

HIPPIE: A Multimodal Deep Learning Model for Electrophysiological Classification of Neurons

Jesus Gonzalez-Ferrer^{a,b,c,g}, Julian Lehrer^{a,b,g}, Hunter E. Schweiger^{a,b,d,h},
Jinghui Geng^{a,b,e,h}, Sebastian Hernandez^{a,b,e}, Francisco Reyes^{a,b,f}, Jess L.
Sevetson^{a,d}, Sofie R. Salama^{a,d}, Mircea Teodorescu^{a,b,c,e}, David Haussler^{a,b,c},
Mohammed A. Mostajo-Radji^{a,b,i}

^aGenomics Institute, University of California Santa Cruz, Santa Cruz, CA, 95060, USA

^bLive Cell Biotechnology Discovery Lab, University of California Santa Cruz, Santa Cruz, CA, 95060, USA

^cDepartment of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, 95060, USA

^dDepartment of Molecular, Cellular and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA, 95060, USA

^eDepartment of Electrical and Computer Engineering, University of California Santa Cruz, Santa Cruz, CA, 95060, USA

^fBiotechnology Program, Berkeley City College, Berkeley, CA, 94704, USA

^gThese authors contributed equally to this work.

^hThese authors contributed equally to this work.

ⁱCorrespondence to mmostajo@ucsc.edu (M.A.M.-R.)

Abstract

Extracellular electrophysiological recordings present unique computational challenges for neuronal classification due to noise, technical variability, and batch effects across experimental systems. We introduce HIPPIE (High-dimensional Interpretation of Physiological Patterns In Extracellular recordings), a deep learning framework that combines self-supervised pretraining on unlabeled datasets with supervised fine-tuning to classify neurons from extracellular recordings. Using conditional convolutional joint autoencoders, HIPPIE learns robust, technology-adjusted representations of waveforms and spiking dynamics. This model can be applied to electrophysiological classification and clustering across diverse biological cultures and technologies. We validated HIPPIE on both *in vivo* mouse recordings and *in vitro* brain slices, where it demonstrated superior performance over other unsupervised methods in cell-type discrimination and aligned closely with anatomically defined classes. Its latent space organizes neurons along electrophysiological

ical gradients, while enabling batch and individual corrected alignment of recordings across experiments. HIPPIE establishes a general framework for systematically decoding neuronal diversity in native and engineered systems.

Keywords: Multielectrode arrays, Electrophysiology, Neuronal classification, Neuroscience data, Neurodevelopment

1. Introduction

A thorough understanding of neuronal cell types, their locations within the brain, and how they connect is essential for decoding the functional properties of neuronal networks and their role in producing behaviors [1, 2]. Recent advances in high-throughput sequencing technologies have enabled researchers to gather extensive gene expression data from single cells, providing unprecedented insights into cellular diversity and function at a molecular level [3, 4]. While these transcriptional profiles provide valuable insights, they fail to capture essential functional characteristics, such as cellular morphology [5] and electrophysiology which are crucial for understanding the roles of cell types within their specific niches and provide physiological insight into cellular states [6].

To date, most cell-type classifications based on electrophysiology rely on low-throughput methods, such as whole-cell patch clamping [7]. However, several high-throughput technologies for electrophysiological characterization have emerged, including planar high-density microelectrode arrays (HD-MEAs) [8, 9, 10], flexible three-dimensional microelectrode array baskets [11], and high-density probes like Neuropixels [12, 13]. These technologies enable extracellular recordings at scale, allowing simultaneous monitoring of multiple units across different areas, thereby facilitating the study of network-level dynamics. High throughput electrophysiological recordings have traditionally distinguished broad cell-type categories using scalar measures, such as action potential width, to segregate putative excitatory "broad-spiking" neurons from "narrow-spiking" inhibitory neurons [14]. Furthermore, integrating these high-throughput approaches with genetic tools like optotagging [15, 16] provides a powerful way of identifying neuronal subtypes within these networks.

Efforts to classify extracellular electrophysiological data have diverged into supervised and unsupervised paradigms. Supervised methods like LOL-CAT [17], which maps spiking patterns to transcriptomic cell types, are con-

strained by sparse labeled datasets. Unsupervised approaches, such as non-linear dimensionality reduction for clustering extracellular waveforms [18] or spiking patterns [19], leverage unlabeled data but operate on single modalities, neglecting complementary information. Multimodal integration has emerged as a priority, exemplified by PhysMAP [20], which adapts single-cell genomics strategies [21] to electrophysiology via UMAP and weighted nearest neighbors. However, UMAP’s inability to model sequential dependencies limits its efficacy for spike-time derived data. Alternative variational autoencoder (VAE) based frameworks [22] unify modalities through separate encoders but prioritize reconstruction over discrimination [23] and require spatial metadata absent in *in vitro* systems like organoids. Self-supervised learning offers a promising middle ground by generating supervisory signals from unlabeled data.

Advances in self-supervised learning, such as scGPT [24], highlight the potential of pretext tasks for representation learning, but domain-specific adaptations are needed. For instance, NEMO [25] employs contrastive learning to align average waveform and autocorrelogram embeddings, circumventing VAEs’ reconstruction bias. However, its assumption of modality-invariant neuronal signatures may not hold for dynamically coupled features (e.g., spike timing vs. waveform). Conditional VAEs (cVAEs) could address these limitations by balancing discriminative and generative objectives [26]. Unlike standard VAEs, cVAEs condition latent representations on auxiliary variables (e.g., experimental parameters), enabling targeted feature extraction [26]. In single-cell genomics, this strategy improves cell-type-specific generation [27], while in computer vision, it achieves attribute-controlled image synthesis [28]. Applied to electrophysiology, cVAEs could disentangle shared and modality-specific neuronal traits without spatial priors [26], making them adaptable to *in vitro* models. By explicitly modeling conditional dependencies (e.g., brain region or cell line), cVAEs may overcome the reconstruction-discrimination trade-off while integrating spike waveforms, temporal dynamics, and metadata into a unified latent space.

Here we present HIPPIE (High-dimensional Interpretation for Physiological Patterns in Intercellular Extracellular recordings), a cVAE framework designed for multimodal neuron classification and clustering by integrating extracellular action potential waveforms with spike-timing derived measurements. We benchmark HIPPIE against PhysMAP, the leading unsupervised method, demonstrating superior performance in cell-type identification and brain region classification across diverse recording technologies. HIPPIE

95 achieves consistent gains in both inductive and transductive learning sce-
 96 narios, highlighting its adaptability to real-world experimental constraints.
 97 By unifying feature extraction, clustering, and classification into a single
 98 framework, HIPPIE advances the interpretability and scalability of electro-
 99 physiological analysis.

100 2. Results

101 2.1. The HIPPIE pipeline

102 We developed HIPPIE, a deep learning framework that uses parallel
 103 cVAEs built from 1dResnet [29] encoder and decoder modules (Figure 1A-B).

104 The HIPPIE pipeline takes as input spike sorted data from extracellular
 105 electrophysiology recordings. From the spike sorted data, the average wave-
 106 form and interspike interval (ISI) distribution are calculated as 1d vectors,
 107 a categorical string representing the recording technology used in the exper-
 108 iment is also used as input to the network. Waveforms are averaged across
 109 spike times for each individual unit, and interspike interval distributions are
 110 computed from the spike train list using 1-millisecond bins, spanning a total
 111 of 100 milliseconds. The average spike waveform is calculated from a sample
 112 of 100 to 500 spikes, a range selected to balance accurate waveform repre-
 113 sentation with efficient memory usage. In our pipeline [30], we employed
 114 Kilosort2 for spike sorting. Kilosort2 uses template matching to compare
 115 individual spike waveforms, resulting in relatively low waveform variability
 116 across spikes. This approach allows us to obtain reliable spike representations
 117 even for low firing rate neurons with around 100 spikes. Given an average
 118 firing rate of 2 Hz, a 10-minute recording would yield approximately 1200
 119 spikes, from which the 500-spike sample provides a representative subset.
 120 The length of the extracted waveform is 2.5 ms, which corresponds to 50
 121 data points when sampled at 20 kHz. For higher sampling rates the number
 122 of points around the spikes can be adjusted to capture the same time window
 123 and then interpolated to 50 points to feed the algorithm. This time window
 124 captures the majority of the action potential duration in most neurons. If
 125 the algorithm were to be adapted for other electrophysiologically active cells,
 126 the time window should be adjusted accordingly. The preprocessing pipeline
 127 to obtain these processed features from raw voltage recordings is described in
 128 more detail in section 4.1. We focused on average waveform and ISI distribu-
 129 tion because they have been shown to be some of the features that correlate
 130 the most with cell type and brain region [18, 31]. This framework could

also be extended with additional measurements, such as autocorrelograms, layer-specific information, or other connectivity features. The outputs from the encoding layers of different modalities are integrated using a mixture module. In this work, the mixture module is implemented as a multilayer perceptron (MLP) [32], which takes the latent embeddings from both unimodal encoders and passes the output to the unimodal decoders. This setup enables the model to learn a shared representation through backpropagation.

2.2. HIPPIE identifies anatomical locations better than other methods

To assess HIPPIE’s capability in generating meaningful brain region-specific embeddings compared to the state-of-the-art method, PhysMAP [20], we conducted benchmark tests using datasets originally employed to validate PhysMAP’s ability to cluster transcriptomic cell types. However, our analysis prioritized the anatomical information within these datasets. Additionally, we supplemented these benchmarks with a new *in vitro* brain slice dataset generated by our group:

1. Juxtacellular Mouse S1 Area Dataset [15]: 293 units recorded in vivo from adult mouse primary somatosensory cortex (S1), spanning layers 2/3, 4, and 5 using custom tetrodes (Figure 2A).

2. CellExplorer Area Dataset [12]: 430 units from in vivo mouse hippocampal regions (CA1, CA3) and visual cortex (V1) recorded with Neuropixel 1.0 probes (Figure 2B).

3. Neonatal Mouse Brain Slice Dataset: This new dataset contains 2,220 units recorded *in vitro* from postnatal day 0–3 mouse brain slices (cerebral cortex, retrosplenial cortex, striatum, diencephalon) using Maxwell Biosystems’ MaxONE platform [33, 30] (Figure 2C).

HIPPIE was pretrained using a leave-one-dataset-out strategy, incorporating the Allen OpenScope dataset [34], in-house recordings, and the above datasets. For each benchmark, the target dataset was excluded during pretraining and used only for evaluation. Following pretraining, models were fine-tuned on 80% of each target dataset (random split) and tested on the remaining 20%. Classification performance was evaluated using a K-nearest neighbors (KNN) classifier with 10-fold cross-validation across three input conditions: unimodal average waveform, unimodal ISI, and a combined multimodal input of average waveform and ISI (hereafter referred to as “joint”).

HIPPIE outperformed other methods in predicting anatomical locations across all datasets, regardless of the deployed modality (Supplemental Tables 1–3, Figure 2), achieving the highest accuracy and macro-F1 scores. While

PhysMAP [20] and PCA-based methods exhibited lower cross-validation variance compared to raw data analyses, their performance did not exceed that of unprocessed recordings. The UMAP representation of the embeddings obtained from the different methods also showcases HIPPIE’s ability to create latent spaces with well differentiated regions that show correspondence to different anatomical location (Supplemental Figure 1 A-C).

These results show that HIPPIE’s self-supervised architecture effectively extracts anatomical discriminative features from ISI distributions and average waveform, surpassing both PhysMAP and modality-agnostic dimensionality reduction approaches.

2.3. HIPPIE enables cross-sample generalization in anatomical classification

Conventional classifiers in electrophysiology often employ transductive splits, where neurons from all experimental samples are pooled before partitioning into training and testing sets [20, 25]. While this approach simplifies evaluation, it risks inflating performance metrics by allowing models to exploit sample-specific artifacts (e.g., electrode drift, session-specific noise) rather than biologically generalizable features [31]. In contrast, inductive splits enforce strict separation between training and testing samples, mimicking real-world scenarios where models must generalize to entirely unseen experimental preparations, a critical benchmark for robustness in neuroscience, where variability in recording conditions (e.g., electrode placement, animal-to-animal differences) remains a persistent challenge [31].

To rigorously assess HIPPIE’s generalizability, we analyzed the Allen Brain Institute’s Visual Coding dataset [34], the largest publicly available repository of *in vivo* Neuropixel 1.0 recordings from the mouse brain. Our analysis includes 16,564 units across 40 anatomically defined brain areas, collected from 47 adult mice. We evaluated HIPPIE under two distinct paradigms:

1. Transductive Split. Using 80-20 stratified splits (balanced across brain areas), HIPPIE achieved near-perfect classification accuracy. The ISI modality outperformed other input types with 99.6% accuracy, closely followed by the waveform (98.4%) and joint (98.4%) modalities. In contrast, baseline methods, including PCA-based dimensionality reduction and PhysMAP [20] failed to exceed 15% accuracy (Supplemental Tables 1–3), indicating their reliance on dataset-specific confounders rather than invariant biological features.

204 **2. Inductive Split.** To simulate real-world generalization, our splits
205 consisted of 37 mice for training and 10 held-out animals for testing. HIP-
206 PIE maintained robust performance, with the waveform (96.5%) and joint
207 modalities (96.1%) modalities exhibiting minimal accuracy loss compared to
208 transductive splits. The ISI modality experienced a sharper decline (99.6%
209 to 91.4%). All baseline methods remained near chance ($\leq 14.2\%$ accuracy),
210 underscoring their inability to disentangle invariant neuronal properties from
211 experimental noise (Figure 3).

212 Visualization of HIPPIE’s latent representations revealed hierarchical clus-
213 tering aligned with anatomical hierarchy even under inductive splits, confirm-
214 ing that its compressed embeddings encode biologically interpretable features
215 rather than session-specific artifacts. These results establish HIPPIE as a
216 model capable of generalizing across distinct experimental preparations while
217 maintaining high anatomical resolution.

218 2.4. HIPPIE accurately classifies neuronal subtypes using waveform features

219 To assess HIPPIE’s ability to resolve cell-type-specific electrophysiological
220 signatures, we analyzed three datasets combining *in vivo* extracellular record-
221 ings with optogenetic tagging of genetically defined neuronal subtypes. These
222 datasets encompassed distinct recording technologies and brain regions:

223 **1. Juxtacellular Mouse S1 Cell Type Dataset (also known as**
224 **Jianing dataset) [15]:** This dataset comprised 224 units recorded juxta-
225 cellularly in the mouse primary somatosensory cortex using whole-cell patch
226 pipettes. Cre-dependent ChR2 expression enabled optogenetic identification
227 of Sst-positive and Pvalb-positive interneurons, with untagged putative exci-
228 tatory neurons. Pvalb-positive interneurons and putative excitatory neurons
229 are further subdivided by cortical layer (L4, L5).

230 **2. Extracellular Mouse Auditory Cortex (A1 Dataset) [35]:**
231 Recorded using NeuroNexus silicon probes, this dataset included 284 units
232 from the auditory cortex. Pvalb-positive and Sst-positive interneurons were
233 optotagged via Cre-mediated ChR2 expression, while putative excitatory
234 neurons were identified by the absence of light response.

235 **3. CellExplorer Cell Type Dataset (Mouse Visual Cortex and**
236 **Hippocampus) [12]:** This Neuropixels 1.0-derived dataset contained 430
237 units spanning visual cortex and hippocampal regions. Vip-positive, Pvalb-
238 positive, and Sst-positive interneurons were optotagged using subtype-specific
239 Cre lines, with untagged neurons classified as putative excitatory.

For each dataset, HIPPIE was pretrained using a leave-one-dataset-out strategy, excluding the target dataset during pretraining. Following fine-tuning on 80% of the target data (stratified by cell type), performance was evaluated on the remaining 20% using KNN classification with 10-fold cross-validation.

We found that HIPPIE’s waveform modality achieved the highest accuracy in cell-type discrimination across all datasets (Juxtacellular Mouse S1 96.42% Extracellular Mouse A1 Dataset 95.04% Cellexplorer Cell Type dataset 97.71%) (Figure 4; Supplemental Tables 4–6). The ISI and joint modalities exhibited slightly lower performance, though still superior to baseline methods (Juxtacellular Mouse S1 93.30% and 90.65%, Extracellular Mouse A1 Dataset 92.20% and 95.04%, Cellexplorer Cell Type dataset 94.13% and 93.64%). PhysMAP and PCA-based approaches demonstrated reduced cross-validation variance but failed to outperform analyses using raw data. Visual inspection of the embeddings produced by HIPPIE and PhysMap also reinforce the superior ability of HIPPIE to create meaningful latent spaces for cell type classification and clustering (Supplemental Figure 2 A-C).

These results establish HIPPIE as a robust framework for identifying neuronal subtypes in high throughput electrophysiology data. Furthermore, HIPPIE’s performance across juxtacellular, silicon probes, and Neuropixels 1.0 recordings highlights its generalizability to multiple experimental paradigms.

2.5. KNN probes are better suited for HIPPIE embeddings than Multilayer Perceptron probes

Although KNN probes were selected as the classification head of our model, alternative approaches, such as MLP probes, have also been employed to assess the discriminative capacity of the latent space in self-supervised models for high-throughput electrophysiological data [25]. To evaluate the effectiveness of our chosen approach, we compared the classification performance of KNN and MLP probes following HIPPIE dimensionality reduction.

Across anatomical datasets, KNN consistently outperformed MLP for single-modality classification. In the CellExplorer Area Dataset [12], ISI embeddings classified with KNN achieved 76.98% accuracy, whereas MLP reached only 69.07%. The difference was even more pronounced for waveform embeddings, where KNN attained 72.79% accuracy compared to just 36.04% for MLP. However, for joint-modality classification, MLP outperformed KNN, achieving 90.23% accuracy versus 76.98%.

277 This trend was also evident in other anatomical datasets. For instance,
278 in the Juxtacellular Mouse S1 Area Dataset, ISI embeddings classified with
279 KNN achieved 93.08% accuracy, whereas MLP reached 91.16%. A similar
280 pattern emerged with waveform embeddings, where KNN obtained 93.97%
281 accuracy compared to MLP’s 72.31%. Yet again, for joint-modality classifi-
282 cation, MLP surpassed both, achieving 94.20% accuracy compared to KNN’s
283 85.69%.

284 Notably, this trend did not hold in the Neonatal Mouse Slice ataset, where
285 KNN outperformed MLP across all modalities. This substantial performance
286 gap suggests that HIPPIE generates a feature space where local neighborhood
287 relationships carry strong discriminative power. The superior performance
288 of KNN implies that samples within the same class cluster more tightly in
289 the HIPPIE-transformed space, making local distance-based methods more
290 effective than the global decision boundaries learned by MLPs. However,
291 for anatomical location classification, KNN’s non-weighted combinations of
292 embeddings may limit its ability to effectively integrate information across
293 multiple modalities.

294 In cell type classification tasks, where waveform features were the most
295 informative, KNN consistently outperformed MLP. In the CellExplorer Cell
296 Type Dataset [12], HIPPIE waveform classified with KNN achieved 97.7%
297 accuracy, substantially surpassing the 81.40% obtained with MLP. This pat-
298 tern held across additional datasets, including the Extracellular Mouse Audi-
299 tory Cortex Dataset [35], where KNN reached 95.04% accuracy compared to
300 85.80% with MLP, and the Juxtacellular Mouse S1 Cell Type Dataset [15],
301 where KNN achieved 96.42% versus MLP’s 71.60%.

302 Overall, these comparisons across ISI, waveform, and joint feature types
303 indicate that HIPPIE’s dimensionality reduction generally produces a feature
304 space where KNN outperforms MLP for single-modality classification, likely
305 due to the strong local neighborhood structure in the transformed space. For
306 joint-modality classification, MLP surpasses KNN in some anatomical clas-
307 sification tasks, but not all, while KNN consistently maintains an advantage
308 in cell type classification. This suggests that MLP’s ability to learn weighted
309 combinations of features provides an advantage for integrating multimodal
310 anatomical information in specific cases, whereas KNN remains more effec-
311 tive when distinguishing cell types. A full set of results for MLP probes is
312 available in Supplemental Table 7 and the boxplots showcasing accuracy
313 and F1 can be found in Supplemental Figure 3.

314 3. Discussion

315 Autoencoders and related dimensionality reduction techniques have emerged
316 as important tools for distilling high-dimensional neuronal recordings into in-
317 terpretable latent representations. By compressing electrophysiological data
318 while preserving biologically meaningful patterns, these methods enable the
319 identification of cell-type-specific features and functional dynamics that are
320 often obscured in raw recordings [22, 25]. The integration of multimodal
321 neuronal data (such as extracellular waveforms and spike timing) into unified
322 embedding spaces improves upon the limitations of single-modality analyses
323 and enables a comprehensive neuronal characterization [20, 25]. Cluster-
324 ing algorithms applied to these latent spaces further refine our capacity to
325 resolve neuronal diversity, uncovering non-linear relationships that may de-
326 lineate novel cell subtypes or functional states.

327 Reproducibility, a persistent challenge in electrophysiology [36], can be
328 tackled through the used of integration techniques and cross-sample align-
329 ment frameworks. This integration would not only enhance the transferabil-
330 ity of analytical pipelines but could also empower validation of brain models,
331 such as organoids, by enabling direct comparisons of neuronal phenotypes
332 across studies. Advances in self-supervised learning are particularly prom-
333 ising for improving reproducibility, as these methods leverage high-throughput
334 datasets (now widely accessible through repositories like DANDI) to learn
335 generalizable representations of neuronal activity [30, 37], and tools such
336 as HIPPIE and others have shown that conserved functional codes can be
337 found across different animals [31]. By prioritizing invariant features across
338 recordings, self-supervised models could overcome batch effects and techni-
339 cal variability inherent to electrophysiological experiments. The development
340 of user-friendly graphical interfaces and web applications for implementing
341 these models further enhances reproducibility by making complex analysis
342 tools accessible to researchers without extensive programming expertise [38].
343 The democratization of computational methods reduces the technical barriers
344 to adopt standardized analysis pipelines, enabling more laboratories to
345 validate and build upon existing findings using consistent methodological
346 approaches.

347 The rapid evolution of multimodal technologies is accelerating progress
348 in neuroscience. Innovations such as *in situ* electro-sequencing [39], which
349 integrates spatial transcriptomic profiling with electrophysiology, are gener-
350 ating rich datasets that link molecular identity to functional properties. Ge-

netically encoded voltage indicators can also link molecular identity, electrophysiology, morphology and anatomical location [40]. Combined with next-generation brain observatories [41], these advancements will produce unprecedented volumes of neuronal data spanning brain regions, behaviors, and developmental stages. These large-scale datasets will enable the profiling of more cell types than ever before. However, they will initially be sparse, creating ideal conditions for self-supervised models like HIPPIE to excel. Initiatives such as the BRAIN Initiative Cell Census Network (BICCN) [1, 42] provide essential ground-truth cell-type atlases to validate machine learning-derived discoveries, while open-data platforms foster collaborative model development.

An algorithm for detecting brain regions and cell types from extracellular neural recordings also represents an important tool for the neurotech field. This tool could enhance brain-computer interfaces by identifying the brain region and cell types generating the signal, providing a better interpretation of user intent [43]. Brain region classification is particularly critical for understanding fundamental physiological differences between brain areas and for accurately targeting regions that are challenging to access via probe insertion. HIPPIE could provide estimates of probe location during experiments, significantly increasing the success rate of reaching target regions. This capability could be especially valuable in primates and human subjects, where histological information is often unavailable, and insertions rely on experimental heuristics that lack standardization across laboratories [44].

4. Methods

4.1. Data processing pipeline

For the dataset produced in house (Neonatal Mouse Brain Slice Dataset), raw recordings were processed following the pipeline described in [30]. The pipeline outputs a compressed ZIP file containing a spike data object structured as a NumPy array and a Python dictionary. This object includes the following components: a spike train list, a neuron data dictionary, the recording's sample rate, and the electrode configuration. The neuron data dictionary provides spatial details, such as channel coordinates, neighboring channels, and spike features like waveforms and amplitudes. The indices of the spike train list correspond directly to those in the neuron data dictionary. Waveforms were averaged across spike times for each unit, and

interspike interval distributions were computed from the spike train list using 1-millisecond bins, spanning a total of 100 milliseconds. The average spike waveform is calculated using a maximum number of 500 raw spikes. The length of the waveform is 2.5 ms, which is 50 data points with a 20 kHz sampling rate. Since HD-MEAs have a high number of electrodes and a small electrode pitch, one neuron’s signal can be recorded simultaneously by multiple electrodes. The neuron’s location on the HD-MEA is considered as the electrode with the maximum signal amplitude.

4.2. HIPPIE’s loss function

The cVAE loss function combines two essential components that work in concert to create a meaningful latent space. The first component is the reconstruction loss, which ensures the decoder can faithfully reconstruct the input data x from the latent representation z . The second component is the Kullback-Leibler (KL) divergence loss, which acts as a regularizer by encouraging the latent space representation to follow a prior distribution, typically a standard normal distribution. This combination is formalized as:

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] + \beta D_{\text{KL}}(q_{\phi}(z|x)||p(z)) \quad (1)$$

where $q_{\phi}(z|x)$ represents the encoder network (also called the recognition model), $p_{\theta}(x|z)$ represents the decoder network (the generative model), and $p(z)$ is the prior distribution. The hyperparameter β controls the trade-off between reconstruction quality and latent space regularity, with $\beta = 1$ recovering the standard VAE formulation.

When using a Gaussian encoder with diagonal covariance matrix, which is the most common choice, the KL divergence term has a convenient analytical form:

$$D_{\text{KL}}(q_{\phi}(z|x)||p(z)) = \frac{1}{2} \sum_{j=1}^J (\sigma_j^2 + \mu_j^2 - \log(\sigma_j^2) - 1) \quad (2)$$

where μ_j and σ_j are the predicted mean and standard deviation for the j th dimension of the latent space.

This formulation ensures that the encoder learns to map inputs to a well-behaved latent space while maintaining the ability to reconstruct the original data accurately. Higher values of β place more emphasis on the regularization term, potentially leading to a more organized latent space at the cost of reconstruction fidelity, while lower values prioritize reconstruction quality.

418 4.3. Classification algorithms

419 The K-Nearest Neighbors (KNN) classification algorithm provides a ro-
 420 bust framework for predictive modeling. K-nearest neighbors classification
 421 is done by majority vote of the labels of training samples closest to the test
 422 sample in feature space. We use euclidean distance for finding the nearest
 423 neighbors. Our approach can easily integrate a confidence scoring mechanism
 424 to evaluate prediction certainty, derived as the proportion of neighboring data
 425 points that share the majority class label, offering a probabilistic measure
 426 of classification certainty. By analyzing the label distribution among the k
 427 neighbors, the confidence score will range from 0 to 1, where higher scores
 428 reflect stronger consensus and better reliability.

429 MLP probing is used to evaluate the quality of VAE embeddings by train-
 430 ing a supervised Multi-Layer Perceptron (MLP) classifier on the latent repre-
 431 sentations. The performance of this classifier indicates how well the learned
 432 embeddings capture task-relevant information. Compared to KNN prob-
 433 ing, which assesses the structure of the latent space by measuring nearest-
 434 neighbor classification accuracy, MLP probing provides a stronger test of
 435 linear separability and feature usefulness by optimizing a parametric model.
 436 While KNN reflects the local organization of embeddings, MLP probing eval-
 437 uates their suitability for structured decision boundaries, making it a more
 438 flexible but potentially less interpretable evaluation method.

439 4.4. Evaluation Metrics

440 We employed KNN and MLP classifiers to assess the performance of di-
 441 mensionality reduction techniques across neuronal cell type classification.

442 4.4.1. Accuracy

443 Accuracy was calculated to evaluate model performance, defined as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Samples}} \quad (3)$$

444 Where the total number of samples is the sum of True Positives, True Neg-
 445 atives, False Positives, and False Negatives.

446 4.4.2. Macro F1 Score

447 The macro-averaged F1 score provides a comprehensive evaluation across
 448 all cell types:

$$\text{Macro F1} = \frac{1}{n} \sum_{i=1}^n F1_i \quad (4)$$

Where the F1 score for each class is:

$$F1_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (5)$$

With Precision and Recall defined as:

$$\text{Precision}_i = \frac{\text{True Positives}_i}{\text{True Positives}_i + \text{False Positives}_i} \quad (6)$$

$$\text{Recall}_i = \frac{\text{True Positives}_i}{\text{True Positives}_i + \text{False Negatives}_i} \quad (7)$$

Performance metrics were estimated using stratified 10-fold cross-validation.

4.5. Code Library Details

The HIPPIE pipeline was designed with an easy to use application programming interface (API) to support a streamlined analysis with minimal code. To achieve this goal, the pipeline was constructed primarily using PyTorch Lightning, a high-level library that aims to improve reproducibility, modularity, and simplicity in PyTorch deep learning code. We used Weights and Biases to visualize training metrics, including accuracy, F1 score, and loss, to facilitate the assessment of model performance.

All the code and instructions to use HIPPIE are available in the Braingeneers GitHub repository: <https://github.com/braingeneers/Hippie>

4.5.1. Training details

For all experiments, models were trained on two phases: the pre-training phase and fine-tuning phase. For the pre-training phase, we train the model on every electrophysiology dataset, excluding whatever dataset we are measuring performance on. We perform a hyperparameter sweep over learning rate and the β parameter which determines the weight between the Kullback-Liebler loss of the latent space against a standard normal distribution and the $L2$ reconstruction error of the decoder. This is because β is a dataset-dependent parameter. During pre-training, we incorporate the recording technology covariate but mask the class label embeddings to zero, even if

labels may be available for some of the pre-training datasets. For fine-tuning and evaluation, we perform a K-fold cross-validation using 10 folds. For each fold, we train the model on the training set (80% of the data) and evaluate it on the remaining 20%. During fine-tuning, we use the class labels available to us and sample such that each mini-batch of training data contains an equal proportion of each label. This is to reduce the tendency of the model to predict only majority class labels while ignoring rare minority classes. We fine-tune with a learning rate equal to $1/10 \times LR$, where LR is the pre-training learning rate. We fine tune for up to 50 epochs with a batch size of 128. For K-nearest neighbors evaluation, we fit the K-nearest neighbors algorithm to the training embeddings and then perform evaluation using the test embeddings compared with the test labels.

4.5.2. Datasets

Juxtacellular Mouse S1 Dataset (Also known as the Jianing dataset): The juxtacellular dataset collected by [45] was downloaded from the PhysMap’s github repository. Recordings were conducted in the primary somatosensory (barrel) cortex of genetically modified mice. Two transgenic lines were used: Pvalb-Cre::Ai32 to identify Pvalb-positive interneurons and Sst-Cre::Ai32 to identify Sst-positive interneurons.

These experiments focused on in vivo juxtacellular electrophysiological recordings using glass micropipettes. Following recording, cells were filled with biocytin/neurobiotin, enabling morphological verification of cell types and alignment of recording depths with cortical layers. Spikes from ChR2-expressing neurons were identified based on laser-evoked responses to brief light pulses. To account for drift, spike waveforms were filtered. Depending on the experiment, we used either the cell type or the area label.

Extracellular Mouse Auditory Cortex Dataset (A1 Dataset): The extracellular data analyzed from [35] was obtained from the PhysMAP GitHub repository. Pvalb-positive and Sst-positive interneurons were identified via opto-tagging in Pvalb-Cre::Ai32 or Sst-Cre::Ai32 mice. According to the dataset description, the identification of ChR2-expressing cells was performed conservatively, with positively identified cells defined as those exhibiting a significant increase in firing rate ($p < 0.001$) within the first 10 ms of stimulation onset.

Recordings were acquired from the primary auditory cortex using silicon probes, with optical stimulation delivered via a fiber positioned at the top of the probe recording sites. Spikes were sorted offline using Klustakwik, and

only units with an ISI violation rate of less than 2% (spike intervals < 2 ms) per cluster were included in further analyses. The same parameters as those published in PhysMAP [20] were applied. Mean waveform for each unit were calculated using the channel with the highest spike amplitude.

CellExplorer Dataset (Mouse Visual Cortex and Hippocampus):

This dataset was downloaded from the PhysMAP github repository and includes recordings from the mouse primary visual cortex (V1), higher visual areas (HVAs), and hippocampus (CA1). Depending on the experiment, either anatomical region labels or cell type classifications were used.

Cell types were identified through opto-tagging in transgenic mouse lines. In the visual cortex, these included Pvalb-positive (Pvalb-IRES-Cre::Ai32), Sst-positive (Sst-IRES-Cre::Ai32), and Vip-positive (Vip-IRES-Cre::Ai32) interneurons. In the hippocampus, excitatory neurons and Pvalb-positive cells were identified, with V1 Pvalb-positive interneurons derived from Pvalb-IRES-Cre::Ai32 mice. Recordings were performed using Neuropixel 1.0 probes with optogenetic stimulation. [46, 47, 48]

Allen Institute’s Visual Coding Dataset: The Allen Institute’s Visual Coding dataset was accessed using Neurodata Without Borders (NWB) files retrieved via AllenSDK. Units were selected for further analysis based on the following default quality criteria: ISI violations < 0.5 , amplitude cutoff < 0.1 , and presence ratio > 0.9 . To ensure robust statistical analysis, brain regions with fewer than 50 units were excluded. The spontaneous stimulus dataset also includes additional units from the “Functional Connectivity” set, complementing those from the “Brain Observatory 1.1” set.

4.5.3. Neonatal Mouse Brain Slice Dataset

This dataset was generated specifically for this manuscript.

Postnatal day 0–3 mouse brains were extracted by a different group following ice-cold anesthetization prior to sacrifice. Immediately upon receipt, the tissue was sectioned into 300 μm slices using a Vibratome Leica VT1200S (amplitude: 1, advancement speed: 0.1 mm/s) in ice-cold sucrose solution (in mm): 72 sucrose, 83 NaCl, 2.5 KCl, 1 NaPO₄, 3.3 MgSO₄, 26.2 NaHCO₃, 22 dextrose, 0.5 CaCl₂. The slices were then transferred onto MaxWell Biosystems MaxOne HD-MEA chips.

Before experiments, the chips were pre-coated overnight at 37°C with 0.1 % Poly-L-ornithine (Millipore Sigma # P4957), followed by three washes with sterile water. Subsequently, they were coated overnight at 37°C with a mixture of 5 $\mu\text{g}/\text{mL}$ Laminin (Millipore Sigma # L2020) and 1 $\mu\text{g}/\text{mL}$

546 Fibronectin (Millipore Sigma DLW354008) in PBS.

The slices were maintained in BrainPhys Neuronal Medium (Stem Cell Technologies # 05790) supplemented with 10% (v/v) Hyclone characterized fetal bovine serum (Cytova # SH30071.03), 1X N-2 Supplement (Thermo Fisher Scientific # 17502048), 1X Chemically Defined Lipid Concentrate (Thermo Fisher Scientific # 11905031), 1X B-27 Supplement (Thermo Fisher Scientific # 17504044), 100 g/mL Primocin (Invivogen # ant-pm-05), and 1% (v/v) Matrigel Growth Factor Reduced (GFR) Basement Membrane Matrix (Corning # 354230). The slices were incubated at 37°C with 5% CO₂.

547 For data processing, the data underwent the standard pipeline developed
548 in the lab and published in the following manuscript [30]. We extracted the
549 average spike waveform for each individual unit. Starting from raw record-
550 ings, these were spike sorted using Kilosort2, followed by auto-curation to get
551 high-quality single-unit activity. Waveforms were then averaged across up to
552 500 spikes within a 5-ms window, using data from the channel capturing
553 the neuron’s largest amplitude. All waveforms were centered at their peak,
554 truncated to 1 ms before and 1.5 ms after the peak. For the ISI distribution,
555 first, the timestamps of detected spikes were extracted for each unit. Then
556 ISIs were computed as the differences between consecutive spike times and
557 then they were binned into 100 equally spaced bins with 100 ms as the max
558 threshold value.

559 4.6. *Methods benchmarked against*

- 560 1. **Raw data:** Raw data does not stand for the raw voltage recording
561 obtained from MEAs but the processed mean waveform and ISI distri-
562 bution for each unit as two 1-dimensional vectors. For the raw joint
563 analysis we concatenated the two vectors into a larger 1-D array. We
564 fed these vectors to the previously described KNN classification probe.
- 565 2. **PCA:** We applied PCA for dimensionality reduction on the normalized
566 1-dimensional vectors, retaining the first 10 principal components for
567 the waveform and ISI modalities. We fed the reduced representation
568 to the previously described KNN classification probe.
- 569 3. **PhysMAP:** PhysMAP is a method for identifying different cell types
570 by integrating multiple physiological data modalities, such as neuronal
571 waveform shape, ISI, peri-stimulus time histogram (PSTH), and de-
572 rived electrophysiological metrics. It achieves this multi-modal integra-
573 tion by combining non-linear dimensionality reduction with a weighted-
574 nearest neighbor (WNN) graph construction, inspired by Seurat v4 [21].

575 In this benchmark we only used it with the average waveform and ISI
576 modalities. The code we used to run PhysMAP can be found in the fol-
577 lowing github repository [https://github.com/EricKenjiLee/Phys](https://github.com/EricKenjiLee/PhysMAP_Manuscript)
578 [MAP_Manuscript](https://github.com/EricKenjiLee/PhysMAP_Manuscript). We fed this reduced representation to the previously
579 described KNN classification probe.

580 *4.6.1. Ethics statement*

581 Experiments were performed according to protocols approved by the In-
582stitutional Animal Care and Use Committee at University of California Santa
583Cruz. Our group had access to tissue already sectioned provided by an inde-
584pendent lab.

585 **5. Declarations**

586 *5.1. Author Contribution Statement*

587 J.G.-F., J.L., and M.A.M.R. conceived the project. S.R.S., M.T., D.H.,
588 and M.A.M.-R. secured funding and supervised the work. J.G.-F., J.L.,
589 H.E.S., S.H., J.G., F.R., and J.L.S. performed the experiments. J.G.-F. J.L.,
590 and M.A.M.-R. wrote the manuscript with contributions from all authors.

591 *5.2. Lead contact*

592 Further information and requests for resources and code should be di-
593rected and will be fulfilled by the lead contact, Mohammed A. Mostajo-Radji
594 (mmostajo@ucsc.edu).

595 *5.3. Material availability*

596 No new materials were produced for this study.

597 *5.4. Data availability*

598 The majority of data used for this study was downloaded from existing
599 datasets. The Neonatal Mouse Brain Slice dataset can be obtained upon
600 request to the corresponding author.

601 *5.5. Code availability*

602 All code for HIPPIE has been deposited in <https://github.com/brain>
603 [geneers/Hippie](https://github.com/braingeneers/Hippie)

604 5.6. Declaration of interests Statement

605 J.G.-F., J.L., and M.A.M.-R. are listed as inventors on provisional patent
606 applications associated with the work presented in this manuscript. The
607 authors declare no conflicts of interest.

608 5.7. Acknowledgments

609 We would like to thank Aidan Schneider for his valuable feedback on this
610 manuscript. In addition, we would like to thank Tommy Finn for providing
611 mouse tissue for this project. This work was supported by Schmidt Futures
612 (SF857) to S.R.S., M.T. and D.H.; National Human Genome Research In-
613 stitute (1RM1HG011543) to S.R.S., M.T. and D.H.; National Science Foun-
614 dation (NSF) (NSF2134955) to S.R.S., M.T. and D.H., (NSF2034037) to
615 M.T.; the National Institute of Mental Health (1U24MH132628) to D.H.
616 and M.A.M.-R.; the California Institute for Regenerative Medicine (CIRM)
617 (DISC4-16285) to S.R.S., M.T. and M.A.M.-R., (DISC4-16337) to M.A.M.-
618 R.; the University of California Office of the President (M25PR9045) to
619 S.R.S., M.T., and M.A.M.-R. H.E.S. was partially supported through the
620 Graduate Research Fellowship Program (GRFP) of the NSF. S.H. was par-
621 tially supported by the UC Doctoral Diversity Initiative (DDI-UCSC-IBSC)
622 program. F.R. was supported by the CIRM Bridges to Stem Cell Research
623 program awarded to Berkeley City College. We are thankful to the Pacific
624 Research Platform, supported by the National Science Foundation under
625 Award Numbers CNS-1730158, ACI-1540112, ACI-1541349, OAC-1826967,
626 the University of California Office of the President, and the University of
627 California San Diego’s California Institute for Telecommunications and In-
628 formation Technology/Qualcomm Institute.

629 References

- 630 [1] Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P.,
631 Polioudakis, D., Regev, A., Sestan, N., Wickersham, I. R., and Zeng,
632 H. The BRAIN Initiative Cell Census Consortium: Lessons Learned
633 Toward Generating a Comprehensive Brain Cell Atlas. *Neuron* **96**(3),
634 542–557 (2017).
- 635 [2] Sabesan, S., Fragner, A., Bench, C., Drakopoulos, F., and Lesica, N. A.
636 Large-Scale Electrophysiology and Deep Learning Reveal Distorted Neu-
637 ral Signal Dynamics After Hearing Loss. *Elife* **12**, e85108 (2023).

- 638 [3] Gonzalez-Ferrer, J., Lehrer, J., O’Farrell, A., Paten, B., Teodorescu, M.,
639 Haussler, D., Jonsson, V. D., and Mostajo-Radji, M. A. SIMS: A Deep-
640 Learning Label Transfer Tool for Single-Cell RNA Sequencing Analysis.
641 *Cell Genomics* **4**(6), 100581 (2024).
- 642 [4] Lee, A. T., Chang, E. F., Paredes, M. F., and Nowakowski, T. J. Large-
643 Scale Neurophysiology and Single-Cell Profiling in Human Neuroscience.
644 *Nature* **630**(8017), 587–595 (2024).
- 645 [5] Michiels, M. Electrophysiology Prediction of Single Neurons Based on
646 their Morphology. *bioRxiv*, <https://doi.org/10.1101/2020.02.04.933697>
647 (2020).
- 648 [6] Camunas-Soler, J. Integrating Single-Cell Transcriptomics with Cellular
649 Phenotypes: Cell Morphology, Ca²⁺ Imaging and Electrophysiology.
650 *Biophysical Reviews* **16**(1), 89–107 (2024).
- 651 [7] Zurek, N. A., Thiyagarajan, S., Ehsanian, R., Goins, A. E., Goyal,
652 S., Shilling, M., Lambert, C. G., Westlund, K. N., and Alles, S. R.
653 Machine Learning Elucidates Electrophysiological Properties Predictive
654 of Multi- and Single-Firing Human and Mouse Dorsal Root Ganglia
655 Neurons. *eNeuro* **11**, ENEURO.0248–24.2024 (2024).
- 656 [8] Spira, M. E., Huang, S.-H., Shmoel, N., and Erez, H. Multisite intracel-
657 lular recordings by MEA. *In Vitro Neuronal Networks: From Culturing*
658 *Methods to Neuro-Technological Applications*, 125–153 (2019).
- 659 [9] Elliott, M. A., Schweiger, H. E., Robbins, A., Vera-Choqueccota, S.,
660 Ehrlich, D., Hernandez, S., Voitiuk, K., Geng, J., Sevetson, J. L., Core,
661 C., et al. Internet-connected cortical organoids for project-based stem
662 cell and neuroscience education. *eneuro* **10**(12), ENEURO.0308–23.2023
663 (2023).
- 664 [10] Robbins, A. S., Schweiger, H. E., Hernandez, S., Spaeth, A., Voitiuk, K.,
665 Parks, D., Van der Molen, T., Geng, J., Sharf, T., Mostajo-Radji, M. A.,
666 Haussler, D., and Teodorescu, M. Goal-Directed Learning in Cortical
667 Organoids. *bioRxiv*, <https://doi.org/10.1101/2024.12.07.627350> (2024).
- 668 [11] Park, Y., Franz, C. K., Ryu, H., Luan, H., Cotton, K. Y., Kim, J. U.,
669 Chung, T. S., Zhao, S., Vazquez-Guardado, A., Yang, D. S., et al. Three-

- 670 Dimensional, Multifunctional Neural Interfaces for Cortical Spheroids
671 and Engineered Assembloids. *Science Advances* **7**(12), eabf9153 (2021).
- 672 [12] Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M.,
673 Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. Neu-
674 ropixels 2.0: A Miniaturized High-Density Probe for Stable, Long-Term
675 Brain Recordings. *Science* **372**(6539), eabf4588 (2021).
- 676 [13] Chung, J. E., Sellers, K. K., Leonard, M. K., Gwilliams, L., Xu, D.,
677 Dougherty, M. E., Kharazia, V., Metzger, S. L., Welkenhuysen, M.,
678 Dutta, B., et al. High-density single-unit human cortical recordings
679 using the Neuropixels probe. *Neuron* **110**(15), 2409–2421 (2022).
- 680 [14] Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. Differen-
681 tial Attention-Dependent Response Modulation Across Cell Classes in
682 Macaque Visual Area V4. *Neuron* **55**(1), 131–141 (2007).
- 683 [15] Ding, L., Balsamo, G., Chen, H., Blanco-Hernandez, E., Zouridis, I. S.,
684 Naumann, R., Preston-Ferrer, P., and Burgalossi, A. Juxtacellular Opto-
685 Tagging of Hippocampal CA1 Neurons in Freely Moving Mice. *Elife* **11**,
686 e71720 (2022).
- 687 [16] Lakunina, A., Socha, K. Z., Ladd, A., Bowen, A. J., Chen, S., Colonell,
688 J., Doshi, A., Karsh, B., Krumin, M., Kulik, P., et al. Neuropixels Opto:
689 Combining high-resolution electrophysiology and optogenetics. *bioRxiv*
690 , 2025–02 (2025).
- 691 [17] Schneider, A., Azabou, M., McDougall-Vigier, L., Parks, D. F., Ensley,
692 S., Bhaskaran-Nair, K., Nowakowski, T., Dyer, E. L., and Hengen, K. B.
693 Transcriptomic Cell Type Structures in Vivo Neuronal Activity Across
694 Multiple Timescales. *Cell Reports* **42**(4) (2023).
- 695 [18] Lee, E. K., Balasubramanian, H., Tsolias, A., Anakwe, S. U., Medalla,
696 M., Shenoy, K. V., and Chandrasekaran, C. Non-linear Dimensionality
697 Reduction on Extracellular Waveforms Reveals Cell Type Diversity in
698 Premotor Cortex. *Elife* **10**, e67490 (2021).
- 699 [19] Stringer, C., Zhong, L., Syeda, A., Du, F., Kesa, M., and Pachitariu,
700 M. Rastermap: A Discovery Method for Neural Population Recordings.
701 *Nature Neuroscience* , 1–12 (2024).

- [20] Lee, E. K., Gul, A., Heller, G., Lakunina, A., Jaramillo, S., Przytycki, P., and Chandrasekaran, C. PhysMAP-Interpretable in vivo Neuronal Cell Type Identification Using Multi-Modal Analysis of Electrophysiological Data. *BioRxiv*, <https://doi.org/10.1101/2024.02.28.582461> (2024).
- [21] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* **184**(13), 3573–3587 (2021).
- [22] Beau, M., Herzfeld, D. J., Naveros, F., Hemelt, M. E., D’Agostino, F., Oostland, M., Sánchez-López, A., Chung, Y. Y., Maibach, M., Kyranakis, S., et al. A Deep-Learning Strategy to Identify Cell Types Across Species from High-Density Extracellular Recordings. *Cell* **0**(0), <https://doi.org/10.1016/j.cell.2025.01.041> (2024).
- [23] Guo, X., Gao, L., Liu, X., and Yin, J. Improved Deep Embedded Clustering with Local Structure Preservation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*, volume 17, 1753–1759, (2017).
- [24] Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI. *Nature Methods*, 1–11 (2024).
- [25] Yu, H., Lyu, H., Xu, E. Y., Windolf, C., Lee, E. K., Yang, F., Shelton, A. M., Olsen, S., Minavi, S., Winter, O., et al. In Vivo Cell-Type and Brain Region Classification Via Multimodal Contrastive Learning. *bioRxiv*, <https://doi.org/10.1101/2024.11.05.622159> (2024).
- [26] Esmaili, B., Wu, H., Jain, S., Bozkurt, A., Siddharth, N., Paige, B., Brooks, D. H., Dy, J., and Meent, J.-W. Structured Disentangled Representations. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 2525–2534. PMLR, (2019).
- [27] Heydari, A. A., Davalos, O. A., Zhao, L., Hoyer, K. K., and Sindi, S. S. ACTIVA: Realistic Single-Cell RNA-Seq Generation with Automatic Cell-Type Identification Using Introspective Variational Autoencoders. *Bioinformatics* **38**(8), 2194–2201 (2022).
- [28] Yan, X., Yang, J., Sohn, K., and Lee, H. Attribute2image: Conditional Image Generation from Visual Attributes. In *Computer Vision–ECCV*

- 735 2016: 14th European Conference, Amsterdam, The Netherlands, Octo-
736 ber 11–14, 2016, *Proceedings, Part IV 14*, 776–791. Springer, (2016).
- 737 [29] He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for
738 Image Recognition. In *Proceedings of the IEEE Conference on Computer*
739 *Vision and Pattern Recognition*, 770–778, (2016).
- 740 [30] Geng, J., Voitiuk, K., Parks, D. F., Robbins, A., Spaeth, A.,
741 Sevetson, J. L., Hernandez, S., Schweiger, H. E., Andrews, J. P.,
742 Seiler, S. T., et al. Multiscale Cloud-based Pipeline for Neu-
743 ronal Electrophysiology Analysis and Visualization. *bioRxiv* ,
744 <https://doi.org/10.1101/2024.11.14.623530> (2024).
- 745 [31] Tolossa, G. B., Schneider, A. M., Dyer, E. L., and Hengen, K. B. A
746 Conserved Code for Anatomy: Neurons Throughout the Brain Embed
747 Robust Signatures of their Anatomical Location into Spike Trains. *eLife*
748 , RP101506 (2024).
- 749 [32] Almeida, L. B. Multilayer perceptrons. In *Handbook of Neural Compu-*
750 *tation*, C1–2. CRC Press (2020).
- 751 [33] Müller, J., Ballini, M., Livi, P., Chen, Y., Radivojevic, M., Shadmani,
752 A., Viswam, V., Jones, I. L., Fiscella, M., Diggelmann, R., et al. High-
753 resolution CMOS MEA platform to study neurons at subcellular, cellu-
754 lar, and network levels. *Lab on a Chip* **15**(13), 2767–2780 (2015).
- 755 [34] de Vries, S. E., Siegle, J. H., and Koch, C. Sharing Neurophysiology
756 Data from the Allen Brain Observatory. *Elife* **12**, e85550 (2023).
- 757 [35] Lakunina, A. A., Nardoci, M. B., Ahmadian, Y., and Jaramillo, S.
758 Somatostatin-Expressing Interneurons in the Auditory Cortex Mediate
759 Sustained Suppression by Spectral Surround. *Journal of Neuroscience*
760 **40**(18), 3564–3575 (2020).
- 761 [36] Nowotny, T., Rospars, J.-P., Martinez, D., Elbanna, S., and Anton, S.
762 Machine Learning for Automatic Prediction of the Quality of Electro-
763 physiological Recordings. *Plos ONE* **8**(12), e80838 (2013).
- 764 [37] van Beest, E. H., Bimbard, C., Fabre, J. M., Dodgson, S. W., Takács,
765 F., Coen, P., Lebedeva, A., Harris, K. D., and Carandini, M. Tracking

- 766 Neurons Across Days with High-Density Probes. *Nature Methods* , 1–10
767 (2024).
- 768 [38] Hornauer, P., Prack, G., Anastasi, N., Ronchi, S., Kim, T., Donner, C.,
769 Fiscella, M., Borgwardt, K., Taylor, V., Jagasia, R., et al. DeePhys:
770 A Machine Learning–Assisted Platform for Electrophysiological Pheno-
771 typing of Human Neuronal Networks. *Stem Cell Reports* **19**(2), 285–298
772 (2024).
- 773 [39] Li, Q., Lin, Z., Liu, R., Tang, X., Huang, J., He, Y., Sui, X., Tian, W.,
774 Shen, H., Zhou, H., et al. Multimodal Charting of Molecular and Func-
775 tional Cell States via In Situ Electro-Sequencing. *Cell* **186**(9), 2002–2017
776 (2023).
- 777 [40] Adam, Y., Kim, J. J., Lou, S., Zhao, Y., Xie, M. E., Brinks, D., Wu, H.,
778 Mostajo-Radji, M. A., Kheifets, S., Parot, V., et al. Voltage imaging
779 and optogenetics reveal behaviour-dependent changes in hippocampal
780 dynamics. *Nature* **569**(7756), 413–417 (2019).
- 781 [41] Koch, C., Svoboda, K., Bernard, A., Basso, M. A., Churchland, A. K.,
782 Fairhall, A. L., Groblewski, P. A., Lecoq, J. A., Mainen, Z. F., Mathis,
783 M. W., et al. Next-Generation Brain Observatories. *Neuron* **110**(22),
784 3661–3666 (2022).
- 785 [42] Bader, F., Bingham, C., David, K. K., Gebrehiwet, H., Lantz, C., Peng,
786 G. C., Rangel-Gomez, M., and Gnadt, J. The NIH BRAIN Initia-
787 tive’s Impacts in Systems and Computational Neuroscience, 2014-2023.
788 *bioRxiv* , <https://doi.org/10.1101/2025.01.30.635684> (2025).
- 789 [43] Nicolas-Alonso, L. F. and Gomez-Gil, J. Brain Computer Interfaces, a
790 Review. *Sensors* **12**(2), 1211–1279 (2012).
- 791 [44] Steinmetz, N. A., Koch, C., Harris, K. D., and Carandini, M. Challenges
792 and Opportunities for Large-Scale Electrophysiology with Neuropixels
793 Probes. *Current Opinion in Neurobiology* **50**, 92–100 (2018).
- 794 [45] Yu, J., Hu, H., Agmon, A., and Svoboda, K. Recruitment of GABAer-
795 gic Interneurons in the Barrel Cortex During Active Tactile Behavior.
796 *Neuron* **104**(2), 412–427 (2019).

- 797 [46] Petersen, P. C., Siegle, J. H., Steinmetz, N. A., Mahallati, S., and
798 Buzsáki, G. CellExplorer: A Framework for Visualizing and Charac-
799 terizing Single Neurons. *Neuron* **109**(22), 3594–3608 (2021).
- 800 [47] Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N.,
801 Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., et al. Survey
802 of Spiking in the Mouse Visual System Reveals Functional Hierarchy.
803 *Nature* **592**(7852), 86–92 (2021).
- 804 [48] Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. Layer-Specific Phys-
805 iological Features and Interlaminar Interactions in the Primary Visual
806 Cortex of the Mouse. *Neuron* **101**(3), 500–513 (2019).

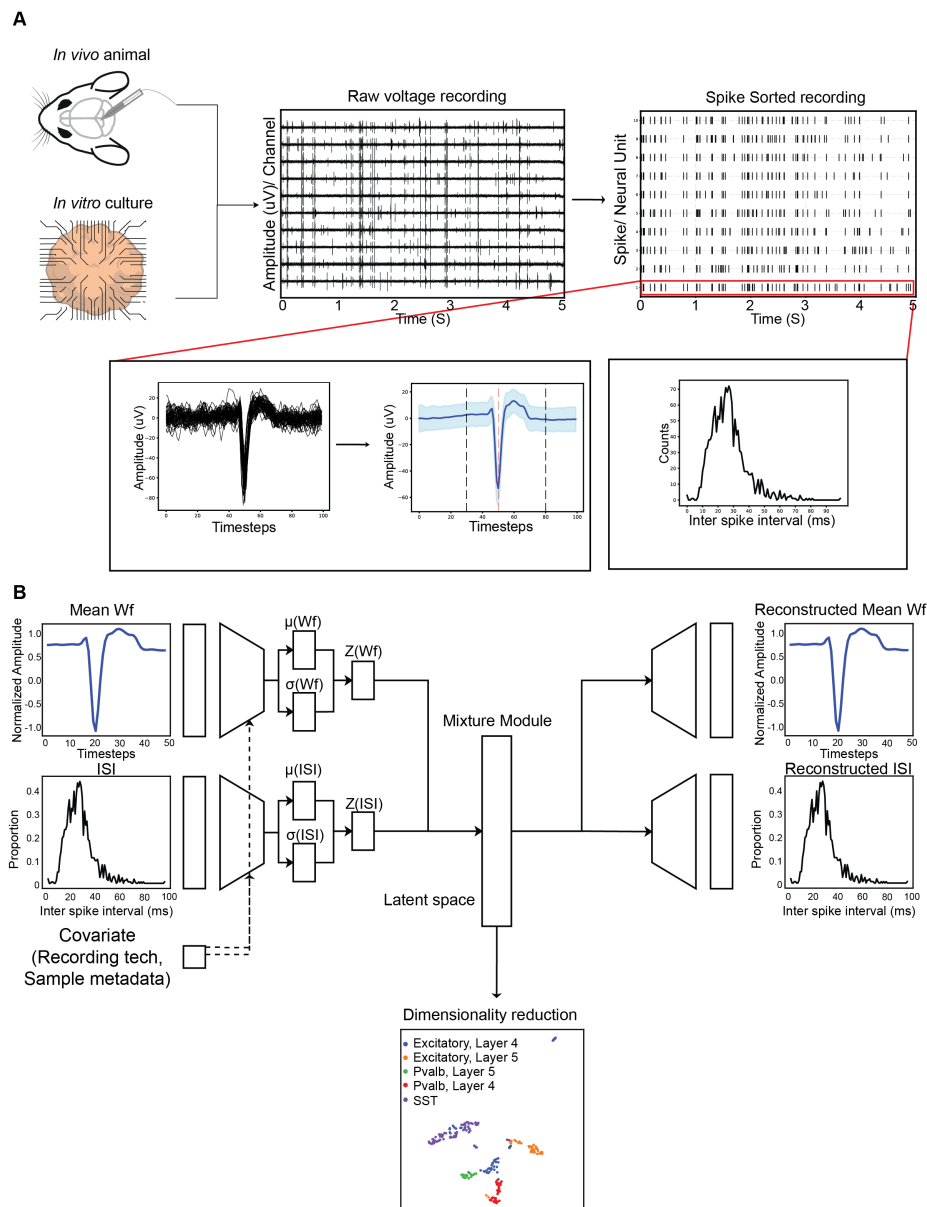


Figure 1: **The HIPPIE pipeline.** **A).** Data from *in vivo* or *in vitro* experiments are spike sorted. From the reconstructed spike trains, all spike timings are collected for each unit. The interspike interval distribution is then computed based on the spike train data. To analyze waveforms, the spike timings are used to retrieve the corresponding raw voltage recordings, extracting a 2.5 ms window centered on the action potential peak. **B).** 1 dimensional vectors representing the mean average waveform and interspike interval (ISI) distributions are fed into paired conditional VAEs. The paired VAEs are trained in parallel, and a mixture module is employed to create a joint representation of the modalities in the embedding space, which will be used for downstream tasks after dimensionality reduction.

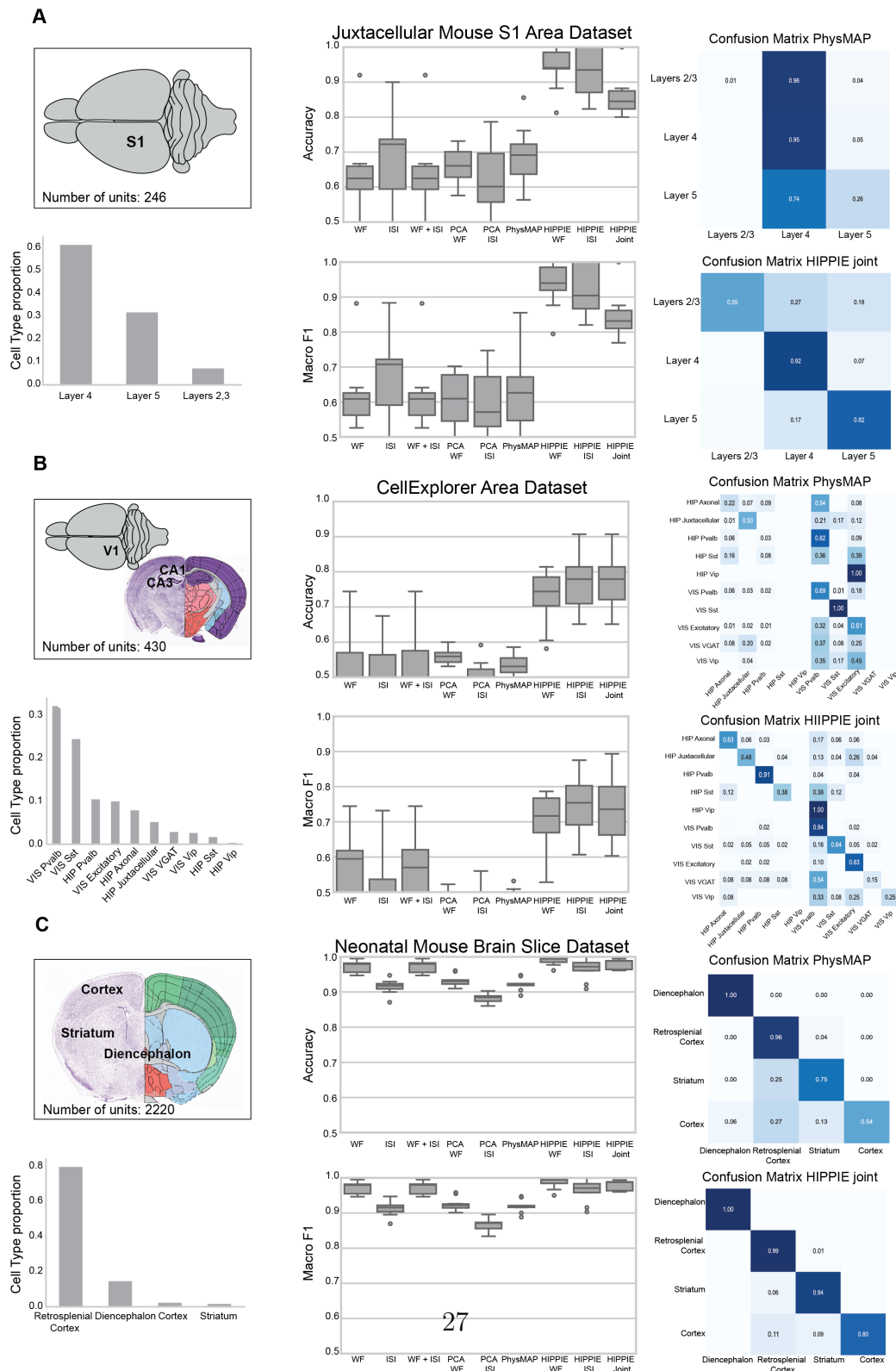


Figure 2: Anatomical locations benchmarks.

Figure 2: **A).** *Juxtacellular Mouse S1 Area Dataset Benchmark*: Left: Experimental location diagram (246 units) and neuronal layer proportions (L2/3, L4, L5). Middle: Boxplots showing Accuracy and Macro F1 across all tested dimensionality reduction methods. Right: Confusion matrices displaying layer-wise accuracy for PhysMAP (top) and HIPPIE joint modalities (bottom). **B).** *CellExplorer Area Dataset Benchmark*: Left: Experimental locations (430 units) and neuronal subtype proportions (Pvalb, Sst, Vgat, Vip, Excitatory, Axonal, Juxtacellular) across Visual (VIS) and Hippocampal (HIP) areas. Middle: Boxplots summarizing Accuracy and Macro F1 for all tested dimensionality reduction methods. Right: Confusion matrices for PhysMAP (top) and HIPPIE joint modalities (bottom). **C).** *Neonatal Mouse Brain Slice Dataset Benchmark*: Left: Experimental locations (2,220 units) and neuronal distribution across brain regions (retrospinal cortex, diencephalon, cortex, striatum). Middle: Boxplots presenting Accuracy and Macro F1 for all dimensionality reduction methods tested. Right: Confusion matrices showing accuracy for PhysMAP (top) and HIPPIE joint modalities (bottom). WF: Average waveform, ISI: Interspike Interval Distribution, Joint: Waveform and Interspike Interval

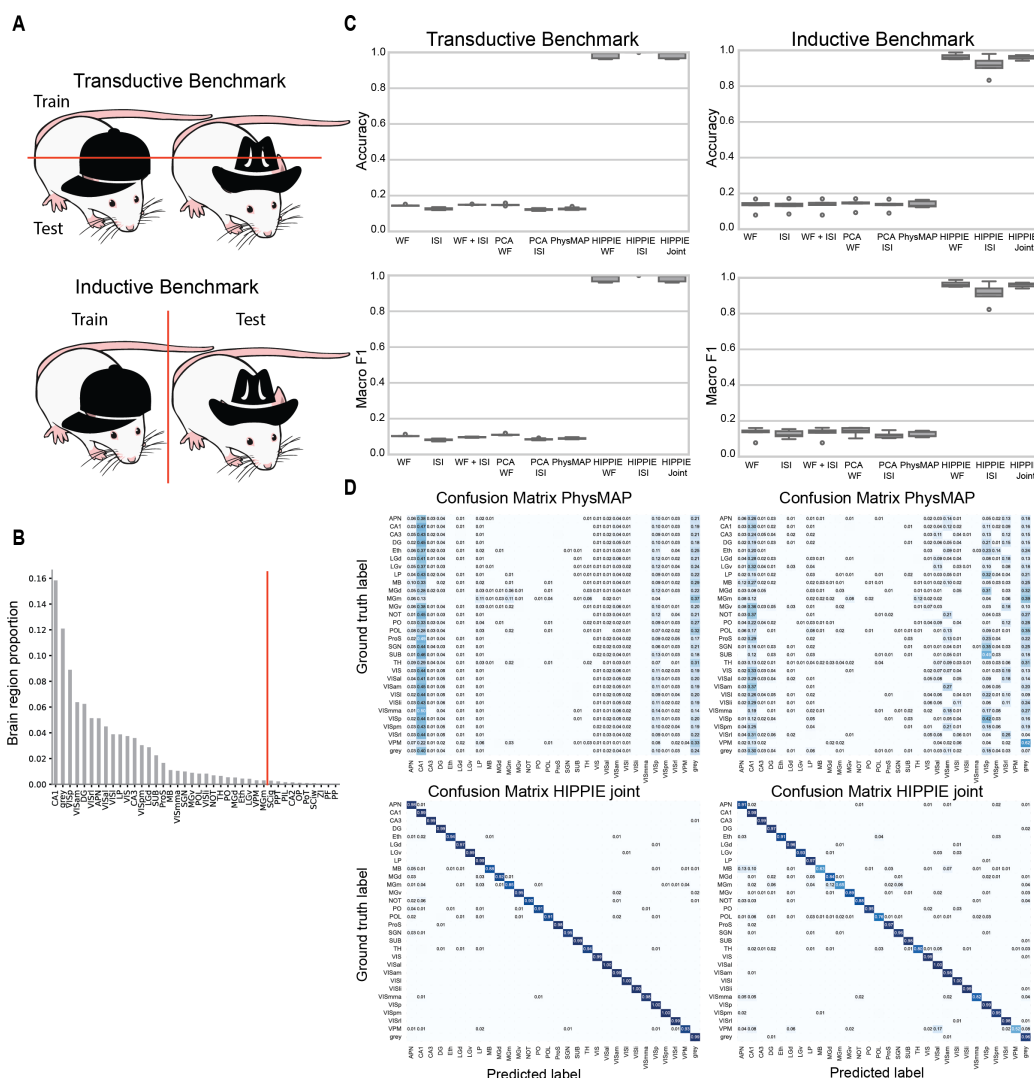


Figure 3: Inductive and transductive benchmarks for anatomical prediction.

Figure 3: **A).** Diagram illustrating the difference between transductive and inductive splits. **B).** Proportion of cells per brain region in the dataset. Red line indicates the cells that were removed from the analysis due to having $N < 50$. **C).** Accuracy and Macro-F1 boxplots comparing the performance of different dimensionality reduction methods using KNN probes. **D).** Confusion matrices for transductive and inductive tests of PhysMAP and HIPPIE joint modality dimensionality reduction methods. WF: Average waveform, ISI: Interspike Interval Distribution, Joint: Average waveform and Interspike Interval

Hippocampal structures include CA1, CA2, CA3, Dentate Gyrus (DG), Prosubiculum (ProS), and Subiculum (SUB). Thalamic structures include the Ethmoid Nucleus (Eth), Dorsal Lateral Geniculate (LGd), Lateral Posterior Nucleus (LP), Ventral Medial Geniculate (MGv), Posterior Complex (PO), Supragenicular Nucleus (SGN), Ventral Posteromedial Nucleus (VPM), Parafascicular Nucleus (PF), Posterior Limiting Nucleus (POL), Ventral Lateral Geniculate (LGv), Posterior Intralaminar Thalamic Nucleus (PIL), Posterior Triangular Thalamic Nucleus (PoT), and the Medial Geniculate Complex, including its medial (MGm) and dorsal (MGd) parts. Visuocortical structures (VIS) include the Anterolateral (VISal), Anteromedial (VISam), Lateral (VISl), Primary (VISp), Posteromedial (VISpm), Rostro-lateral (VISrl), and Laterointermediate (VISli) areas. Midbrain structures include the Anterior Pretectal Nucleus (APN), Superior Colliculus – Motor Related, Intermediate Gray Layer (SCig), Nucleus of the Optic Tract (NOT), Superior Colliculus – Motor Related, Intermediate White Layer (SCiw), Posterior Pretectal Nucleus (PPT), and Olivary Pretectal Nucleus (OP). The only hypothalamic structure included is the Zona Incerta (ZI). Cells labeled as grey represent units without an assigned brain region.

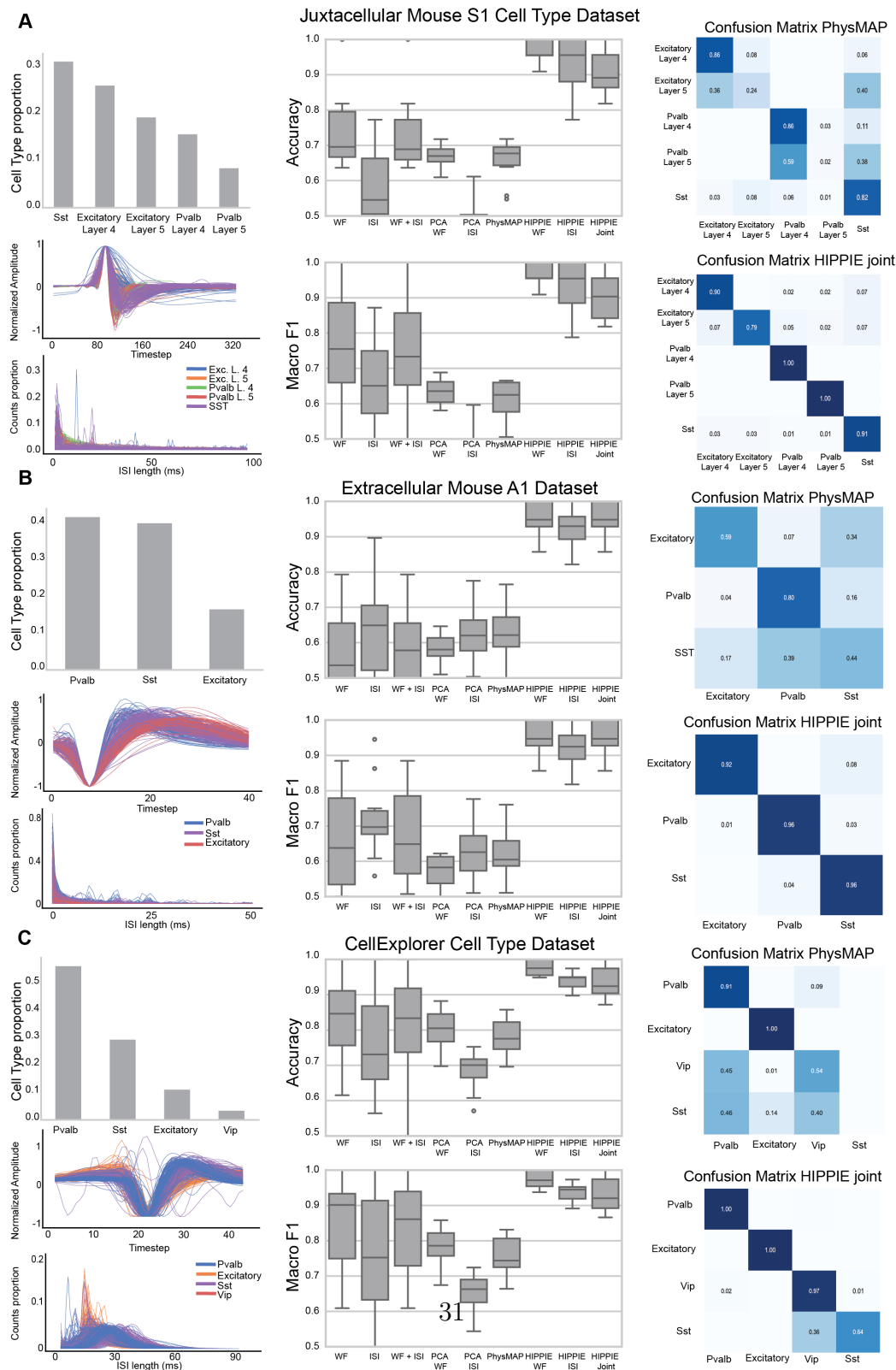


Figure 4: Cell types classification benchmarks.

Figure 4: **A).** *Juxtacellular Mouse S1 Area Dataset Benchmark*: Left: Neuronal subtype proportions (Sst, Pvalb Layer 4, Pvalb Layer 5, Excitatory Layer 4, Excitatory Layer 5) (top), amplitude (middle) and ISI distribution (bottom). Middle: Boxplots showing Accuracy and Macro F1 across all tested dimensionality reduction methods. Right: Confusion matrices displaying layer-wise accuracy for PhysMAP (top) and HIPPIE joint modalities (bottom). **B).** *Extracellular Mouse Auditory Cortex dataset Benchmark*: Left: Neuronal subtype proportions (Sst, Pvalb Layer 4, Pvalb Layer 5, Excitatory Layer 4, Excitatory Layer 5) (top), amplitude (middle) and ISI distribution (bottom). Middle: Boxplots showing Accuracy and Macro F1 across all tested dimensionality reduction methods. Right: Confusion matrices displaying layer-wise accuracy for PhysMAP (top) and HIPPIE joint modalities (bottom). **C).** *CellExplorer Cell Type Dataset Benchmark*: Left: Neuronal subtype proportions (Sst, Pvalb, Vip, Excitatory) (top), amplitude (middle) and ISI distribution (bottom). Middle: Boxplots showing Accuracy and Macro F1 across all tested dimensionality reduction methods. Right: Confusion matrices displaying layer-wise accuracy for PhysMAP (top) and HIPPIE joint modalities (bottom). WF: Average waveform, ISI: Interspike Interval Distribution, Joint: Waveform and Interspike Interval