frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 April 2026
DOI 10.3389/fnins.2026.1605209

Check for updates

OPEN ACCESS

EDITED BY
Mehdi Adibi,
Monash University, Australia

REVIEWED BY
Răzvan Gămănuț,
Okinawa Institute of Science and Technology Graduate University, Japan
Anguo Zhang,
Fuzhou University, China

*CORRESPONDENCE
Steffen Albrecht
✉ salbrec@uni-mainz.de
Maik C. Stüttgen
✉ maik.stuettgen@uni-mainz.de
Heiko J. Luhmann
✉ luhmann@uni-mainz.de
Illia Horenko
✉ horenko@rptu.de

†PRESENT ADDRESSES
Steffen Albrecht,
Department of General Practice and Primary Healthcare, University of Auckland, Auckland, New Zealand

RECEIVED 03 April 2025
REVISED 05 March 2026
ACCEPTED 13 March 2026
PUBLISHED 01 April 2026

CITATION
Albrecht S, Vandevelde JR, Vecchi E, Berra G, Bassetti D, Stüttgen MC, Luhmann HJ and Horenko I (2026) Spiking neural networks provide accurate and time-efficient models for whisker stimulus classification of the awake mouse.
Front. Neurosci. 20:1605209.
doi: 10.3389/fnins.2026.1605209

COPYRIGHT
© 2026 Albrecht, Vandevelde, Vecchi, Berra, Bassetti, Stüttgen, Luhmann and Horenko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Spiking neural networks provide accurate and time-efficient models for whisker stimulus classification of the awake mouse

Steffen Albrecht[1*†], Jens R. Vandevelde[1], Edoardo Vecchi[2], Gabriele Berra[2], Davide Bassetti[3], Maik C. Stüttgen[4*], Heiko J. Luhmann[1*] and Illia Horenko[3*]

[1]Institute of Physiology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany, [2]Undergraduate Education (FBDTI), Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Lugano-Viganello, Switzerland, [3]Department of Mathematics, Artificial Intelligence in Mathematics, TU Kaiserslautern, Kaiserslautern, Germany, [4]Institute of Pathophysiology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

Machine learning algorithms have great potential for classifying brain activity, and lightweight classifier algorithms, requiring little computational resources, can be used on low-energy neuromorphic hardware designed for implantable neuroprosthetics. One of these efficient algorithms, the Liquid State Machine, implements the concept of Spiking Neural Networks and has been shown to achieve outstanding results on the task of whisker stimulus detection from the mouse barrel cortex, a widely used model system. While this is promising for neuroprosthetics, it has been unclear how a Spiking Neural Network or other machine learning algorithms perform on data recorded from awake mice and how trained models generalize across individuals, the latter being relevant to transferring trained models to new hardware. Using laminar multi-electrode local field potential recordings obtained from four mice performing a single-whisker detection task, we benchmarked the performance of a collection of lightweight classification algorithms. We found that the Liquid State Machine, a generalized linear model, and the time series classifier ROCKET are the most accurate for stimulus detection. Among those, the Liquid State Machine achieved the fastest model training and inference runtime and provided robust accuracy across individual mice. Additional analyses show that there is no significant improvement in using multiple cortical layers as input for the model and that 40 ms of stimulus recording is sufficient to maintain high detection accuracy.

KEYWORDS
electrophysiology, local field potential, machine learning, neural prosthesis, neuroprosthetics, spiking neural networks, whisker stimulus classification

Introduction

The development and improvement of brain-machine interfaces (BMI) strongly advanced during the last decades to process brain signals for the control of an external computer or machine (Rapeaux and Constandinou, 2021; Sitaram et al., 2017). BMIs help paralyzed patients suffering from severe brain damage or disrupted connectivity between different brain regions, as typically caused by, for example, spinal cord damage, stroke, or neurodegenerative diseases

such as amyotrophic lateral sclerosis (Semprini et al., 2018). BMIs can read out and interpret brain activity to control external actuators such as robotic arms or to function as a spelling interface, thus enabling patients to write text and communicate with their social environment (Rezeika et al., 2018; Sunny et al., 2016).

Instead of turning neural activity into an action performed by a computer or machine, the interface can also create a signal to be sent back to the brain. This would result in a closed-loop system that creates artificial stimulation for the brain based on the signal it receives from it. Such systems are called brain-machine-brain interfaces (BMBIs) and are promising for neural prostheses to replace impaired or even missing biological functionality (Krucoff et al., 2016). For instance, it has been shown that such closed-loop systems can bridge damaged neural pathways in the rat brain after traumatic brain injury by interpreting action potentials captured by multielectrode arrays (MEA) implanted in the cerebral cortex (Guggenmos et al., 2013).

MEAs present several advantages when considered for a neural prosthesis, as the shape of the MEA probe can be tailored to each individual case and is flexible enough to adapt to different brain regions. They provide electrophysiological recordings at a high spatio-temporal resolution that can be further processed differently. One option is to use the local field potential (LFP), i.e., the down-sampled and low-pass filtered raw signal, which includes frequency components below 300 Hz (Petschenig et al., 2022). The LFP reflects the gross spatially weighted average of membrane potential fluctuations of thousands of neurons within a few hundred microns around the electrode (Buzsáki et al., 2012). Another option is to extract action potentials or spikes from the high-frequency signal, usually band-pass filtered between 300 and 3,000 Hz. In the post-processing of the signal (spike sorting), spikes are categorized as single units if the waveform can be clearly identified as action potentials of a single neuron. Otherwise, if spikes are overlapping, they are accumulated as linear combinations of action potentials from small neuron populations in the vicinity of the recording electrode, called multi-unit activity (Rossant et al., 2016). In both cases, computational approaches implemented on the neural prosthesis are challenged to interpret the incoming data accurately. An additional requirement is that these approaches must operate on small-scale and low-powered hardware that is suitable for chronic implantation but provides limited computational resources.

To explore the capability of computational methods to comply with these constraints in a BMBI scenario, the rodent barrel cortex provides an advantageous experimental model (Feldmeyer et al., 2013). Its prominent organization in cortical columns provides a one-to-one topographic representation of single whiskers, sensory organs on the animal's snout that can be mechanically stimulated in a well-controlled manner (Staiger and Petersen, 2021). Whisker stimulation evokes a localized response in the barrel cortex that can be captured with MEAs and analyzed with the appropriate computational approaches like machine learning (ML) classification algorithms. Based on such recordings, it has been shown that ML classification algorithms, or classifiers, can accurately identify the cortical depth of the recording electrode and the deflection intensity of the whisker stimulus (Wang et al., 2018a; Wang et al., 2018b). While these authors showed that a pre-trained model can be applied for classification on a small microchip within a reasonable amount of time, they did not investigate if it is feasible to train ML models on such hardware, which would be required if a model needs recalibration due to small fluctuations in the positioning of the recording electrodes. However, retraining the model is more challenging in this context as it requires more computational resources than using a pre-trained for inference, i.e., classifying brain activity. Petschenig et al. (2022) addressed this challenge by benchmarking algorithms for classifying whisker deflection amplitudes based on MEA recordings from the barrel cortex of an anesthetized rat. Among the benchmarked methods, the Liquid State Machine (LSM) algorithm turned out to be highly accurate, and this is particularly relevant since it implements a Spiking Neural Network (SNN) developed explicitly for neuromorphic hardware. Hardware of this type are small, brain-inspired computing architectures designed for processing brain signals in a highly efficient manner, representing an appropriate option for chronically implanted prostheses (Buccelli et al., 2019; Werner et al., 2016).

The barrel cortex has, therefore, served as an ideal experimental model for investigating which computational approaches are appropriate for a neural prosthesis. The above-summarized developments in the classification of the whisker stimulus intensity based on stimulus-evoked activity from the barrel cortex are highly promising for the field of neuroprosthetics, as they demonstrate the ability to implement BMIs or BMBIs in low-powered hardware and their capability to decode information from neural activity. However, these results are based on recordings under anesthesia, which is known to affect the pattern of neural activity (Steriade et al., 1993). Also, even in the awake state, cortical responses are known to differ depending on whether an animal is actively engaged in a sensory processing task or merely exposed to sensory stimulation without the need to react to it (Otazu et al., 2009; Carcea et al., 2017; De Franceschi and Barkat, 2021). Thus, it remains unclear if ML classifiers also perform well on data recorded from awake and behaving animals. Such data is expected to contain both a higher level of and stronger variability in spontaneous activity compared to the anesthetized state, and therefore, it could be more challenging for the algorithms to detect or classify stimulus-evoked activity. Moreover, whether such models can be transferred from one individual to another is unknown. Such generalizable models would allow for training models on larger datasets composed of recordings from more than one individual, making the models more robust.

In this study, we benchmark lightweight ML algorithms on the task of whisker stimulus detection based on an electrophysiological dataset obtained from the barrel cortex of four mice while they performed a behavioral task. Using recordings from several sessions with different mice allowed us to clearly separate data from one individual for model training and then use this data for model evaluation, providing a clear picture of how these models generalize across individuals.

Relevant to the application of such models, we benchmark their runtime and investigate the impact of including data from multiple cortical layers. For the Liquid State Machine, the fastest and one of the best-performing algorithms, we analyzed how detection accuracy changes for detecting stimuli with different intensities and shorter MEA recordings.

Results

Dataset

The dataset was obtained from the barrel cortex of four head-fixed water-restricted mice while performing a go/no-go whisker stimulus

detection task. Mice were trained to respond to whisker stimulation by licking at a water spout (Figure 1A). During experimental sessions, a silicon probe with two shanks, 32 electrodes each, was used to obtain electrophysiological recordings from cortical layers (L) 2/3, 4, 5, and 6. Based on a current-source-density (CSD) analysis, we identified different cortical layers and assigned all electrodes to their corresponding layer, selecting the electrode that was most central to derive a layer-specific signal from the MEA recordings (see Methods). Using the signals from different layers allowed us to investigate which layer provides the most important information for stimulus detection and whether integrating the signals from all layers leads to more accurate detection.

In total, 19 experimental sessions were performed with 300 trials each. Each trial of the behavioral task consisted of different epochs (Figure 1B). Before whisker stimulation, the mice had to refrain from licking for at least 400 ms (silence period). Since stimuli were only presented when animals refrained from licking for at least 0.4–3 s (varying randomly from trial to trial), we were able to use the 400 ms before stimulus onset, called *pre-stimulus*, to characterize spontaneous activity in the barrel cortex. The stimulus itself consisted of a single-whisker 60-Hz sinusoidal vibration of 100 ms duration, delivered in the rostrocaudal direction to the C1 whisker. We used this *peri-stimulus* time window to analyze stimulus-evoked activity. If the mouse responded within 500 ms after stimulus onset, it was rewarded with a water droplet. Each recording session included 50 catch-trials without stimulation (0% intensity to assess spontaneous licking) and stimulus trials with different intensities ranging from 20 to 100% of maximum amplitude (corresponding to amplitudes between 93 and 463 µm and angular velocities ranging from 324 to 1,600 °/s at 2 mm from the whisker base, respectively; see Vandevelde et al., 2023, for more

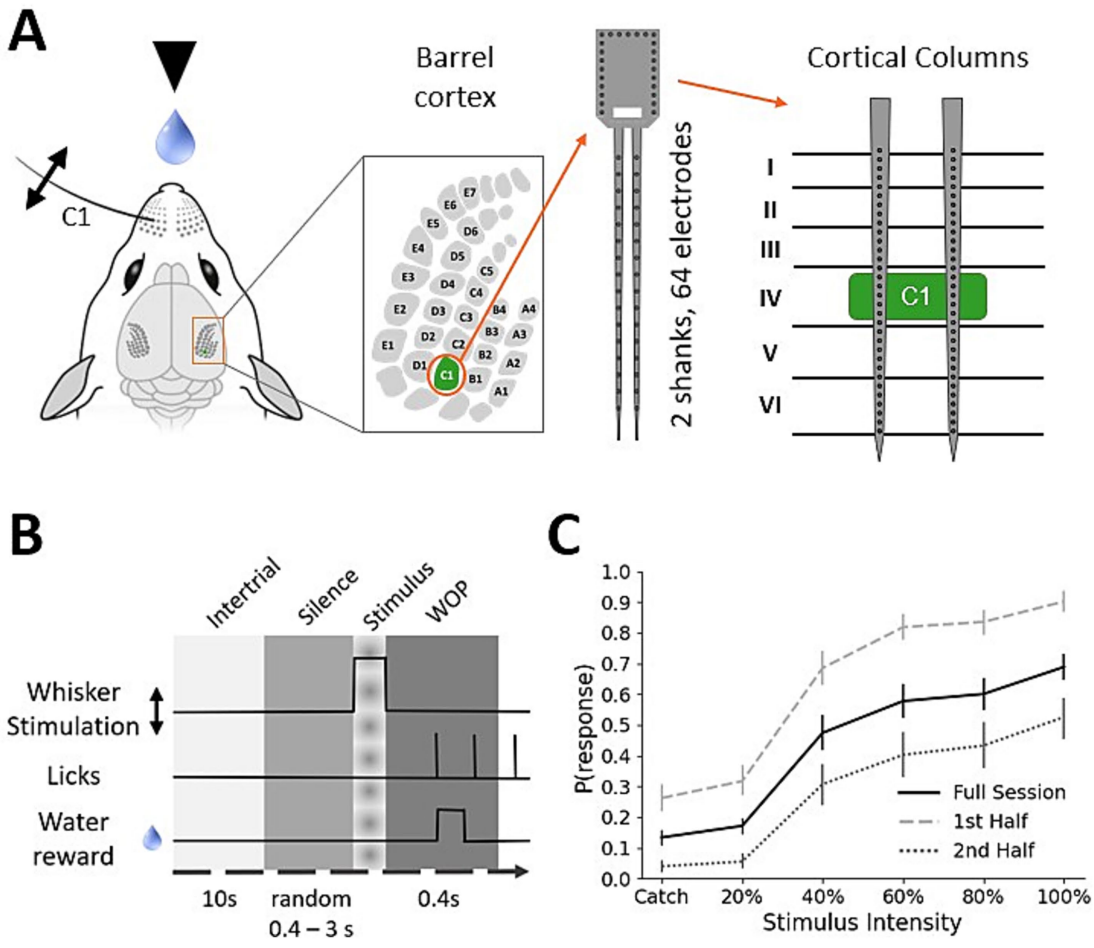


FIGURE 1
Experimental setup and psychometric curves. **(A)** Head-fixed mouse within reach of a water spout. Multi-electrode-array recordings are obtained from the cortical column associated with the stimulated whisker C1. A silicon probe with 64 electrodes, 32 on each of two shanks, was used to measure activity in all cortical layers, as shown in the sketch on the right-hand side panel, with barrels indicated in L4. For histological image examples, we refer to our previous experiments, in which shanks were only inserted after their location had been marked (Yeganeh et al., 2022). **(B)** Overview of the behavioral task. Mice were conditioned to respond to whisker stimulation by licking at the water spout. Responding to stimulus trials (but not catch trials) within 500 ms was rewarded with a drop of water. We refer to the time window from −400 ms to 0, relative to stimulus onset, as the pre-stimulus, or spontaneous, activity. The peri-stimulus or evoked activity is represented by the time window 0 to +100 ms as the stimulus is a sinusoidal whisker vibration of 100 ms duration. A hit trial, a stimulus trial with a successful mouse response, is defined as a lick within the window of opportunity (WOP) from +100 to +500 ms. Trials with licks between 0 and +100 ms were excluded as licking activity caused strong artifacts in the LFP. **(C)** Psychometric curve averaged across 19 sessions from 4 mice. Only trials that remain after all filtering criteria are applied are considered (see Methods). Whisker stimulus intensity ranged from 0% (no stimulus) over 20, 40, 60, and 80–100% intensity (percentage of maximum deflection amplitude). Trials without stimulus (0% intensity) are called catch trials and were included to estimate the response rates achieved by random licks. Licks within the WOP in catch trials are called false alarms. Error bars show the standard error of the mean (SEM) computed over sessions. Subpanels in A were modified from (Yamashita et al., 2018).

details). Stimuli in this range have been found to induce weak to moderately strong activity in barrel cortex and therefore span the range of barely to easily detectable stimuli (Yang et al., 2016; Vandevelde et al., 2023). Each stimulus was presented 50 times, and the stimulus sequence was randomized. The use of different intensities allowed us to investigate the response rate of mice during the behavioral experiment, as well as investigate the shape and strength of the evoked response in the barrel cortex. Using a randomized sequence was important to prevent mice from learning any patterns in whisker stimulation, which could potentially bias the response. The response rate increased monotonically with increasing stimulus intensity (Figure 1C), which is expected considering similar experiments (Stüttgen et al., 2006; Stüttgen and Schwarz, 2008; Stüttgen and Schwarz, 2010).

For more details about the behavioral experiment, we refer to our previous publication (Vandevelde et al., 2023).

Setup of machine learning classifier benchmark

As we integrated data from different recording sessions, we could not use the spiking activity as each session provides a different set of single- and multi-units, which precludes the assembly of a common feature vector to describe trials across sessions, a prerequisite for the ML algorithms. Besides this, it has been demonstrated on a single-session dataset that (i) LFP features are as informative as features derived from multi-unit activity and (ii) it is not necessary to derive more sophisticated features from the LFP, such as the response peak amplitude or response onset latency, because using the raw signal provides sufficient information for the ML methods used (Petschenig et al., 2022). In general, the LFP recording is more stable in comparison to action potentials, and therefore, it is recommended to use the LFP in the context of neuroprosthetics (Andersen et al., 2004; Markowitz et al., 2011). Following these previous studies, we use the raw LFP traces at 1,000 Hz resolution, low-pass filtered at 150 Hz but without further feature engineering, referring to them as *RAW* features, hereafter. Additionally, we investigate *FFT* features, consisting of the frequency components we obtain by applying the Fast Fourier Transform to the filtered signal (Brigham, 1988). Including the FFT features was inspired by Sederberg et al. (2019), who demonstrated that field potentials in the barrel cortex can be informative with respect to how whisker stimuli are perceived in the barrel cortex of awake mice.

We chose to use the LFP for two purposes: first, stimulus detection (SD), i.e., to decode whether a stimulus was presented to the animal, and second, response prediction (RP), i.e., to predict whether the animal was responding to the whisker stimulus. To that end, we benchmarked six ML classification algorithms, also called *classifiers*, known for good performance in various applications (see Table 1 for an overview of the selected classifiers). The Decision Tree (DT) achieves short runtimes due to its simplicity (Loh, 2011). Compared to a single decision tree, Random Forest (RF), which makes use of an ensemble of decision trees, and XGBoost (XGB), which uses stacked decision trees, are known to provide more accurate models while maintaining low runtime requirements (Breiman, 2001; Chen and Guestrin, 2016). Similar characteristics are associated with generalized linear models (GLM) that are, once trained, very fast during inference as they rely on a linear combination of weighted variables, so the logistic regression, a GLM with underlying logit function, has been included in the benchmark (Kleinbaum and Klein, 2010). Interestingly, comparable studies investigating brain activity classification did not explore algorithms developed for time series classification. We, therefore, included ROCKET (RCKT), a classification algorithm dedicated to time series input (Dempster et al., 2020; Pantiskas et al., 2022). As mentioned above, spiking neural networks are promising due to their ability to run efficiently on low-powered neuromorphic hardware, which motivated the inclusion of the Liquid State Machine (LSM). More complex artificial neural network architectures, such as long short-term memory (LSTM), have not been explored due to their high computational requirements and because they did not provide better models in comparable studies (Petschenig et al., 2022).

All algorithms were compared based on different classification scenarios. The term *classification scenario* is here used to describe a specific combination of which classification task (SD or RP) is considered and which features are used, further specifying the scenario by type (RAW, FFT), cortical layer, and time window from which the recording was derived (Figure 2A). For each algorithm, hyperparameter tuning was applied by evaluating a grid of parameter settings, also called grid search, using 3-fold cross-validation on the training data (see Supplementary Table T1 for details). Furthermore, the model evaluation was carefully done, splitting up experimental trials based on individual mice to ensure cross-individual validation in a way that trials from one individual are separated by the training set, used for hyperparameter tuning and training the classification model and

TABLE 1 Overview of ML classifiers used in the benchmark.

Algorithm	Runtime requirements	Developed for time series	Application examples related to neuroprosthetics
DT	Low		Srimaharaj and Chaisricharoen (2021) and Chikkudu and Annamalai (2024)
RF	High		Bhadra et al. (2025) and Giri et al. (2024)
XGB	Medium		Tiwari and Chaturvedi (2019) and Khanday et al. (2024)
RCKT	Medium	✓	Menon et al. (2023)
GLM	Low		Giri et al. (2024) and Chikkudu and Annamalai (2024)
LSM	Low		Petschenig et al. (2022) and Behrenbeck et al. (2019)

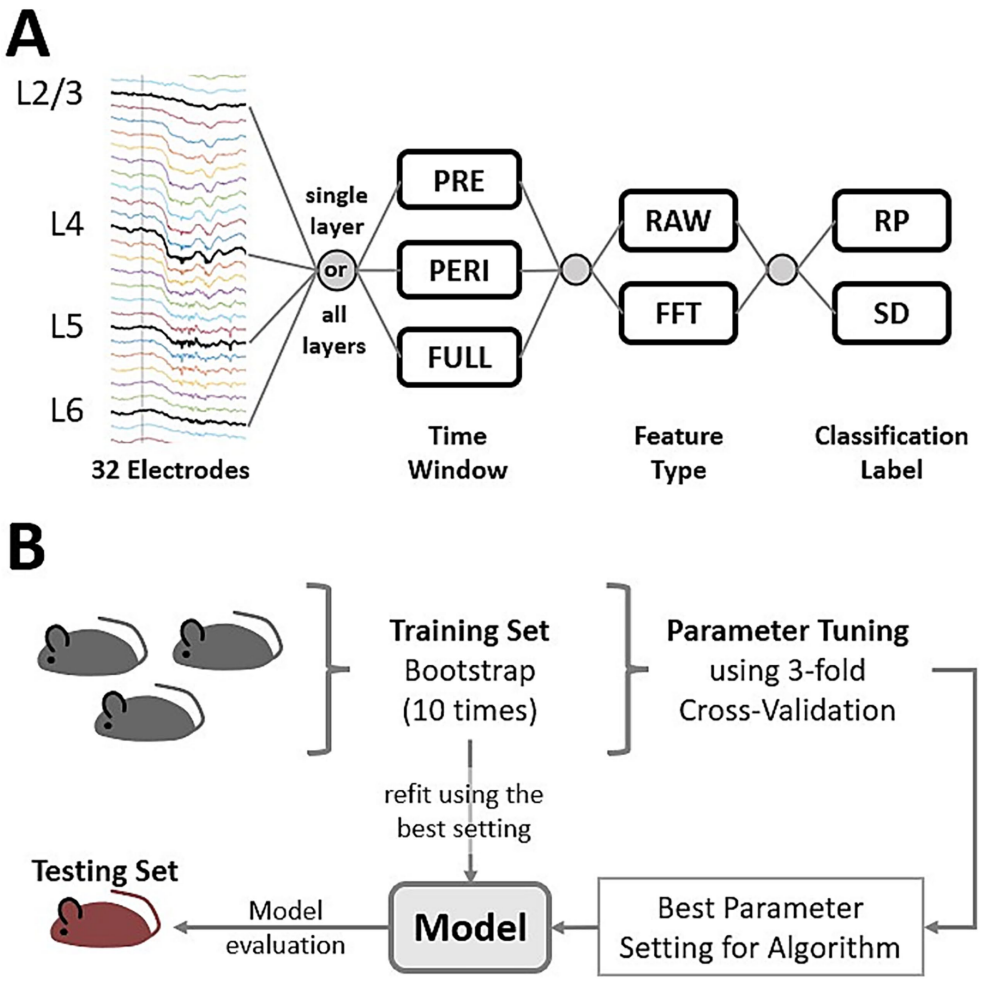


FIGURE 2
Machine learning features and setup of the benchmark. **(A)** Data flow for the machine learning classification scenarios. Four electrodes were selected to derive the LFP for different cortical layers. The time windows, PRE-stimulus, PERI-stimulus, or both (FULL), and the feature type specified the ML features. The classification label specifies the task of the algorithms, which was either response prediction (RP) or stimulus detection (SD), the latter being the focus of this study. Using different combinations of layer, time window, feature type, and classification label results in 24 different classification scenarios. **(B)** Essential for the benchmark was splitting experimental trials into training and testing sets. The training set was used to explore the hyperparameter settings of the machine learning algorithms. Given one algorithm, multiple hyperparameter settings were evaluated by a grid search using a three-fold cross-validation. The setting providing the best results in the grid search was used to train a model based on the full training set and eventually validated using the testing set. The testing set was created using data from only one mouse to ensure that trials from the same individual or experimental session were not included in both training and testing sets. This was repeated for four mice. Given one training–testing split, 10 bootstrap samples were drawn from the training set to train and evaluate multiple models. Given one classification scenario and one algorithm, this results in 40 model evaluations, increasing the statistical power of aggregated measurements for model accuracy. Feature values for RAW and FFT were linearly scaled using Min-Max Scaling (scaling values to the range [0, 1]). The minimum and maximum were computed from the training set only, then used to scale feature values in the training and testing sets.

testing set, used to evaluate the model performance (Figure 2B). Training–testing splits have been repeated four times, using four individuals and for each split, 10 bootstraps were drawn to increase the reliability of the aggregated accuracy measures. For each bootstrapping, the size of the training and testing set differs because the number of experimental sessions available differs between mice (see Supplementary Table T2 for details about the set sizes).

In all classification scenarios, we used perfectly balanced datasets, meaning each subset contained the same number of positive trials (stimulus trial or response trial) and negative trials (catch trial or no-response). This was achieved by down-sampling trials from the over-represented, and due to the expected 50% positives in the data, we can assume that the baseline accuracy achieved by random predictions converges to 50%. Note that including some scenarios was mainly motivated by creating a baseline scenario in which a low accuracy (close to 50%) is expected. For instance, we would not expect pre-stimulus LFPs to encode information useful for stimulus detection. Hence, an accuracy of 50% would be expected for this scenario from any classifier. For response prediction, the expectation can be slightly different. Even though the barrel cortex is not involved in decision-making, its spontaneous activity might encode different states related to attention and excitement, which is relevant to the behavioral experiment (Cowley et al., 2020). Hence, it might be possible for classifiers to leverage weak patterns predictive for the response, even from pre-stimulus (spontaneous) activity.

Machine learning classifier benchmark

The first analysis addresses the question of the extent to which ML classification algorithms, also called *classifiers*, can leverage

10.3389/fnins.2026.1605209

information from LFP features to train accurate classification models for response prediction (RP) or stimulus detection (SD) and which classifier trains the most accurate models.

Figure 3 provides an overview of the accuracy obtained for all classifiers (innermost *y*-axis) applied to the classification scenarios, specified by the two classification tasks (SD, RP), the time window for the recording (PRE, PREI, FULL), feature type (RAW, FFT), cortical layers (innermost *x*-axis), and the stimulus intensities used for trial selection (see subtitles). Brighter colors represent higher accuracy observed for SD in general. Using the PERI and FULL time window results in higher accuracy for both tasks, SD and PR. Regarding the feature types, the RAW features are more informative than the FFT, shown by higher accuracy for both SD and RP. The lower half of the heatmap shows accuracy measures obtained for SD, for which a higher level of accuracy can be observed in general. The accuracy for SD reached up to 90% when trials with 0% (catch-trials) and maximum stimulus intensity of 100% were contrasted.

Features derived from L4 are more informative than those derived from other layers. Considering a feature vector combining all layers (ALL on the *x*-axis) slightly improves the accuracy over the performance of L4 alone. As expected, the peri-stimulus activity (PERI), describing the stimulus-evoked response in the LFP, is more informative for both classification tasks in comparison to the pre-stimulus activity. Classification models trained on pre-stimulus activity features are unable to perform accurately for SD, as expected. For RP, however,

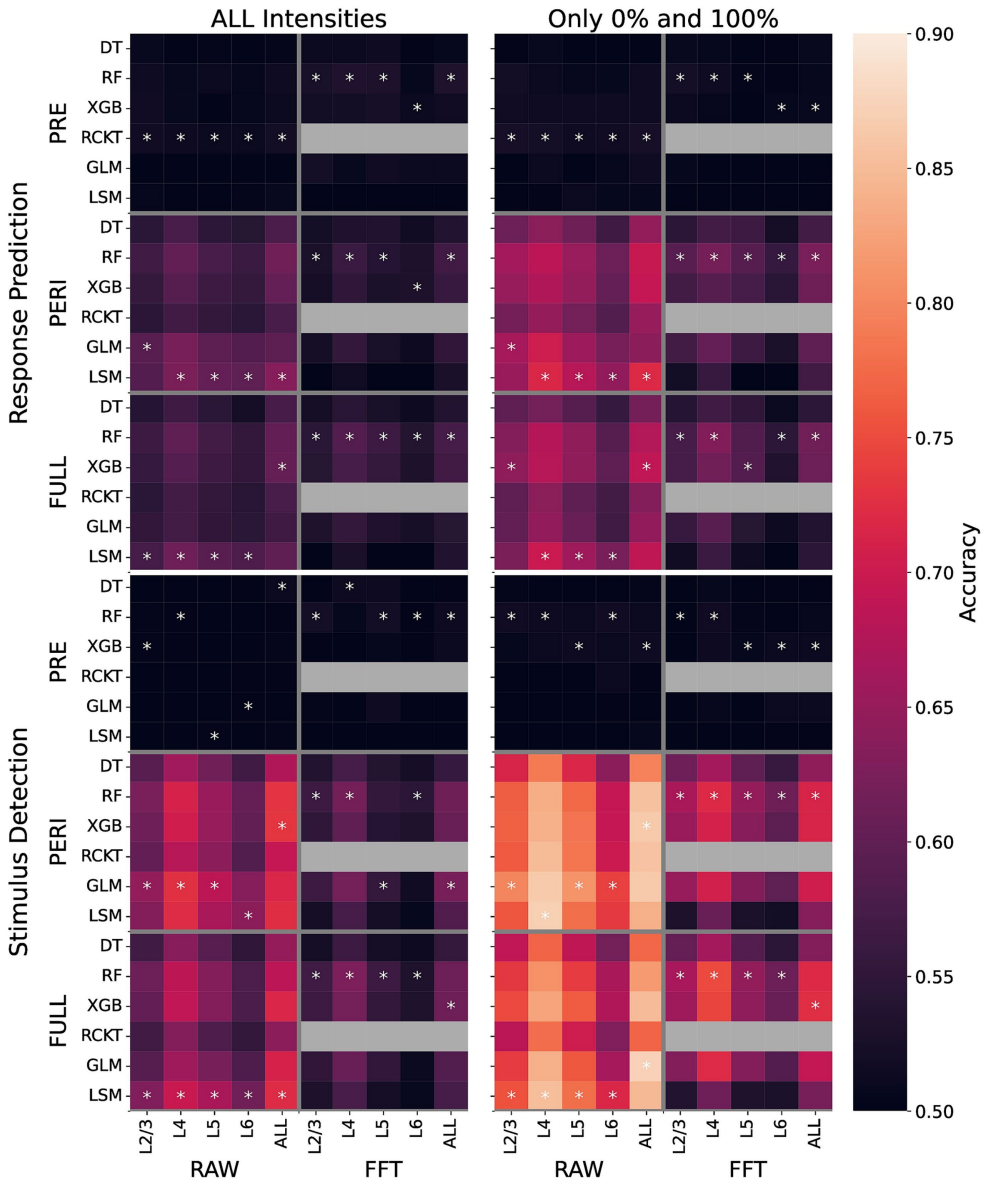


FIGURE 3
Machine learning benchmark. Accuracy represents the classification performance of six different algorithms in several supervised classification scenarios, mainly defined by the task of stimulus detection (SD) and response prediction (RP). These are specified by the time window before (PRE) or after (PERI) stimulus onset or the complete time window (FULL), concatenating the PRE- and PERI-stimulus LFP signal. Furthermore, there are different feature types, RAW and FFT, which were derived from the LFP measured in different cortical layers (L2/3, L4, L5, L6) or using a concatenation of feature vectors from all layers (ALL). The underlying datasets contain either trials from all intensities or only catch trials (0%) and full intensity (100%) trials. The asterisks indicate which classifier achieved the best accuracy.

the accuracy slightly increased over the 50% baseline when RF is used on pre-stimulus FFT features (Figure 3).

Regarding the classifiers, RF performs well in scenarios exploring FFT features. For the RAW features that are more informative for the SD task, we observed that the GLM and LSM achieved the highest performance, with RCKT slightly lower. We do not see an improvement in model performance for SD when integrating the FULL time window over using peri-stimulus alone (Figure 3).

However, from the median accuracies discussed in this section, it is not clear if it is worth integrating all layers, expected to increase model training time, over using solely L4. Moreover, a runtime analysis is required for the best algorithms, considering the practical application of these models. Therefore, a more detailed comparison of the most accurate algorithms follows, focusing on SD based on the peri-stimulus RAW features.

Detailed comparison of the most accurate algorithms

For this comparison of the best-performing classifiers, the SD classification task has been selected with 0 and 100% intensity trials only, because the highest classification accuracy was achieved in this scenario. Considering GLM, RCKT, and LSM as the top-performing classifiers in terms of accuracy, the LSM is the strongest competitor based on the runtime comparison (Figure 4A). RCKT was faster than the GLM for model training but much slower during inference. In general, the runtime for model training and inference was higher when all layers were used, which is expected as algorithms are processing four times more ML features than using only a single layer. Focusing on LSM and GLM as the fastest algorithms, we investigated the difference in accuracy for the SD task using all layers and only L4 (Figure 4B). In this comparison, the LSM achieved the lowest performance when all layers were integrated, which suggests that the GLM is more efficient in dealing with high-dimensional data as it achieved the highest median accuracy overall when it integrated all layers. However, as the accuracy achieved by the GLM with all layers is not significantly higher than that of the LSM with only L4, we conclude that the LSM is the best choice for stimulus detection, especially due to its superior performance in the runtime comparison with the lowest model training and inference runtimes using only L4. Furthermore, it achieved slightly more robust accuracy distributions across all individuals than the GLM, for which more outliers and lower median accuracy measures were observed overall (Figure 4C). To assess whether the accuracy scores of the GLM and LSM (Figure 4B) differed significantly, the Wilcoxon signed-rank test was applied on 40 pairs (four mice times ten bootstraps). A pair of values consists of the accuracy values from the two classifiers, obtained on the same mouse-specific testing set and the same bootstrap sample, using random seeds to ensure bootstrap samples are equal.

Stimulus detection for different intensities using shorter peri-stimulus LFP signals

For practical applications, it is important to know how accurate the detection is for stimuli of different intensities (Petschenig et al., 2022; Wang et al., 2018b). Furthermore, it is crucial to investigate how long the LFP recordings need to be to maintain accuracy at the highest level possible. The shorter this recording time is, the earlier the model can be applied after stimulus onset. To address these two aspects, an additional SD experiment was performed using subsets of the data

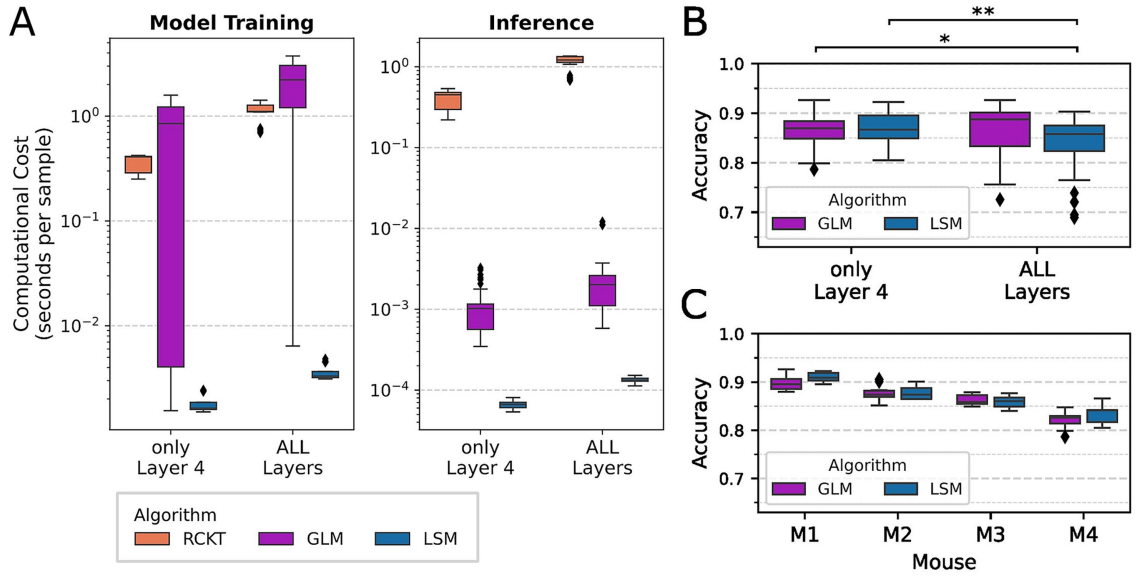


FIGURE 4
More detailed comparison of the most accurate and fastest algorithms. The scenario selected for these detailed visualizations is the SD using RAW features based on 0 and 100% intensity trials only. **(A)** The computational time was measured during model training and when a trained model was applied to samples from the testing set to classify those (inference). As training and testing sets differ in size depending on which mouse was used to produce the split, the runtime, measured in seconds, was normalized using the size of the training and testing sets, respectively. Note that the range of the *y*-axis differs between the subplots and that the *y*-axis is on a logarithmic scale. **(B)** Comparison of the accuracy distribution achieved by GLM and LSM using the raw LFP (see RAW in Figure 3) derived from layer L4 or concatenating all layers. The asterisks indicate significant differences with * →*p* value < 0.05 and ** →*p* value < 0.01 using the Wilcoxon signed-rank test. Distributions are not significantly different if nothing is indicated. **(C)** Classification performance of GLM and LSM for the different training, testing splits based on four mice. Boxplots describe the distribution of accuracies measured on 10 bootstrap samples. Diamonds show the outliers of the boxplot.

with catch-trials and stimulus trials of intensities from 20 to 100%. Note that this includes the results from the benchmark obtained for 100% stimulus intensity and the complete 100 ms of peri-stimulus activity. These SD experiments were further specified by different lengths of the RAW LFP signal as input for the classification model, using L4 as the most informative layer. All additional classification experiments were done for the LSM using the validation strategy applied during the initial ML benchmark analysis (Figure 2B).

We observed that LSM's detection accuracy is lower for stimuli of low intensities (Figure 5). In comparison to the accuracy achieved on the complete (100 ms) peri-stimulus LFP, approximately 35–40 ms is required to achieve a high level of accuracy for the higher intensities greater than or equal to 60%. For 40% intensity, even though the overall accuracy was much lower than for higher intensities, the LFP can be as short as 15 ms to achieve high accuracy for this case. Note that the RAW features were used for these tests, representing the LFP after low-pass and band-stop filtering, which further delays this process by a total of 0.2 ms (measured on a common CPU, expected to be slightly higher on neuromorphic hardware).

Imbalanced data and model calibration

To mitigate potential impacts of class imbalance (over- or under-representation of the positive class), datasets have been downsampled to train models on data with 50% positive (stimulus present) and 50% negative (stimulus absent) samples. For deployment, however, it is important to investigate how algorithms perform on imbalanced data. Using L4 as the most informative layer and RAW as the most informative feature type for stimulus detection, we benchmarked the GLM and LSM again and added comparisons on data without downsampling. We used the dataset with all stimulus intensities, resulting in 83% stimulus trials, and the dataset containing only 0% stimulus intensity (catch) trials and 100% stimulus intensity trials, for which downsampling had almost no impact, as the fraction of positives changed from 50 to 49.3%. Due to the class imbalance in the data, we added ROC and Precision-Recall curves to evaluate model performance (Saito and Rehmsmeier, 2015). Additionally, we used the Brier loss score to evaluate model calibration and studied the impact of Platt Calibration on the model probabilities from GLM and LSM (Platt, 1999; Niculescu-Mizil and Caruana, 2005).

The results from this analysis (see Supplementary Figure S1) show that (i) the GLM is well-calibrated and does not benefit from Platt Calibration, (ii) the LSM needs Platt Calibration to improve its calibration performance, especially for imbalanced data, and (iii) the overall performance of GLM and LSM is very similar, both achieving area under ROC and Precision-Recall curve clearly better than a random guess or naïve model.

Note that testing data is not used for Platt Calibration. A default Logistic Regression (scikit-learn) is trained as a calibration model on the classification probabilities and class labels of a 33% hold-out dataset from the training set. Consequently, the model evaluated on the mouse-specific testing set has been fitted on only 67% of the training set when using Platt Calibration. The fitted calibration model is applied to transform the classification probabilities from the testing set before computing the evaluation metrics.

Discussion

Three ML classification algorithms were identified as the best-performing ones, achieving the highest accuracy for stimulus detection. In a subsequential runtime analysis, the LSM has been superior in achieving the lowest runtimes for model training and inference.

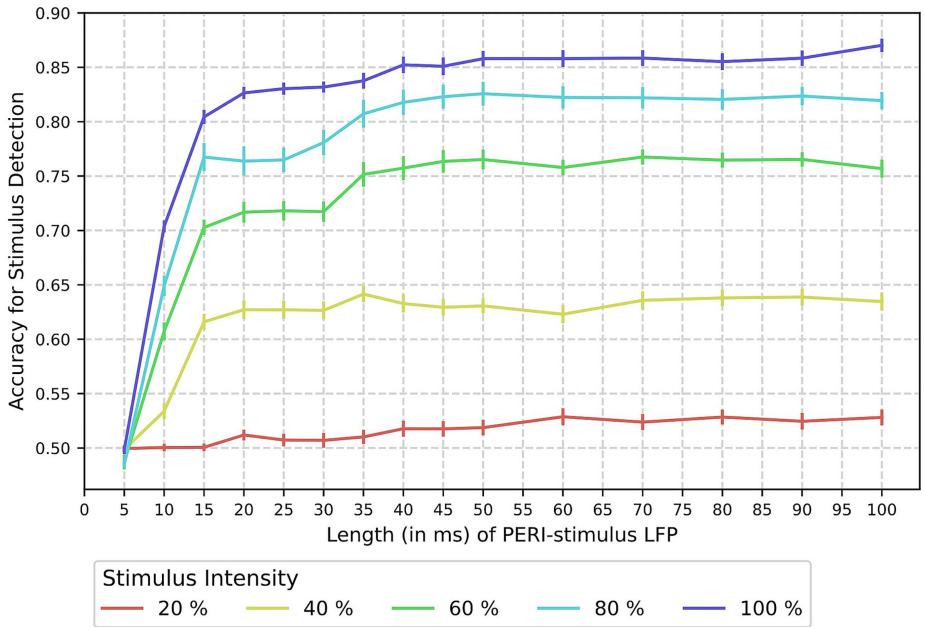


FIGURE 5
Accuracy of stimulus detection for different intensities using shorter peri-stimulus LFP signals. Accuracy for stimulus detection when the LSM is trained and evaluated to differentiate catch trials (0% intensity) from stimulus trials of different intensities (20–100%). Each classification experiment was repeated using shorter LFP recordings derived from the peri-stimulus signal, as shown on the *x*-axis. The lines describe the mean accuracy based on 40 training–testing splits, and the error bars describe the standard error of the mean.

Hence, the LSM has not only been faster within the runtime analysis performed on an ordinary CPU but it is also expected to be more efficient when deployed on neuromorphic hardware as it, in contrast to other machine learning algorithms, mirrors the physical architecture of low-powered chips used in such hardware. Accordingly, the LSM could even be used to train a model on the prosthesis, which is essential for recalibrating a model by retraining after it has been implanted (Petschenig et al., 2022). Recalibration of a model can be necessary if the electrodes measuring the input signal are slightly moved and small shifts in the signal are expected.

Furthermore, the results of this study show that the classification models are robust across four mice from which the data was derived, with the LSM being slightly more robust than the GLM. This is promising as models could potentially be trained on data collected from a large cohort of patients and then deployed in new patients. When the model is used for classification on low-powered hardware, the low inference times are paramount. While the LSM and GLM were much faster when applying the model (inference) than training it, the ROCKET classifier showed similar runtime requirements during training and inference. This is explained by ROCKET's algorithmic strategy of applying convolutional kernels to the time series signal before using a sample as input. This signal preprocessing is required during training and when a trained model is used. Hence, this requires additional computational resources during inference.

The runtime analysis also showed that all algorithms require less computational resources when only one layer is used. Even though the GLM achieved the highest stimulus detection accuracy when the signal of all layers was integrated, its accuracy was not significantly higher than that of the LSM when applied with only L4. Observing L4 as the most informative layer within the barrel cortex is not unexpected, as it is the main recipient of the thalamic input about whisker deflection, and multi-unit activity in this layer features higher neurometric detection sensitivity than in other layers (Vandevelde et al., 2023; Reyes-Puerta et al., 2015; Reyes-Puerta et al., 2015). The signal is expected to become noisier as it is forwarded to other layers of barrel cortex, which also receive information from other cortical areas, and may integrate this information with texture information provided by consecutive whisker-object contacts for routing to action-related cortical areas (Zuo and Diamond, 2019; Gămănuţ et al., 2018). The redundancy of whisker information represented in different cortical layers, along with (apparent) noise added from other brain areas, could explain why the LFP in layers other than L4 does not strongly enhance stimulus detection (Arabzadeh et al., 2006). Additionally, using a concatenation of feature vectors from all layers results in high-dimensional data, characterized by a high number of features but a low sample size, which can lead to less accurate stimulus detection models, as ML algorithms are less effective in capturing meaningful patterns in the large number of features (Horenko, 2020; Vecchi et al., 2022). We conclude that layers other than L4 can provide additional information to slightly improve the SD accuracy, as shown by the GLM, but that these improvements are too small to justify choosing the GLM over the LSM, being much faster and only slightly less accurate, when using solely L4.

Besides the runtime of the machine learning model itself, it is also important to know how long the LFP signal needs to be for accurate stimulus detection. We, therefore, analyzed the classification performance for stimulus detection using shorter RAW feature vectors and found that high accuracy, as observed for 100 ms of the peri-stimulus signal, can be maintained using only 35–40 ms of the signal. For practical applications, this means that a delay by this time frame should be expected before a stimulus can be detected and processed in a BMI or BMBI device. This analysis has also been done for different intensities and reveals that accuracy decreases with decreasing intensity. This is a critical observation, indicating that only strong stimuli are detected reliably. At the same time, it reflects the perception of mice as they also showed a lower response rate for weaker stimuli during the behavioral experiment (Figure 1C), and because generally, whisker stimulus intensities were relatively low compared to the full range of possible contacts during unrestrained locomotion (Ritt et al., 2008).

Generally, an LFP signal describing the evoked response of a whisker stimulus in the barrel cortex is characterized by a strong drop in the current approximately 5–10 ms post-stimulus onset (see example sessions in Supplementary Figure S2). This change in the signal seems to be strong and consistent enough that a classifier does not need to leverage complex, non-linear patterns, which is implied by the strong performance of the linear model (GLM) used in the benchmark. This might also explain why the RAW LFP signal is more informative than oscillatory patterns encoded by the FFT features. Comparing the LFP and FFT features of two sessions from our dataset revealed an increased amplitude for the frequency components describing 60 Hz, and its harmonics can exist (Supplementary Figure S2). However, these FFT patterns are inconsistent and only present if the electrode has been close to a cluster of velocity-sensitive neurons activated by the peak velocity for the whisker movement during stimulation. Remember that the vibrotactile stimulus was a cosine of 60 Hz, moving the whisker in the anterior–posterior direction. Importantly, in the primary afferent neurons of the trigeminal ganglion (the first station of the ascending whisker sensory pathway), activity in response to whisker deflection is strongly determined by stimulus velocity (Stüttgen et al., 2006). This is also true for whisker-responsive neurons in the thalamus (Temereanca and Simons, 2003) and barrel cortex (Pinto et al., 2000).

Response prediction and the patterns in spontaneous activity

Besides investigating neural activity with respect to how whisker stimuli are processed in the barrel cortex, the dataset allowed us to search for patterns predictive of the behavioral response. This is a critical aspect as we record from a somatosensory brain region believed to be involved in the process of evaluating an incoming stimulus and provide the basis for sensory-driven decision-making (Staiger and Petersen, 2021). Considering the response prediction analysis as part of the ML benchmark, we observe a moderate accuracy for response prediction models trained on the RAW features from the peri-stimulus or full signal (Figure 3). However, an additional analysis based on the predictions of the corresponding models strongly suggests that the response prediction models actually perform stimulus detection, which results in quite accurate models due to the strong correlation between the presence and absence of stimulus and the occurrence of behavioral response (see Supplementary Figure S3).

Interestingly, our benchmark revealed that patterns exist in the harmonic oscillations derived from spontaneous (pre-stimulus) activity that are slightly predictive toward the response of mice (see RP with pre-stimulus FFT features, all intensities, Figure 3). For the FFT features derived from L2/3, L4 or L5 we observed accuracy slightly, but significantly better than the random baseline when all stimulus

intensities are considered in the data (p-value from Fischer's test < $1e^{-20}$ in all three cases). Even though this difference in accuracy is small, these results indicate that neuronal activity in the barrel cortex is also related to attention or anticipation of the expected whisker stimulus (Lee and Dan, 2012; Stüttgen and Schwarz, 2018). Such weak patterns in the spontaneous activity could reflect brain states in which the mouse is more attentive to perceive the stimulus, which impacts the response, especially when not only the full-intensity trials but also those with a small or intermediate intensity are analyzed. While these are indications relevant to the role of the barrel cortex in the behavioral response of mice, future work is required to further investigate the FFT features, which is out of scope for this study and might require further behavioral experiments.

Practical application and limitations

Previous *in vivo* recordings in the rat barrel cortex have identified inhibitory interneurons in L4 as the units carrying the highest amount of sensory stimulus-related information (Reyes-Puerta et al., 2015; Reyes-Puerta et al., 2015). Thus, neuronal activity in L4 represents a good target to obtain information on the spatio-temporal properties of the sensory input. Measuring the activity from L4 is not only possible but also very reliable with invasive technologies as used in our experimental model. However, in a practical application, it will be more beneficial to use non-invasive technologies to avoid an intervention in the patient's brain. Resolving activity in a specific portion of tissue without having direct access could be realized by laminar inference based on MEG recordings, although MEG is not applicable in neuroprosthetics (Bonaiuto et al., 2018).

Our study confirms that ML models can be highly accurate on whisker stimulus classification, even for LFP signals recorded from awake mice. However, the datasets used for such explorations, including the one we used, are based on precisely controlled single-whisker stimuli, certainly different from how whiskers bend and move when animals are actively palpating during unrestrained exploration (Petschenig et al., 2022; Wang et al., 2018a; Wang et al., 2018b). Therefore, for future work, it would be relevant to apply and test the suggested algorithms on recordings from microchips implanted in freely moving animals. The hardware required to do this could use wireless and battery-free devices (Martinez et al., 2018; Won et al., 2023). Recordings from such experiments might result in LFP traces that are less distinguishable, as we would expect that whisker deflections from freely moving animals are less distinct.

Additionally, our benchmark for imbalanced data covers only the case of overrepresentation of the positive class, which is the opposite of a real-life scenario in which stimulus events are less frequent compared to non-stimulus recordings. Our dataset provided well-defined trials for events with an induced stimulus, with five different intensities over catch trials without a stimulus. Thus, using all trials of the dataset without downsampling results in overrepresentation of the positive class. Nevertheless, this analysis demonstrates that both algorithms perform well on imbalanced data in general and reveals that additional calibration of classification probabilities is required when using the LSM, particularly on imbalanced data. Interestingly, on the imbalanced data, the GLM and LSM achieved accuracy values close to those expected from a naïve model that always returns the positive class (stimulus present). However, both algorithms yielded high area-under-the-curve scores when using the ROC and Precision-Recall curves as threshold-independent evaluation metrics. We therefore conclude that, in imbalanced deployment settings, both algorithms require decision threshold optimization, which is left for future work.

To realize the closed-loop system, the output of the ML classifier could be used to control an electrode located in another brain region to induce intracortical microstimulation that can convey artificial signals that may be as complex as the somatic perception of tactile properties from different objects (Jackson et al., 2006; O'Doherty et al., 2011). However, a complete experimental setup would be required to verify whether the whole process of reading in the MEA recording, applying the ML classifier, and creating a feedback signal is fast enough to restore brain functionality, establishing timely signal processing with other cortical areas.

Model robustness and generalizability

Another important aspect is the robustness of the findings concluded from our ML-based analyses. The sample size is a key factor in statistical analyses, and machine learning models tend to be more meaningful and generalizable when trained on large datasets (Carlini and Wagner, 2017; Rice et al., 2020). Even though electrophysiological recordings are challenging from different perspectives, we could explore a comparatively large dataset using >3,400 trials for six stimulus intensities. The final subsets used for the ML-based analysis were restricted by a selection of trials for the different intensities, resulting in a sample size of at least 600 trials (see Supplementary Table T2), which is approximately three times higher than in comparable studies using only one recording session, e.g., (Petschenig et al., 2022; Wang et al., 2018a; Wang et al., 2018b).

In summary, based on a large electrophysiological dataset enabling a robust model evaluation, we demonstrated that SNNs are accurate and efficient in interpreting neural activity toward neuroprosthetics applications related to the barrel cortex. Using the LSM classifier, model training could be realized on a small microchip to be implemented on the prosthesis, which can be beneficial to retrain a model for adjusting to small drifts of the implanted electrode, for instance. While our tests were done on an infrastructure with common CPUs with the main perspective to compare the model training time between the different classifiers, it has been shown recently that LSMs can efficiently be implemented and executed on microchips, as demonstrated on a DYNAP-SE neuromorphic processor (Petschenig et al., 2022). Another opportunity is to train a model on an external machine, transferring the model and only applying it on the prosthesis (Wang et al., 2018b). If future technologies enabled a more stable implantation of the electrodes, retraining on the prosthesis might not be necessary. Our generalizability analysis confirms that using a pretrained model is possible even if the underlying data was derived from an individual different from the patient who needs the prosthesis.

Methods

Electrophysiological dataset

The data investigated for this study were obtained from experiments with head-fixed mice performing a go/no-go whisker stimulus detection task (Vandevelde et al., 2023). Mice were water-restricted and learned that they were rewarded by a drop of water if licking within 500 ms after the onset of a 100-ms sinusoidal vibration of a

single whisker. Before mice were trained to perform this task, they were habituated to the head-fixed setting, which was important to enable the recording of neural activity in the cortical column in the barrel cortex associated with the stimulated whisker. This was done with multi-electrode-array (MEA) silicon probes through all cortical layers using 2-shank-64-channel probes with a distance between the shanks of either 200 or 250 µm. (Neuro Nexus, Ann Arbor, United States, or Cambridge Neurotech, Cambridge, United Kingdom, respectively). Each shank has 32 electrodes, and the distance between electrodes is 25 µm, which enables the observation of neural activity over an extent of ~0.8 mm.

For more details about the mouse line, behavioral task training, habituation, surgery, MEA recordings, and the layer-electrode association using current source density (CSD) plots, we refer to Vandevelde et al. (2023). After inspecting the CSD plots, we used one electrode per layer to derive layer-specific LFPs from the MEA recordings. The location and extent of layer 4 could be easily identified by a prominent short-latency current sink in the CSD, computed from the average LFP response to repeated high-intensity C1 whisker deflections, and therefore could be separated from layers 2/3 above and layer 5 below. The border between layers 5 and 6 was estimated based on recording depth (for more details, see Vandevelde et al., 2023, as well as Reyes-Puerta et al., 2015; Yang et al., 2017, 2018). Psychometric curves (Figure 1C) differ slightly from those shown in our previous study (Vandevelde et al., 2023) because we excluded trials with a lick response within the first 100 ms after stimulus onset as the muscle movement related to licking caused artifacts in the recordings.

LFP preprocessing and LFP features

To exclude LFP traces contaminated by artifacts, we applied trial filters for the full range from −420 ms to +120 ms according to the following criteria. Trials in which animals were presumably moving and/or licking randomly had large voltage fluctuations and sometimes burst-like oscillations in activity. Trials in which the signal reached saturation were identified by applying a cutoff of ±2 mV at any time point within the time range specified above. Bursts were identified by moving time windows of 20 ms in which all electrodes strongly correlated, computed by Pearson's correlation (Schober et al., 2018). Electric shorts were partly captured by the MEA electrodes, resulting in a continuous 50 Hz sinusoidal signal, and trials with a strong 50 Hz frequency power were removed. Trials for which a lick was detected within the above-mentioned time window were also excluded.

The RAW features describe the LFP trace at 1,000 Hz resolution, resulting in 400 features describing, for instance, the pre-stimulus activity of 400 ms from −400 ms to 0 ms (stimulus onset). Hence, the peri-stimulus activity from 0 to +100 ms is described by a feature vector of length 100, while 500 features represent the full time window from −400 ms to +100 ms. Using causal Butterworth filtering, the LFP traces were low-pass filtered at 150 Hz, and the 47–53 Hz band was removed, using a first-order low-pass and a sixth-order band-stop filter (Selesnick and Burrus, 2002). Using the same time-windows, the Fast-Fourier-Transform (FFT) was applied to derive the strength of the frequency components (Brigham, 1988). According to the length of the signal derived from the investigated time windows, the resolution is 2.5 Hz and 10 Hz going up to 147.5 Hz and 140 Hz for the pre- and peri-stimulus activity, respectively. The first FFT feature (0 Hz) represents the constant, which is simply the average over the signal. Eventually, 60 pre-stimulus and 15 peri-stimulus features describe the FFT feature vectors, resulting in 75 features for the full range. Due to the strong differences between ongoing and evoked activity, the FFT was applied to those time windows separately, and the resulting FFT features were merged afterward in order to create the full (pre- and peri-stimulus) FFT feature vector.

Soft- and hardware specifications

We used MOGON II, the high-performance computing system from the University of Mainz. The whole analysis was implemented in Python. We ran DT, RF, and GLM from the *scikit-learn* package (Pedregosa et al., 2011) and XGB using a separate library called *xgboost* (Chen and Guestrin, 2016). The algorithm ROCKET (RCKT) was integrated via the *sktime* package (Löning et al., 2019), implementing its univariate and multivariate variants (Dempster et al., 2020; Pantiskas et al., 2022). The LSM algorithm was used from the GitHub repository https://github.com/IGITUGraz/LSM, published by Maass et al. (2002). The CPU specifications on the SMP compute nodes on MOGON are Intel® Xeon® E5 v4, 2.2GHz.

Data availability statement

The datasets used, and the Python code performing the ML benchmark are publicly available in a gitlab repository: https://gitlab.rlp.net/salbrec/sdrpml.git.

Ethics statement

The animal study was approved by a local ethics committee (#23 177-07/G14-1-080) following the European and German national regulations (European Communities Council Directive, 86/609/EEC). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

SA: Data curation, Visualization, Writing – original draft, Formal analysis, Validation, Methodology, Software, Investigation, Conceptualization, Writing – review & editing. JV: Validation, Conceptualization, Formal analysis, Data curation, Investigation, Writing – review & editing. EV: Formal analysis, Visualization, Software, Writing – review & editing, Validation. GB: Software, Formal analysis, Writing – review & editing, Visualization, Validation. DB: Formal analysis, Visualization, Software, Writing – review & editing, Validation, Investigation. MS: Resources, Funding acquisition, Writing – review & editing, Formal analysis, Data curation, Methodology, Conceptualization, Project administration, Supervision, Investigation. HL: Writing – review & editing, Conceptualization, Investigation, Funding acquisition, Supervision, Resources, Data curation, Project administration, Formal analysis, Methodology. IH: Resources, Formal analysis, Supervision, Writing – review & editing, Investigation, Methodology, Conceptualization, Funding acquisition, Software, Project administration.

Funding

The author(s) declared that financial support was received for this work and/or its publication. We gratefully acknowledge the funding from the Carl-Zeiss Foundation (0563-2.8/738/2) initiative "Emergent Algorithmic Intelligence." This work was supported by grants from the Deutsche Forschungsgemeinschaft (LU375/11-1 to HL and STU544/3-1 and STU544/4-1 to MS).

Acknowledgments

Parts of this research were conducted using the supercomputer MOGON 2 and/or advisory services offered by Johannes Gutenberg-University Mainz (hpc.uni-mainz.de), which is a member of the AHRP (Alliance for High Performance Computing in Rhineland-Palatinate, www.ahrp.info) and the Gauss Alliance e.V. The authors gratefully acknowledge the computing time granted on the supercomputer MOGON 2 at Johannes Gutenberg-University Mainz (hpc.uni-mainz.de).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2026.1605209/full#supplementary-material

References

Andersen, R. A., Musallam, S., and Pesaran, B. (2004). Selecting the signals for a brain–machine interface. *Curr. Opin. Neurobiol.* 14, 720–726. doi: 10.1016/j.conb.2004.10.005

Arabzadeh, E., Panzeri, S., and Diamond, M. E. (2006). Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *J. Neurosci.* 26, 9216–9226. doi: 10.1523/jneurosci.1491-06.2006

Behrenbeck, J., Tayeb, Z., Bhiri, C., Richter, C., Rhodes, O., Kasabov, N., et al. (2019). Classification and regression of spatio-temporal signals using NeuCube and its realization on SpiNNaker neuromorphic hardware. *J. Neural Eng.* 16:026014. doi: 10.1088/1741-2552/aafabc

Bhadra, K., Giraud, A.-L., and Marchesotti, S. (2025). Learning to operate an imagined speech brain-computer interface involves the spatial and frequency tuning of neural activity. *Commun. Biol.* 8:271. doi: 10.1038/s42003-025-07464-7

Bonaiuto, J. J., Rossiter, H. E., Meyer, S. S., Adams, N., Little, S., Callaghan, M. F., et al. (2018). Non-invasive laminar inference with MEG: comparison of methods and source inversion algorithms. *NeuroImage* 167, 372–383. doi: 10.1016/j.neuroimage.2017.11.068

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Brigham, E. O. (1988). *The Fast Fourier Transform and its Applications*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, Inc.

Buccelli, S., Bornat, Y., Colombi, I., Ambroise, M., Martines, L., Pasquale, V., et al. (2019). A neuromorphic prosthesis to restore communication in neuronal networks. *IScience* 19, 402–414. doi: 10.1016/j.isci.2019.07.046

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241

Carcea, I., Insanally, M. N., and Froemke, R. C. (2017). Dynamics of auditory cortical activity during behavioural engagement and auditory perception. *Nat. Commun.* 8:14412. doi: 10.1038/ncomms14412

Carlini, N., and Wagner, D. (2017). "Towards evaluating the robustness of neural networks," in *2017 IEEE Symposium on Security and Privacy (SP)*, (New York City, USA: IEEE), 39–57.

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York City, USA: ACM), 785–794.

Chen-Bee, C. H., Zhou, Y., Jacobs, N. S., Lim, B., and Frostig, R. D. (2012). Whisker array functional representation in rat barrel cortex: transcendence of one-to-one topography and its underlying mechanism. *Front. Neural Circuits* 6:93. doi: 10.3389/fncir.2012.00093

Chikkudu, S., and Annamalai, S. (2024). "Motor-imagery signal preprocessing using the SMOTE technique and neuromuscular disorder detection," in *2024 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, (New York City, USA: IEEE), 324–329.

Cowley, B. R., Snyder, A. C., Acar, K., Williamson, R. C., Yu, B. M., and Smith, M. A. (2020). Slow drift of neural activity as a signature of impulsivity in macaque visual and prefrontal cortex. *Neuron* 108, 551–567.e8. doi: 10.1016/j.neuron.2020.07.021

De Franceschi, G., and Barkat, T. R. (2021). Task-induced modulations of neuronal activity along the auditory pathway. *Cell Rep.* 37, 1–23. doi: 10.1016/j.celrep.2021.110115

Dempster, A., Petitjean, F., and Webb, G. I. (2020). Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Min. Knowl. Discov.* 34, 1454–1495. doi: 10.1007/s10618-020-00701-z

Feldmeyer, D., Brecht, M., Helmchen, F., Petersen, C. C., Poulet, J. F., Staiger, J. F., et al. (2013). Barrel cortex function. *Prog. Neurobiol.* 103, 3–27. doi: 10.1016/j.pneurobio.2012.11.002

Gămănuț, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., van Essen, D. C., Knoblauch, K., et al. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. *Neuron* 97, 698–715.e10. doi: 10.1016/j.neuron.2017.12.037

Giri, H. R., Negi, P. C. B. S., Sharma, S., and Sharma, N. (2024). "An intuitive real-time brain control interface based on motor imagery and execution," in *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, (New York City, USA: IEEE), 1–6.

Guggenmos, D. J., Azin, M., Barbay, S., Mahnken, J. D., Dunham, C., Mohseni, P., et al. (2013). Restoration of function after brain damage using a neural prosthesis. *Proc. Natl. Acad. Sci.* 110, 21177–21182. doi: 10.1073/pnas.1316885110

Horenko, I. (2020). On a scalable entropic breaching of the overfitting barrier for small data problems in machine learning. *Neural Comput.* 32, 1563–1579. doi: 10.1162/neco

Jackson, A., Mavoori, J., and Fetz, E. E. (2006). Long-term motor cortex plasticity induced by an electronic neural implant. *Nature* 444, 56–60. doi: 10.1038/nature05226

Khanday, O. M., Ouellet, M., and Pérez Córdoba, J. L. (2024). Decoding the Mind: Neural Differences and Semantic Representation in Perception and Imagination across Modalities.

Kleinbaum, D. G., and Klein, M. (2010). *Logistic Regression*. New York, NY: Springer.

Krucoff, M. O., Rahimpour, S., Slutzky, M. W., Edgerton, V. R., and Turner, D. A. (2016). Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation. *Front. Neurosci.* 10:584. doi: 10.3389/fnins.2016.00584

Lee, S.-H., and Dan, Y. (2012). Neuromodulation of brain states. *Neuron* 76, 209–222. doi: 10.1016/j.neuron.2012.09.012

Loh, W. (2011). Classification and regression trees. *WIREs Data Min. Knowl. Discov.* 1, 14–23. doi: 10.1002/widm.8

Löning, M., Bagnall, A., Ganesh, S., Kazakov, V., Lines, J., Király, F. J., et al (2019). Sktime: a unified interface for machine learning with time series. arXiv preprint arXiv:1909.07872. doi: 10.48550/arXiv.1909.07872

Maass, W., Natschläger, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* 14, 2531–2560. doi: 10.1162/089976602760407955

Markowitz, D. A., Wong, Y. T., Gray, C. M., and Pesaran, B. (2011). Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J. Neurosci.* 31, 18412–18422. doi: 10.1523/JNEUROSCI.4165-11.2011

Martinez, D., Clément, M., Messaoudi, B., Gervasoni, D., Litaudon, P., and Buonviso, N. (2018). Adaptive quantization of local field potentials for wireless implants in freely moving animals: an open-source neural recording device. *J. Neural Eng.* 15:025001. doi: 10.1088/1741-2552/aaa041

Menon, A., Olascoaga, L. I. G., Balanaga, V., Natarajan, A., Ruffing, J., Ardalan, R., et al. (2023). "Shared control of assistive robots through user-intent prediction and hyperdimensional recall of reactive behavior," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, (New York City, USA: IEEE), 12638–12644.

Niculescu-Mizil, A., and Caruana, R. (2005). "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning - ICML '05*, (New York City, USA: ACM Press), 625–632.

O'Doherty, J. E., Lebedev, M. A., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., et al. (2011). Active tactile exploration using a brain–machine–brain interface. *Nature* 479, 228–231. doi: 10.1038/nature10489

Otazu, G. H., Tai, L.-H., Yang, Y., and Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* 12, 646–654. doi: 10.1038/nn.2306

Pantiskas, L., Verstoep, K., Hoogendoorn, M., and Bal, H. (2022). "Taking ROCKET on an efficiency mission: multivariate time series classification with LightWaveS," in *2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, (New York City, USA: IEEE), 149–152.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Petschenig, H., Bisio, M., Maschietto, M., Leparulo, A., Legenstein, R., and Vassanelli, S. (2022). Classification of whisker deflections from evoked responses in the somatosensory barrel cortex with spiking neural networks. *Front. Neurosci.* 16:838054. doi: 10.3389/fnins.2022.838054

Pinto, D. J., Brumberg, J. C., and Simons, D. J. (2000). Circuit dynamics and coding strategies in rodent somatosensory cortex. *J. Neurophysiol.* 83, 1158–1166. doi: 10.1152/jn.2000.83.3.1158

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classifiers* 10, 61–74.

Rapeaux, A. B., and Constandinou, T. G. (2021). Implantable brain machine interfaces: first-in-human studies, technology challenges and trends. *Curr. Opin. Biotechnol.* 72, 102–111. doi: 10.1016/j.copbio.2021.10.001

Reyes-Puerta, V., Kim, S., Sun, J.-J., Imbrosci, B., Kilb, W., and Luhmann, H. J. (2015). High stimulus-related information in barrel cortex inhibitory interneurons. *PLoS Comput. Biol.* 11:e1004121. doi: 10.1371/journal.pcbi.1004121

Reyes-Puerta, V., Sun, J.-J., Kim, S., Kilb, W., and Luhmann, H. J. (2015). Laminar and columnar structure of sensory-evoked multineuronal spike sequences in adult rat barrel cortex in vivo. *Cereb. Cortex* 25, 2001–2021. doi: 10.1093/cercor/bhu007

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain–computer interface spellers: a review. *Brain Sci.* 8:57. doi: 10.3390/brainsci8040057

Rice, L., Wong, E., and Kolter, Z. (2020). "Overfitting in adversarially robust deep learning," in International Conference on Machine Learning, PMLR, (Granada, Spain), 8093–8104.

Ritt, J. T., Andermann, M. L., and Moore, C. I. (2008). Embodied information processing: vibrissa mechanics and texture features shape micromotions in actively sensing rats. *Neuron* 57, 599–613. doi: 10.1016/j.neuron.2007.12.024

Rossant, C., Kadir, S. N., Goodman, D. F., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., et al. (2016). Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* 19, 634–641. doi: 10.1038/nn.4268

Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10:e0118432. doi: 10.1371/journal.pone.0118432

Schober, P., Boer, C., and Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesth. Analg.* 126, 1763–1768. doi: 10.1213/ANE.0000000000002864

Sederberg, A. J., Pala, A., Zheng, H. J., He, B. J., and Stanley, G. B. (2019). State-aware detection of sensory stimuli in the cortex of the awake mouse. *PLoS Comput. Biol.* 15:e1006716. doi: 10.1371/journal.pcbi.1006716

Selesnick, I. W., and Burrus, C. S. (2002). Generalized digital Butterworth filter design. *IEEE Trans. Signal Process.* 46, 1688–1694. doi: 10.1109/78.678493

Semprini, M., Laffranchi, M., Sanguineti, V., Avanzino, L., de Icco, R., de Michieli, L., et al. (2018). Technological approaches for neurorehabilitation: from robotic devices to brain stimulation and beyond. *Front. Neurol.* 9:212. doi: 10.3389/fneur.2018.00212

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164

Srimaharaj, W., and Chaisricharoen, R. (2021). A novel processing model for P300 brainwaves detection. *J. Web Eng.* 20, 2545–2570. doi: 10.13052/jwe1540-9589.20815

Staiger, J. F., and Petersen, C. C. (2021). Neuronal circuits in barrel cortex for whisker sensory perception. *Physiol. Rev.* 101, 353–415. doi: 10.1152/physrev.00019.2019

Steriade, M., Nunez, A., and Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *J. Neurosci.* 13, 3252–3265. doi: 10.1523/JNEUROSCI.13-08-03252.1993

Stüttgen, M. C., Rüter, J., and Schwarz, C. (2006). Two psychophysical channels of whisker deflection in rats align with two neuronal classes of primary afferents. *J. Neurosci.* 26, 7933–7941. doi: 10.1523/JNEUROSCI.1864-06.2006

Stüttgen, M. C., and Schwarz, C. (2008). Psychophysical and neurometric detection performance under stimulus uncertainty. *Nat. Neurosci.* 11, 1091–1099. doi: 10.1038/nn.2162

Stüttgen, M. C., and Schwarz, C. (2010). Integration of vibrotactile signals for whisker-related perception in rats is governed by short time constants: comparison of neurometric and psychometric detection performance. *J. Neurosci.* 30, 2060–2069. doi: 10.1523/JNEUROSCI.3943-09.2010

Stüttgen, M. C., and Schwarz, C. (2018). Barrel cortex: what is it good for? *Neuroscience* 368, 3–16. doi: 10.1016/j.neuroscience.2017.05.009

Sunny, T., Aparna, T., Neethu, P., Sunny, T. D., Venkateswaran, J., Vishnupriya, V., et al. (2016). Robotic arm with brain–computer interfacing. *Procedia Technol.* 24, 1089–1096. doi: 10.1016/j.protcy.2016.05.241

Temereanca, S., and Simons, D. J. (2003). Local field potentials and the encoding of whisker deflections by population firing synchrony in thalamic barreloids. *J. Neurophysiol.* 89, 2137–2145. doi: 10.1152/jn.00582.2002

Tiwari, A., and Chaturvedi, A. (2019). "A multiclass EEG signal classification model using spatial feature extraction and XGBoost algorithm," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (New York City, USA: IEEE), 4169–4175.

Vandevelde, J. R., Yang, J.-W., Albrecht, S., Lam, H., Kaufmann, P., Luhmann, H. J., et al. (2023). Layer-and cell-type-specific differences in neural activity in mouse barrel cortex during a whisker detection task. *Cereb. Cortex* 33, 1361–1382. doi: 10.1093/cercor/bhac141

Vecchi, E., Pospíšil, L., Albrecht, S., O'Kane, T. J., and Horenko, I. (2022). eSPA+: scalable entropy-optimal machine learning classification for small data problems. *Neural Comput.* 34, 1220–1255. doi: 10.1162/neco_a_01490

Wang, X., Magno, M., Cavigelli, L., Mahmud, M., Cecchetto, C., Vassanelli, S., et al. (2018a). "Rat cortical layers classification extracting evoked local field potential images with implanted multi-electrode sensor," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, (New York City, USA: IEEE), 1–6.

Wang, X., Magno, M., Cavigelli, L., Mahmud, M., Cecchetto, C., Vassanelli, S., et al. (2018b). "Embedded classification of local field potentials recorded from rat barrel cortex with implanted multi-electrode array," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, (New York City, USA: IEEE), 1–4.

Werner, T., Vianello, E., Bichler, O., Garbin, D., Cattaert, D., Yvert, B., et al. (2016). Spiking neural networks based on OxRAM synapses for real-time unsupervised spike sorting. *Front. Neurosci.* 10:474. doi: 10.3389/fnins.2016.00474

Won, S. M., Cai, L., Gutruf, P., and Rogers, J. A. (2023). Wireless and battery-free technologies for neuroengineering. *Nat. Biomed. Eng.* 7, 405–423. doi: 10.1038/s41551-021-00683-3

Yamashita, T., Vavladeli, A., Pala, A., Galan, K., Crochet, S., Petersen, S. S. A., et al. (2018). Diverse long-range axonal projections of excitatory layer 2/3 neurons in mouse barrel cortex. *Front. Neuroanat.* 12:33. doi: 10.3389/fnana.2018.00033

Yang, H., Kwon, S. E., Severson, K. S., and O'connor, D. H. (2016). Origins of choice-related activity in mouse somatosensory cortex. *Nat. Neurosci.* 19, 127–134. doi: 10.1038/nn.4183

Yang, J.-W., Prouvot, P.-H., Reyes-Puerta, V., Stüttgen, M. C., Stroh, A., and Luhmann, H. J. (2017). Optogenetic modulation of a minor fraction of parvalbumin-positive interneurons specifically affects spatiotemporal dynamics of spontaneous and sensory-evoked activity in mouse somatosensory cortex in vivo. *Cereb. Cortex* 27, 5784–5803. doi: 10.1093/cercor/bhx261

Yang, J.-W., Prouvot, P.-H., Stroh, A., and Luhmann, H. J. (2018). "Combining optogenetics with mea, depth-resolved LFPs and assessing the scope of optogenetic network modulation," in *Optogenetics: A Roadmap*, ed. A. Stroh (New York, NY: Springer), 133–152.

Yeganeh, F., Knauer, B., Guimarães Backhaus, R., Yang, J. W., Stroh, A., Luhmann, H. J., et al. (2022). Effects of optogenetic inhibition of a small fraction of parvalbumin-positive interneurons on the representation of sensory stimuli in mouse barrel cortex. *Sci. Rep.* 12:19419. doi: 10.1038/s41598-022-24156-y

Zuo, Y., and Diamond, M. E. (2019). Texture identification by bounded integration of sensory cortical signals. *Curr. Biol.* 29, 1425–1435.e5. doi: 10.1016/j.cub.2019.03.017