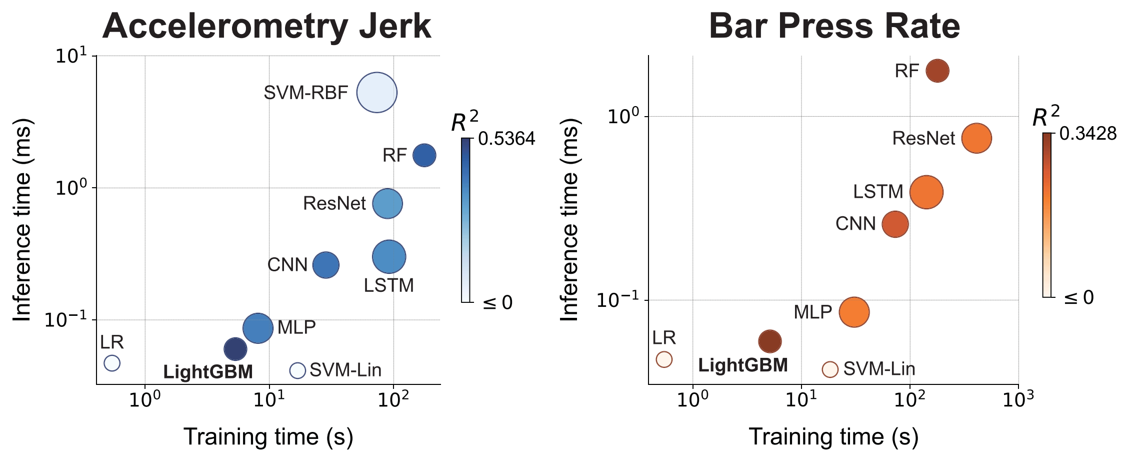

Figure 2. Comparative performance of ML models for decoding accelerometry jerk (Left) and bar press rate (Right) in accuracy, training time, inference time, and memory size, averaged across subjects and sessions. Each circle represents a model, with the size of the circle proportional to the memory size (on a logarithmic scale), and with the color of the circle proportional to the accuracy evaluated using R^2 (in an exponential scale).

R^2 suggests that the decoded behavior captures less variation in the real behavior than a constant value equivalent to the average of the ground truth, indicating relatively poor decoding performance. Our findings in table 2 indicate that by using our extracted neuro-markers as inputs, LightGBM outperformed other ML models in decoding both accelerometry jerk and bar press rate in 14 out of 16 comparisons. We also employed ANNs to test their capability of automatic neural feature extraction for decoding defensive behaviors. LSTM-Raw, WaveNet-Raw (an advanced CNN model with a multi-temporal-scale structure) [127], and ResNet-Raw [126] used raw LFP signals as inputs for the decoding tasks appendix B. The results from this experimental setup, as delineated in table A2, were systematically compared with the decoding performance of models that incorporate neuro-markers (table 2). This analysis reveals that manual feature extraction consistently outperforms the ANN models using raw LFP signals, and it suggests that these neuro-markers introduce a level of specificity and relevance that current ANN architectures struggle to achieve autonomously when working with neural data.

The performance superiority of LightGBM for decoding defensive behaviors is further evident from the comparative analysis of accuracy, training and inference times, as well as memory usage across different models. As illustrated in figure 2, LightGBM achieved the highest decoding accuracy among all tested ML models for both behavioral metrics, as indicated by the R^2 values in the scatter plots (0.5364 for accelerometry jerk and 0.3428 s for bar press rate). Moreover, LightGBM not only shows a reduced memory footprint (28.30 kB) but also exhibits significantly lower training times (5.336 s for accelerometry jerk and 5.130 s for bar press rate) and inference times (0.06 006 ms for accelerometry jerk and 0.05 966 ms for bar press rate), compared with other models that achieve relatively lower decoding accuracy (SVM-RBF, RF, MLP, LSTM, CNN, and ResNet), presented in figure 2 and table A3. Training time is the duration in clock time for training a model using all training samples, and inference time refers to the time used for each model to provide a prediction for one test sample. These attributes of LightGBM suggest its significant advantages in predicting avoidance behaviors in a closed-loop application where accuracy, speed, and memory efficiency are all critical [39]. Therefore, we selected LightGBM as the basis model for subsequent analyses in our study.

All ML models were conducted on a single core of AMD Ryzen Threadripper PRO 5995WX CPU by Python 3.8.16, scikit-learn 1.2.2, and Pytorch 2.0.1, and they were run on the same platform Red Hat Enterprise Linux 7.9. The implementation of LR, SVM-Lin, SVM-RBF, and RF was conducted using scikit-learn, while MLP, LSTM, CNN, ResNet, LSTM-Raw, WaveNet-Raw, and ResNet-Raw were implemented via Pytorch. LightGBM was implemented using the LightGBM package provided by Microsoft.

2.7. Model training and evaluation

Figure 1(d) illustrates the model training process using the dataset configuration detailed in figure 1(c). LightGBM models were trained in a subject-specific and session-specific manner, premised on the hypothesis that neural representations of defensive behaviors exhibit inter-subject variability. Furthermore, we fitted models separately for each recording session because we expected the neural encoding to shift over time. Tone-shock conditioning and extinction learning both involve significant plasticity in the IL-BLA circuit, and thus defensive behaviors might be driven by different activity patterns before vs. after a given stage of learning. For each subject and session, 5 LightGBM models were trained and assessed using the 5 folds designated for the training and validation sets, which were subsequently used for the selection of top-ranked features based on high feature importance values. A final LightGBM model was then trained using the aggregated training and validation sets and evaluated against the held-out test set, incorporating either band power features, selected top-ranked features (named as LightGBM-Top in the following analyses), or the entire set of extracted features.

The model's decoding performance was quantified using R^2 and r to compare ground truth with predicted behavioral measurements. The loss in R^2 served to evaluate the neuro-markers' contribution to decoding performance by their exclusion from the model. R^2 was also applied in the validation set's performance analysis to guide the selection of a specific number of top-ranked features. Additionally, r was also utilized to compare the similarities between feature importance matrices. The evolution of the training curves, delineated by the percentage change in L2 Loss (Mean Squared Error Loss) with increasing iterations, provided further insight into training dynamics.

2.8. Feature selection

Integrating an increased number of neuro-markers across spectral, temporal, and connectivity domains may enhance the decoding accuracy for defensive behaviors. However, this augmentation results in a proliferation of features, increasing computation time and memory requirements. Furthermore, some features may be uninformative or redundant within the ML framework, complicating the derivation of neuroscientific insights from models based on an extensive array of features. In our study, we extracted 17 types of neuro-markers, totaling 1296 features as input into the model, which inflated the computational costs unnecessarily. Consequently, we implemented a feature selection method to reduce computational demands, mitigate the risk of model overfitting, and identify which LFP features were most informative and, thereby, potentially causal to behavior.

We utilized SHapley Additive exPlanations (SHAP) for the assessment of feature importance among neuro-markers [128]. SHAP is a comprehensive measure of feature importance based on the Shapley values from a conditional expectation function of the original model. These values offer a unique feature importance metric that adheres to three desirable properties including local accuracy, missingness, and consistency when evaluating the additive attribution of one feature to the prediction output [128].

For each defensive behavior across every recording session, we assessed the SHAP values for every feature across the 5 LightGBM models, each trained using a distinct fold. Subsequently, for each feature, we computed the mean of its absolute SHAP values across all data instances and folds, establishing this as the cumulative contribution of the feature within that session. To identify the top-ranked features that are both subject- and behavior-specific and exhibit consistency across different days, these calculated attributions were further averaged over 4 recording sessions to determine the ultimate feature importance, as shown below:

$$\mathrm{imp}_i = \frac{1}{5NS}\sum_{s}\sum_{f}\sum_{n}\left|\phi_{n,f,s,i}\right| \qquad (2)$$

where imp_i is the importance of feature $i \in \{1,2,\ldots,M\}$, $\phi_{n,f,s,i}$ is the SHAP value for data instance $n \in \{1,2,\ldots,N\}$, fold $f \in \{1,2,\ldots,5\}$, session $s \in \{1,2,\ldots,S\}$, and feature i, and M, N, S are the numbers of features, samples, and sessions, respectively.

Features were subsequently ranked based on their SHAP importance in the training set. Then in the order of their rankings, we sequentially added the features into the feature set, as the input to the model. All models with increasing amounts of ranked features were subsequently trained on the training set and evaluated on the validation set using R^2. The peak validation performance is the highest average R^2 across 5 folds among these models. We implemented a paired-sample t-test between the average R^2 of these models and the peak performance, and identified the feature subset with a p-value no less than 0.05 in the significance test and with the minimum number of features as the selected feature set. The feature selection process therefore can be written as follows:

$$\begin{aligned}
\theta^*_{i,f} &= \arg\min_{\theta} \ L\left(h\left(x_{j_1}^{\mathrm{train}_f},\ldots,x_{j_i}^{\mathrm{train}_f}|\theta\right), y^{\mathrm{train}_f}\right)\\
R_i^2 &= \frac{1}{5}\sum_f R^2\left(h\left(x_{j_1}^{\mathrm{val}_f},\ldots,x_{j_i}^{\mathrm{val}_f}|\theta^*_{i,f}\right), y^{\mathrm{val}_f}\right)\\
R^2_{\max} &= \max_i \ R_i^2\\
i^* &= \min_i \ \left\{i : g\left(R_i^2, R^2_{\max}\right) \geqslant 0.05\right\}\\
\boldsymbol{X}^* &= \left[x_{j_1},\ldots,x_{j_{i^*}}\right] \qquad (3)
\end{aligned}$$

where $x_i^{\mathrm{train}_f}$ is the i^{th} feature in the training set of the f^{th} fold with $i \in \{1,\ldots,M\}$ and $f \in \{1,\ldots,5\}$. $\{x_{j_i}^{\mathrm{train}_f}\}$ are the sorted features such that $\mathrm{imp}_{j_1}^{\mathrm{train}_f} > \ldots > \mathrm{imp}_{j_M}^{\mathrm{train}_f}$, and $\{x_{j_i}^{\mathrm{val}_f}\}$ are the features in the validation set of the f^{th} fold sorted as in the training set. y^{train_f} and y^{val_f} are the target variables in the training and validation set of the f^{th} fold respectively. $h(\theta)$ is the model with parameters θ. $L(\bullet,\bullet)$ is the loss function, $R^2(\bullet,\bullet)$ is the coefficient of determination function, and $g(\bullet,\bullet)$ is the function of calculating p-value in paired-sample t-test. i^* is the amount of selected features, $\{x_i\}$ is the set of features, and $\boldsymbol{X}^*$ is the selected feature set.

2.9. Statistical analysis

We conducted paired-sample t-tests to assess the differences in decoding performance between accelerometry jerk, bar press rate, and freezing score. We applied the same method to determine whether decoding performance using selected top-ranked features was significantly different from the peak performance identified during feature selection. Additionally, paired-sample t-tests compared the SHAP values of features without or with various temporal delays, by employing neural features not only from the current window, but also from the preceding windows lagged by up to 20 s, across all recording sessions and subjects. Because there is an inherent motor delay between the perception of threat and the emission of defensive behavior, decoding might perform better if that delay were taken into account using this lagged method.

Independent-sample t-tests were utilized to determine the statistical significance of overall SHAP feature importance within specific frequency bands relative to all other bands, across all recording sessions and subjects. This test was also applied to evaluate the significance of feature contributions to decoding performance within specific frequency bands in comparison with contributions from all other frequency bands. The Wilcoxon signed-rank test was employed to compare decoding performance when using band power features, selected top-ranked features, and the entire set of extracted features across all subjects. To account for multiple comparisons, Bonferroni corrections were applied to adjust p-values, tailored to the number of comparisons conducted. The implementation of paired-sample t-tests, independent-sample t-tests, and the Wilcoxon signed-rank tests were carried out using the SciPy package.

2.10. Comparison with existing work using LightGBM for neural decoding

LightGBM is a well-established method in various fields, but its application within the context of predicting defensive responses for closed-loop systems in neuropsychiatric modulation presents unique challenges and opportunities. This section compares our innovative approach to the LightGBM application with existing methods, emphasizing the advancements we have made in decoding avoidance behaviors.

Most existing applications of LightGBM in neural decoding have typically focused on limited datasets, often emphasizing either spectral or temporal features but seldom integrating extensive connectivity measures [39, 98, 129–132]. Such studies include more straightforward applications in general classification tasks within neuroscience but lack the depth of neural correlates integration. Our study uniquely integrates a broad array of neuro-markers from three different domains-spectral, temporal, and connectivity. This comprehensive integration allows for a deeper understanding and modeling of neural dynamics associated with defensive behaviors, enhancing the model’s ability to predict and interact with complex neural phenomena.

While other studies may use LightGBM, they often employ more traditional feature selection techniques, such as recursive feature elimination or principal component analysis, which do not provide the same level of interpretability or direct linkage to neural dynamics [114, 129, 130, 133, 134]. Conversely, our approach utilizes SHAP to perform feature selection, enhancing interpretability and focusing on the most impactful features for model prediction. This methodology is innovative in its application, providing clear insights into how specific neuro-markers influence model outputs, thereby advancing the field towards more interpretable and effective neural decoding strategies.

Comparative analyses in existing works often focus solely on the neural decoding accuracy of ML models including LightGBM [131, 133, 135]. Our research, however, not only demonstrates the superior performance of LightGBM over traditional and some ANN models but also provides a detailed comparative analysis showing this advantage across multiple metrics (accuracy, training time, inference time, and memory size). These metrics are all critical factors for real-time decoding in closed-loop neuromodulation, highlighting our model's suitability for high-stakes, real-time applications in clinical settings.

In summary, while LightGBM is used across various domains for decoding, our application in the context of neuropsychiatric closed-loop systems leverages unique dataset characteristics, advanced feature selection methods, and rigorous comparative performance evaluation to significantly enhance the utility and efficacy of this tool in neuroscientific research.

3. Results

3.1. Comparison of decodability of defensive behaviors using proposed ML framework

The comparison of training processes and decoding performances for accelerometry jerk, bar press rate, and freezing score is depicted in figure 3. Figure 3(a) depicts the training curves, showcasing the L2 loss changes, averaged across subjects and recording sessions. The models underwent training using the training set, with the percentage change in L2 loss from the initial untrained state evaluated on both the training and validation sets. For all three behaviors, the L2 loss for training sets exhibited a consistent decline with additional iterations. However, the validation set loss for the freezing score demonstrated minimal improvement (−9.5%), in contrast to accelerometry jerk (−53.6%) and bar press rate (−34.6%).

There were large differences in the degree to which the different forms of defensive behavior could be decoded from the IL/BLA LFPs (i.e. in the degree to which these behaviors were encoded within the LFPs in that brain circuit). Specifically, the freezing score was only marginally decodable across sessions, with the mean coefficient of determination (R^2) averaged across subjects never surpassing 0.12 in all recording sessions, as shown in figure 3(b). While the bar press rate showed a higher degree of decodability, performances were slightly diminished during the conditioning session, attributed to strong bar press suppression resulting from foot shocks. Accelerometry jerk emerged as the most reliably decodable behavior, with the mean R^2 values across subjects consistently exceeding 0.46 in all recording sessions. Overall, decoding accuracy varied significantly among different defensive behaviors, following a descending order from accelerometry jerk to bar press rate to freezing score. These findings remained consistent when evaluated using both R^2 and the Pearson correlation coefficient (r) for performance assessment.

The variation in decoding performance may arise in part from the distinct characteristics of each behavioral signal, as illustrated in figure 3(c). Accelerometry jerk is characterized by a smoothly fluctuating signal that remains predominantly non-zero. In contrast, bar press rate often drops to zero but then has sharp deviations from baseline during bouts of pressing. Freezing score exhibits some local deviations even after smoothing. Regarding the freezing score in figure 3(c), the model succeeds in tracking the global trend, resulting in a relatively high r. Nevertheless, it struggles to capture local variations, leading to an R^2 of 0.071 for the freezing score. This indicates that the decoded behavior scarcely captures variance from the actual behavior, offering only marginal predictive improvement over the expected value of the true behavior. In light of these findings, subsequent analyses concentrated on accelerometry jerk and bar press rate, given that interpretations derived from the non-predictive models of freezing score could potentially be misleading.

3.2. Importance and contribution of neuro-markers to the decoding performance

Subsequently, our focus shifted towards understanding the importance of each neuro-marker type in decoding defensive behaviors. Figures 4(a) and (b) delineate the importance of three feature domains, diverse neuro-markers, and frequency bands in decoding defensive behaviors. A notable observation is that spectral, temporal, and connectivity features all play a crucial role in decoding defensive behaviors. Specifically, temporal (43.0%) and spectral (41.1%) features outweigh connectivity features (15.9%) for the prediction of accelerometry jerk. In contrast, for bar press rate prediction, connectivity (39.8%) emerges as the predominant domain, surpassing spectral (34.7%) and temporal (25.5%) features. Among the individual types of neuro-markers for accelerometry jerk decoding, band power (BP) (33.2%), line length (LL) (21.0%), and band power ratio between channels (BPRC) (10.2%) stand out as the most influential features within the spectral, temporal, and connectivity domains, respectively. This holds true despite the availability of a larger number of connectivity features compared to spectral or temporal features, owing to connectivity's reliance on the squared number of channels. For bar press rate, BP (18.9%) and BPRC (15.5%) consistently rank as critical, with other spectral and connectivity features like band power ratio between bands (BPRB) (11.2%) and band Pearson correlation (BCorr) (8.2%) also contributing substantially to predictions, unlike other temporal features. Notably, for the leading contributors (BP and BPRC) as well as other neuro-markers

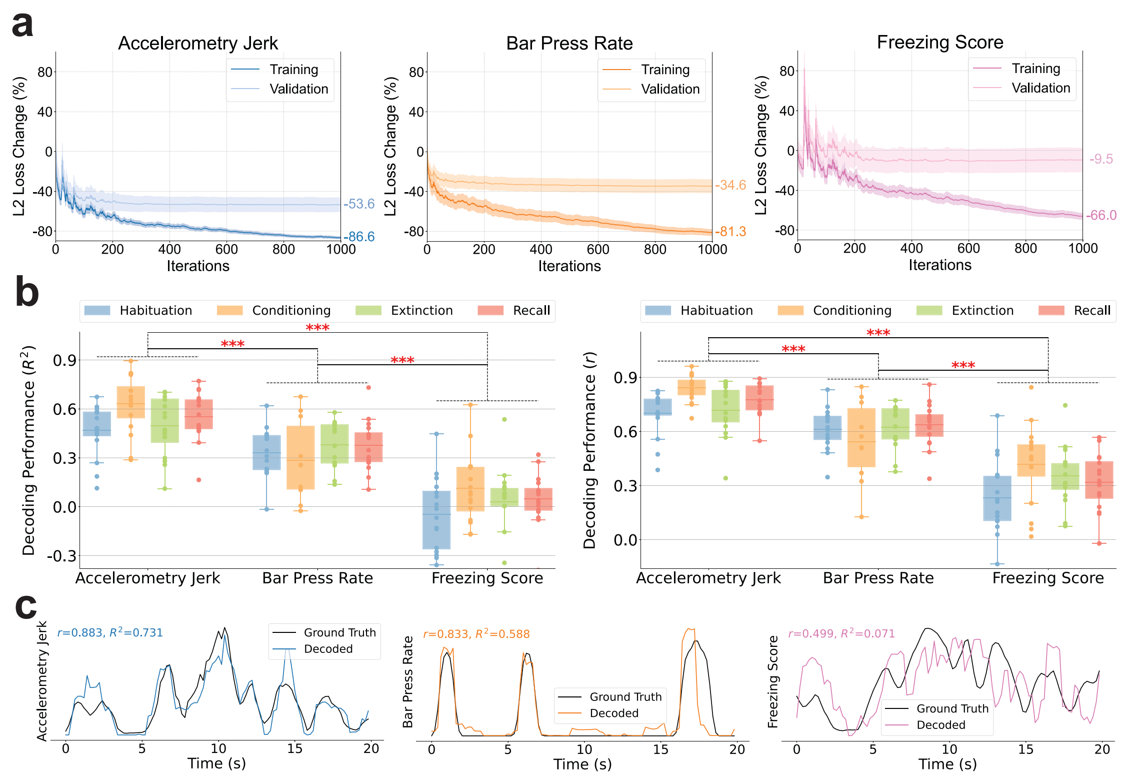

Figure 3. Decoding performances for accelerometry jerk, bar press rate, and freezing score. (a) Comparisons of training curves of our ML model for decoding accelerometry jerk (Left), bar press rate (Middle), and freezing score (Right). The percentages of the change of L2 Loss on the training and validation sets are shown with the increasing number of iterations during the training process. The shading areas indicate standard errors across subjects. (b) Decoding performance for accelerometry jerk, bar press rate, and freezing score in four recording sessions averaged across subjects. The line in the boxplot shows the average performance across subjects, and each dot indicates the result of an individual subject. The performances were evaluated using the coefficient of determination (Left, R^2) and the Pearson correlation coefficient (Right, r). The asterisks denote the significant difference in the decoding performance of two defensive behaviors. (Paired-sample t-test; $***: p < 0.001$) (c) Decoding examples for accelerometry jerk (Left), bar press rate (Middle), and freezing score (Right), from a single rat and session (OB44, Habituation), with performance evaluated using R^2 and r.

that span seven frequency bands, including BCorr, coherence (Coh), and phase locking value (PLV), their high-gamma components are identified as crucial for decoding defensive behaviors, except Coh for accelerometry jerk and PLV for bar press rate, which prominently feature alpha and gamma components, respectively.

Beyond quantifying feature importance by evaluating their attribution to the prediction, we also explored their impact on decoding performance, as illustrated in figures 4(c) and (d). For accelerometry jerk, BP, LL, and BPRC were identified as principal contributors, aligning with their established predictive importance in figure 4(a). The order of neuro-marker contributions to accelerometry jerk decoding performance as shown in figure 4(c) mirrors their predictive significance as depicted in figure 4(a). In the case of bar press rate, BCorr, BPRC, and BPRB maintain a substantial impact on decoding performance, consistent with figure 4(b). However, BP's contribution appears noticeably diminished relative to the aforementioned features, underscoring its reduced spectral significance in comparison with BPRB for bar press rate decoding.

This analytical approach to feature importance in both prediction and decoding performance elucidates the substantial importance of neuro-markers across all three domains. BP and BPRC emerge as common key contributors for decoding both defensive behaviors, with LL for accelerometry jerk and BCorr and BPRB for bar press rate also deemed important in terms of prediction and performance.

3.3. Importance and contribution of band powers in different frequency bands to the decoding performance

In figure 4, band power emerged as one of the most influential features. We delved deeper into its importance in terms of prediction and decoding performance across seven frequency bands, including delta, theta, alpha, beta, low-gamma, gamma, and high-gamma, extracted from both IL and BLA, for the decoding of accelerometry jerk and bar press rate, as detailed in figure 5. The importance matrices in figures 5(a)–(d) highlight the importance of band power in these frequency bands across recording sessions, brain regions, and targeted behaviors. Collectively, these matrices consistently

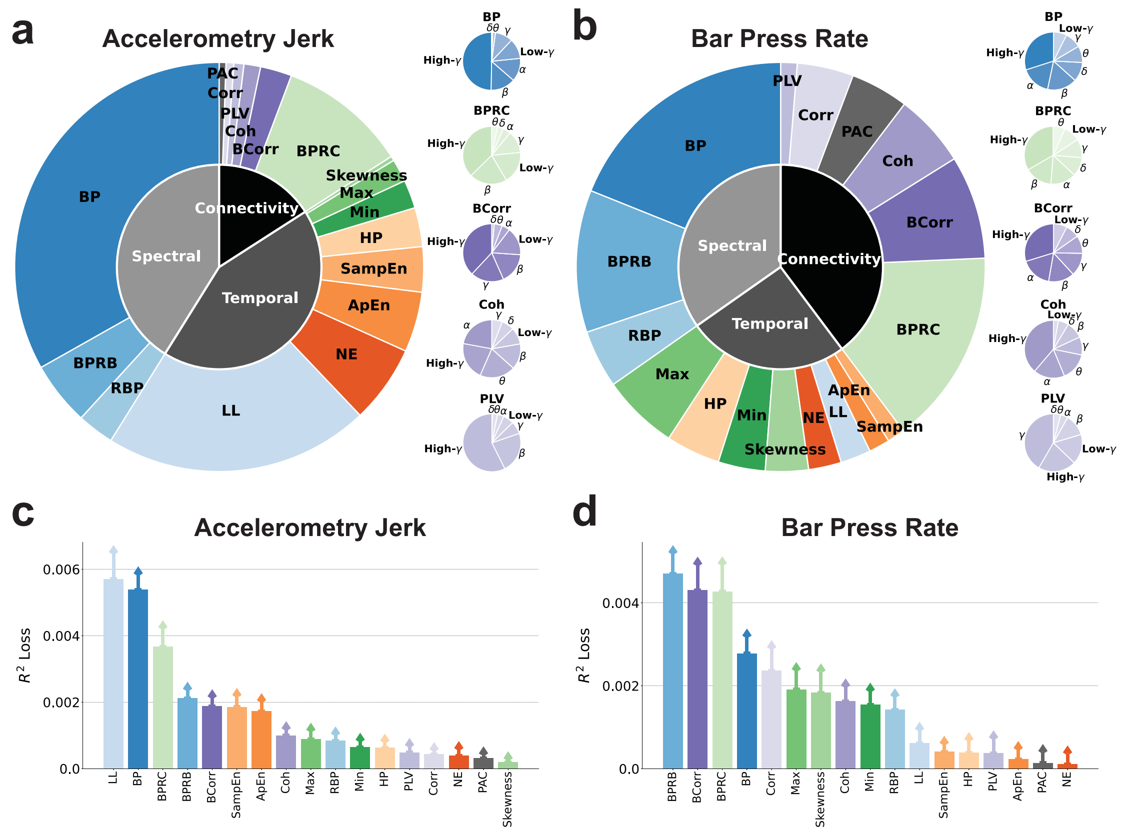

Figure 4. The importance of neuro-markers during decoding process. (a) (Left) The importance of various types of neuro-markers for decoding the accelerometry jerk is illustrated in the outer chart. The importance of markers in the spectral, temporal, and connectivity domains is shown in the inner chart. (Right) The importance of band power, band power ratio between channels, band Pearson correlation, coherence, and the phase locking value in different frequency bands is shown in the small charts. (b) Same as in (a), but for the bar press rate. (c) Contribution of various types of neuro-markers to the decoding performance for accelerometry jerk, averaged across subjects and recording sessions and evaluated using the R^2 loss after removing each type of markers from the ML input. Error bars indicate the standard errors across subjects and sessions. (d) Same as in (c), but for bar press rate.

reveal that high-gamma power in the IL and BLA is more important for predicting behavior than all other frequency bands, and that this is true across different phases of aversive learning and extinction. Figures 5(e)–(h) compare the pairwise similarities among the elements of the importance matrices from figures 5(a)–(d), examining either the significance of spectral power within identical bands and sessions across different brain regions or in decoding diverse defensive behaviors. These importance matrices exhibit substantial correlation with each other ($r > 0.61$, $p < 6.0e-4$), with the high-gamma components invariably displaying elevated importance values. This pattern suggests that high-gamma power maintains a consistent association with defensive behavior across various contexts.

Expanding our analysis to consider the band powers from another angle, we explored their impact on decoding performance across seven frequency bands, as depicted in figures 5(i) and (j). Notably, the exclusion of high-gamma power leads to a significantly stronger decline in model performance across subjects and sessions compared with all other bands, aligning with observations from figures 5(a)–(h). Therefore, the comprehensive findings of figure 5 underscore the pivotal role of high-gamma power as the spectral band most closely linked to defensive behavior, both in terms of attribution to prediction and decoding performance.

3.4. Importance and contribution of cross-region neuro-markers in different frequency bands to the decoding performance

In figure 4, the band power ratio between IL and BLA emerged as a pivotal feature, especially in the context of bar press rate decoding. We dissected the relative contribution of different frequency bands as depicted in figure 6. Here again, high gamma features were identified as the most influential encoders of defensive behaviors. Additionally, beta band ratios from BLA to IL exhibited marginal significance for accelerometry jerk, as illustrated in figure 6(c). Figures 6(e)–(h) explore the pairwise similarities among the elements of the importance matrices from figures 6(a)–(d), assessing

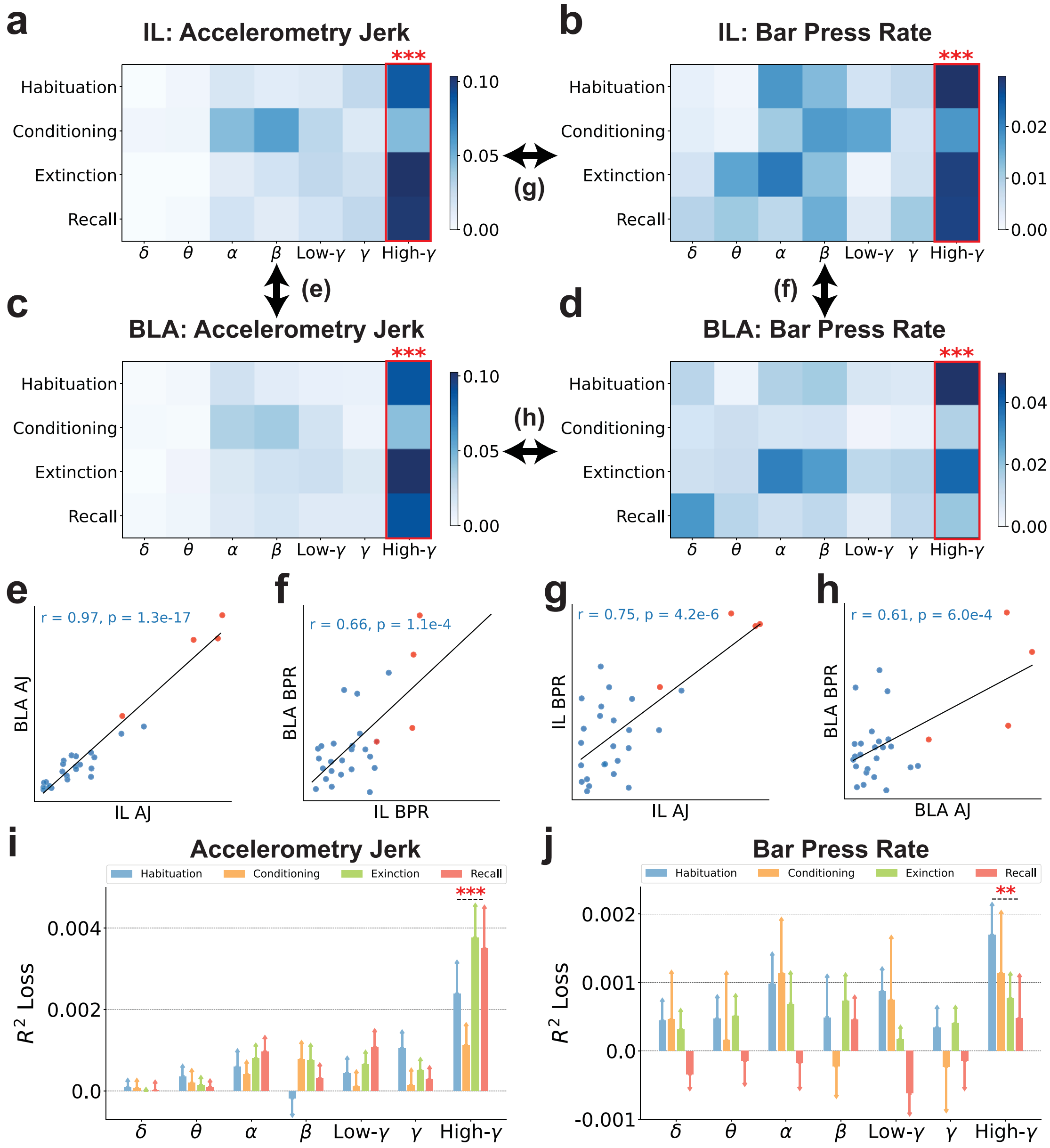

Figure 5. The high-gamma band powers are generally more predictive than other bands for decoding defensive behaviors. (a)–(d) The importance of band powers (BP) in seven frequency bands and four recording sessions is illustrated in the importance matrices, where each element is the SHAP values averaged across subjects. Bands with significantly higher importance than the other bands are marked with red asterisks. (Independent-sample t-test; $* * * : p < 0.001$). (a) BP in IL for decoding accelerometry jerk. (b) BP in IL for decoding bar press rate. (c) BP in BLA for decoding accelerometry jerk. (d) BP in BLA for decoding bar press rate. (e)–(h) The similarities between the importance matrices were evaluated using the Pearson correlation coefficient (r) and p-value (p). Each dot indicates its importance in the same band and the same session between different brain regions or defensive behaviors. Red points denote the importance of high-gamma powers. (e) BP in IL and BLA for decoding accelerometry jerk (AJ). (f) BP in IL and BLA for decoding bar press rate (BPR). (g) BP in IL for decoding AJ and BPR. (h) BP in BLA for decoding AJ and BPR. (i)–(j) The contribution of BP in seven frequency bands to the decoding performance, averaged across subjects, was evaluated using the R^2 loss after removing each band from the ML inputs. The error bars indicate the standard errors across subjects. Bands with significantly higher contributions than the other bands are marked with asterisks. (Independent-sample t-test; $** : p < 0.01, * * * : p < 0.001$). (i) The contribution of BP to the decoding performance for accelerometry jerk. (j) The contribution of BP to the decoding performance for bar press rate.

either the importance of spectral power ratios within identical bands and sessions across two reciprocal ratios (IL/BLA and BLA/IL) or in decoding various defensive behaviors. These comparisons revealed significant similarities ($r > 0.67,\ p < 1.1e-4$). High-gamma power ratios were distinctly more important than other bands in various analyses presented in figures 6(e)–(h). This comprehensive analysis indicates a clear concordance in the significance of high-gamma power ratios between IL and BLA, aligning with the patterns of importance outlined in figures 6(a)–(d).

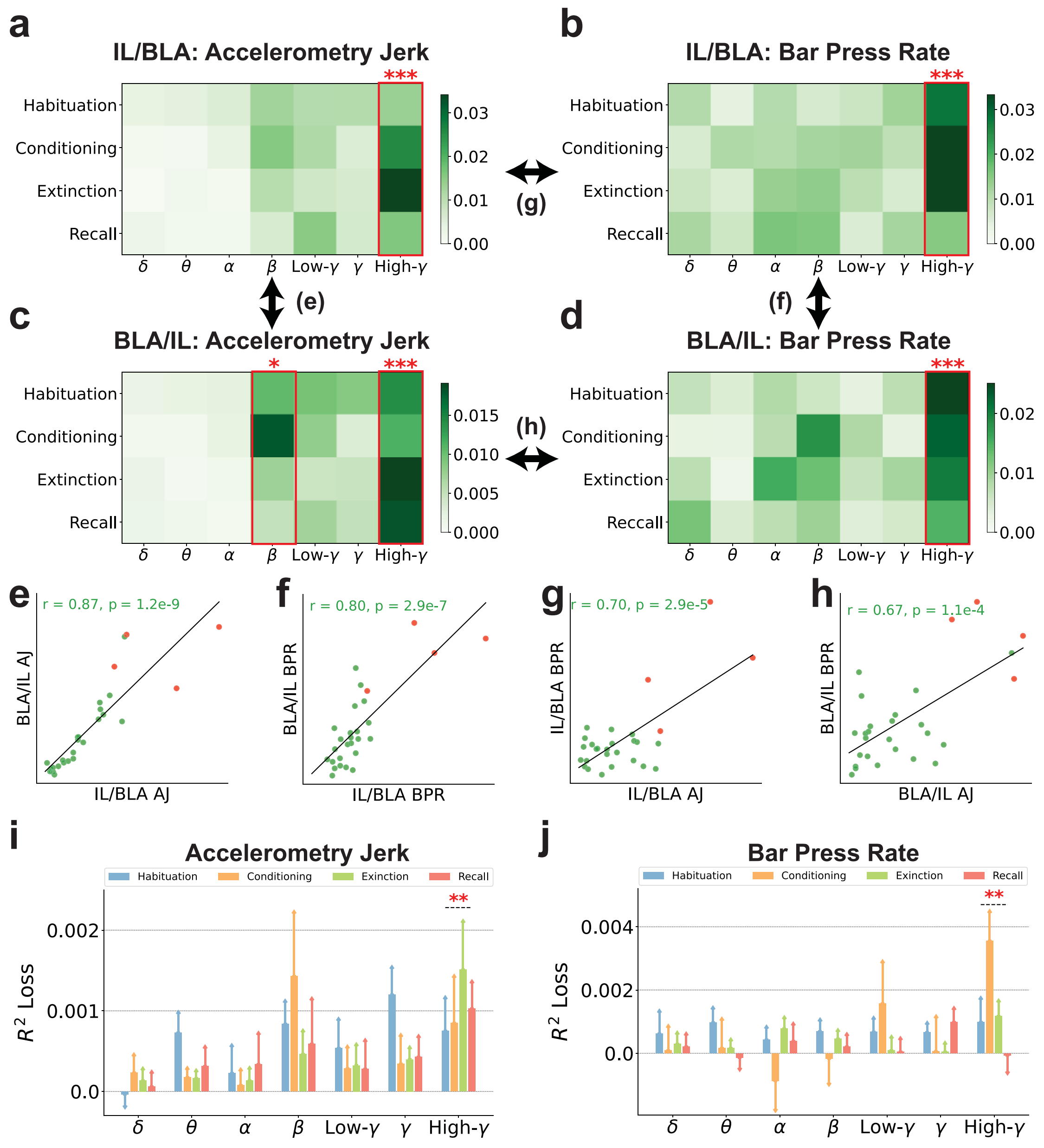

Figure 6. The high-gamma band power ratios between IL and BLA are generally more predictive than other bands for decoding defensive behaviors. (a)–(d) The importance of band power ratios between IL and BLA (BPRC) in seven frequency bands and four recording sessions is illustrated in the importance matrices, where each element is the SHAP values averaged across subjects. Bands with significantly higher importance than the other bands are marked with red asterisks. (Independent-sample t-test; $* : p < 0.05, *** : p < 0.001$). (a) BPRC of IL to BLA (IL/BLA) for decoding accelerometry jerk. (b) IL/BLA for decoding bar press rate. (c) BPRC of BLA to IL (BLA/IL) for decoding accelerometry jerk. (d) BLA/IL for decoding bar press rate. (e)–(h) The similarities between the importance matrices were evaluated using the Pearson correlation coefficient (r) and p-value (p). Each dot indicates its importance in the same band and the same session between different brain regions or defensive behaviors. Red points denote the importance of high-gamma power ratios. (e) BPRC of IL to BLA (IL/BLA) and BPRC of BLA to IL (BLA/IL) for decoding accelerometry jerk (AJ). (f) IL/BLA and BLA/IL for decoding bar press rate (BPR). (g) IL/BLA for decoding AJ and BPR. (h) BLA/IL for decoding AJ and BPR. (i)–(j) The contribution of BPRC in seven frequency bands to the decoding performance, averaged across subjects, was evaluated using the R^2 loss after removing each band from the ML inputs. The error bars indicate the standard errors across subjects. Bands with significantly higher contribution than the other bands are marked with asterisks. (Independent-sample t-test; $** : p < 0.01$). (i) The contribution of BPRC to the decoding performance for accelerometry jerk. (j) The contribution of BPRC to the decoding performance for bar press rate.

To gain further insights into the band power ratios, we examined their impact on decoding performance, as illustrated in figures 6(i) and (j). High-gamma power ratios consistently led to the most substantial decrease in performance across subjects and sessions when excluded from the ML model. Thus, high-gamma power ratios are critical to decoding performance for both accelerometry jerk and bar press rate, surpassing the impact of all other frequency bands.

The band power ratios reveal variations in the activation levels between IL and BLA, offering

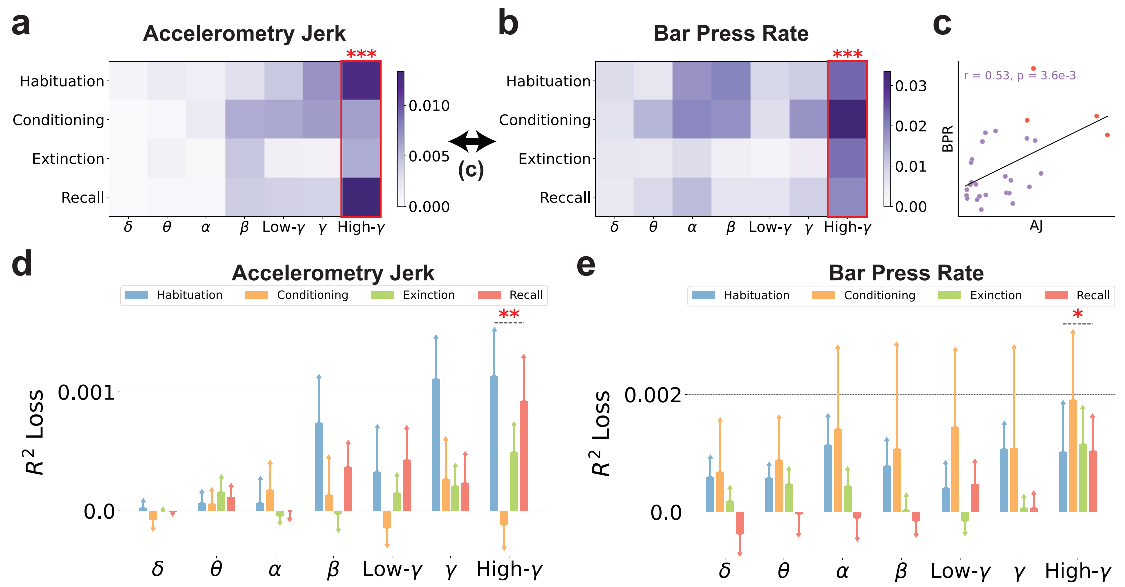

Figure 7. The high-gamma band Pearson correlations between IL and BLA are generally more predictive than other bands for decoding defensive behaviors. (a) and (b) The importance of band Pearson correlations between IL and BLA (BCorr) in seven frequency bands and four recording sessions is illustrated in the importance matrices, where each element is the SHAP values averaged across subjects. Bands with significantly higher importance than the other bands are marked with red asterisks. (Independent-sample t-test; $*** : p < 0.001$). (a) BCorr for decoding accelerometry jerk. (b) BCorr for decoding bar press rate. (c) The similarity between the importance matrices for decoding accelerometry jerk and bar press rate was evaluated using the Pearson correlation coefficient (r) and p-value (p). Each dot indicates its importance in the same band and the same session between different defensive behaviors. Red points denote the importance of high-gamma correlations. (d) and (e) The contribution of BCorr in seven frequency bands to the decoding performance, averaged across subjects, was evaluated using the R^2 loss after removing each band in the ML inputs. The error bars indicate the standard errors across subjects. Bands with significantly higher contributions than the other bands are marked with asterisks. (Independent-sample t-test; $* : p < 0.05$, $** : p < 0.01$). (d) The contribution of BCorr to the decoding performance for accelerometry jerk. (e) The contribution of BCorr to the decoding performance for bar press rate.

insights into their differential engagement during defensive behaviors. These ratios allow researchers to deduce the degree of synchronization and the dynamic interactions between IL and BLA. However, it is important to note that band power ratios alone do not directly quantify the functional connectivity of these regions. Consequently, we further explored the Pearson correlations between neural signals of IL and BLA, evaluating their significance for prediction and impact on decoding performance across various frequency bands, as depicted in figure 7. In this analysis, correlations within the high-gamma frequency band emerged as the most informative features, outperforming those of other frequency bands in decoding both accelerometry jerk and bar press rate, as shown in figures 7(a) and (b). Figure 7(c) demonstrates the similarity between the elements of the importance matrices from figures 7(a) and (b), revealing a significant correlation ($r = 0.53, p = 3.6e-3$). We further explored the impact of band Pearson correlations on decoding performance, as depicted in figures 7(d) and (e). High-gamma correlations consistently led to the most significant decline in performance when excluded from the model.

Collectively, band power ratios and band Pearson correlations elucidate the neural representations between IL and BLA through distinct lenses. Therefore, the findings presented in figures 5–7 together show that, across spectral and connectivity domains, oscillations in the high-gamma range within and between IL and BLA are the most reliable encoder of defensive behaviors. Thus, within the scope of our study, these features appear to be the most reliable among the spectral, temporal, and connectivity neuro-markers used for decoding avoidance behaviors in a closed-loop paradigm.

3.5. Feature selection chooses important neuro-markers and maintains decoding performance

In this study, we introduced 17 types of neuro-markers as features, yielding a total of 1296 features for inclusion in our ML framework. Incorporating all these features would lead to increased computational and memory demands. Feature selection is a widely recognized strategy for mitigating the computational burden of cognitive decoders [69, 136]. Figure 8 explores the impact of feature dimensionality on decoding performance and the proportion of various types of neuro-markers among the selected features. Figure 8(a) presents the feature selection process based on feature importance as quantified by SHAP values. Here, the decoding accuracy on the validation sets, averaged across subjects, is

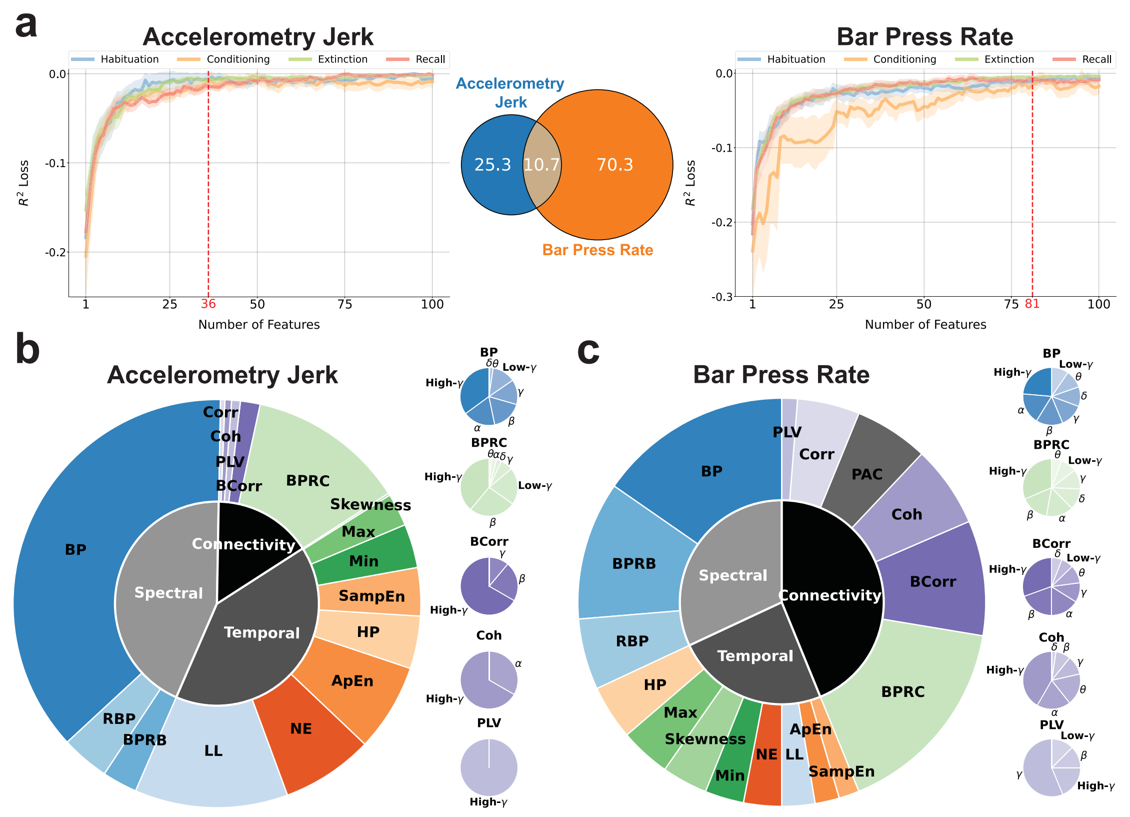

Figure 8. Neuro-markers with higher feature importance and contribution to the decoding performance were chosen through feature selection. (a) Feature selection using feature importance evaluated by SHAP values for decoding the accelerometry jerk (Left) and bar press rate (Right). Features were iteratively selected according to the decreasing order of SHAP feature importance. Shading areas indicate the standard errors across subjects. 36 and 81 top-ranked features were selected from all 1296 features respectively. The performance saturates after 100 features. Venn diagram represents the overlap of top-ranked features for the accelerometry jerk and bar press rate (Middle). (b) (Left) The proportion of different types of neuro-markers among the selected top-ranked features for decoding the accelerometry jerk is illustrated in the outer chart. The proportion of selected features in the spectral, temporal, and connectivity domains is shown in the inner chart. (Right) The proportion of band power, band power ratio between channels, band Pearson correlation, coherence, and the phase locking value in different frequency bands are shown in the small charts. (c) Same as in (b), but for decoding the bar press rate.

depicted in relation to the quantity of top-ranked features. Notably, performance improves with an increasing number of selected features, reaching a plateau at approximately 100 features. By selecting only 36 and 81 top-ranked features, we observed that decoding accuracy on validation sets across all recording sessions was comparable to, and not significantly inferior to, the peak performance identified through an exhaustive exploration of all possible counts of top-ranked features (Paired-sample t-test; accelerometry jerk: $p = 1.1e-1$, bar press rate: $p = 6.1e-2$. See section 2.8). These findings underscore the feasibility of dramatically reducing feature dimensionality by 97.2% (36 out of 1296) and 93.8% (81 out of 1296) without significantly compromising decoding efficacy. Within the subset of 36 and 81 top-ranked features selected for the decoding of accelerometry jerk and bar press rate, respectively, an average of 10.7 features are concordant and can predict both defensive behaviors across subjects.

Figure 8(b) and (c) delineate the distribution of different types of neuro-markers within the selected features, aligning with the previously established importance of these markers regarding prediction and performance as depicted in figure 4. In the case of accelerometry jerk, spectral (43.8%) and temporal (40.6%) features were more frequently selected over connectivity features (15.6%). BP (37.2%), LL (12.2%), and BPRC (12.5%) emerged as the predominant feature groups within the spectral, temporal, and connectivity domains, respectively, as shown in figure 8(b). Conversely, for bar press rate, as illustrated in figure 8(c), the model exhibited a preference for selecting connectivity features (43.8%) over spectral (31.9%) and temporal (24.2%), with BPRC (16.2%), BP (15.4%), BPRB (11.0%), and BCorr (9.1%) identified as leading predictors. Across all neuro-markers that span 7 frequency bands, including BP, BPRC, BCorr, Coh, and PLV, high-gamma components were most frequently chosen for decoding both defensive behaviors, with the exception of PLV for bar press rate, where the gamma component was more prominently featured.

Since there can be a lag between decisions and manifested behavior, we also evaluated the decoding

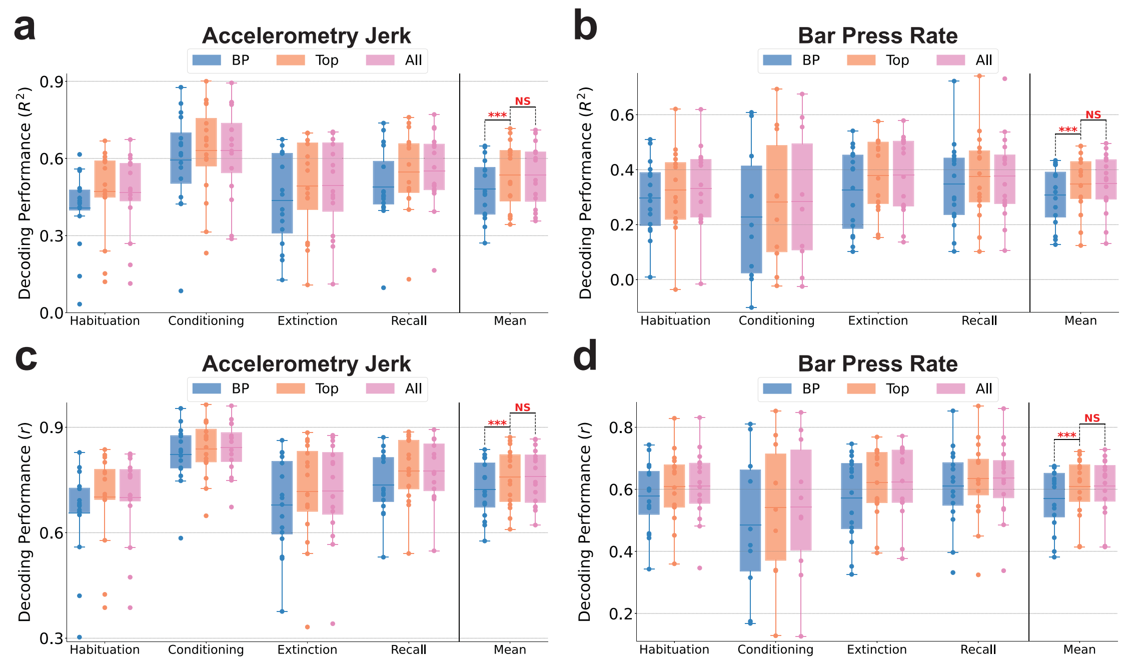

Figure 9. The decoding performance was improved by using more than one type of neuro-marker, but remained good when only high-importance features were used for training. (a) The comparisons of decoding performance for the accelerometry jerk using band powers (BP), selected top-ranked features (Top: LightGBM-Top), and all features (All) in four recording sessions averaged across subjects, evaluated using the coefficient of determination (R^2). Mean denotes the performance averaged across sessions. The line in the boxplot shows the average performance across subjects, and each dot indicates the result of an individual subject. The comparisons in Mean were marked with asterisks if one averaged performance was significantly greater than the other. (Wilcoxon signed-rank test; $*** : p < 0.001$, $NS : p \geqslant 0.05$) (b) Same as in (a), but for decoding the bar press rate. (c) Same as in (a), but evaluated using the Pearson correlation coefficient (r). (d) Same as in (c), but for decoding the bar press rate.

results using lagged neural data, as presented in figure A1. Figures A1(a) and (b) illustrate that neural features temporally close to the current time point yield superior decoding accuracy for both behaviors, suggesting these features encapsulate a richer neural representation regarding defensive responses than those from earlier time windows. As shown in figure A1(c), the inclusion of features from preceding time windows together with current features does not markedly enhance the decoding performance for accelerometry jerk. In contrast, figure A1(d) indicates a modest improvement in the decoding of bar press rate when previous time window features are incorporated. Furthermore, figures A1(e) and (f) indicate that the predictive power of features for both behaviors is predominantly concentrated in recent time windows, with a significant decline in predictivity as the temporal gap widens. The findings indicate that neural representations closest to the event of interest are most informative for decoding both defensive behaviors, with immediate past features contributing more significantly to model accuracy than older ones. We thus have emphasized the importance, in preceding and subsequent analyses, of features aligned to behavior with zero lag. The demonstrated temporal gradient in feature predictivity could inform the development of a more refined real-time decoder for neuropsychiatric interventions.

In figure 9, we explore the dependency of decoding performance on the diverse types of neuro-markers employed and the dimensionality of the feature set. This comparison is made between decoding outcomes utilizing only conventional band powers, decoding with a selected group of features as identified in figure 8, and decoding with the entire set of extracted features. Figures 9(a)–(d) present the decoding performance for accelerometry jerk and bar press rate using band power features, selected top-ranked features (LightGBM-Top), and all features, assessed by R^2 and r metrics, respectively. The addition of other neuro-markers beyond only band power, coupled with feature selection, significantly enhances performance across sessions (for accelerometry jerk, R^2 from 0.4815 to 0.5357, r from 0.7229 to 0.7579; for bar press rate, R^2 from 0.3073 to 0.3476, r from 0.5708 to 0.6092). Moreover, employing the limited feature set as delineated in figure 8 does not lead to a significant reduction in performance when compared to the utilization of all features. This observation holds true for the decoding of both defensive behaviors evaluated by both R^2 and r metrics. Notably, the adoption of feature selection exceptionally reduces the model's training time (182.3 ms for accelerometry jerk, 309.6 ms for bar press rate), inference time (0.05 079 ms for accelerometry jerk, 0.05 072 ms for bar press rate), and memory usage (16.6 kB) across all subjects and sessions (table A3).

Collectively, the results presented in figures 8 and 9 underscore that the feature selection process effectively identifies important features with significant additive attribution to the prediction output and remarkable contribution to the performance in decoding defensive behaviors. Through this process, a select group of top-ranked features not only sustains decoding performance with remarkably reduced training/inference time and memory usage but also enhances performance in comparison to relying exclusively on band power features.

4. Discussion

We developed a machine learning framework for accurately decoding defensive behaviors from multi-channel local field potentials recorded from the infralimbic cortex and basolateral amygdala. Critically, accelerometry jerk and bar press rate exhibited higher decodability compared to the freezing score, as evidenced by both the training dynamics and performance evaluations on the test set (figure 3). These two decodable behaviors were encoded by distinct sets of highly informative features (figures 4 and 8).

This research builds upon our previous work, which underscored that these metrics each capture unique facets of defensive behavior [89]. The variation in encoding between behaviors suggests that they may have distinct neural substrates, i.e. that a closed-loop system designed to modulate defensive processes might need to control different aspects of cortical/amygdala physiology depending on the exact process being targeted. The challenge in accurately decoding the freezing score—conceptually the inverse of freezing and calculated from video frame changes to approximate the rat's horizontal velocity—is intriguing. Given its mathematical relationship with accelerometry jerk, which essentially represents a higher derivative of movement than freezing score, this difficulty is unexpected. On the other hand, considering that mammalian motor control often optimizes for minimum jerk [137], it stands to reason that such dynamics are more directly encoded in the neural circuitry that eventually affects motor planning. Further, as noted in section 2.3, accelerometry may simultaneously capture passive defense (freezing) and active defense (darting behaviors [92]), and thus might be more directly correlated to signals in threat/defense-related circuits.

Freezing, as derived from the freezing score, is probably the single most common behavior used to study the IL, BLA, and the broader circuits of the extended amygdala [85, 138, 139]. Its association with various LFP processes, particularly emphasizing local oscillations and cross-regional synchrony within the theta band, is well-documented [83, 93]. Therefore, our inability to decode this behavior accurately presents a notable discrepancy. One possible explanation for this difference could be our focus on decoding second-to-second changes in behavior, in contrast to previous studies that typically examined longer timescales, such as the percentage of a cue tone spent in freezing versus other behaviors [80]. As illustrated in figure 3(c), our decoders demonstrated better performance in capturing these broader timescales (trends or global means) than short-term variability in freezing score. This observation aligns with our previous behavioral research, which indicated that the mean freezing score across subjects correlated more closely with the mean accelerometry jerk and bar press rate, than when analyzing individual subjects [89]. This may be attributed to the averaging process across subjects, which effectively smoothed away local variance while preserving global trends, thereby rendering freezing score more comparable with other measured behaviors.

Beyond the conventional use of band power features for decoding cognitive and emotional processes [38, 69, 73], our model incorporates a broader array of neuro-markers across spectral, temporal, and connectivity domains. Temporal features demonstrated a particularly significant contribution to decoding accelerometry jerk over bar press rate, as evidenced in figures 4(a) and (b). This disparity likely stems from the capability of temporal-domain features to capture changes over very short intervals, reflecting the dynamic and swift variations in the defensive behaviors that define the accelerometry jerk data. Interestingly, connectivity features played a more pronounced role in decoding bar press rate compared to accelerometry jerk. This distinction may reflect the difference in behavioral characterization underpinning these behaviors; unlike accelerometry jerk, bar press rate involves the suppression of a reward-seeking response, diverging from motion-based defensive behaviors like freezing. Hence, prior studies linking defensive behaviors with theta oscillations and cross-regional LFP connectivity may more accurately depict variations in reward-related processes. A noteworthy finding is that, alongside coherence (Coh), significant decoding insights were derived from the band power ratio between channels (BPRC) and band Pearson correlation (BCorr). Thus, BPRC and BCorr warrant increased consideration over Coh in subsequent fear regulation research. We have demonstrated that these features encompass unique information not captured by band power alone [38, 74, 110, 140].

The high-gamma band was particularly important for decoding accelerometry jerk and bar press rate in BP, BPRC, and BCorr (figures 5–7). This finding contrasts with earlier research, where fear-related behavior was primarily correlated with theta band power and sycnhrony [83, 93]. The divergence in findings could stem from our distinct analytical methodology. Whereas previous studies often explored categorical differences, such as contrasting animals showing low versus high freezing behavior in a dichotomized analysis, our approach aimed to directly predict behaviors within individual animals and sessions. Within this shorter timescale, the involvement of faster processes, like those within the high-gamma range, may become more pivotal. We also used different electrodes, with tighter spacing that emphasizes local signals within IL and BLA. This again would emphasize more spatially local high-frequency components over more spatially distributed low-frequency LFPs. However, this emphasis on local signals more realistically models a clinical scenario, where electrodes would be implanted within a relatively small brain region.

Other studies have attempted to decode/predict defensive behavior from similar circuits, but with very different goals or methods. For instance, [64] attempted to predict freezing from LFP, but at a trial-to-trial level, i.e. predicting the percentage of freezing within a multi-second window from the 4 Hz LFP power during that same time window. This is a much different and more forgiving problem than we attempted here, because the wide temporal windows will smooth away noise. Further, the ability to predict behavior at fine timescales may be clinically relevant. A second study [59] predicted active escape (shuttle runs) using ensemble analyses of recorded single neurons (potentially more informative but more costly than LFP; see Introduction). When the authors attempted to inhibit defensive behavior through optogenetic manipulation, they found that shuttle runs could only be blocked if light were delivered during the period when an animal actively decided to respond. Delivering light alongside the threat cue, but not during behavior preparation, delayed but could not eliminate the behavior. Further, neither of those studies focused on optimizing a decoder for maximum performance with a minimal set of features. They used decoding methods as a means towards mechanistic explanation, as opposed to our emphasis on showing performance as a step towards clinical utility.

Through our feature selection process, we strategically chose a limited subset of features to minimize the computation time and memory demands of our ML framework. Utilizing only 36 and 81 top-ranked features, as depicted in figure 8, we not only significantly surpassed the decoding performance achieved with 56 BP features but also matched the performance obtained with the full set of 1296 features, as demonstrated in figure 9. This indicates that neuro-markers other than BP encode unique information critical for decoding. The analytical findings from figures 4–7 further support that incorporating a broader spectrum of neural representations enhances decoding effectiveness, offering a more nuanced insight into neuro-markers' roles in modulating defensive behaviors. Additionally, our results imply the existence of a considerable number of features that are either non-predictive or redundant within the model. The feature selection process effectively eliminates less informative features for each subject, thereby significantly reducing computational expenses during training and inference phases and lowering memory requirements. These efficiencies, combined with the high decoding accuracy, underscore the importance of an optimized feature selection strategy for neural decoders in neuropsychiatric brain-machine interfaces (BMIs).

Advanced machine learning models have been shown to markedly enhance neural decoding performance over conventional approaches. In our investigation, we assessed the decoding capabilities of state-of-the-art models in neural decoding tasks using our extracted neuro-markers, including LR, SVM-Lin and SVM-RBF, RF, MLP, LSTM, CNN, ResNet, and Light Gradient Boosting Machine (LightGBM). Building on our prior research on seizure detection [97, 98, 141], mental fatigue prediction [75], finger movement classification [114, 142], and tremor detection from electrophysiological signals [104, 143], gradient-boosted decision tree models (GBDT) including LightGBM were found to outperform traditional ML models, including SVM and linear discriminant analysis (LDA). Our findings further reveal that LightGBM was the best-performing model in 14 out of 16 comparisons across decoding tasks, as shown in 2. Although RF performed slightly better than LightGBM in decoding accelerometry jerk during the conditioning session as per R^2 and in decoding bar press rate as per r, LightGBM demonstrated significantly shorter training times (accelerometry jerk: 5.336 s, bar press rate: 5.130 s) and inference times (accelerometry jerk: 0.06 006 ms, bar press rate: 0.05 966 ms) compared to RF (training times: accelerometry jerk: 176.8 s, bar press rate: 179.3 s; inference times: accelerometry jerk: 1.765 ms, bar press rate: 1.772 ms), as presented in figure 2 and table A3. Overall, LightGBM achieved superior efficiency in training/inference and memory usage compared with other models with promising but lower decoding accuracy, as shown in figure 2. These results underscore LightGBM's capability to deliver both precise decoding outcomes and remarkably rapid decoding speeds with limited memory resources, which are the key qualities for a decoder within a closed-loop BMI system. Furthermore, LightGBM offers additional advantages over other ML models: it supports parallel computation, greatly speeding up training and inference processes. Importantly, it exhibits low hardware complexity, as demonstrated in recent low-power hardware implementations of closed-loop neuromodulation systems [97, 105, 144]. Collectively, these attributes underscore LightGBM's potential applicability in future fully-implantable and closed-loop psychiatric BMIs.

Although we demonstrated that LightGBM outperforms artificial neural networks (ANNs) in accuracy for decoding defensive behaviors using neuromarkers, this comparison may overlook the ANNs' capabilities for automatic feature extraction and the critical information in raw LFP signals. Recurrent neural networks such as LSTM could identify concealed temporal dependencies, while CNNs excel in decoding the mixed spatial and temporal information embedded in neural representations. Consequently, we utilized several modern ANNs, including LSTM-Raw, WaveNet-Raw, and ResNet-Raw, with raw LFP signals as inputs for decoding defensive behaviors. These architectures provide unique approaches to feature extraction, unbounded by predefined neuromarkers. Findings in tables 2 and A2 indicate that the overall performance of these ANN-Raw models is generally inferior to that of models employing manually selected neuro-markers. This discrepancy underscores a critical insight: although ANNs can autonomously extract features from complex neural data, the relevance and utility of these features for specific decoding tasks may be limited without the guided feature selection afforded by manual methods. Models that use neuro-markers, such as LightGBM, benefit from the integration of domain-specific knowledge and established neuroscientific findings during feature extraction. This strategy inherently directs the model's focus toward the most informative predictors of defensive behaviors. Therefore, while the exploration of ANNs for decoding from raw LFP signals provides valuable insights into data-driven feature discovery, current evidence strongly supports the continued use of neuro-marker-based models for higher accuracy and reliability in decoding defensive behaviors. This comparison also highlights an important consideration for future research: the development of hybrid models that integrate the strengths of both manual and automatic feature extraction methods, leveraging GBDT's decision strategy, domain-specific insights, and nuanced neural encodings. Additionally, designing ANN architectures specifically tailored to capture the temporal dynamics and spatial configurations of neural data could bridge the gap between traditional ML models and ANN approaches. These strategies could further enhance decoding accuracy and the generalization capacity of our models.

In our study, we employed two metrics to assess feature importance: SHapley Additive exPlanations (SHAP) and the loss of R^2 upon feature removal. While SHAP values elucidate each feature's additive attribution to prediction, they do not explicitly evaluate the necessity of features for decoding performance. Conversely, the loss of R^2 quantifies a feature's impact on performance, yet this metric might yield ambiguous interpretations in cases of high feature correlation. Additionally, it fails to satisfy the three desirable properties of additive feature attribution methods outlined by SHAP, namely local accuracy, missingness, and consistency [128]. Thus, there is a compelling opportunity for researchers to explore alternative metrics for evaluating feature importance in terms of prediction and performance that both minimize computational complexity and embody the aforementioned properties. These metrics also highlight a specific limitation of the LightGBM approach: although we can identify which bands/features are most important for a given analysis (here, high-gamma), we cannot directly use that importance for a simple, biomarker-driven intervention. Tree-based methods focus on dichotomizing a given feature at a specific value, but can select that feature again at deeper tree levels if needed. Thus, they can model complex non-linear and non-smooth relationships between neural signals and behavior. Unlike a simpler model such as a LR, however, tree-based methods do not produce clear or simple relations such as 'to decrease defensive behaviors, it would be desirable to reduce BLA high-gamma power'. Inferring and testing such potential causalities would require different approaches, e.g. permuting the model's inputs in a systematic way and measuring the outputs. On the other hand, the superior decoding accuracy, feasibility for hardware implementation, and substantial pruning potential of tree-based models, as demonstrated in [98, 105, 141] could enable more efficient and effective closed-loop interventions compared to conventional approaches that rely solely on individual biomarkers [145, 146].

In this research, we evaluated our model using an offline paradigm on a dataset aimed at decoding defensive behaviors. To ascertain the robustness of our model across a wider array of neuropsychiatric applications, it would be beneficial to validate our model design using additional datasets, encompassing either identical or divergent tasks. Moreover, transitioning from offline to online neural decoding represents a significant challenge. In our future work, we intend to deploy our decoding framework within an online paradigm, thereby facilitating an assessment of our model's performance in real-time applications.

5. Conclusion

In this study, we analyzed LFP signals from IL and BLA of rats subjected to a tone-shock protocol to extract neuro-markers. These markers were subsequently utilized in our ML decoding framework, which incorporates SHAP-based feature selection and LightGBM for decoding defensive behaviors. Notably, the accelerometry jerk and bar press rate proved to be more decodable than the freezing score. We achieved an average decoding performance of $R^2 = 0.5357$ and $r = 0.7579$ for the accelerometry jerk, and $R^2 = 0.3476$ and $r = 0.6092$ for the bar press rate, with exceptionally low training/inference time and memory usage: less than 310 ms for training, less than 0.051 ms for inference, and 16.6 kB of memory on a single core of AMD Ryzen Threadripper PRO 5995WX CPU. BP and BPRC emerged as significant neuro-markers for prediction and decoding performance. The high-gamma band within BP, BPRC, and BCorr was consistently identified as crucial for decoding both defensive behaviors across both brain regions. The selection of top-ranked features not only surpassed the performance achieved using only BP features but also maintained the performance level of models utilizing the entire feature set. Our findings underscore the efficacy of developing an accurate and low-latency model for decoding defensive behavior based on LFP features from circuits strongly linked to these behaviors. This work lays the groundwork for future development of an implantable closed-loop psychiatric BMI, showcasing the potential of our framework in advancing neuropsychiatric treatment modalities.

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

This work was supported by the National Institute of Mental Health Grant R01-MH-123634.

Ethical statement

All experimental details were approved by the Institutional Animal Care and Use Committee at the University of Minnesota and performed in compliance with the Guide for the Care and Use of Animals. Research facilities were accredited by the American Association for the Accreditation of Laboratory Animal Care.

Appendix A. Detailed size of dataset among subjects and sessions

Table A1 presents a comprehensive overview of the dataset sizes distributed across the four experimental sessions for training, validation, and test sets. For each session, the dataset was split into training, validation, and test sets with a ratio of 16:4:9. The variability in dataset size across different sessions underscores the diverse conditions under which the models were trained, validated, and tested, contributing to a robust evaluation of the predictive models developed from this data.

Table A1. Size of training, validation, and test sets averaged over subjects in each recording session (Mean$_{\pm \mathrm{std}}$).

Session	Training	Validation	Test
Habituation	2219$_{\pm 8}$	555$_{\pm 2}$	1248$_{\pm 5}$
Conditioning	1974$_{\pm 568}$	493$_{\pm 142}$	1110$_{\pm 319}$
Extinction	7518$_{\pm 13}$	1879$_{\pm 3}$	4229$_{\pm 7}$
Recall	2302$_{\pm 43}$	575$_{\pm 11}$	1295$_{\pm 24}$

Appendix B. Decoding performance of alternative artificial neural network models on raw LFP data

Both LSTM and CNN models excel in automatic feature extraction, which involves identifying the most relevant features from raw data without human intervention. Therefore, we also implemented LSTM, WaveNet, and ResNet on raw LFP data, named LSTM-Raw, WaveNet-Raw, and ResNet-Raw, for the same decoding tasks.

The comparative analysis of decoding performance between models trained on manually selected neuro-markers (table 2) and models trained on raw LFP data (table A2) shows a marked superiority in the models employing neuro-markers. The neuro-markers likely encapsulate more relevant information for predicting behavioral outcomes, thus allowing the models to establish stronger and more predictive relationships with the observed behaviors. This advantage underscores the importance of our feature extraction and selection process in enhancing model accuracy in predicting avoidance behaviors.

Table A2. Performance of LSTM-Raw and WaveNet-Raw for decoding defensive behaviors averaged across subjects in each recording session. The performance was evaluated using the coefficient of determination (R^2) and the Pearson correlation coefficient (r).

Behavior	Metric	Session	LSTM-Raw	WaveNet-Raw	ResNet-Raw
Accelerometry jerk	R^2	Habituation	−0.0006282	0.2393	0.2385
		Conditioning	−0.3068	0.1180	−0.8459
		Extinction	0.08186	0.3871	0.3855
		Recall	−0.03857	0.3853	0.3205
	r	Habituation	0.05207	0.5927	0.5844
		Conditioning	0.01866	0.7395	0.6863
		Extinction	0.1754	0.6711	0.6678
		Recall	0.08446	0.6722	0.6817
Bar press rate	R^2	Habituation	0.05732	0.0164	−2.299
		Conditioning	−0.01080	−16.3277	−28.53
		Extinction	−0.06147	0.2014	0.1864
		Recall	−0.09371	−0.09922	−1.657
	r	Habituation	0.2886	0.4765	0.3224
		Conditioning	0.2326	0.1642	0.2427
		Extinction	0.2984	0.5565	0.5125
		Recall	0.2473	0.5117	0.3555

Table A3. Memory size, training time, and inference time of ML models for decoding defensive behaviors. Training and inference time were averaged across subjects and sessions.

		Accelerometry jerk		Bar press rate	
Model	Memory (kB)	Training (s)	Inference (ms)	Training (s)	Inference (ms)
LR	5.066	0.5426	0.04699	0.5436	0.04755
SVM-Lin	5.066	16.90	0.04125	18.39	0.04192
SVM-RBF	39280	73.36	5.280	——[a]	——[a]
RF	29.26	176.8	1.765	179.3	1.772
MLP	365.0	8.108	0.08635	30.66	0.08605
LSTM	1387	91.69	0.3010	141.9	0.3869
CNN	88.50	28.51	0.2602	73.14	0.2589
ResNet	336.8	89.33	0.7599	411.1	0.7608
LSTM-Raw	228.5	1589	3.210	3540	3.211
WaveNet-Raw	67.22	45.29	0.8177	199.3	0.8117
ResNet-Raw	263.9	384.9	1.412	1537	1.402
LightGBM	28.30	5.336	0.06006	5.130	0.05966
LightGBM-Top[b]	16.60	0.1823	0.05079	0.3096	0.05072

[a] Not applicable due to the fact that the SVM-RBF failed to decode bar press rate across subjects and sessions.

[b] LightGBM-Top refers to LightGBM using selected top-ranked features after feature selection.

Appendix C. Model training procedure and hyperparameter selection

LightGBM, LR, SVM-Lin, SVM-RBF, RF, MLP, LSTM, CNN, and ResNet used neuro-markers, $X_{\text{neuro}} \in \mathbb{R}^M$, as input for these models, where M is the number of neuro-markers. LSTM-Raw, WaveNet-Raw, and ResNet-Raw used raw LFP signals, $X_{\text{LFP}} \in \mathbb{R}^{C\times T}$, where C is the number of channels and T is the sequence length.

LightGBM and RF were trained for 1000 iterations, and artificial neural network (ANN) models (MLP, LSTM, CNN, ResNet, LSTM-Raw, WaveNet-Raw, and ResNet-raw) were trained for 100 epochs, all using MSE Loss. The training processes were early-stopped when the validation loss did not decrease for five consecutive iterations/epochs. All ANN models were followed by a prediction structure composed of three fully connected layers with the dimensions of 64, 32, and 1 for prediction. ANN models were trained with a batch size of 64 samples and Adam optimizer [147].

The detailed architectures of ML models are introduced below, and we performed the following hyperparameter search based on the lowest mean squared error on the validation set:

LightGBM: The model employed a traditional Gradient Boosting Decision Tree for boosting. We optimized the hyperparameters including (a) the maximum number of leaves in each tree within the range {3, 5, 13, 29}; (b) the ratio of data instances randomly sampled for training each tree within range {0.7, 0.8, 0.9, 1.0}; (c) the frequency of sampling

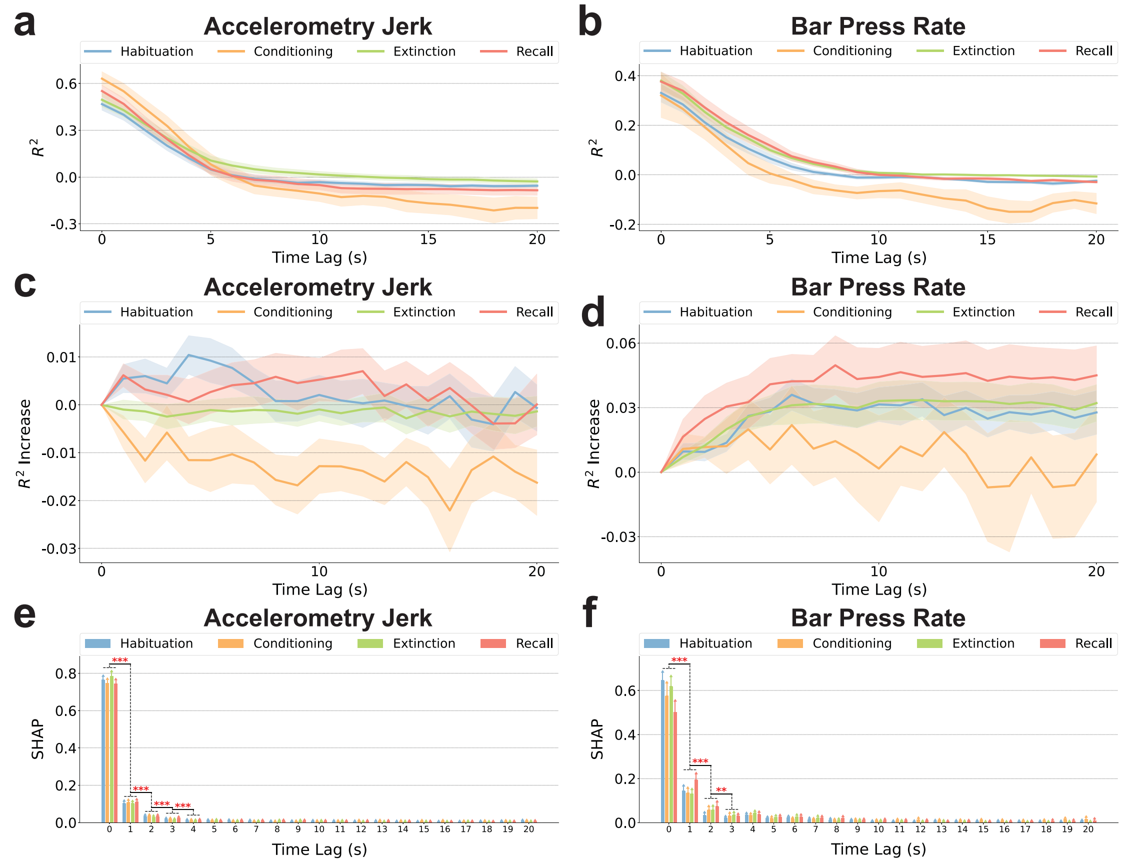

Figure A1. Time-dependent decoding performance and feature importance for two defensive behaviors. (a) Decoding performance for accelerometry jerk using neural features either aligned at zero lag to behavior or lagged by up to 20 s with a step size of 1 s for each time lag, in four recording sessions averaged across subjects, evaluated using R^2. The shaded areas indicate the standard errors across subjects. (b) Same as (a), but for decoding bar press rate. (c) Decoding performance for accelerometry jerk when using features from more than one time lag, assessed using an increase in the R^2 compared to a baseline without temporal integration, in four recording sessions averaged across subjects. Neural features from both the current and preceding time windows, extending up to 20 s, were incorporated. The shaded areas indicate the standard errors across subjects. (d) Same as (c), but for decoding bar press rate. (e) SHAP importance of features integrating data from the current and preceding time windows up to 20 seconds, averaged across subjects in four sessions for decoding accelerometry jerk. The error bars indicate the standard errors across subjects. The comparisons between features from consecutive time windows are marked with asterisks if one time window's SHAP values are significantly greater than those of the preceding window across sessions and subjects (Paired-sample t-test; $** : p < 0.01$, $*** : p < 0.001$). (f) Same as (e), but for decoding bar press rate.

data instances {0, 4, 8, 12}; (d) the ratio of features randomly sampled for training each tree within range {0.7, 0.8, 0.9, 1.0}, (e) the learning rate within the range [1e-3, 1e-1];. After hyperparameter selection, it was trained and evaluated using: (a) 5; (b) 0.9; (c) 8; (d) 1.0; (e) 1e-1. **LightGBM-Top** (which is the LightGBM using selected top-ranked features after feature selection) optimized the same hyperparameters within the same range, and it was trained and evaluated using: (a) 5; (b) 0.8; (c) 8; (d) 1.0; (e) 1e-1 after hyperparameter selection.

LR: We did not perform a hyperparameter search for this model.

SVM-Lin: We optimized the hyperparameters including (a) the epsilon in the epsilon-tube of SVM within the range {0.01, 0.05, 0.1, 0.5}, (b) the tolerance in loss function for stopping criterion within the range {1e-4, 1e-3, 1e-2}. After hyperparameter selection, it was trained and evaluated using: (a) 0.1; (b) 1e-3.

SVM-RBF: We optimized the hyperparameters including (a) the epsilon in the epsilon-tube of SVM within the range {0.01, 0.05, 0.1, 0.5}, (b) the tolerance in loss function for stopping criterion within the range {1e-4, 1e-3, 1e-2}. After hyperparameter selection, it was trained and evaluated using: (a) 0.1; (b) 1e-3.

RF: We optimized the hyperparameters including (a) the maximum depth of the tree within the range {2, 3, 4, 5}; (b) the ratio of data instances randomly sampled for training each tree within range {0.7, 0.8, 0.9, 1.0}; (c) the ratio of features randomly sampled for each split within range {0.7, 0.8, 0.9, 1.0}. After hyperparameter selection, it was trained and evaluated using: (a) 3; (b) 0.8; (c) 0.9.

MLP: The model employs multiple stacked fully connected layers, each followed by a ReLU and a Dropout layer. We optimized the hyperparameters including (a) the number of hidden layers within the range {1, 2, 3, 4}; (b) the dimension of hidden units within the range {32, 64, 128}; (c) the dropout rate within the range {0, 0.1, 0.2, 0.3}; (d) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 2; (b) 64; (c) 0; (d) 1e-3.

LSTM: The model employs multiple stacked LSTM layers, each followed by a Dropout layer. The final hidden state of the model was used as input for the prediction structure. We optimized the hyperparameters including (a) the number of recurrent layers within the range {1, 2, 3}; (b) the dimension of hidden units within the range {32, 64, 128}; (c) the input sequence length within the range {5, 10, 15}; (d) the dropout rate within the range {0, 0.1, 0.2, 0.3}; (e) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 1; (b) 64; (c) 5; (d) 0; (e) 1e-3.

CNN: The model employs multiple stacked 2-D convolutional layers with kernel size $(1,K)$ and stride $(1,2)$, each followed by a 2-D batch normalization, a ReLU, and a Dropout layer. The output of the model was flattened as input for the prediction structure. We optimized the hyperparameters including (a) the number of convolution layers within the range {2, 4, 6, 8}; (b) the kernel size K within the range {2, 4, 6, 8}; (c) the number of channels within the range {1, 2, 4}; (d) the dropout rate within the range {0, 0.1, 0.2, 0.3}; (e) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 4; (b) 4; (c) 4; (d) 0; (e) 1e-3.

ResNet: The model employs ResNet18 with 18 stacked 2-D convolutional layers with kernel size $(1,K)$, each followed by a 2-D batch normalization, and a ReLU layer [126]. The output of the model was flattened as input for the prediction structure. We optimized the hyperparameters including (a) the kernel size K within the range {3, 5, 7}; (b) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 3; (b) 1e-3.

LSTM-Raw: The model employs multiple stacked LSTM layers, each followed by a Dropout layer. The final hidden state of the model was used as input for the prediction structure. We optimized the hyperparameters including (a) the number of recurrent layers within the range {1, 2, 3}; (b) the dimension of hidden units within the range {32, 64, 128}; (c) the dropout rate within the range {0, 0.1, 0.2, 0.3}; (d) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 2; (b) 64; (c) 0.2; (d) 1e-3.

WaveNet-Raw: The model employs an architecture close to EEGWaveNet [127], using depth-wise 2-D convolutional layers with kernel size $(1,2)$ and stride $(1,2)$, and spatial-temporal 2-D convolutional layers with kernel size $(1,4)$ and stride $(1,2)$. Each spatial-temporal 2-D convolutional layer was followed by a 2-D batch normalization, and a LeakyReLU layer. The output of each scale went through global average pooling and they were collectively concatenated as input for the prediction structure. We optimized the hyperparameters including (a) the number of channels for spatial-temporal 2-D convolution within the range {8, 16, 32}; (b) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 16; (b) 1e-3.

ResNet-Raw: The model employs ResNet18 with 18 stacked 2-D convolutional layers with kernel size $(1,K)$, each followed by a 2-D batch normalization, and a ReLU layer [126]. The output of the model went through global average pooling and then was used as input for the prediction structure. We optimized the hyperparameters including (a) the kernel size K within the range {3, 5, 7}; (b) the learning rate within the range [1e-5, 1e-3]. After hyperparameter selection, it was trained and evaluated using: (a) 3; (b) 1e-3.

ORCID iDs

Jinhan Liu https://orcid.org/0000-0002-7887-8169
Sumedh S Nagrale https://orcid.org/0000-0002-0039-9125
Shreya Yadav https://orcid.org/0009-0009-8644-9223
Alik S Widge https://orcid.org/0000-0001-8510-341X
Mahsa Shoaran https://orcid.org/0000-0002-6426-4799

References

[1] Adolphs R 2013 The biology of fear *Curr. Biol.* **23** R79–R93

[2] Barlow D H 2004 *Anxiety and its Disorders: The Nature and Treatment of Anxiety and Panic* (Guilford Press)

[3] LeDoux J E 1998 *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (Simon and Schuster)

[4] Stein D J and Nesse R M 2011 Threat detection, precautionary responses and anxiety disorders *Neurosci. Biobehav. Rev.* **35** 1075–9

[5] LeDoux J E 2000 Emotion circuits in the brain *Annu. Rev. Neurosci.* **23** 155–84

[6] Baxter A J, Scott K M, Vos T and Whiteford H A 2013 Global prevalence of anxiety disorders: a systematic review and meta-regression *Psychol. Med.* **43** 897–910

[7] Shin L M and Liberzon I 2010 The neurocircuitry of fear, stress and anxiety disorders *Neuropsychopharmacology* **35** 169–91

[8] Bandelow B and Michaelis S 2015 Epidemiology of anxiety disorders in the 21st century *Dialog. Clin. Neurosc.* **17** 327–35

[9] Vigo D, Thornicroft G and Atun R 2016 Estimating the true global burden of mental illness *Lancet Psychiatry* **3** 171–8

[10] Duvarci S and Pare D 2014 Amygdala microcircuits controlling learned fear *Neuron* **82** 966–80

[11] Fenster R J, Lebois L A M, Ressler K J and Suh J 2018 Brain circuit dysfunction in post-traumatic stress disorder: from mouse to man *Nat. Rev. Neurosci.* **19** 535–51

[12] Tovote P, Paul Fadok J and Lüthi A 2015 Neuronal circuits for fear and anxiety *Nat. Rev. Neurosci.* **16** 317–31

[13] Adhikari A 2014 Distributed circuits underlying anxiety *Front. Behav. Neurosci.* **8** 112

[14] Janak P H and Tye K M 2015 From circuits to behaviour in the amygdala *Nature* **517** 284–92

[15] LeDoux J E and Pine D S 2016 Using neuroscience to help understand fear and anxiety: a two-system framework *Am. J. Psychiatry* **173** 1083–93

[16] Mobbs D, Hagan C C, Dalgleish T, Silston B and Prévost C 2015 The ecology of human fear: survival optimization and the nervous system *Front. Neurosci.* **9** 55

[17] Adhikari A, Topiwala M A and Gordon J A 2010 Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety *Neuron* **65** 257–69

[18] Poulos A M, Mehta N, Lu B, Amir D, Livingston B, Santarelli A, Zhuravka I and Fanselow M S 2016 Conditioning-and time-dependent increases in context fear and generalization *Learn. Mem.* **23** 379–85

[19] Roelofs K 2017 Freeze for action: neurobiological mechanisms in animal and human freezing *Phil. Trans. R. Soci.* B **372** 20160206

[20] Campos A C, Fogaça M V, Aguiar D C and Guimaraes F S 2013 Animal models of anxiety disorders and stress *Braz. J. Psychiatry* **35** S101–11

[21] Colom-Lapetina J, Li A J, Pelegrina-Perez T C and Shansky R M 2019 Behavioral diversity across classic rodent models is sex-dependent *Front. Behav. Neurosci.* **13** 45

[22] Deslauriers J, Toth M, Der-Avakian A and Risbrough V B 2018 Current status of animal models of posttraumatic stress disorder: behavioral and biological phenotypes and future challenges in improving translation *Biol. Psychiatry* **83** 895–907

[23] Robinson O J, Pike A C, Cornwell B and Grillon C 2019 The translational neural circuitry of anxiety *J. Neurol. Neurosurg. Psychiatry* **90** 1353–60

[24] Terburg D, Scheggia D, Triana Del Rio R, Klumpers F, Cristian Ciobanu A, Morgan B, Montoya E R, Bos P A, Giobellina G and van den Burg E H 2018 The basolateral amygdala is essential for rapid escape: a human and rodent study *Cell* **175** 723–35

[25] Langevin J-P, Koek R J, Schwartz H N, Chen J W Y, Sultzer D L, Mandelkern M A, Kulick A D and Krahl S E 2016 Deep brain stimulation of the basolateral amygdala for treatment-refractory posttraumatic stress disorder *Biol. Psychiatry* **79** e82–e84

[26] Langevin J-P, De Salles A A F, Kosoyan H P and Krahl S E 2010 Deep brain stimulation of the amygdala alleviates post-traumatic stress disorder symptoms in a rat model *J. Psychiatr. Res.* **44** 1241–5

[27] Holtzheimer P E and Mayberg H S 2011 Deep brain stimulation for psychiatric disorders *Annu. Rev. Neurosci.* **34** 289–307

[28] Luyten L, Hendrickx S, Raymaekers S, Gabriëls L and Nuttin B 2016 Electrical stimulation in the bed nucleus of the stria terminalis alleviates severe obsessive-compulsive disorder *Molecular Psychiatry* **21** 1272–80

[29] Widge A S 2023 Closing the loop in psychiatric deep brain stimulation: physiology, psychometrics and plasticity *Neuropsychopharmacology* **49** 1–12

[30] Widge A S, Ellard K K, Paulk A C, Basu I, Yousefi A, Zorowitz S, Gilmour A, Afzal A, Deckersbach T, Cash S S, Kramer M A 2017 Treating refractory mental illness with closed-loop brain stimulation: progress towards a patient-specific transdiagnostic approach *Exp. Neurol.* **287** 461–72

[31] Shin U, Ding C, Woods V, Widge A S and Shoaran M 2023 A 16-channel low-power neural connectivity extraction and phase-locked deep brain stimulation soc *IEEE Solid-State Circuits Lett.* **6** 21–24

[32] Shin U, Ding C, Somappa L, Woods V, Widge A S and Shoaran M 2022 A 16-channel 60 μw neural synchrony processor for multi-mode phase-locked neurostimulation *2022 IEEE Custom Integrated Circuits Conf. (CICC)* (IEEE) pp 01–02

[33] Widge A S, Dougherty D D and Moritz C T 2014 Affective brain-computer interfaces as enabling technology for responsive psychiatric stimulation *Brain-Comput. Interfaces* **1** 126–36

[34] Shanechi M M 2019 Brain–machine interfaces from motor to mood *Nat. Neurosci.* **22** 1554–64

[35] Shoaran M 2023 Next-generation closed-loop neural interfaces: circuit and AI-driven innovations *IEEE Solid-State Circuits Mag.* **15** 41–49

[36] Yoo J and Shoaran M 2021 Neural interface systems with on-device computing: Machine learning and neuromorphic architectures *Curr. Opin. Biotechnol.* **72** 95–101

[37] Widge A S, Zorowitz S, Basu I, Paulk A C, Cash S S, Eskandar E N, Deckersbach T, Miller E K and Dougherty D D 2019 Deep brain stimulation of the internal capsule enhances human cognitive control and prefrontal cortex function *Nat. Commun.* **10** 1536

[38] Sani O G, Yang Y, Lee M B, Dawes H E, Chang E F and Shanechi M M 2018 Mood variations decoded from multi-site intracranial human brain activity *Nat. Biotechnol.* **36** 954–61

[39] Zhu B, Shin U and Shoaran M 2021 Closed-loop neural prostheses with on-chip intelligence: A review and a low-latency machine learning model for brain state detection *IEEE Trans. Biomed. Circuits Syst.* **15** 877–97

[40] Sani O G, Yang Y and Shanechi M M 2023 Brain-machine interfaces for closed-loop electrical brain stimulation in neuropsychiatric disorders *Handbook of Neuroengineering* (Springer) pp 1317–42

[41] Shenoy K V and Carmena J M 2014 Combining decoder design and neural adaptation in brain-machine interfaces *Neuron* **84** 665–80

[42] Serruya M, Hatsopoulos N, Fellows M, Paninski L and Donoghue J 2003 Robustness of neuroprosthetic decoding algorithms *Biol. Cybern.* **88** 219–28

[43] Ethier C, Oby E R, Bauman M J and Miller L E 2012 Restoration of grasp following paralysis through brain-controlled stimulation of muscles *Nature* **485** 368–71

[44] Baeg E H, Kim Y B, Huh K, Mook-Jung I, Kim H T and Jung M W 2003 Dynamics of population code for working memory in the prefrontal cortex *Neuron* **40** 177–88

[45] Ibos G and Freedman D J 2017 Sequential sensory and decision processing in posterior parietal cortex *eLife* **6** 23743

[46] Zhang K, Ginzburg I, McNaughton B L and Sejnowski T J 1998 Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells *J. Neurophysiol.* **79** 1017–44

[47] Davidson T J, Kloosterman F and Wilson M A 2009 Hippocampal replay of extended experience *Neuron* **63** 497–507

[48] Hung C P, Kreiman G, Poggio T and DiCarlo J J 2005 Fast readout of object identity from macaque inferior temporal cortex *Science* **310** 863–6

[49] Rich E L and Wallis J D 2016 Decoding subjective decisions from orbitofrontal cortex *Nat. Neurosci.* **19** 973–80

[50] Dekleva B M, Ramkumar P, Wanda P A, Kording K P and Miller L E 2016 Uncertainty leads to persistent effects on reach representations in dorsal premotor cortex *eLife* **5** 14316

[51] Raposo D, Kaufman M T and Churchland A K 2014 A category-free neural population supports evolving demands during decision-making *Nat. Neurosci.* **17** 1784–92

[52] Quian Quiroga R, Snyder L H, Batista A P, Cui H and Andersen R A 2006 Movement intention is better predicted than attention in the posterior parietal cortex *J. Neurosci.* **26** 3615–20

[53] Weygandt M, Blecker C R, Schäfer A, Hackmack K, Haynes J-D, Vaitl D, Stark R and Schienle A 2012 fmri pattern recognition in obsessive–compulsive disorder *Neuroimage* **60** 1186–93

[54] Drevets W C 2001 Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders *Curr. opin. Neurobiol.* **11** 240–9

[55] Mayberg H S 2003 Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment *Br. Med. Bull.* **65** 193–207

[56] Kupfer D J, Frank E and Phillips M L 2012 Major depressive disorder: new clinical, neurobiological and treatment perspectives *Lancet* **379** 1045–55

[57] Mayberg H S 1997 Limbic-cortical dysregulation: a proposed model of depression *J. Neuropsychiatry Clin. Neurosci.* **9** 471–81

[58] Hochberg L R, Serruya M D, Friehs G M, Mukand J A, Saleh M, Caplan A H, Branner A, Chen D, Penn R D and Donoghue J P 2006 Neuronal ensemble control of prosthetic devices by a human with tetraplegia *Nature* **442** 164–71

[59] Jercog D, Winke N, Sung K, Martin Fernandez M, Francioni C, Rajot D, Courtin J, Chaudun F, Jercog P E and Valerio S 2021 Dynamical prefrontal population coding during defensive behaviours *Nature* **595** 690–4

[60] Likhtik E, Stujenske J M, Topiwala M A, Harris A Z and Gordon J A 2014 Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety *Nat. Neurosci.* **17** 106–13

[61] Stujenske J M, Likhtik E, Topiwala M A and Gordon J A 2014 Fear and safety engage competing patterns of theta-gamma coupling in the basolateral amygdala *Neuron* **83** 919–33

[62] Lesting J, Narayanan R T, Kluge C, Sangha S, Seidenbecher T and Pape H-C 2011 Patterns of coupled theta activity in amygdala-hippocampal-prefrontal cortical circuits during fear extinction *PLoS One* **6** e21714

[63] Wang D, Zhang Q, Li Y, Wang Y, Zhu J, Zhang S and Zheng X 2014 Long-term decoding stability of local field potentials from silicon arrays in primate motor cortex during a 2D center out task *J. Neural Eng.* **11** 036009

[64] Karalis N, Dejean C, Chaudun F, Khoder S, Rozeske R R, Wurtz H'ene, Bagur S, Benchenane K, Sirota A and Courtin J *et al* 2016 4-hz oscillations synchronize prefrontal–amygdala circuits during fear behavior *Nat. Neurosci.* **19** 605–12

[65] Brendan Ritchie J, Michael Kaplan D and Klein C 2019 Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience *Br. J. Phil. Sci.* **70** 581–607

[66] Lydon-Staley D M, Cornblath E J, Sizemore Blevins A and Bassett D S 2021 Modeling brain, symptom and behavior in the winds of change *Neuropsychopharmacology* **46** 20–32

[67] Glaser J I, Benjamin A S, Farhoodi R and Kording K P 2019 The roles of supervised machine learning in systems neuroscience *Prog. Neurobiol.* **175** 126–37

[68] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44

[69] Basu I, Yousefi A, Crocker B, Zelmann R, Paulk A C, Peled N, Ellard K K, Weisholtz D S, Rees Cosgrove G and Deckersbach T *et al* 2023 Closed-loop enhancement and neural decoding of cognitive control in humans *Nat. Biomed. Eng.* **7** 576–88

[70] Provenza N R, Paulk A C, Peled N, Restrepo M I, Cash S S, Dougherty D D, Eskandar E N, Borton D A and Widge A S 2019 Decoding task engagement from distributed network electrophysiology in humans *J. Neural Eng.* **16** 056015

[71] Avvaru S, Provenza N R, Widge A S and Parhi K K 2021 Spectral features based decoding of task engagement: The role of theta and high gamma bands in cognitive control *2021 43rd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 6062–5

[72] Alagapan S, Sueng Choi K, Heisig S, Riva-Posse P, Crowell A, Tiruvadi V, Obatusin M, Veerakumar A, Waters A C and Gross R E *et al* 2023 Cingulate dynamics track depression recovery with deep brain stimulation *Nature* **622** 1–9

[73] Bijanzadeh M, Khambhati A N, Desai M, Wallace D L, Shafi A, Dawes H E, Sturm V E and Chang E F 2022 Decoding naturalistic affective behaviour from spectro-spatial features in multiday human IEEG *Nat. Human Behav.* **6** 823–36

[74] Hultman R, Ulrich K, Sachs B D, Blount C, Carlson D E, Ndubuizu N, Bagot R C, Parise E M, Vu M-A T, Gallagher N M, Wang J 2018 Brain-wide electrical spatiotemporal dynamics encode depression vulnerability *Cell* **173** 166–80

[75] Yao L, Baker J L, Schiff N D, Purpura K P and Shoaran M 2021 Predicting task performance from biomarkers of mental fatigue in global brain activity *J. Neural Eng.* **18** 036001

[76] Yao L, Baker J L, Ryou J-W, Schiff N D, Purpura K P and Shoaran M 2020 Mental fatigue prediction from multi-channel ecog signal *ICASSP 2020-2020 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE) pp 1259–63

[77] Sellers K K, Khambhati A N, Stapper N, Fan J M, Rao V R, Scangos K W, Chang E F and Krystal A D 2023 Closed-loop neurostimulation for biomarker-driven, personalized treatment of major depressive disorder *J. Vis. Exp.* e65177

[78] Wu W *et al* 2020 An electroencephalographic signature predicts antidepressant response in major depression *Nat. Biotechnol.* **38** 439–47

[79] Avvaru S, Provenza N R, Widge A S and Parhi K K 2021 Decoding human cognitive control using functional connectivity of local field potentials *2021 43rd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 451–4

[80] Dejean C, Courtin J, Rozeske R R, Bonnet M C, Dousset V, Michelet T and Herry C 2015 Neuronal circuits for fear expression and recovery: recent advances and potential therapeutic strategies *Biol. Psychiatry* **78** 298–306

[81] Caroline Blanchard D, Hynd A L, Minke K A, Minemoto T and Blanchard R J 2001 Human defensive behaviors to threat scenarios show parallels to fear-and anxiety-related defense patterns of non-human mammals *Neurosci. Biobehav. Rev.* **25** 761–70

[82] Mobbs D and Kim J J 2015 Neuroethological studies of fear, anxiety and risky decision-making in rodents and humans *Curr. Opin. Behav. Sci.* **5** 8–15

[83] Bocchio M, Nabavi S and Capogna M 2017 Synaptic plasticity, engrams and network oscillations in amygdala circuits for storage and retrieval of emotional memories *Neuron* **94** 731–43

[84] Milad M R, Igoe S and Orr S P 2011 Fear conditioning in rodents and humans *Animal Mod. Behav. Anal.* 111–32

[85] McDannald M A 2023 Pavlovian fear conditioning is more than you think it is *J. Neurosci.* **43** 8079–87

[86] Reis F M C V, Liu J, Schuette P J, Lee J Y, Maesta-Pereira S, Chakerian M, Wang W, Canteras N S, Kao J C and Adhikari A 2021 Shared dorsal periaqueductal gray activation patterns during exposure to innate and conditioned threats *J. Neurosci.* **41** 5399–420

[87] Reis F M C V, Lee J Y, Maesta-Pereira S, Schuette P J, Chakerian M, Liu J, La-Vu M Q, Tobias B C, Ikebara J M and Hiroaki Kihara A 2021 Dorsal periaqueductal gray ensembles represent approach and avoidance states *eLife* **10** e64934

[88] Li Y, Dong X, Li S and Kirouac G J 2014 Lesions of the posterior paraventricular nucleus of the thalamus attenuate fear expression *Front. Behav. Neurosci.* **8** 94

[89] Younk R and Widge A 2022 Quantifying defensive behavior and threat response through integrated headstage accelerometry *J. Neurosci. Methods* **382** 109725

[90] Siegle J H, Hale G J, Newman J P and Voigts J 2015 Neural ensemble communities: open-source approaches to hardware for large-scale electrophysiology *Curr. Opin. Neurobiol.* **32** 53–59

[91] Lo M-C, Younk R and Widge A S 2020 Paired electrical pulse trains for controlling connectivity in emotion-related brain circuitry *IEEE Trans. Neural Syst. Rehab. Eng.* **28** 2721–30

[92] Gruene T M, Flick K, Stefano A, Shea S D and Shansky R M 2015 Sexually divergent expression of active and passive conditioned fear responses in rats *eLife* **4** e11352

[93] Totty M S and Maren S 2022 Neural oscillations in aversively motivated behavior *Front. Behav. Neurosci.* **16** 936036

[94] Logesparan L, Casson A J and Rodriguez-Villegas E 2012 Optimal features for online seizure detection *Med. Biol. Eng. Comput.* **50** 659–69

[95] Maling N and McIntyre C 2016 Local field potential analysis for closed-loop neuromodulation *Closed loop neuroscience* (Academic) 67–80

[96] Stavisky S D, Kao J C, Nuyujukian P, Ryu S I and Shenoy K V 2015 A high performing brain–machine interface driven by low-frequency local field potentials alone and together with spikes *J. Neural Eng.* **12** 036009

[97] Shoaran M, Allahgholizadeh Haghi B, Taghavi M, Farivar M and Emami-Neyestanak A 2018 Energy-efficient classification for resource-constrained biomedical applications *IEEE J. Emerg. Sel. Top. Circuits Syst.* **8** 693–707

[98] Zhu B, Farivar M and Shoaran M 2020 Resot: resource-efficient oblique trees for neural signal classification *IEEE Trans. Biomed. Circuits Syst.* **14** 692–704

[99] Bandarabadi M, Teixeira C A, Rasekhi J and Dourado A 2015 Epileptic seizure prediction using relative spectral power features *Clin. Neurophys.* **126** 237–48

[100] Zhang Z and Parhi K K 2015 Low-complexity seizure prediction from IEEG/sEEG using spectral power and ratios of spectral power *IEEE Trans. Biomed. Circuits Syst.* **10** 693–706

[101] Bandarabadi M, Teixeira C A, Netoff T I, Parhi K K and Dourado A 2014 Robust and low complexity algorithms for seizure detection *2014 36th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society* (IEEE) pp 4447–50

[102] Koolen N, Jansen K, Vervisch J, Matic V, De Vos M, Naulaers G and Van Huffel S 2014 Line length as a robust method to detect high-activity events: automated burst detection in premature EEG recordings *Clin. Neurophys.* **125** 1985–94

[103] Majid Mehmood R M and Jong Lee H J 2015 Eeg based emotion recognition from human brain using hjorth parameters and svm *Int. J. Bio-Sci. Bio-Technol.* **7** 23–32

[104] Yao L, Brown P and Shoaran M 2020 Improved detection of parkinsonian resting tremor with feature engineering and kalman filtering *Clin. Neurophys.* **131** 274–84

[105] Shin U, Ding C, Zhu B, Vyza Y, Trouillet A, Revol E C M, Lacour S P and Shoaran M 2022 Neuraltree: A 256-channel 0.227-μj/class versatile neural activity classification and closed-loop neuromodulation soc *IEEE J. Solid-State Circuits* **57** 3243–57

[106] Mukhopadhyay S and Ray G C 1998 A new interpretation of nonlinear energy operator and its efficacy in spike detection *IEEE Trans. Biomed. Eng.* **45** 180–7

[107] Xiang J, Maue E, Fan Y, Qi L, Mangano F T, Greiner H and Tenney J 2020 Kurtosis and skewness of high-frequency brain signals are altered in paediatric epilepsy *Brain Commun.* **2** fcaa036

[108] Srinivasan V, Eswaran C and Sriraam N 2007 Approximate entropy-based epileptic eeg detection using artificial neural networks *IEEE Trans. Inf. Technolo. Biomed.* **11** 288–95

[109] Jie X, Cao R and Li. Li 2014 Emotion recognition based on the sample entropy of eeg *Bio-Med. Mater. Eng.* **24** 1185–92

[110] Kirkby L A, Luongo F J, Lee M B, Nahum M, Van Vleet T M, Rao V R, Dawes H E, Chang E F and Sohal V S 2018 An amygdala-hippocampus subnetwork that encodes variation in human mood *Cell* **175** 1688–1700.e14

[111] Munia T T K and Aviyente S 2019 Time-frequency based phase-amplitude coupling measure for neuronal oscillations *Sci. Rep.* **9** 12441

[112] Nandi B, Swiatek P, Kocsis B and Ding M 2019 Inferring the direction of rhythmic neural transmission via inter-regional phase-amplitude coupling (ir-PAC) *Sci. Rep.* **9** 6933

[113] Dasdemir Y, Yildirim E and Yildirim S 2017 Analysis of functional brain connections for positive–negative emotions using phase locking value *Cogn. Neurodyn.* **11** 487–500

[114] Yao L, Zhu B and Shoaran M 2022 Fast and accurate decoding of finger movements from ECoG through riemannian features and modern machine learning techniques *J. Neural Eng.* **19** 016037

[115] Arbabshirani M R, Plis S, Sui J and Calhoun V D 2017 Single subject prediction of brain disorders in neuroimaging: promises and pitfalls *Neuroimage* **145** 137–65

[116] Olsen S T, Basu I, Taha Bilge M T, Kanabar A, Boggess M J, Rockhill A P, Gosai A K, Hahn E, Peled N, Ennis M, Shiff I 2020 Case report of dual-site neurostimulation and chronic recording of cortico-striatal circuitry in a patient with treatment refractory obsessive compulsive disorder *Front. Human Neurosci.* **14** 569973

[117] Zhu B, Cruz-Garza J G, Yang Q, Shoaran M and Kalantari S 2022 Identifying uncertainty states during wayfinding in indoor environments: an EEG classification study *Adv. Eng. Inf.* **54** 101718

[118] Koppe G, Meyer-Lindenberg A and Durstewitz D 2021 Deep learning for small and big data in psychiatry *Neuropsychopharmacology* **46** 176–90

[119] Durstewitz D, Koppe G and Meyer-Lindenberg A 2019 Deep neural networks in psychiatry *Mol. Psychiatry* **24** 1583–98

[120] Kuhlmann L, Karoly P, Freestone D R, Brinkmann B H, Temko A, Barachant A, Li F, Titericz G, Lang B W and Lavery D 2018 Epilepsyecosystem ORG: CROWD-sourcing reproducible seizure prediction with long-term human intracranial EEG *Brain* **141** 2619–30

[121] Ke G, Meng Q, Finley T, Wang T, Chen W, dong Ma W, Ye Q and Liu T-Y 2017 Lightgbm: a highly efficient gradient boosting decision tree *Advances in Neural Information Processing Systems* p 30

[122] Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97

[123] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

[124] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80

[125] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems* p 25

[126] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8

[127] Thuwajit P, Rangpong P, Sawangjai P, Autthasan P, Chaisaen R, Banluesombatkul N, Boonchit P, Tatsaringkansakul T, Sudhawiyangkul T and Wilaiprasitporn T 2021 Eegwavenet: Multiscale cnn-based spatiotemporal feature extraction for EEG seizure detection *IEEE Trans. Ind. Inf.* **18** 5547–57

[128] Lundberg S M and Lee S-I 2017 A unified approach to interpreting model predictions *Advances in Neural Information Processing Systems* p 30

[129] Pan H, Li Z, Tian C, Wang Li, Fu Y, Qin X and Liu F 2023 The lightgbm-based classification algorithm for chinese characters speech imagery BCI system *Cogn. Neurodyn.* **17** 373–84

[130] Dhar J 2022 An adaptive intelligent diagnostic system to predict early stage of parkinson's disease using two-stage dimension reduction with genetically optimized lightgbm algorithm *Neural Comput. Appl.* **34** 4567–93

[131] Aggarwal S, Aggarwal L, Singh Rihal M and Aggarwal S 2018 EEG based participant independent emotion classification using gradient boosting machines *2018 IEEE 8th Int. Advance Computing Conf. (IACC)* (IEEE) pp 266–71

[132] Zhu B and Shoaran M 2021 Unsupervised domain adaptation for cross-subject few-shot neurological symptom detection *2021 10th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* (IEEE) pp 181–4

[133] Zeng H, Yang C, Zhang H, Wu Z, Zhang J, Dai G, Babiloni F and Kong W 2019 A lightgbm-based EEG analysis method for driver mental states classification *Comput. Intell. Neurosci.* **2019** 3761203

[134] Abenna S, Nahid M and Bajit A 2022 Motor imagery based brain-computer interface: improving the EEG classification using delta rhythm and lightgbm algorithm *Biomed. Signal Process. Control* **71** 103102

[135] Jia H, Yu S, Yin S, Liu L, Yi C, Xue K, Li F, Yao D, Xu P and Zhang T 2023 A model combining multi branch spectral-temporal cnn, efficient channel attention and lightgbm for mi-bci classification *IEEE Trans. Neural Syst. Rehab. Eng.* **31** 1311–20

[136] Mitchell T M, Hutchinson R, Niculescu R S, Pereira F, Wang X, Just M and Newman S 2004 Learning to decode cognitive states from brain images *Mach. Learn.* **57** 145–75

[137] Flash T and Hogan N 1985 The coordination of arm movements: an experimentally confirmed mathematical model *J. Neurosci.* **5** 1688–703

[138] LeDoux J and Daw N D 2018 Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour *Nat. Rev. Neurosci.* **19** 269–82

[139] Dunsmoor J E, Niv Y, Daw N and Phelps E A 2015 Rethinking extinction *Neuron* **88** 47–63

[140] Jackson A D, Cohen J L, Phensy A J, Chang E F, Dawes H E and Sohal V S 2024 Amygdala-hippocampus somatostatin interneuron beta-synchrony underlies a cross-species biomarker of emotional state *Neuron* (https://doi.org/10.1016/j.neuron.2023.12.017)

[141] Zhu B, Taghavi M and Shoaran M 2019 Cost-efficient classification for neurological disease detection *2019 IEEE Biomedical Circuits and Systems Conf. (BioCAS)* (IEEE) pp 1–4

[142] Yao L and Shoaran M 2019 Enhanced classification of individual finger movements with ecog *2019 53rd Asilomar Conf. on Signals, Systems and Computers* (IEEE) pp 2063–6

[143] Yao L, Brown P and Shoaran M 2018 Resting tremor detection in parkinson's disease with machine learning and kalman filtering *2018 IEEE Biomedical Circuits and Systems Conf. (BioCAS)* (IEEE) pp 1–4

[144] Zhu B, Shin U and Shoaran M 2020 Closed-loop neural interfaces with embedded machine learning *2020 27th IEEE Int. Conf. on Electronics, Circuits and Systems (ICECS)* (IEEE) pp 1–4

[145] Scangos K W, Khambhati A N, Daly P M, Makhoul G S, Sugrue L P, Zamanian H, Liu T X, Rao V R, Sellers K K and Dawes H E 2021 Closed-loop neuromodulation in an individual with treatment-resistant depression *Nat. Med.* **27** 1696–700

[146] Sun F T, Morrell M J and Wharen R E 2008 Responsive cortical stimulation for the treatment of epilepsy *Neurotherapeutics* **5** 68–74

[147] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)