

Multitask Learning of Biophysically-Detailed Neuron Models

Jonas Verhellen^{1†*}, Kosio Beshkov^{1†*}, Sebastian Amundsen¹, Torbjørn V. Ness², and Gaute T. Einevoll^{1,2*}

¹CINPLA, University of Oslo

²Department of Physics, Norwegian University of Life Sciences

†These authors contributed equally to this work

*For correspondence: jverhell@gmail.com, kosio.neuro@gmail.com, gaute.einevoll@nmbu.no

Abstract

The human brain operates at multiple levels, from molecules to circuits, and understanding these complex processes requires integrated research efforts. Simulating biophysically-detailed neuron models is a computationally expensive but effective method for studying local neural circuits. Recent innovations have shown that artificial neural networks (ANNs) can accurately predict the behaviour of these detailed models in terms of spikes, electrical potentials, and optical readouts. While these methods have the potential to accelerate large network simulations by several orders of magnitude compared to conventional differential equation based modelling, they currently only predict voltage outputs for the soma or a select few neuron compartments. Our novel approach, based on enhanced state-of-the-art architectures for multitask learning (MTL), allows for the simultaneous prediction of membrane potentials in each compartment of a neuron model, at a speed of up to two orders of magnitude faster than classical simulation methods. By predicting all membrane potentials together, our approach not only allows for comparison of model output with a wider range of experimental recordings (patch-electrode, voltage-sensitive dye imaging), it also provides the first stepping stone towards predicting local field potentials (LFPs), electroencephalogram (EEG) signals, and magnetoencephalography (MEG) signals from ANN-based simulations. It further presents a challenging benchmark for MTL architectures due to the large amount of data involved, the presence of correlations between neighbouring compartments, and the non-Gaussian distribution of membrane potentials.

1 Introduction

In the seven decades since Hodgkin and Huxley first described the action potential in terms of ion channel gating [25, 22, 21, 23, 24, 18, 6], the scientific community has gained a comprehensive understanding of how individual neurons process information. In contrast, however, the behaviour of large networks of neurons remains poorly understood. Experimental studies provide qualitative insights through statistical correlations between recorded neural activity and sensory stimulation or animal behaviour [26, 27, 29, 28], but statistical modelling offers little information on how networks perform neural computation or give rise to neural representations. Mechanistic modelling, in which detailed neuron models or networks of detailed neuron models are simulated on a computer, offers an alternative approach to studying the network dynamics of neural circuits [9, 11].

Thanks to recent pioneering efforts facilitated by large supercomputers, we are now able to construct simulations containing tens of thousands of model neurons that mimic specific cortical columns in mammalian sensory cortices [61, 55, 43, 58, 5]. Even more recent advances have reduced the need for supercomputers by distilling the output of biophysically-detailed neuron models into easier-to-evaluate artificial neural networks (ANN) [4, 51].

In the present paper we explore accelerated techniques for biophysically-detailed neuron models with full electrophysiological detail, a task that was previously deemed impractical due to the high computational cost involved. Our approach builds on previous work that focused on predicting outgoing action potentials or other experimental variables in a limited number of compartments. However, our approach is unique in that it allows for the simultaneous prediction of the membrane potentials (and membrane currents) for all compartments, see Figure 1. This allows for future comparison of model output with a wider range of experimental recordings of membrane potentials (dendritic patch electrodes, voltage-sensitive dye imaging), and also the calculation of extracellular signals such as local field potentials (LFPs), electroencephalogram (EEG) signals, and magnetoencephalography (MEG) signals [17]. This approach is an important step forward in our ability to simulate and eventually better understand the functioning of the brain in health and disease.

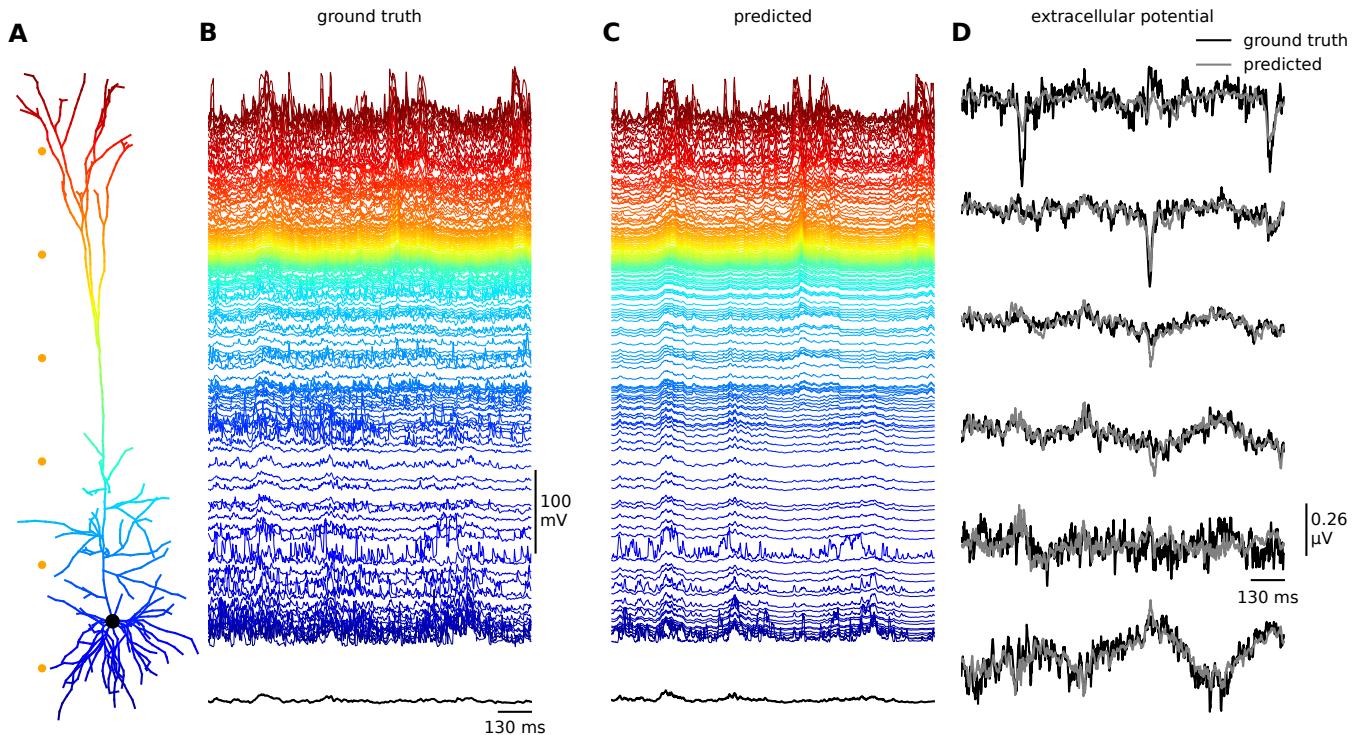


Figure 1: A) Illustration of the biophysically-detailed model with 639 compartments of a cortical layer V pyramidal cell model [19], which is the main object of study in this paper, color-coded by the compartment depth (consistent across all panels). B) Membrane voltages as calculated by a biophysically-detailed simulations of the multi-compartment model, used as the ground truth throughout this paper. C) Membrane voltages as predicted by our best-performing multi-task learning architecture, one time step at a time with a the time resolution of 1 ms for the prediction. D) Comparison between the ground truth and predicted extracellular potentials as calculated at six points outside the neuron (orange dots in panel A).

By predicting all membrane potentials across a biophysically-detailed neuron model simultaneously, rather than in the soma only as recently done by [4], we have made the transition from single-task learning to multi-task learning (MTL). In contrast to single-task learning that requires a separate model to be trained for each target, MTL optimises a single artificial neural architecture to predict multiple (heterogeneous) targets simultaneously. MTL approaches aim to improve generalisation and efficiency across tasks by leveraging statistical relationships between multiple targets [13, 62, 57, 34, 42].

In our work, statistical relationships between tasks, i.e. membrane potential predictions, arise from the correlations between compartments due to the biophysical mechanisms of the ion currents running through a neuron. To capture shared patterns that could be missed in single-task learning, multi-task learning generally relies on either one of two categories of neural architectures, respectively known as hard parameter [7] or soft parameter sharing models [10].

Hard parameter sharing models have a shared bottom layer in the neural network, while the output branches are task-specific. This means that the lower layers of the neural network are identical for all tasks, and only the final layers are tailored to each individual task. In contrast, soft parameter sharing models use dedicated sets of learning parameters and feature mixing mechanisms, allowing each task to have its own set of features and learning parameters while still efficiently sharing information. The choice between soft and hard parameter sharing models often depends on the nature of the tasks being learned. For example, if there is a substantial amount of overlap between the features necessary for each task, a hard parameter sharing model may be more appropriate, while if each (or any) task requires its own unique set of features for effective learning, a soft parameter sharing model may be more effective.

In this paper, we explore the capacity of various MTL architectures for distilling the full electrophysiology of a multi-compartment, biophysically-detailed layer 5 pyramidal neuron model [19]. Specifically, we compare a single type of hard parameter sharing model with two novel versions of state-of-the-art soft parameter sharing models, known as Multi-gate Mixture-of-Experts (MMoE) [41] and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx) [2]. Notably, in our computational experiments, we observed that the soft parameter sharing models substantially outperformed the hard parameter sharing model in the knowledge distillation of the electrophysiology of a biophysically-detailed neuron model. These findings highlight the potential benefits of soft-parameter sharing models for distilling complex electrophysiological data from biophysically-detailed neuron models, which we expect to have broad implications for computational neuroscience research.

For a comprehensive understanding of the MTL architectures utilised in this study, along with corresponding network diagrams, please refer to the detailed information provided in the Methods and Materials section. To make the following sections more accessible however, we briefly highlight that MMoE and MMoEEx, rely on a learnable feature mixing mechanism that controls the contribution of each learned data representation to the prediction of each task. The amount of learnable parameters in the feature mixing mechanisms grows linearly with the flattened input data size (batch size excluded). For the neural input data considered in this paper, the memory requirements regarding the amount of learnable parameters becomes prohibitively large. As a solution, we extend the MMoE and MMoEEx architectures to contain a compressor module that fixes the amount of learnable parameters in the feature mixing mechanisms to a predetermined number.

2 Results

To test the capability of different MTL architectures in accurately representing the full dynamic membrane potential of each of the 639 compartments of the large biophysically-detailed neuron model, we trained the models on a balanced dataset of simulated neural activity. For each target - a collection of membrane potentials (and the presence or absence of an outgoing action potential) - in the dataset, we provide a 100 ms history of neural activity and synaptic inputs from all compartments of the biophysically-detailed neuron model to the MTL models. Further details regarding the dataset and pre-processing steps can be found in the Methods and Materials section. To facilitate a fair comparison between MTL methods, we constructed a hard parameter sharing model (14 million trainable parameters) that is close in size to the soft parameter sharing models (12 million trainable parameters). Similarly, we used the same training procedures for each of the models – Adam [30] with or without task-balancing [38] – and evaluated the inference speed in a single session on publicly available hardware.

2.1 Multitask Prediction of Membrane Potential Dynamics in Biophysically-Detailed Neuron Models

After conducting comprehensive training on all three Multi-Task Learning (MTL) models, namely Hard Parameter Sharing (MH), Multi-gate Mixture-of-Experts (MMoE), and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx), we observed that the soft parameter sharing models exhibited a superior performance compared to

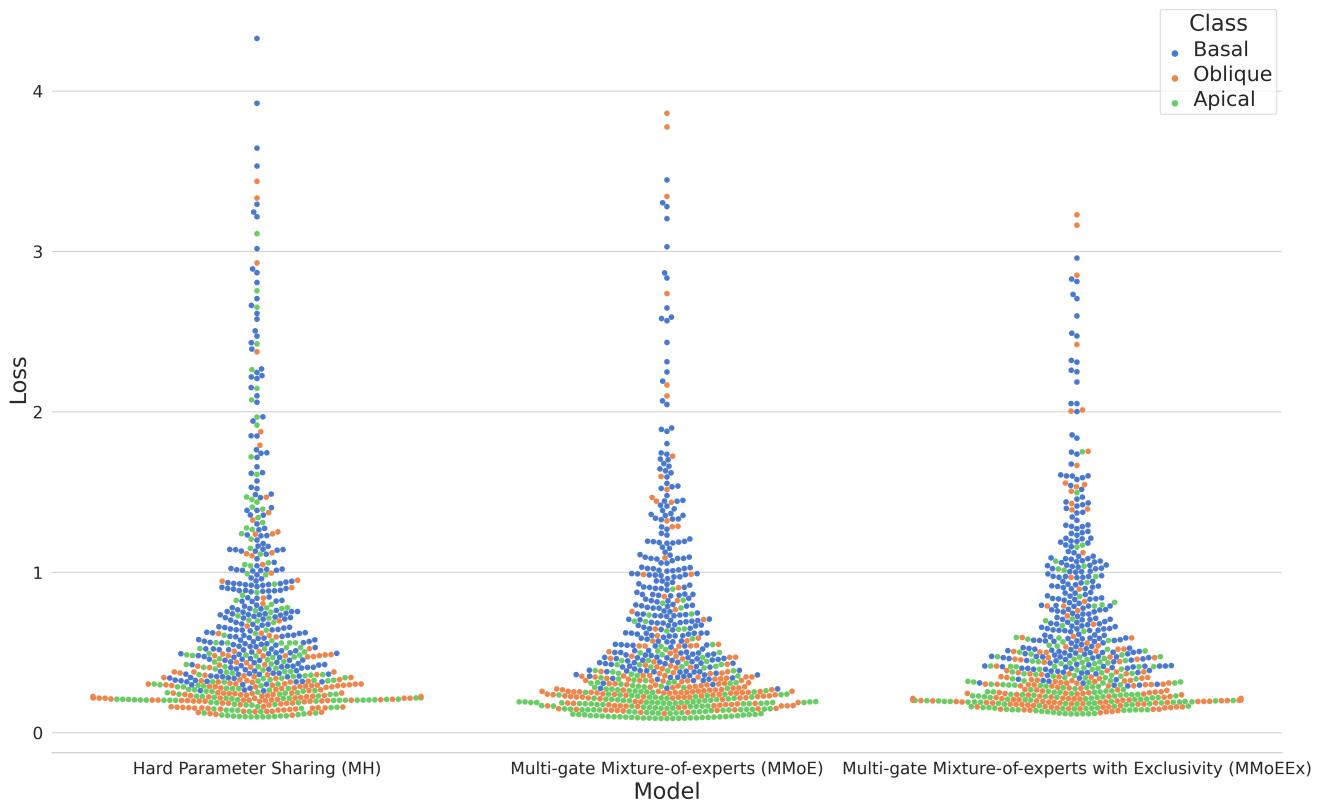


Figure 2: Swarm plots of the generalisation loss for each of the compartments (basal, oblique, apical) of the neuron model for each of the three models: Hard Parameter Sharing (MH), Multi-gate Mixture-of-experts (MMoE), and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEEx). Note that basal dendrites of a neuron receive incoming signals from other neurons and convey them towards the cell body, while the apical dendrite extends from the cell body to integrate signals from distant regions, and oblique dendrites play a role in the integration of synaptic inputs at various angles away from the other dendrites. The vertical axis represents the numeric loss values while the horizontal axis, the different ANN models are indicated. The density of the data points in a specific region indicates how many compartments have a similar loss value for each model.

the hard parameter sharing model, not only in terms of training but also in terms of generalization loss, see Figure 2 which showcases the training progress and the respective losses for each model. During the training process, we closely monitored the performance of each model through Tensorboard [1]. The hard parameter sharing model achieved a minimal training loss of 0.439 and a minimal validation loss of 0.680. The MMoE model, on the other hand, demonstrated better progress, reaching a minimal training loss of 0.241 and a minimal validation loss of 0.533. Similarly, the MMoEEEx model exhibited better training results than the hard parameter sharing model, with a minimal recorded training loss of 0.262 and a minimal validation loss of 0.551.

For further evaluation, we focused on the MMoE model, as it showcased the best performance based on the validation results. By utilising the model weights saved at its optimal validation performance, we were able to predict the membrane potential of a compartment within the biophysically-detailed neuron model with a root mean squared error of 3.78 mV compared to a standard deviation of 13.20 mV across the compartments and batches of the validation data. As explained in more detail in the Discussion section, the electrophysiological data upon which the biophysically-detailed neuron model is built, has an experimental standard deviation of around 5 mV, depending on the type of neural activity. It is worth noting that the network models we explored primarily prioritised learning the membrane potentials of the neuron model's compartments rather than accurately predicting spike generation in the soma. As a consequence, the presented results omit the performance analysis related to spike generation in the soma.

2.2 The Importance of Expert Diversity in MMoE and MMoEEx trained on Neural Data

Previous studies have suggested that higher diversity among experts in terms of the data representation they generate, see the Methods and Materials section, could potentially improve training and generalisation outcomes in MTL [41, 2], particularly for soft parameter sharing models. Motivated by these findings, we conducted a thorough investigation into the effect of expert diversity within the training process of the MMoE and MMoEEx models. To quantify and analyse expert diversity, we employed various diversity metrics. In addition to the diversity score, previously introduced by [2], which is based on the standardised distance matrix between experts, we extended our analysis to also include its determinant and permanent, see Methods. By considering these metrics, we were hoping to gain deeper insight into the level of diversity present among the experts throughout the training procedure. However, all three metrics exhibited the same temporal pattern. For a visual representation of these correlations and their implications on expert diversity, refer to Figure 1 in Appendix 1.

Interestingly, despite MMoEEx originally having been proposed in the literature as a means to enhance expert diversity, we made the intriguing observation that expert diversity exhibited significant variation across different training runs of the same model (MMoE or MMoEEx) with identical training data. This observation suggests that a soft parameter sharing model's architecture alone does not guarantee any consistent trend in expert diversity, see the appendices for more details, and that multiple training runs are necessary for complex deep learning models to learn about their properties. Furthermore, we studied the strength of the expert weights used to predict each compartment in both MMoE and MMoEEx, see Figure 3. While larger average weight does not necessarily imply a better prediction, it corresponds to a larger contribution to the prediction coming from a particular expert. As one can see in both models, different experts focus on different parts of the neuron, with relatively higher mean weights in higher apical and basal regions. One noticeable difference is that in MMoEEx, the oblique parts of the neuron are more strongly represented, whereas somatic compartments are visibly different from zero in one of the experts.

2.3 Fast Simulation of Full Membrane Potential Dynamics of Multiple Neurons

Traditional simulation environments such as NEURON rely on numerical integration of compartment-specific differential equations that represent the active and passive biophysical mechanisms of the modeled neuron. This approach requires a significant amount of computational resources. One of the main attractive features of distilling biophysically-detailed neuron models into ANNs is the significant speed-up deep learning architectures can provide. Previous work from the literature, where only the output of the soma compartment was predicted, has shown that NEURON simulations of biophysically-detailed neuron model networks, consisting of up to 5000 model neuron instances, are up to five orders of magnitude slower than their ANN counterparts [51]. Note that, in network simulations, accelerators like GPUs can be used to compute independent timesteps - for different instances of the same model neuron - of the ANN model in parallel. In practice, the batch-size of the MTL models during inference can directly be interpreted as the number of model neurons being evaluated at the same time.

We recorded the mean and standard deviation runtimes of seven independent predictions for each of the three MTL models for different amounts of neurons and simulation times. In Table 1, we present the results of 100 neurons simulated for 1 ms, 10 ms and 100 ms, and for a 100ms simulation of 1000 neurons. To provide context, we also recorded the mean and standard deviation runtimes of seven independent 100 ms simulations of a single neuron in NEURON. Note that in the NEURON simulation, each timestep is passed to the next internally, while for the runtime tests of the deep learning models we are limited to running each timestep from Python causing an overhead for the MTL models. Clearly, predictions by MTL models are significantly faster than single-core simulations in NEURON. For the set-up with 1000 neurons, the hard parameter sharing model reaches an acceleration of 2 orders of magnitude over the NEURON simulation. Being able to predict voltages in parallel has a profound effect on efficiency, as a 1ms simulation of 100 neurons can be predicted in 0.45 seconds which is nearly half the time necessary for NEURON to simulate a single neuron for 100 ms, while requiring the same amount of update steps.

Also note that while prediction times for the MTL models scale roughly linearly with simulation times, they do not scale linearly with batch size. Within the maximum batch size that can be accommodated by the GPU memory, the number of simulated neurons should only affect the evaluation time of a neural network through the previously discussed Python overhead, as is seen in the recorded prediction times for the hard parameter sharing model. However, for the soft parameter sharing models which contain matrix multiplications in the gates we observe a steep increase in runtime for large batch sizes. In general, it is worth remarking that MMoE and MMoEEx demonstrate inference runtimes which are significantly slower compared to the hard parameter sharing model. This discrepancy can be attributed to the hard parameter sharing model having a lower number of

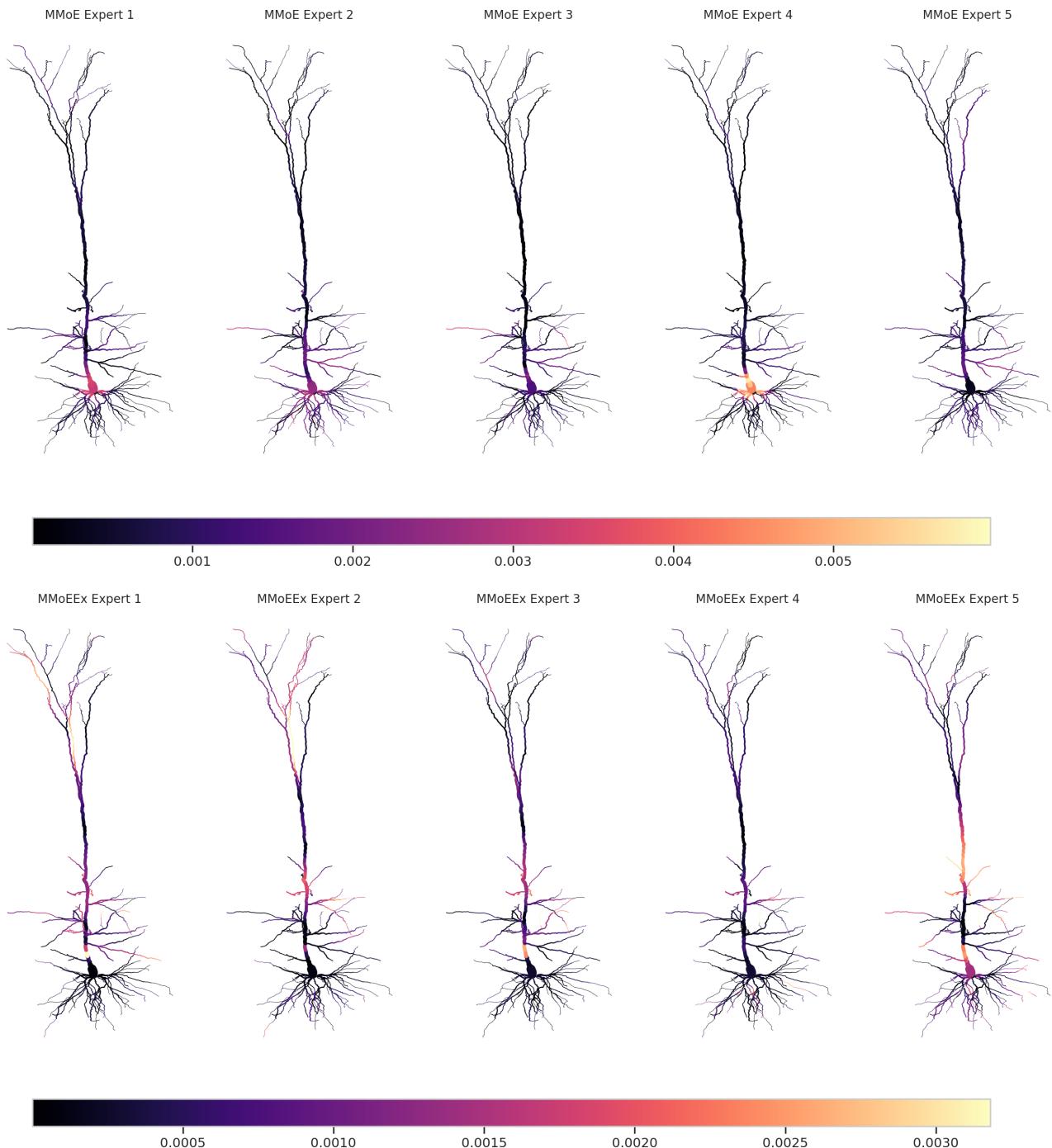


Figure 3: Mean weights from the experts in both Multi-gate Mixture-of-experts (MMoE; top row) and Multi-gate Mixture-of-experts with Exclusivity (MMoEEx; bottom row) after training the models on neural data, projected onto the different compartments of the cortical layer V biophysically-detailed neuron model. Each row consists of five subplots, representing different experts within the models. The color scale (normalised to the mean weights of the first expert) is indicated by the horizontal color bar located below each row.

Runtime (s)	Std. Dev. (ms)	Model	Hardware	Simulation (ms)	Batch Size
1.01	29.7	Neuron	CPU	100	1
0.07	0.85	MH	GPU	1	100
0.45	1.03	MMoE	GPU	1	100
0.44	1.43	MMoEEx	GPU	1	100
0.70	25.4	MH	GPU	10	100
4.36	35.7	MMoE	GPU	10	100
4.35	15.0	MMoEEx	GPU	10	100
7.19	145	MH	GPU	100	100
43.6	58.6	MMoE	GPU	100	100
43.5	82.4	MMoEEx	GPU	100	100
10.9	153	MH	GPU	100	1000
365.0	34.2	MMoE	GPU	100	1000
365.0	33.5	MMoEEx	GPU	100	1000

Table 1: Inference speed of all three MTL architectures - hard parameter sharing (MH), Multi-gate Mixture-of-experts (MMoE), and Multi-gate Mixture-of-experts with Exclusivity (MMoEEx) - after training on neural data generated by a biophysically-detailed model of a cortical layer 5 pyramidal neuron model compared to the classical NEURON simulation.

convolution filters and the soft parameter sharing models containing large and computationally expensive matrix multiplications.

3 Discussion

3.1 Multitask Prediction of Membrane Potential Dynamics in Biophysically-Detailed Neuron Models

In our investigation, we evaluated three advanced multi-task learning (MTL) neural network architectures [7, 41, 2], including MMoE and MMoEx which were augmented with a novel compressor module, to tackle the challenging task of capturing intricate membrane potential dynamics in a complex, multi-compartment, biophysically-detailed model of a layer 5 pyramidal neuron [19]. Given the considerable complexity of the model comprising 639 compartments, each generating distinct timeseries data potentially containing spikes, this MTL problem presented an exceptionally difficult distillation task. Nevertheless, our findings are encouraging, with the MMoE model achieving the lowest validation loss and demonstrating a root mean squared error of 3.78 mV in predicting membrane potential across all compartments. In comparison, the root mean squared error for the MMoEEx and MH models was 3.90 mV and 4.81 mV respectively. Notably, these errors are considerably smaller than both the standard deviation of our test set (13.20 mV) and the experimentally measured standard deviations of the peak membrane voltages during perisomatic step current firing (4.97 mV - 6.93 mV) or back-propagating action potential Ca^{2+} firing (5 mV) in layer 5 pyramidal neurons [33, 19].

Experimental recordings of the after-hyperpolarization depth of the membrane potential in the soma of layer 5 pyramidal neurons have a slightly lower standard deviation (3.58 mV - 5.82 mV) than electrophysiological measurements of action potentials [31, 32, 19], indicating that a more sensitive evaluation of distilled neuron models could be based on their performance in specific neuronal scenario's. Notably, thus far, all trained models did not manage to learn the binary somatic spike prediction task, a challenge that could potentially be mitigated through a computationally intensive hyperparameter search for γ (explained in the Methods and Materials section) to assign higher importance to this specific task during the learning process. Although task balancing methods for soft parameter sharing models were explored in accordance with the procedures detailed in the Methods and Materials section, our findings indicate no improvement in performance, as outlined in Appendix 2. Furthermore, it is essential to highlight that these MTL models are anticipated to find utility in accelerating LFP calculations, where the significance of subthreshold components in membrane potential dynamics is well-established [52, 54, 35]. For further insights into calculating extracellular potentials based on the deep learning model's membrane potential predictions and an example of such a downstream prediction, please refer to Appendix 3.

3.2 Measuring Experts Diversity in MMoE and MMoEEx Trained on Neural Data

To test the conjecture that in soft parameter-sharing MTL models, high diversity between experts can be beneficial in training and generalisation, we computed several diversity measures. Indeed, initially it looked like the model with highest diversity (MMoE) performed the best, despite the fact that unlike its counterpart (MMoEEx), it has no explicit inductive bias towards expert diversity. However, a subsequent retraining of the models showed that the known diversity metrics are neither robust nor correlated with training or generalisation performance. Further exploration of the exclusivity hyperparameter α might lead to more desirable results for the MMoEEx model. Additionally, increasing the number of experts should only lead to improvements in training and generalisation if a substantial amount of independence between the tasks had not yet been incorporated. Future work can make these statements more precise by studying the contribution of individual experts to the prediction of membrane potentials in morphologically distinct neuronal compartments.

3.3 Fast Simulation of Full Membrane Potential Dynamics of Multiple Neurons

Biophysically-detailed neuron models distilled into ANNs can be evaluated at significantly higher speeds than their classical counterparts. Because a single instance of a deep learning model can be used to predict outputs for multiple instances of the same neuron model in parallel, on accelerators such as GPUs, these models are particularly suited to accelerating large networks of model neurons. Previous results from the literature [51] have shown that using deep learning models that only predict the output voltage of the soma, could result in a five order magnitude speed-up for a network model of 5000 neurons. In this paper, we have shown that a speed-up of two orders of magnitude can be obtained for a network of 1000 neurons by making use of MTL deep learning models, when the voltage traces for each compartment of the underlying biophysically-detailed neuron model need to be predicted. The highest speedup was achieved with the MH model, which has a slightly higher root mean squared error (4.81 mV) compared to the other two models, but still performs within experimentally observed variability [19].

The presented MTL models did not succeed in the binary somatic spike prediction task. As such, the models are not particularly suited for running recurrently connected neural network simulations. They can, however, be used for investigating, for example, LFP, EEG, or MEG signals from different types of predetermined synaptic input to neural populations, which has commonly been used in the literature to study the origin and information content of these brain signals [53, 36, 37, 59, 40, 45, 14, 48, 15, 39, 47, 60, 50, 44, 16]. Furthermore, in cases where the synaptic input is predetermined, the individual timesteps can be treated independently and in parallel. For example, if we assume a batch size of 1000, this can either correspond to simultaneously simulating 1000 timesteps of a single neuron, or one timestep for a population of 1000 neurons.

An important application of the biophysically-detailed neuron model is the calculation of LFPs and downstream EEG and MEG signals. Recent work has shown the importance of LFP recordings in validating large computational models of brain tissue [56], such as the mouse V1 cortical area, developed by the Allen Institute [5]. The Allen Institute mouse V1 model contains 114 distinct model neuron types, and could be accelerated (by several orders of magnitude) by 114 separate deep learning models each representing one such neuron type. The current biophysically-detailed Allen V1 model makes use of neuron models with passive dendrites which should be significantly easier to distil into an MTL architecture than the layer 5 pyramidal neuron model represented here. Further acceleration of the MTL models discussed in this paper could be achieved by running model inference on multiple GPUs or more advanced accelerators such as TPUs and IPUs.

4 Methods and Materials

4.1 Multicompartmental NEURON Simulations and Data Balancing

As a baseline for training and testing, we used an existing dataset of electrophysiological data generated in NEURON [4] based on a well-known biophysical-detailed and multi-compartment model of cortical layer V pyramidal cells [19]. This model contains a wide range of dendritic (Ca^{2+} -driven) and perisomatic (Na^+ and K^+ -driven) active properties which are represented by ten key active ionic currents that are unevenly distributed over different dendritic compartments. The data was generated in response to presynaptic spike trains sampled from a Poisson process. For the purposes of this paper, it is important to note that the biophysically-detailed model contains 639 compartments and 1278 synapses and that 128 simulations of the complete model for 6 seconds of biological time each were included before data balancing.

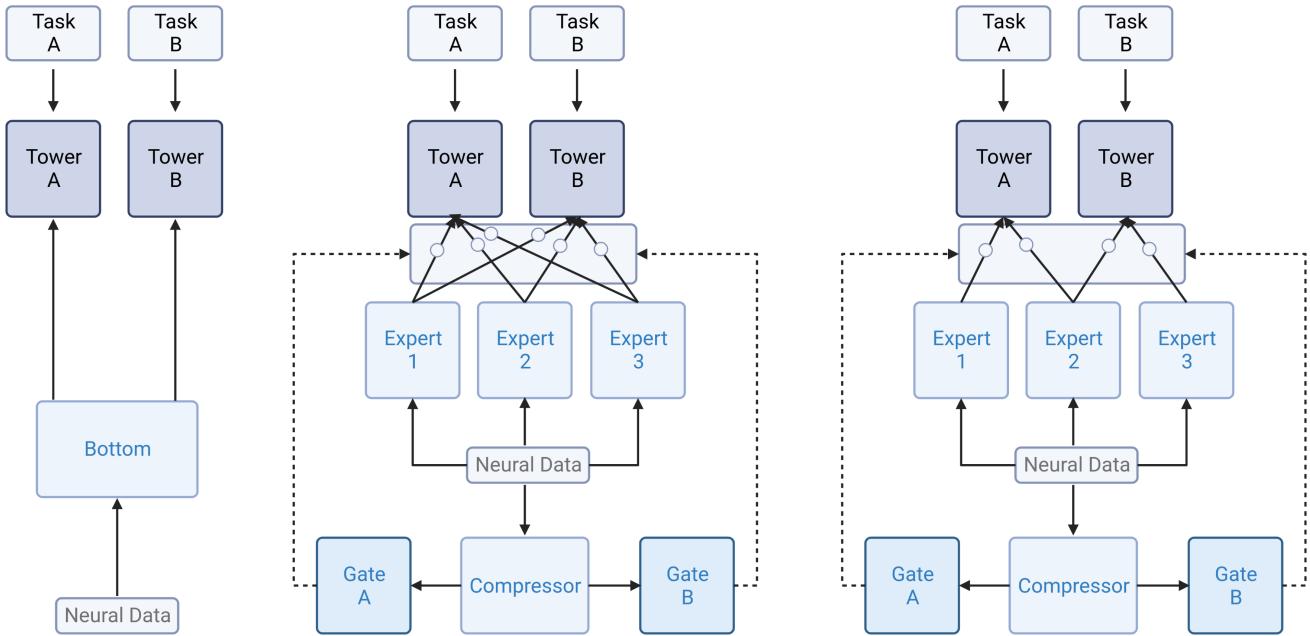


Figure 4: Schematic representation of the architectures of the hard parameter model (left), the MMoE model (middle), and the MMoEEx model (right).

The subthreshold dynamics of the membrane potential in a compartment have small variations and as a result one would expect them to be easy to predict. However, suprathreshold deviations generated by action potentials propagating through the dendrites can be more problematic to predict. To address this issue, we implemented a form of data balancing. We first identified the time points at which somatic spikes occurred and afterwards we used them to create a dataset in which one third of the targets included a spiking event. Additionally, we standardised the membrane potential through z-scoring. We used input data consisting of membrane potentials and incoming synaptic events across all compartments during a 100 ms time window, and target data consisting of 1 ms of membrane potentials and a binary value for the presence or absence of a somatic spike.

4.2 Hard and Soft Parameter Sharing Architectures for Multitask Learning

We implemented a temporal convolutional network (TCN) architecture [3] to learn spatiotemporal relationships between inputs (synaptic events and membrane potentials) and outputs (membrane potentials), for both hard and soft parameter-sharing architectures for multi-task learning. While recurrent neural networks [20, 8] such as GRU's and LSTM's are commonly used for sequence tasks, research has shown that convolutional architectures can perform just as well or better on tasks like audio synthesis, language modelling, and machine translation [3]. A TCN is a one-dimensional convolutional network with a causal structure, which guarantees causality layer by layer through the use of causal convolutions, padding and dilation. TCNs preserve the temporal ordering of the input data, making it impossible for the network to use information from the future to make predictions about the past. This is important because it ensures that the network can be applied in real-world scenarios where only past information is available.

In early studies of modern multi-task learning, hard parameter sharing was used to share the initial layers ("the bottom") and task-specific top layers ("the towers") of the neural network architecture. This approach has the advantage of being scalable with increasing numbers of tasks, but it can result in a biased shared representation that favours tasks with dominant loss signals. To address this issue, soft parameter sharing architectures have been developed, which utilise dedicated representations for each task. In this study, we employ two soft parameter sharing models, the multi-gate mixture-of-experts (MMoE) [41] and the multi-gate mixture-of-experts with exclusivity (MMoEEx) [2]. These models combine representations learned by multiple shared bottoms ("the

experts") through gating functions that apply linear combinations using learnable and data-dependent weights. MMoEEx is an extension of MMoE that encourages diversity among expert representations by randomly setting a subsection of weights to zero.

In our study, we employed a three-layered TCN with varying channel sizes of 32, 16, and 8 and a kernel size of 10. We also implemented a dropout of 0.2 for both the bottom architecture (hard parameter sharing) and the expert architectures (soft parameter sharing). For each tower, we used a three-layered feed-forward network with ELU activation functions and 10 or 25 hidden nodes, depending on whether it was a soft or hard parameter sharing model, respectively. To ensure increased stability, we replaced the final RELU activation function in the original TCN implementation with a sigmoid activation function. In addition, we included an extra TCN expert as a compressor module in the soft parameter sharing models to reduce the size of the data before feeding it into the gating functions, see Figure 4, which helped to avoid quadratic growth of the trainable parameters with respect to the input data size. To improve efficiency, we implemented mixed precision and multi-GPU training for all three models. Our models were trained on up to six parallel RTX2080Ti GPUs, and we performed inference experiments on publicly available Tesla P100 GPUs via Kaggle.

4.3 Task Balancing and Expert Diversity

Effective multi-task learning sometimes requires some form of task balancing to reduce negative transfer or to prevent one or more tasks from dominating the optimisation procedure. To avoid these issues we make use of loss-balanced task weighting (LBTW) [38] which dynamically updates task weights in the loss function during training. For each batch, LBTW calculates loss weights based on the ratio between the current loss and the initial loss for each task, and a hyperparameter α . As α goes to 0, LBTW approaches standard multitask learning training. All taken together, the loss function can be summarised as

$$L(\hat{y}, y) = L_{spike}(\hat{y}_{spike}, y_{spike}) + \sum_{i=0}^N L_i(\hat{y}_i, y_i) = w_{spike} \cdot \text{BCE}_{logits}(\hat{y}_{spike}, y_{spike}) + \sum_{i=0}^N w_i \cdot \text{MSE}(\hat{y}_i, y_i), \quad (1)$$

where \hat{y} and y are the target and the predicted data respectively, w_{spike} and w_i are the task weight, BCE_{logits} is the binary cross entropy loss combined with a sigmoid activation function, and MSE is the mean square error. The summation index i runs over all N compartments of the biophysically-detailed neuron model, and the LBTW task weights are recalculated every epoch E , for each batch B , according to

$$w_{spike} = \left(\frac{L_{spike,E,B}(\hat{y}_{spike}, y_{spike})}{L_{spike,E,0}(\hat{y}_{spike}, y_{spike})} \right)^\alpha \quad \text{and} \quad w_i = \left(\frac{L_{i,E,B}(\hat{y}_i, y_i)}{L_{i,E,0}(\hat{y}_i, y_i)} \right)^\alpha. \quad (2)$$

To measure the diversity in expert representations in MMoE and MMoEEx, we made use of diversity measurement proposed in the original MMoEEx paper. In that paper, the diversity between two experts n and m is calculated as a (real-valued) distance $d(n, m)$ between the learned representations f_n and f_m , as defined by

$$d(n, m) = \sqrt{\sum_{i=0}^N (f_n(x_i) - f_m(x_i))^2}, \quad (3)$$

where N is the number of samples x_i in the validation set, which is used to probe the diversity of the experts. The diversity matrix D of a trained MMoE or MMoEEx model is defined by calculating all pairwise distances between expert representations as described above and normalising the matrix. In this normalised matrix, a pair of experts with distance close to 0 are considered near-identical, and experts with a distance close to 1 are considered to be highly diverse. To compare two different models in terms of overall diversity, we respectively define the first, second, and third diversity scores of a model as the mean entry (\bar{d}_1) and the determinant (\bar{d}_2), and the permanent (\bar{d}_3) of its diversity matrix D .

5 Acknowledgments

This research was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 945371, the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement No. 945539 Human Brain Project(HBP) SGA3 and by UiO:Life Science through the 4MENT convergence environment. Figure 3 has been made with Biorender.com through an academic license.

6 Code Availability

The LFDeep: Multitask Learning of Biophysically-Detailed Neuron Models project code is hosted on GitHub. The repository includes the implementation of the paper's three multitask learning architectures for the desitllation of biophysically-detailed neuron models. To access the latest version of the code and contribute to the project, please visit the official GitHub repository at <https://github.com/Jonas-Verhellen/LFDeep>.

References

- [1] ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., GOODFELLOW, I., HARP, A., IRVING, G., ISARD, M., JIA, Y., JOZEFOWICZ, R., KAISER, L., KUDLUR, M., LEVENBERG, J., MANÉ, D., MONGA, R., MOORE, S., MURRAY, D., OLAH, C., SCHUSTER, M., SHLENS, J., STEINER, B., SUTSKEVER, I., TALWAR, K., TUCKER, P., VANHOUCKE, V., VASUDEVAN, V., VIÉGAS, F., VINYALS, O., WARDEN, P., WATTENBERG, M., WICKE, M., YU, Y., AND ZHENG, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [2] AOKI, R., TUNG, F., AND OLIVEIRA, G. L. Heterogeneous multi-task learning with expert diversity. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 19, 6 (2022), 3093–3102.
- [3] BAI, S., KOLTER, J. Z., AND KOLTUN, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).
- [4] BENIAGUEV, D., SEGEV, I., AND LONDON, M. Single cortical neurons as deep artificial neural networks. *Neuron* 109, 17 (2021), 2727–2739.
- [5] BILLEH, Y. N., CAI, B., GRATIY, S. L., DAI, K., IYER, R., GOUWENS, N. W., ABBASI-ASL, R., JIA, X., SIEGLE, J. H., OLSEN, S. R., ET AL. Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron* 106, 3 (2020), 388–403.
- [6] BROWN, A. The hodgkin and huxley papers: still inspiring after all these years, 2022.
- [7] CARUANA, R. Multitask learning: A knowledge-based source of inductive bias1. In *Proceedings of the Tenth International Conference on Machine Learning* (1993), Citeseer, pp. 41–48.
- [8] CHUNG, J., GULCEHRE, C., CHO, K., AND BENGIO, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
- [9] DAYAN, P., AND ABBOTT, L. F. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.
- [10] DUONG, L., COHN, T., BIRD, S., AND COOK, P. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers)* (2015), pp. 845–850.
- [11] EINEVOLL, G. T., DESTEXHE, A., DIESMANN, M., GRÜN, S., JIRSA, V., DE KAMPS, M., MIGLIORE, M., NESS, T. V., PLESSER, H. E., AND SCHÜRMANN, F. The scientific case for brain simulations. *Neuron* 102, 4 (2019), 735–744.
- [12] EINEVOLL, G. T., KAYSER, C., LOGOTHETIS, N. K., AND PANZERI, S. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Reviews Neuroscience* 14, 11 (2013), 770–785.
- [13] FINN, C., ABBEEL, P., AND LEVINE, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (2017), PMLR, pp. 1126–1135.
- [14] HAGEN, E., DAHMEN, D., STAVRINOU, M. L., LINDÉN, H., TETZLAFF, T., ALBADA, S. J. V., GRÜN, S., DIESMANN, M., AND EINEVOLL, G. T. Hybrid scheme for modeling local field potentials from point-neuron networks. *Cerebral Cortex* 26 (2016), 4461–4496.
- [15] HAGEN, E., FOSSUM, J. C., PETTERSEN, K. H., ALONSO, J. M., SWADLOW, H. A., AND EINEVOLL, G. T. Focal local field potential signature of the single-axon monosynaptic thalamocortical connection. *Journal of Neuroscience* 37 (2017), 5123–5143.
- [16] HAGEN, E., MAGNUSSON, S. H., NESS, T. V., HALNES, G., BABU, P. N., LINSSEN, C., MORRISON, A., AND EINEVOLL, G. T. Brain signal predictions from multi-scale networks using a linearized framework. *PLOS Computational Biology* 18 (2022), 2022.02.28.482256.
- [17] HAGEN, E., NÆSS, S., NESS, T. V., AND EINEVOLL, G. T. Multimodal modeling of neural network activity: computing lfp, ecog, eeg, and meg signals with lfp 2.0. *Frontiers in neuroinformatics* 12 (2018), 92.

- [18] HÄUSSER, M. The hodgkin-huxley theory of the action potential. *Nature neuroscience* 3, 11 (2000), 1165–1165.
- [19] HAY, E., HILL, S., SCHÜRMANN, F., MARKRAM, H., AND SEGEV, I. Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS computational biology* 7, 7 (2011), e1002107.
- [20] HOCHREITER, S., AND SCHMIDHUBER, J. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [21] HODGKIN, A. L., AND HUXLEY, A. F. The components of membrane conductance in the giant axon of loligo. *The Journal of physiology* 116, 4 (1952), 473.
- [22] HODGKIN, A. L., AND HUXLEY, A. F. Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *The Journal of physiology* 116, 4 (1952), 449.
- [23] HODGKIN, A. L., AND HUXLEY, A. F. The dual effect of membrane potential on sodium conductance in the giant axon of loligo. *The Journal of physiology* 116, 4 (1952), 497.
- [24] HODGKIN, A. L., AND HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology* 117, 4 (1952), 500.
- [25] HODGKIN, A. L., HUXLEY, A. F., AND KATZ, B. Measurement of current-voltage relations in the membrane of the giant axon of loligo. *The Journal of physiology* 116, 4 (1952), 424.
- [26] HUBEL, D. H., AND WIESEL, T. N. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology* 148, 3 (1959), 574.
- [27] HUBEL, D. H., AND WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology* 160, 1 (1962), 106.
- [28] HUBEL, D. H., AND WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* 195, 1 (1968), 215–243.
- [29] HUBEL, D. H., AND WIESEL, T. N. Ferrier lecture-functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 198, 1130 (1977), 1–59.
- [30] KINGMA, D. P., AND BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [31] LARKUM, M. E., ZHU, J. J., AND SAKMANN, B. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398, 6725 (1999), 338–341.
- [32] LARKUM, M. E., ZHU, J. J., AND SAKMANN, B. Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons. *The Journal of physiology* 533, 2 (2001), 447–466.
- [33] LE BÉ, J.-V., SILBERBERG, G., WANG, Y., AND MARKRAM, H. Morphological, electrophysiological, and synaptic properties of corticocallosal pyramidal cells in the neonatal rat neocortex. *Cerebral cortex* 17, 9 (2007), 2204–2213.
- [34] LEE, S., AND SON, Y. Multitask learning with single gradient step update for task balancing. *Neurocomputing* 467 (2022), 442–453.
- [35] LINDÉN, H., PETTERSEN, K. H., AND EINEVOLL, G. T. Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. *Journal of computational neuroscience* 29 (2010), 423–444.
- [36] LINDÉN, H., PETTERSEN, K. H., AND EINEVOLL, G. T. Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. *Journal of computational neuroscience* 29 (12 2010), 423–44.
- [37] LINDÉN, H., TETZLAFF, T., POTJANS, T. C., PETTERSEN, K. H., GRÜN, S., DIESMANN, M., AND EINEVOLL, G. T. Modeling the spatial reach of the lfp. *Neuron* 72 (12 2011), 859–72.
- [38] LIU, S., LIANG, Y., AND GITTER, A. Loss-balanced task weighting to reduce negative transfer in multi-task learning. In *Proceedings of the AAAI conference on artificial intelligence* (2019), vol. 33, pp. 9977–9978.
- [39] LUO, J., MACIAS, S., NESS, T. V., EINEVOLL, G. T., ZHANG, K., AND MOSS, C. F. Neural timing of stimulus events with microsecond precision. *PLoS biology* 16 (2018), 1–22.
- [40] LEŠKI, S., LINDÉN, H., TETZLAFF, T., PETTERSEN, K. H., AND EINEVOLL, G. T. Frequency dependence of signal power and spatial reach of the local field potential. *PLoS computational biology* 9 (7 2013), 1–23.
- [41] MA, J., ZHAO, Z., YI, X., CHEN, J., HONG, L., AND CHI, E. H. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (2018), pp. 1930–1939.

- [42] MALLYA, A., DAVIS, D., AND LAZEBNIK, S. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 67–82.
- [43] MARKRAM, H., MULLER, E., RAMASWAMY, S., REIMANN, M. W., ABDELLAH, M., SANCHEZ, C. A., AILAMAKI, A., ALONSO-NANCLARES, L., ANTILLE, N., ARSEVER, S., ET AL. Reconstruction and simulation of neocortical microcircuitry. *Cell* 163, 2 (2015), 456–492.
- [44] MARTÍNEZ-CAÑADA, P., NESS, T. V., EINEVOLL, G. T., FELLIN, T., AND PANZERI, S. Computation of the electroencephalogram (eeg) from network models of point neurons. *PLOS Computational Biology* 17 (2021), e1008893.
- [45] MAZZONI, A., LINDÉN, H., CUNTZ, H., LANSNER, A., PANZERI, S., AND EINEVOLL, G. T. Computing the local field potential (lfp) from integrate-and-fire network models. *PLOS Computational Biology* 11 (2015), e1004584.
- [46] NESS, T. V., HALNES, G., NÆSS, S., PETTERSEN, K. H., AND EINEVOLL, G. T. Computing extracellular electric potentials from neuronal simulations. In *Computational Modelling of the Brain: Modelling Approaches to Cells, Circuits and Networks*. Springer, 2021, pp. 179–199.
- [47] NESS, T. V., REMME, M. W., AND EINEVOLL, G. T. h-type membrane current shapes the local field potential from populations of pyramidal neurons. *Journal of Neuroscience* 38 (2018), 6011–6024.
- [48] NESS, T. V., REMME, M. W. H., AND EINEVOLL, G. T. Active subthreshold dendritic conductances shape the local field potential. *Journal of Physiology* 594 (2016), 3809–3825.
- [49] NUNEZ, P. L., AND SRINIVASAN, R. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.
- [50] NÆSS, S., HALNES, G., HAGEN, E., HAGLER, D. J., DALE, A. M., EINEVOLL, G. T., AND NESS, T. V. Biophysically detailed forward modeling of the neural origin of eeg and meg signals. *NeuroImage* 225 (2021), 2020.07.01.181875.
- [51] OLAH, V. J., PEDERSEN, N. P., AND ROWAN, M. J. Ultrafast simulation of large-scale neocortical microcircuitry with biophysically realistic neurons. *eLife* 11 (2022), e79535.
- [52] PETTERSEN, K. H., AND EINEVOLL, G. T. Amplitude variability and extracellular low-pass filtering of neuronal spikes. *Biophysical journal* 94, 3 (2008), 784–802.
- [53] PETTERSEN, K. H., HAGEN, E., AND EINEVOLL, G. T. Estimation of population firing rates and current source densities from laminar electrode recordings. *Journal of computational neuroscience* 24 (6 2008), 291–313.
- [54] PETTERSEN, K. H., LINDÉN, H., DALE, A. M., AND EINEVOLL, G. T. Extracellular spikes and csd. *Handbook of neural activity measurement* 1 (2012), 92–135.
- [55] POTJANS, T. C., AND DIESMANN, M. The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebral cortex* 24, 3 (2014), 785–806.
- [56] RIMEHAUG, A. E., STASIK, A. J., HAGEN, E., BILLEH, Y. N., SIEGLE, J. H., DAI, K., OLSEN, S. R., KOCH, C., EINEVOLL, G. T., AND ARKHIPOV, A. Uncovering circuit mechanisms of current sinks and sources with biophysical simulations of primary visual cortex. *eLife* 12 (jul 2023), e87169.
- [57] RUDER, S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098* (2017).
- [58] SCHMIDT, M., BAKKER, R., HILGETAG, C. C., DIESMANN, M., AND VAN ALBADA, S. J. Multi-scale account of the network structure of macaque visual cortex. *Brain Structure and Function* 223, 3 (2018), 1409–1435.
- [59] SCHOMBURG, E. W., ANASTASSIOU, C. A., BUZSAKI, G., AND KOCH, C. The spiking component of oscillatory extracellular potentials in the rat hippocampus. *Journal of Neuroscience* 32 (8 2012), 11798–11811.
- [60] SKAAR, J.-E. W., STASIK, A. J., HAGEN, E., NESS, T. V., AND EINEVOLL, G. T. Estimation of neural network model parameters from local field potentials (lfps). *PLoS Computational Biology* 16 (2020), e1007725.
- [61] TRAUB, R. D., CONTRERAS, D., CUNNINGHAM, M. O., MURRAY, H., LEBEAU, F. E., ROOPUN, A., BIBBIG, A., WILENT, W. B., HIGLEY, M. J., AND WHITTINGTON, M. A. Single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts. *Journal of neurophysiology* 93, 4 (2005), 2194–2232.
- [62] ZHANG, Y., AND YANG, Q. An overview of multi-task learning. *National Science Review* 5, 1 (2018), 30–43.

7 Supporting Information

7.1 Measuring Diversity in MMoE and MMoEEx

In principle, MTL problems are expected to benefit from higher representational diversity, by which we mean that the representations provided by the different experts capture different aspects of the input data. At least in theory, the MMoEEx architecture was constructed to promote higher diversity among representations, however as mentioned before, we did not observe any clear relationship between diversity and improved performance as measured by the expert diversity metrics discussed in the Methods and Materials section. Moreover, after retraining our models we observed that none of the diversity metrics were consistent between runs, see figure 5. Also, note that all three metrics follow the same pattern within a training session. This result supports the point of view that, at least for the task at hand, a large amount of diversity is not required to find a solution.

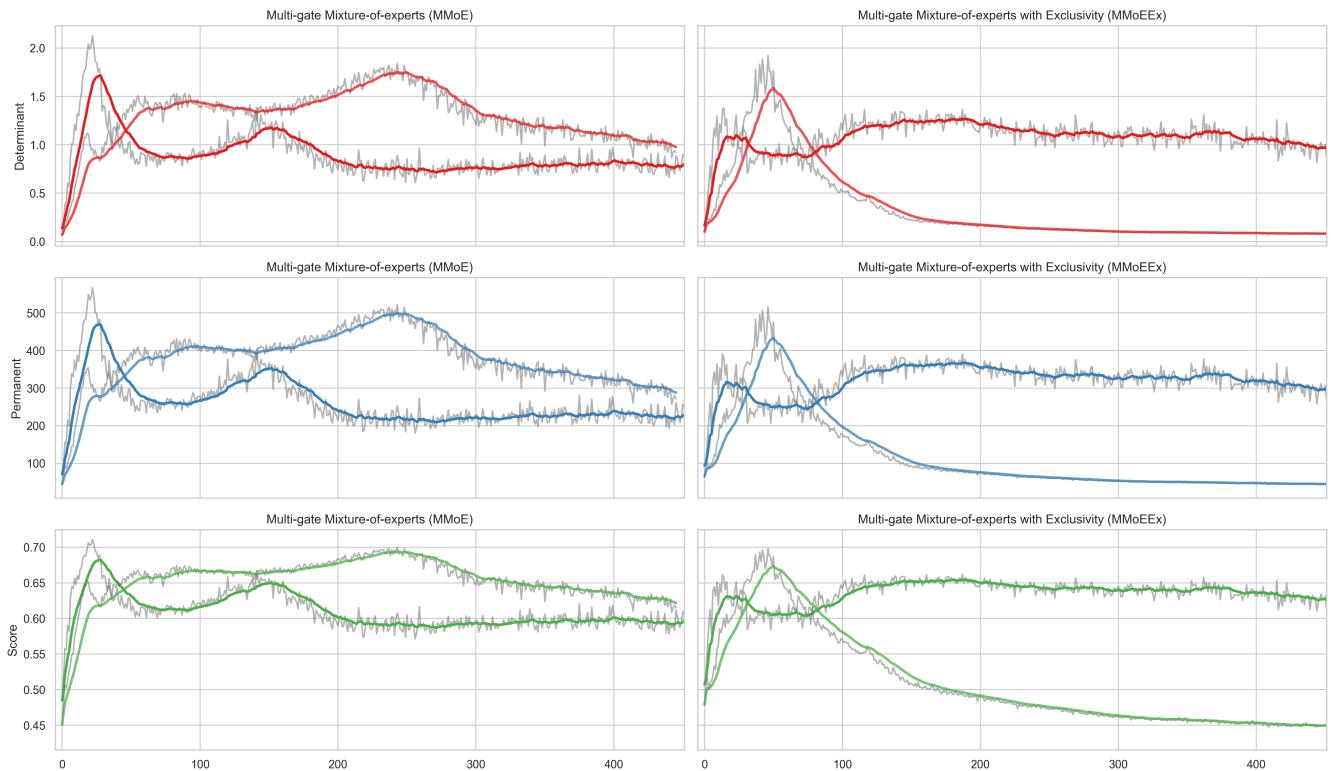


Figure 5: Smoothed time series plot of the evolution of the three diversity metrics (determinant, permanent and score) for two different training runs for both the MMoE (left) and the MMoEEx models (right).

7.2 Loss-Balanced Task Weighing for MMoE and MMoEEx

Despite the overall effectiveness of soft parameter models in achieving high average task performance, it is important to acknowledge the potential for negative transfer learning which could lead to a multi-task model underperforming on specific tasks. To address this issue, Loss-Balanced Task Weighting (LBTW) was proposed in the literature as a valuable tool for mitigating potential negative transfer effects. LBTW applies a dynamic update to the task weights in the loss function used during model training. Detailed information regarding the specific loss function and weight updates used by LBTW can be found in the Methods and Materials section. In this appendix, we present results obtained by applying LBTW to the MMoE and MMoEEx architectures, and compare them to the previously described training and generalisation performance of the MMoE and MMoEEx architectures without task-balancing.

The hard parameter model trained with LBTW achieved a minimum validation loss of 0.718, which is higher than the loss of 0.680 obtained through the standard training procedure. Because LBTW was originally designed for soft parameter sharing models only, the lack of improvement in this case was not unexpected. However, the



Figure 6: Swarm plots of the generalisation loss for each of the compartments (basal, oblique, apical) of the for each of the three LBTW-trained models: Hard Parameter Sharing (MH), Multi-gate Mixture-of-experts (MMoE), and Multi-gate Mixture-of-Experts with Exclusivity (MMoEEx).

LBTW-trained MMoE and MMoEEx models unfortunately also exhibited higher validation losses compared to the standard training procedures. More specifically, the LBTW-trained MMoE model achieved a validation loss of 0.626, while the standard training procedure resulted in a loss of 0.533. Similarly, the LBTW-trained MMoEEx model had a validation loss of 0.634, whereas the standard training procedure yielded a lower validation loss of 0.551. For a visual representation of the individual validation losses at the end of the training run, refer to Figure 1 of this appendix.

7.3 Calculation of Extracellular Potentials

The membrane currents passing through cellular membranes in the vicinity of an electrode give rise to measurable extracellular potentials. One of the downstream applications of the distilled biophysically-detailed models we have presented in this paper is predicting such extracellular potentials, as was shown in Figure 1 of the main text. In this appendix we discuss the calculation of the extracellular potential based on the prediction of our multi-task learned models in more detail. To calculate them we rely on volume conductor theory [49], which dictates that assuming an infinite homogeneous and isotropic extracellular medium, the potential ϕ resulting from a multi-compartment neuron model is given by,

$$V_e(\mathbf{r}_e, t) = \frac{1}{4\pi\sigma} \sum_{n=0}^N \frac{I_n(t)}{|\mathbf{r}_e - \mathbf{r}_n|}. \quad (4)$$

In this equation [46], $I_n(t)$ denotes the transmembrane currents in a compartment positioned at \mathbf{r}_n , and \mathbf{r}_e indicates the the position of the electrode. The sum runs over all N compartments of the biophysically-detailed neuron model, and σ is the extracellular conductivity which is experimentally determined to be around $0.3Sm^{-1}$ for cortical grey matter [12]. The multi-task learning models predicted membrane potentials, and not membrane currents, but given full knowledge of both the membrane potentials and the cell model itself, the corresponding membrane currents could be directly inferred (see the LFPy documentation regarding the method "get_transformation_matrix_vmem_to_imem"). Using the inferred membrane currents and the software package LFPy [17], we calculated the resulting extracellular potentials.

Extracellular potentials are a compound result of the contributions of different neuron compartments, and due to the fact that we do not explicitly train our models to predict them, accurate prediction of extracellular potentials is not necessarily a given. While on average our predictions are about as accurate as the experimental variance of the Hay model, some compartments or specific patterns of neural activity may be more or less accurately predicted. For instance as shown in Figure 7 of this appendix, the predictions of membrane voltages in apical and oblique compartments match the ground truth closely, whereas predictions for basal compartments are generally less accurate. As a result of this spatial variance, extracellular potentials near the soma are less accurately predicted in this specific case. Therefore, the use of these distilled biophysically-detailed neuron models in downstream prediction of extracellular potential should be handled with care.

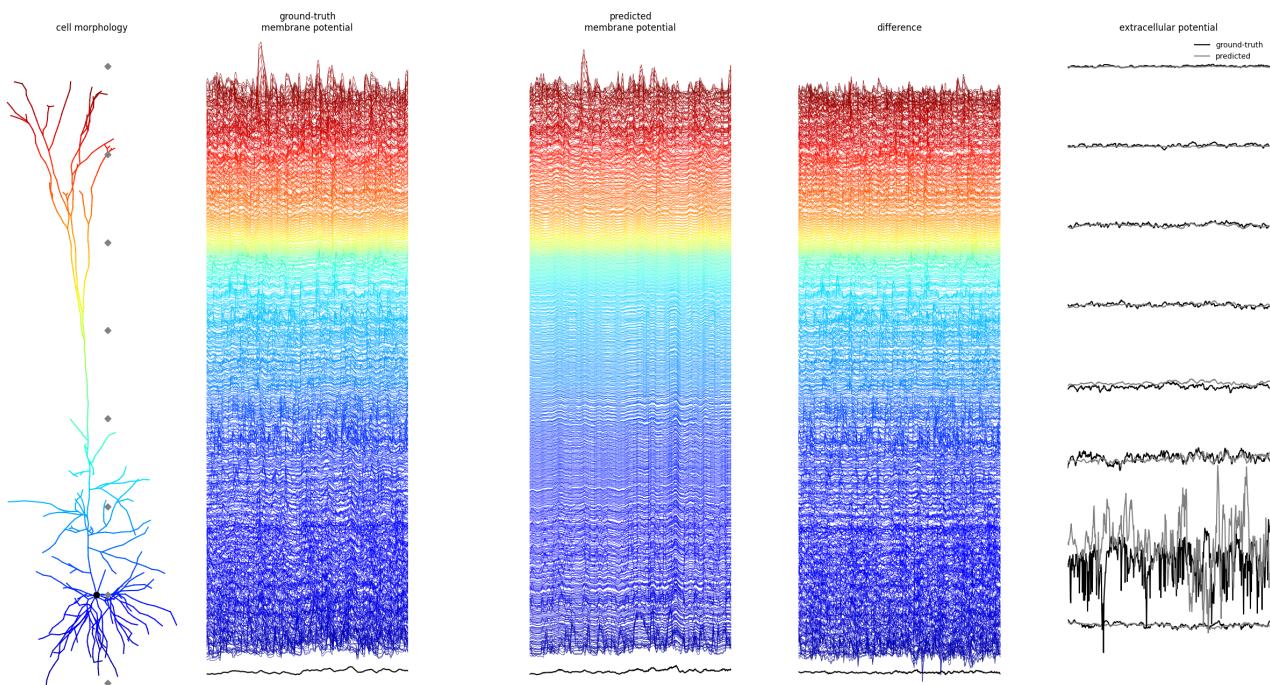
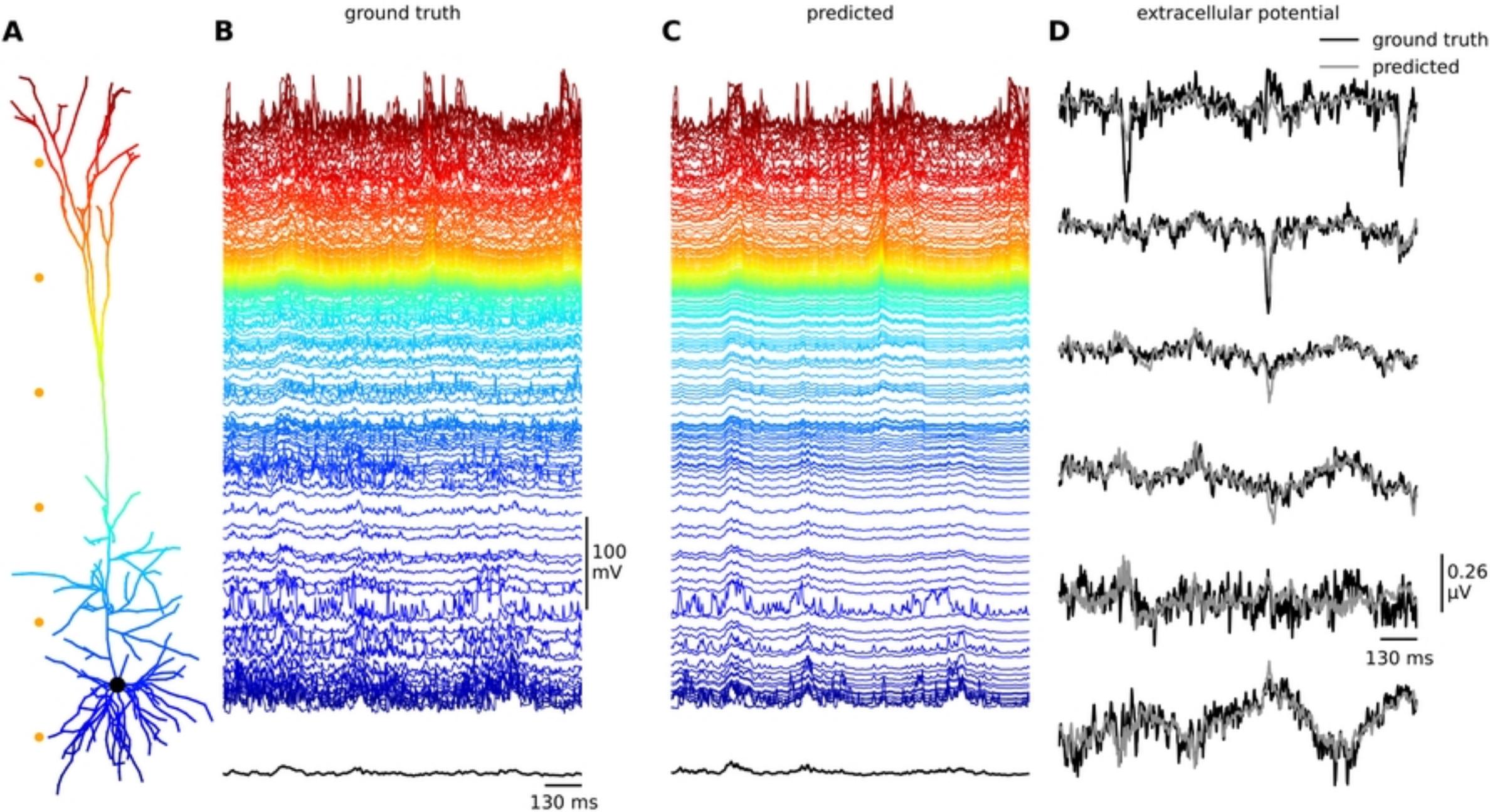
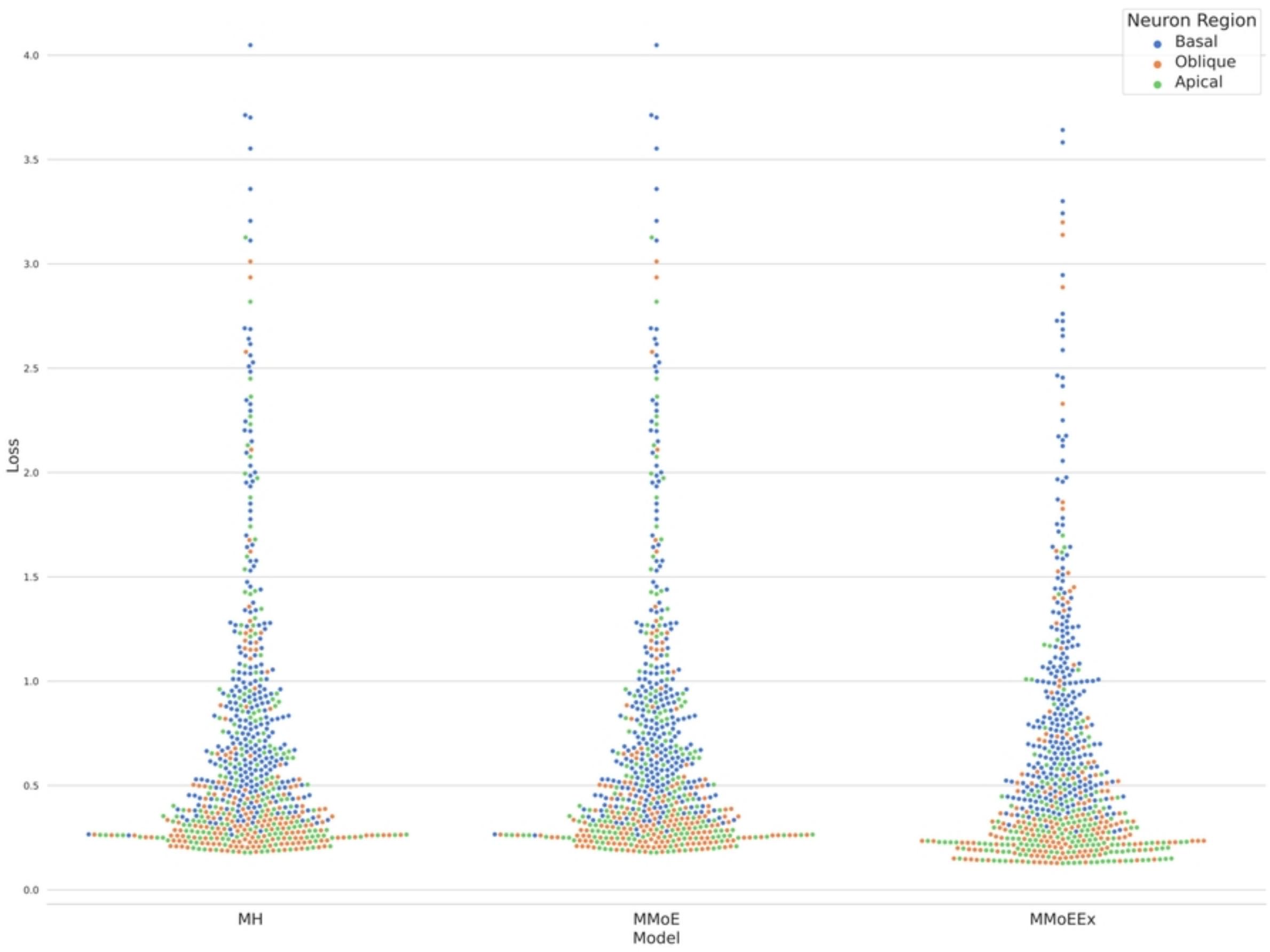
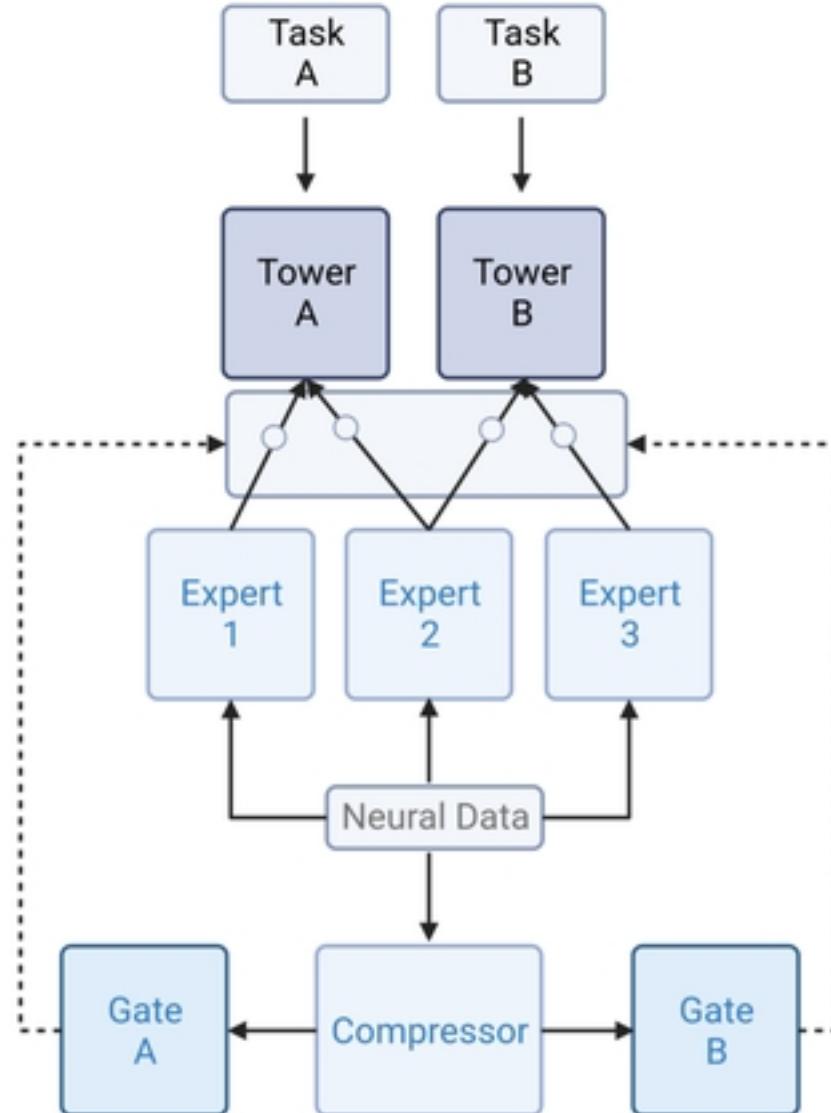
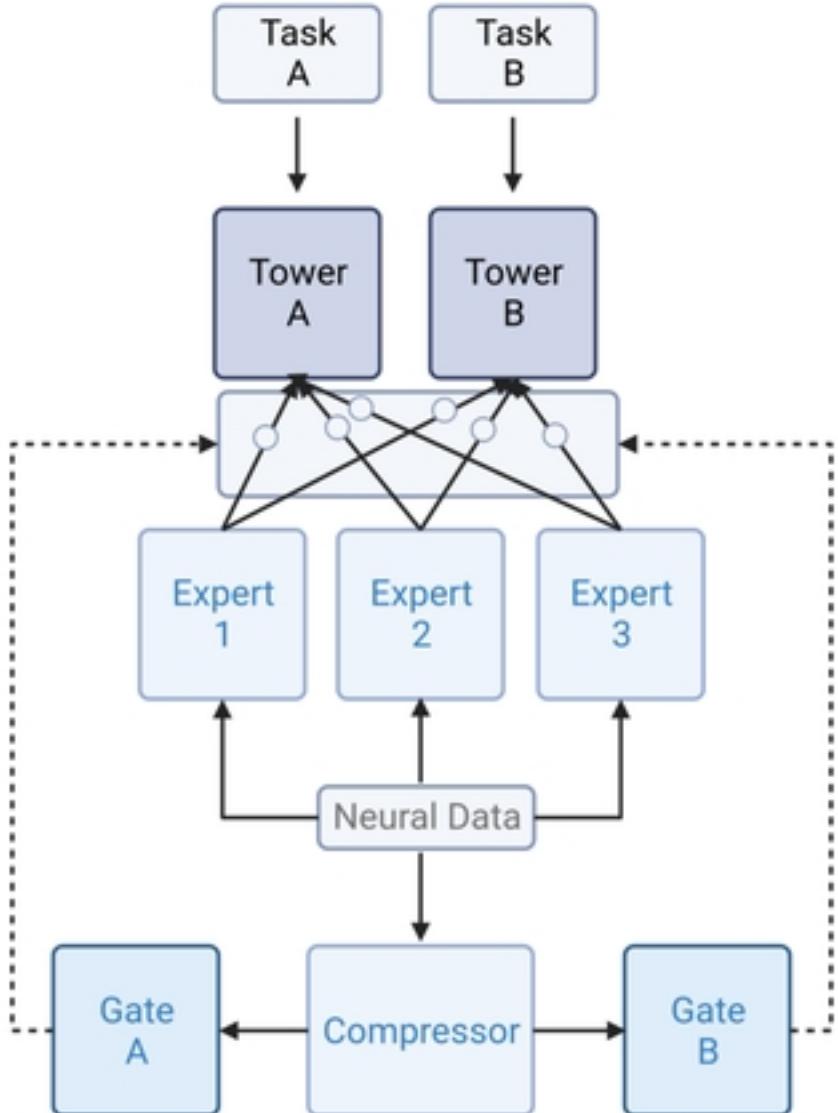
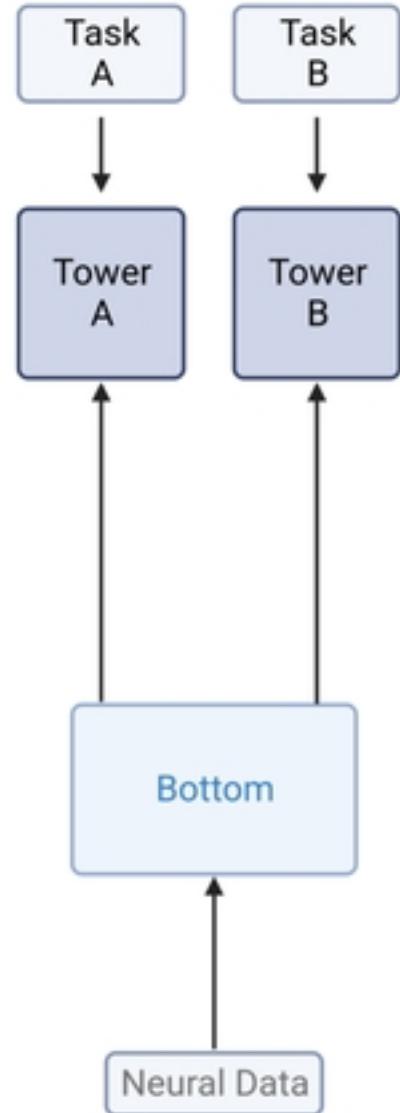


Figure 7: From left to right. Illustration of a biophysically-detailed model of a multi-compartment cortical layer V pyramidal cell model. Membrane voltages as calculated by a biophysically-detailed simulations of the multi-compartment model, used as the ground truth throughout this paper. Membrane voltages as predicted by our best-performing multi-task learning architecture, one time step at a time (1 ms). Comparison between the ground truth and predicted extracellular potentials calculated at eight points representing the position of the electrodes.







MMoE Diversity Score



MMoEEx Diversity Score



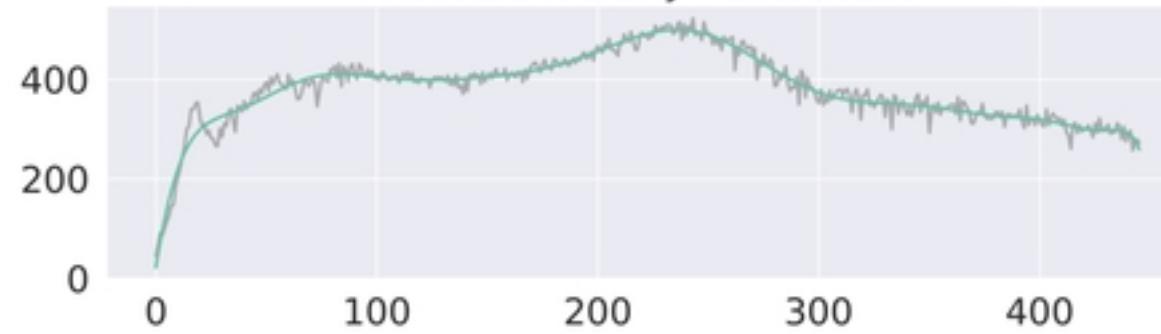
MMoE Diversity Determinant



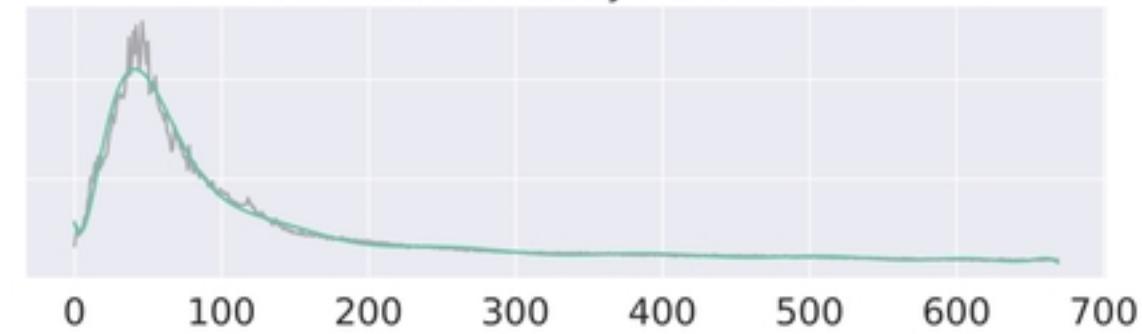
MMoEEx Diversity Determinant

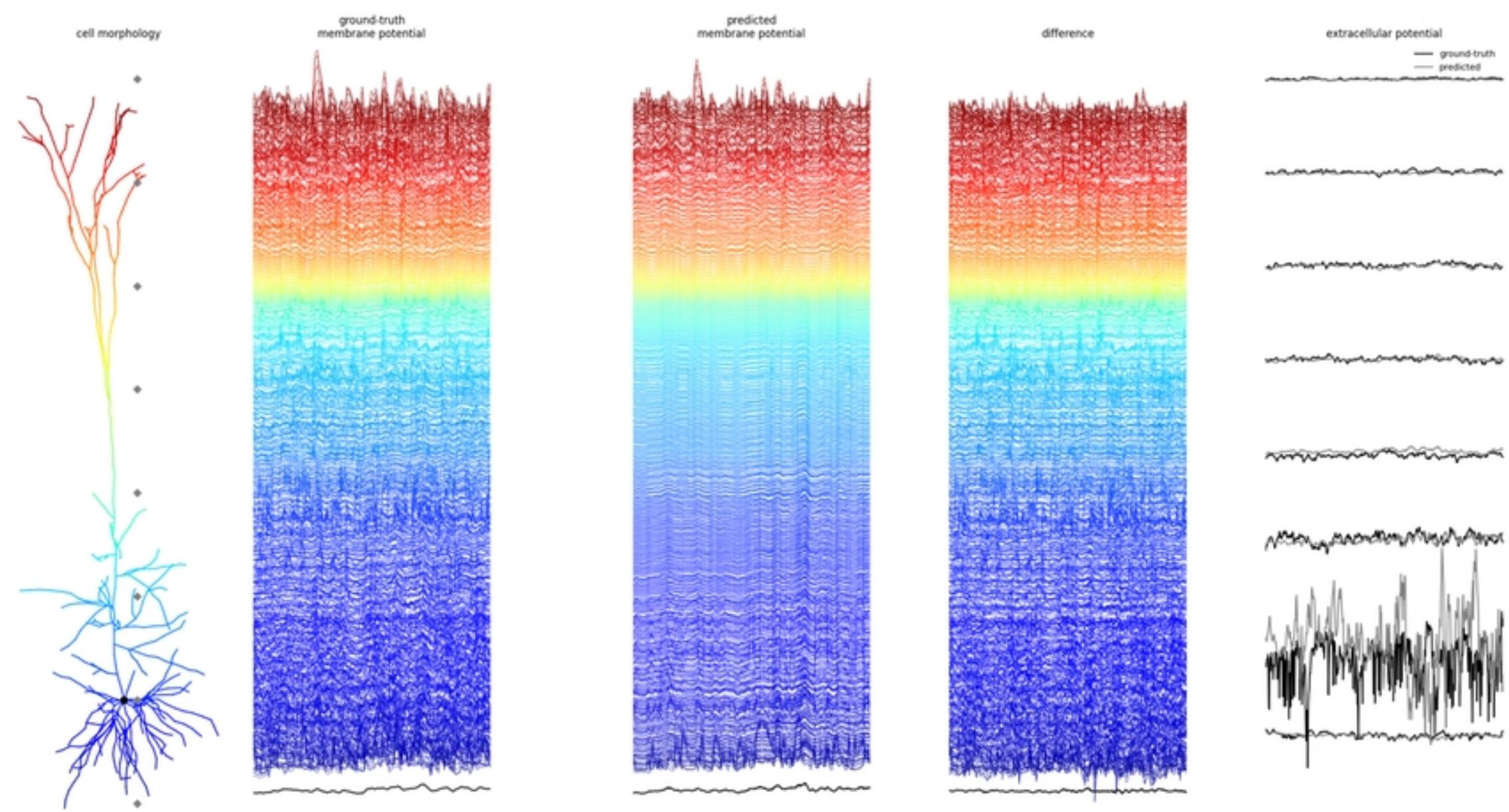


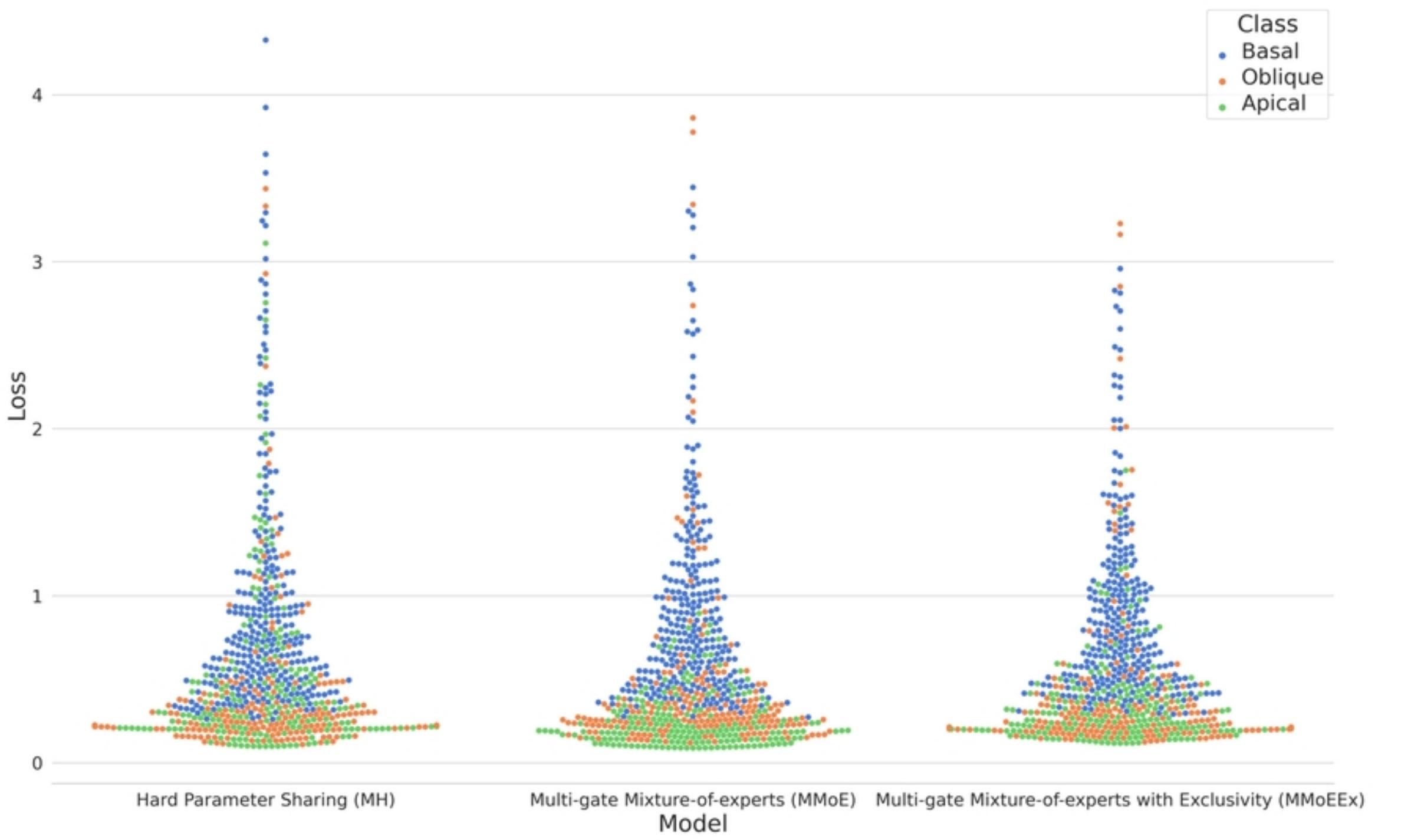
MMoE Diversity Permanent



MMoEEx Diversity Permanent







MMoE Expert 1

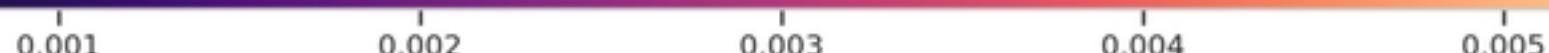
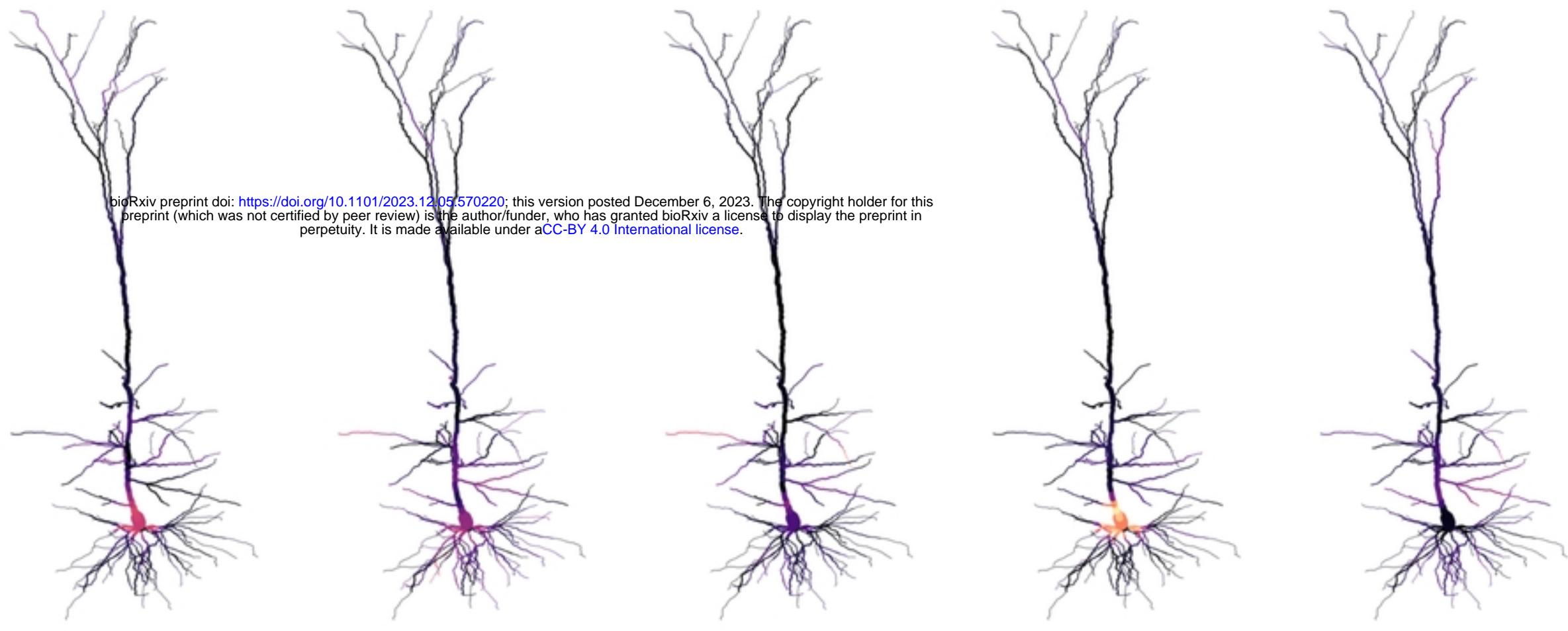
MMoE Expert 2

MMoE Expert 3

MMoE Expert 4

MMoE Expert 5

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.05.570220>; this version posted December 6, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



MMoEEEx Expert 1

MMoEEEx Expert 2

MMoEEEx Expert 3

MMoEEEx Expert 4

MMoEEEx Expert 5

