



**HAL**  
open science

# Machine learning and computational modeling approaches towards detection of pathological activity markers in the cortex

Ivan Lazarevich

► **To cite this version:**

Ivan Lazarevich. Machine learning and computational modeling approaches towards detection of pathological activity markers in the cortex. *Neurons and Cognition [q-bio.NC]*. Université Paris sciences et lettres; Université d'Etat Lobatchevski de Nijni Nograd (Russie), 2021. English. NNT: 2021UPSLE071 . tel-04099086

**HAL Id: tel-04099086**

**<https://theses.hal.science/tel-04099086>**

Submitted on 16 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à Ecole normale supérieure  
Dans le cadre d'une cotutelle avec Lobachevsky  
State University of Nizhny Novgorod

**Approches d'apprentissage automatique et de modélisation  
informatique pour la détection de marqueurs d'activité  
pathologique dans le cortex**

Machine learning and computational modeling approaches  
towards detection of pathological activity markers in the cortex

Soutenue par  
**Ivan LAZAREVICH**  
Le 23 novembre 2021

Ecole doctorale n°158  
Cerveau,  
cognition,  
comportement

Spécialité  
**Neuroscience**

Composition du jury :

Uwe, MASKOS  
Directeur de recherche, Institut Pasteur *Président*

Jorge, MEJÍAS  
Assistant Professor,  
University of Amsterdam *Rapporteur*

Alexander, HRAMOV  
Full Professor, Innopolis University *Examineur*

Alexey, OSSADTCHI  
Full Professor, National Research University Higher  
School of Economics *Rapporteur*

Victor, KAZANTSEV  
Associate Professor, Lobachevsky State  
University of Nizhny Novgorod *Co-directeur de thèse*

Boris, GUTKIN  
Directeur de recherche,  
Ecole normale supérieure *Directeur de thèse*

## RÉSUMÉ

---

Une question centrale dans le décodage de l'activité cérébrale est de savoir comment détecter des schémas d'activité anormale ou pathologique. Dans cette thèse, nous appliquons des méthodes d'apprentissage automatique et des approches de modélisation informatique pour analyser les données d'activité neuronale dans le cortex et apprendre à détecter des marqueurs d'activité anormale dans des modèles animaux de plusieurs maladies neurodégénératives courantes. Tout d'abord, nous identifions les méthodes d'apprentissage automatique qui fonctionnent bien dans les tâches de classification sur des données d'activité au niveau d'un seul neurone. Nous établissons une référence pour la classification des trains de pointes neuronales et trouvons les caractéristiques de séries chronologiques hautement prédictives de l'état du circuit neuronal dans différentes tâches et zones cérébrales. En utilisant les approches établies, nous analysons un ensemble de données d'activité neuronale dans le cortex préfrontal (PFC) chez des animaux présentant des mutations entraînant des dysfonctionnements du système de signalisation cholinergique couramment associés à des maladies telles que la schizophrénie et la maladie d'Alzheimer. Nous utilisons également la modélisation informatique de la dynamique des circuits locaux dans le PFC pour expliquer mécaniquement les origines des changements d'activité observés chez ces animaux. Enfin, nous testons les approches d'apprentissage automatique sur un ensemble de données multimodales d'activité neuronale dans un modèle animal de sclérose latérale amyotrophique (SLA) précoce. Nous montrons que pour concevoir un système capable de détecter avec précision l'activité pathologique inhérente à la SLA précoce, il faut entraîner le modèle à extraire des informations de l'interaction entre les modalités de l'activité corticale et le mouvement animal. Dans l'ensemble, nous avons fourni des informations sur la façon dont la modélisation informatique et l'apprentissage automatique pourraient fournir des outils pour détecter la pathologie à partir des enregistrements d'activité des circuits neuronaux.

## MOTS CLÉS

---

activité neurale, apprentissage automatique, maladie neurodégénérative, exploration de données, activité corticale

## ABSTRACT


---

A central question in brain activity decoding is how to detect patterns of abnormal or pathological activity. In this thesis, we apply machine learning methods and computational modelling approaches to analyze neural activity data in the cortex and to learn how to detect markers of abnormal activity in animal models of several common neurodegenerative diseases. First, we identify the machine learning methods that perform well in classification tasks on single-neuron level activity data. We establish a benchmark for neuronal spike train classification and find the time-series features that are highly predictive of the neural circuit state across different tasks and brain areas. Using the established approaches, we analyze a data set of neural activity in the prefrontal cortex (PFC) in animals with mutations leading to dysfunctions of the cholinergic signalling system commonly associated with diseases such as schizophrenia and Alzheimer's disease. We also use computational modelling of the local circuit dynamics in the PFC to mechanistically explain the origins of the activity changes observed in these animals. Finally, we test the machine learning approaches on a multimodal data set of neural activity in an animal model of early amyotrophic lateral sclerosis (ALS). We show that in order to design a system that is capable of accurately detecting the pathological activity inherent to early ALS, one has to train the model to extract information from the interaction between the modalities of cortical activity and animal movement. Overall, we have provided insights into how computational modelling and machine learning could provide tools for detecting pathology from neural circuit activity recordings.

## KEYWORDS

---

neural activity, machine learning, neurodegenerative disease, data mining, cortical activity



# Contents

<b>1</b>	<b>Introduction</b>	<b>18</b>
1.1	Thesis summary . . . . .	18
1.2	Background . . . . .	20
<b>2</b>	<b>Neural activity classification with machine learning models trained on inter-spike interval time-series data</b>	<b>23</b>
2.1	Introduction . . . . .	25
2.2	Materials and methods . . . . .	28
2.2.1	Overview of time series classification methods . . . . .	28
2.2.2	The proposed spike train classification benchmark . . . . .	33
2.2.3	Cross-validation scheme and data preprocessing . . . . .	35
2.3	Results . . . . .	38
2.3.1	Visual stimulus type classification from retinal spike trains . . . . .	38
2.3.2	Nearest-neighbor models for spike train classification . . . . .	38
2.3.3	Hand-crafted feature extraction for time-series classification . . . . .	40
2.3.4	Prediction accuracy from quantized spike trains . . . . .	45
2.3.5	Unsupervised spike train temporal structure recognition . . . . .	46
2.3.6	The set of predictive spike train features . . . . .	48
2.4	Discussion . . . . .	53
2.5	Conclusion . . . . .	53
<b>3</b>	<b>Local circuit modeling of schizophrenia and Alzheimer’s disease related cholinergic system pathologies in the PFC</b>	<b>55</b>
3.1	Introduction . . . . .	57
3.2	Methods . . . . .	59
3.2.1	Neural population rate model . . . . .	59
3.2.2	Model parameter search procedure . . . . .	62
3.2.3	Modeling nAChRs . . . . .	64
3.3	Results . . . . .	65
3.3.1	Summary of the experimental approach and results . . . . .	65
3.3.2	Modelling formalism and strategy . . . . .	68
3.3.3	Bistable layer II/ III local PFC circuit firing rate dynamics replicate ultraslow fluctuations in WT mice . . . . .	72
3.3.4	Heuristic analysis of impact of inhibitory population activity variation on network state stability. . . . .	75

3.3.5	Impact of the nAChR genetic manipulations on the temporal structure of the ultraslow activity fluctuations . . . . .	76
3.3.6	Layer II/III circuit model accounts for the VIP and SOM neuron firing rate changes under schizophrenia-associated $\alpha 5$ pathology. . . . .	80
3.3.7	Nicotine re-normalizes $\alpha 5$ SNP PFC network activity through desensitization and upregulation of SOM $\beta 2$ nAChRs . . . . .	81
3.3.8	Population model of the PFC fitted to data in APP-expressing mice predicts PYR hyper-activity reduction by galantamine . . . . .	86
3.4	Discussion . . . . .	91
3.4.1	Summary of the results. . . . .	91
3.4.2	Predictions of the model. . . . .	92
3.4.3	Limitations of the model and future direction. . . . .	94
3.4.4	Implications for nicotine withdrawal in schizophrenia . . . . .	95
3.5	Conclusions . . . . .	96
<b>4</b>	<b>Discovery of the cholinergic system pathologies in the PFC from cortical activity using machine learning</b>	<b>97</b>
4.1	Analysing prefrontal cortex activity in mice with nAChR dysfunctions	98
4.1.1	Two-photon imaging data pre-processing . . . . .	99
4.1.2	PFC activity structure revealed with time-series features . . . . .	100
4.1.3	Detecting nAChR dysfunction from single-neuron activity data with machine learning . . . . .	104
4.2	Detection of $\alpha 5$ subunit containing nAChR dysfunction from interneuronal activity and the effect of nicotine application . . . . .	113
4.3	Detecting the effect of beta-amyloid expression on the activity of PFC neurons . . . . .	115
4.4	Conclusions . . . . .	117
<b>5</b>	<b>Deep learning models detect markers of early amyotrophic lateral sclerosis from neural activity and movement data</b>	<b>119</b>
5.1	Introduction . . . . .	120
5.2	Neural activity data in the model of the early amyotrophic lateral sclerosis . . . . .	121
5.3	Model validation scheme . . . . .	122
5.4	Visualizing data set structure by hand-crafted time-series feature encoding . . . . .	123
5.5	Pathological activity detection with deep convolutional neural networks	124
5.6	Choice of DNN architecture: an ablation study . . . . .	130
5.7	Prediction accuracy as a function of input data modality and animal state . . . . .	134
5.8	Conclusions . . . . .	140
<b>6</b>	<b>Discussion</b>	<b>143</b>
	<b>Bibliography</b>	<b>145</b>

# List of Figures

2-1	(A) Examples of spiking activity recordings in the CRCNS fcx-1 dataset in the WAKE state. Left: spike train raster of a random subset of excitatory cells (red) and inhibitory cells (blue). Right: examples of ISI series produced from spike train chunks of inhibitory/excitatory cells in the fcx-1 dataset. (B) Interspike interval value distribution histograms generated from the aggregated spike trains of retinal ganglion cells in response to a "white noise checkerboard" visual stimulus (blue) and a "randomly moving bar" stimulus (red). (C) Interspike interval value distribution histograms generated from the aggregated PFC spike trains (fcx-1 dataset) corresponding to the WAKE (blue) or SLEEP (red) state of the rat. . . . .	29
2-2	Spike train classification accuracy values for the retinal neuron activity dataset for nearest-neighbor models with different distance metrics . The task is defined as binary classification of the stimulus type ("white noise checkerboard" or "randomly moving bar"), with the test set balanced in class distribution (that is, accuracy=0.5 corresponds to chance level). . . . .	37
2-3	(Caption next page.) . . . . .	39
2-3	(Previous page.) Spike train classification metric values for the retinal neuron activity dataset on a range of models. The task is defined as binary classification of the stimulus type ("white noise checkerboard" or "randomly moving bar"), with the test set balanced in class distribution (that is, accuracy=0.5 corresponds to chance level). Accuracy is shown on the left and AUC-ROC on the right for the same set of models and train/test dataset splits. The "simple" model tag corresponds to spike trains encoded with 6 basic distribution statistics (representing a simple baseline), the "raw ISI" tag implies that the model has been directly trained on ISI time-series data without encoding. The "tsfresh" tag corresponds to encoding with the full set of time-series features. "ISIE" stands for interspike-interval encoding of the spike train, "SCe" stands for spike-count encoding. "ISIE + SPe" means that feature vectors corresponding to both types of encoding are concatenated. . . . .	40

2-4	Spike train classification metric values for the WAKE/SLEEP state prediction dataset (left) and VIP/SST interneuron type prediction dataset (right) on a range of models. The task is defined as binary classification, with the test set balanced in class distribution for both datasets (that is, accuracy=0.5 corresponds to chance level). Accuracy is shown on the top panes and AUC-ROC on the bottom panes for the same set of models and train/test dataset splits. The "simple" model tag corresponds to spike trains encoded with 6 basic distribution statistics (representing a simple baseline), the "raw ISI" tag implies that the model has been directly trained on ISI time-series data without encoding. The "tsfresh" tag corresponds to encoding with the full set of time-series features. "ISIE" stands for interspike-interval encoding of the spike train, "SCe" stands for spike-count encoding. "ISIE + SCe" means that feature vectors corresponding to both types of encoding are concatenated. . . . .	41
2-5	Spike train feature embeddings for WAKE (points marked red) vs. SLEEP (points marked blue) activity states of the neural circuit. Two-dimensional embeddings of the (20-dimensional) selected- <i>tsfresh</i> -feature space using (A) unsupervised UMAP and (B) supervised UMAP embedding algorithms for spike trains corresponding to WAKE vs. SLEEP activity states. . . . .	45
2-6	Classification accuracy for the fcx-1 WAKE vs. SLEEP task in the case of multiple randomly sampled same-class spike train chunks per prediction (with prediction done via majority voting). The model trained in these trials is a random forest classifier on the full set of <i>tsfresh</i> features. The boxplots reflect the median accuracy and the variance between different train/test splits as done in the main text for the fcx-1 data set. The red crosses correspond to the theoretical estimate under the assumption of independently sampled spike train chunks. . . . .	51
2-7	Boxplots of <i>tsfresh</i> -extracted feature distributions for features with high discriminative power as detected by the trained decision tree ensemble classifiers in the retinal stimulus type prediction task. Two-sided Mann-Whitney-Wilcoxon test with Bonferroni correction is performed to assess statistical significance; **** denotes $p < 1e - 4$ . . . .	52
3-1	(Caption next page.) . . . . .	66



3-1	(Previous page.) Bistable firing rate dynamics of interconnected neural populations replicates ultra-slow fluctuations recorded in the PFC of WT mice. (A1) Two-photon image of GCaMP6f expressing neurons, from Koukouli et al. [2017]. (Scale bar: 20 $\mu m$ ). (A2) Inferred events of a population of simultaneously recorded cells in a WT mouse, obtained by deconvolving the spontaneous Ca <sup>2+</sup> transients. 80% of the recorded cells are PYR neurons Koukouli et al. [2017]. (A3) Time varying population mean activity of the neurons shown in A2. The dashed red line delineates the threshold between high and low activity states (H-states and L-states, respectively). Red periods correspond to H-states and blue periods to L-states. See Koukouli et al. [2016a] for more info on the methods. (A4) H-state and L-state durations recorded in the different neuron types. H-state durations are similar between neuron types (mean of 3 seconds, no statistical differences), as well as L-state durations (mean of 20 seconds, no statistical differences). (B1) Schematic of the studied circuitry. (B2) A simulated rastergram of neuronal activity, generated for illustration purposes from the population rate model, using a Poisson process with $\lambda(t)=mean\_population\_rate(t)$ . (B3) Time varying mean population activity of pyramidal neurons, computed from the network model. We use the same method as Koukouli et al. [2016a] to delineate H-states (in red) and L-states (in blue). (B4) The model reproduces the similar H-state and L-state durations across neural types. Distribution of a total of 500 state durations collected are plotted. . . . .	67
3-2	(Caption next page.) . . . . .	71
3-3	(Previous page.) The distribution of fitting (WT) and validation (KO) errors along with some of the parameter values for the set of candidate models found during parameter optimization. The main model candidate chosen in the paper is denoted with the red color. (A) Scatter plot for a range of candidate model parameter sets whereby every parameter set is represented by its WT activity statistics fitting error (MAPE of WT high-low activity statistics) and the KO activity prediction error (MAPE of PYR firing rates in $\alpha 5$ , $\alpha 7$ and $\beta 2$ knockout states, see Fig. 4A1), normalized to the error values of the selected parameter set (denoted with red color), (B) scatter plot of $\omega_{ee}$ , $\omega_{pe}$ parameter values for the candidate parameter sets normalized to the selected values, (C) same for $I_{ext-s}$ , $I_{ext-v}$ parameters, (D) same for $I_{ext-e}$ , $J_{adapt}$ parameters, (E) distributions of the WT fitting errors and nicotine treatment activity level prediction errors (error in predicted change in WT and $\alpha 5$ SNP PYR firing rates after nicotine treatment), along with the $\beta 2$ nAChR enhancement factors required for the candidate models to reproduce the normalization of $\alpha 5$ SNP activity to WT levels under nicotine treatment. . . . .	72

3-4	Evolution of WT activity statistics fitting error (normalized to the error value at $k_d=0.8$ ) for different values of the divisive-to-subtractive inhibition ratio $k_d$ while the other parameter values in the model are fixed (corresponding to the main chosen parameter set). The default value of $k_d$ is 0.8 for the fitted model. . . . .	73
3-5	Evolution of the normalized WT statistics fitting and KO activity level prediction errors as a function of the SOM-PV synaptic connection strength added to the model with the main chosen parameter set. Note that model outputs are not significantly affected in terms of the fitting and validation errors for low values of the SOM-PV connection strength. The maximal value of the synaptic strength parameters that we considered in the model was equal to about 50 (in arbitrary units). . . . .	74
3-6	(Caption next page.) . . . . .	79
3-6	(Previous page.) Effects of changing external inputs to inhibitory populations on network state stability in the fully connected network (A1) H-state (red line) and L-state (blue line) PYR activities as a function of the external input current to VIP interneurons population. The dashed green line shows the selected parameter value to reproduce WT mice neural dynamics. KO of nAChRs is associated with a decrease of external currents (black arrow). (A2) H-state (red line) and L-state (blue line) durations as a function of the external input current to PV interneurons population. The shaded areas delineate $\pm$ sem. The dashed green line shows the selected parameter value to reproduce WT mice neural dynamics. (B1) Same as (A1), but for the external input current to SOM INs population. (B2) Same as (A2), but for the external input current to SOM INs population. (C1) Same as (A1), but for the external input current to VIP INs population. (C2) Same as (A2), but for the external input current to VIP INs population. . . . .	80
3-7	(Caption next page.) . . . . .	82

3-7	(Previous page.) Accounting for the nAChR KO and mutation impact on ultraslow fluctuations (A1) Mean of H-state durations, for WT and mutant mice, modified from Fig 3C in Koukouli et al. [2016a] for $\alpha 7$ KO and $\beta 2$ KO mice. For $\alpha 5$ KO and $\alpha 5$ SNP mice, we use the same method as in Koukouli et al. [2016a]. All mutant mice distributions are significantly different from WT (Kruskal-Wallis, $P < 0.001$ ), except for $\alpha 5$ SNP mice. The error bars are $\pm$ sem. (A2) Mean of L-state durations, for WT and mutant mice, modified from Fig. 3B in Koukouli et al. [2016a] for $\alpha 7$ KO and $\beta 2$ KO mice. $\beta 2$ KO mice L-state durations are significantly lower compared to WT (Kruskal-Wallis, $P < 0.001$ ). The error bars are $\pm$ sem. (A3) Mean % of populations with H-states and L-states transitions. The error bars are $\pm$ sem. The circle shows the proportions computed for single mice. $\beta 2$ KO mice exhibit significantly lower % of populations with H-states and L-states transitions (ANOVA, $P < 0.05$ ), compared to WT mice. Modified from Fig. 3D in Koukouli et al. [2016a] for $\alpha 7$ KO and $\beta 2$ KO mice. (B1-2-3) Same as (A1-2-3), computed from simulations. (C) H-states (red bars) and L-states (blue bars) activity levels from simulations (filled bars) and experiments (empty bars). Error bars are $\pm$ sem. Experimental data described in Koukouli et al. [2016a]. A total of 200 simulation repetitions with 500 state transitions each were carried out to produce the simulated data. . . . .	83
3-8	(Caption next page.) . . . . .	85
3-8	(Previous page.) Reproduction of the effects of KO and mutations of nAChRs on neural firing rates (A) Boxplots of PYR neurons firing rates for WT and mutant mice, computed from simulations. All distributions are significantly different (Kruskal-Wallis, $P < 0.001$ ). (B) Boxplots of VIP interneurons baseline activities for WT and CRISPR mice. CRISPR mice exhibit lower neural activities compared to WT mice (Kruskal-Wallis, $P < 0.001$ ). (C) Boxplots of SOM interneurons baseline activities for WT and CRISPR simulated mice. CRISPR mice exhibit higher neural activities compared to WT mice (Kruskal-Wallis, $P < 0.001$ ). (D) Boxplots of PV interneurons baseline activities for WT and mutant mice affected by the CRISPR technology for the deletion of the $\alpha 5$ subunits, according to simulations. CRISPR mice exhibit similar levels of neural activity compared to WT mice. A total of 200 simulation repetitions with 500 state transitions each were carried out to produce the simulated data. . . . .	86
3-9	Summary of activity levels in different groups of APP-expressing mice vs. control groups. (A) Representative recordings of spontaneous $\text{Ca}^{2+}$ transients in WT, WT APP, $\alpha 7$ KO, $\alpha 7$ KO APP, $\beta 2$ KO, $\beta 2$ KO APP, $\alpha 5$ KO and $\alpha 5$ KO APP mice. The detected calcium transients are indicated in red. (B) Median frequency of spontaneous $\text{Ca}^{2+}$ transients/min of WT, WT APP, $\alpha 7$ KO, $\alpha 7$ KO APP, $\beta 2$ KO, $\beta 2$ KO APP, $\alpha 5$ KO and $\alpha 5$ KO APP mice. (C) Same, but for mean transient duration in seconds. . . . .	87

3-10	Summary of interneuron activity levels in different groups of APP-expressing mice. (Left column) Representative recordings of spontaneous Ca <sup>2+</sup> transients in different interneuron populations in control and APP mice. (Middle left column) Median frequency of spontaneous Ca <sup>2+</sup> transients/min in different interneuron populations in control and APP mice. (Middle right column) Cumulative distribution of firing frequency (transient/min) in different interneuron populations in control and APP mice. (Right column) Median transient durations in different interneuron populations in control and APP mice.	88
3-11	Galantamine restores PYR neuron hyperactivity early in AD. Top: Representative recordings of spontaneous Ca <sup>2+</sup> transients in WT APP vehicle and WT APP galantamine treated mice. Bottom left: Median frequency of spontaneous Ca <sup>2+</sup> transients/min of WT APP vehicle (3.969±0.11 transients/min; n= 4 mice) and WT APP galantamine treated mice (3.22±0.09 transients/min; n= 4 mice). p=0.0286, Mann-Whitney test. Bottom center: Cumulative distribution of firing frequency (transients/min) of WT APP vehicle and WT APP galantamine treated mice. Bottom right: Median transient durations of WT APP vehicle (5.078±0.27 secs) and WT APP galantamine treated mice (5.269±0.18 secs). p=0.6857, Mann-Whitney test.	89
3-12	Median simulated activity levels for the fitted population model in pyramidal neurons and interneurons corresponding to different animal groups with nAChR knock-outs and APP expression.	90
3-13	Predictions on activity level change in the model fitted on APP data upon a block of the $\beta$ 2-containing nAChRs or enhanced activity of the $\alpha$ 5-containing nAChRs.	90
3-14	(Caption next page.)	93

3-14	(Previous page.) Desensitization and upregulation of $\beta 2$ nAChRs normalizes $\alpha 5$ SNP mice network activity to WT levels after chronic nicotine application (A1) Distribution of PYR neurons firing rates for WT mice, computed from simulations of nicotine effects on $\alpha 7$ , $\beta 2$ , and $\alpha 5$ nAChRs. All distributions are significantly different (Kruskal-Wallis, $P < 0.001$ ). (A2) Distribution of PYR neurons firing rates for WT and $\alpha 5$ SNP mice, computed from simulations of nicotine effects on $\alpha 7$ , $\beta 2$ , and $\alpha 5$ nAChRs in $\alpha 5$ SNP mice. All distributions are significantly different (Kruskal-Wallis, $P < 0.001$ ). (B) Distribution of PYR neurons firing rates for WT and $\alpha 5$ SNP mice, control and treated with 2 days of chronic nicotine application, obtained from simulations. All distributions are significantly different (Kruskal-Wallis, $P < 0.001$ ). (C) Distribution of PYR neurons firing rates for WT and $\alpha 5$ SNP simulated mice, control and treated with 7 days of chronic nicotine application. All distributions are significantly different (Kruskal-Wallis, $P < 0.001$ ). The model predicts an upregulation of $\beta 2$ nAChRs. (D) Distribution of PYR neurons firing rates for WT and $\alpha 5$ SNP mice, control and after nicotine removal following 7 days of chronic nicotine administration, predicted from simulations. All distributions are significantly different (Kruskal-Wallis, $P < 0.001$ ). A total of 200 simulation repetitions with 500 state transitions each were carried out to produce the simulated data. . . . .	94
4-1	(Caption next page.) . . . . .	101
4-1	Top: activity traces obtained by two-photon imaging of GCaMP6f expressing neurons (data from Koukouli et al. [2017]), middle: a set of neural activity traces from the same data set with pre-processing applied to the extracted signals, bottom: value distribution histograms obtained from the pre-processed time-series for two separate groups of animals – the wild-type control group and the group with the $\alpha 5$ -containing nAChR knockout mutation (aggregated over neurons over animals). . . . .	102
4-2	Two-dimensional tSNE projection of the top-25 discriminative <i>tsfresh</i> features for recorded neuronal activity traces in the imaging experiments for normal and nAChR knockout states. Each point in the scatter plot corresponds to a trace from a single neuron; activity states are color-coded. . . . .	103
4-3	Two-dimensional projection by a VAE model trained on vectors of the top-25 discriminative <i>tsfresh</i> features for recorded neuronal activity traces in the imaging experiments for normal and nAChR knockout states. Each point in the scatter plot corresponds to a trace from a single neuron; activity states are color-coded. . . . .	103
4-4	Graph representation of the Wasserstein distance matrix between the distributions of the first tSNE mapping component in different knockout states. Note the proximity of $\alpha 7$ KO and $\beta 2$ KO states. . . . .	104

4-5	Test set accuracy and AUC-ROC score distributions in the $\alpha 5$ nAChR KO detection task depending on the feature representation of the time-series and its pre-processing. Distributions are computed over different train/test splits of the data set and undersampling is performed to keep the class balance in the training and testing data sets.	106
4-6	Two-dimensional PCA (top) and t-SNE (bottom) embeddings of the WT vs. $\alpha 5$ KO neuronal signals data set encoded with top-30 most important <i>tsfresh</i> features. Animal group is color-coded.	108
4-7	Feature importance scores extracted from a random forest classifier trained on the <i>tsfresh</i> embeddings in the WT vs. $\alpha 5$ KO animal classification task.	109
4-8	Confusion matrix (left panel) and ROC curve (right panel) of a random forest classifier trained on <i>tsfresh</i> embeddings in the WT vs. $\alpha 5$ KO animal classification task.	110
4-9	Training results from the shapelet learning algorithm applied to the WT vs. $\alpha 5$ KO single-neuron activity classification task. Bottom: evolution of cross-entropy loss value during shapelet learning. Top left: Learned shapelet waveforms after the end of the training. Top right: FFT spectra of the shapelet waveforms. Note the peaks in the 0.75-1.25 Hz frequency band (the delta band) in many shapelets.	110
4-10	Classification accuracy (top panel) and AUC-ROC scores (bottom panel) in a WT vs. $\alpha 5$ KO neural activity classification task with Catch22 + Random Forest classification models trained on band-pass filtered signals depending on the frequency range in the band.	112
4-11	Accuracy scores of a random forest classifier model trained on <i>tsfresh</i> -encoded data in the WT vs. $\alpha 5$ KO classification task when predictions are made for several samples at a time and averaged depending on the number of samples. Samples are independently taken from the training set.	114
4-12	Accuracy (left panel) and AUC-ROC (right panel) scores in the WT vs. $\alpha 5$ KO animal group classification task from the imaging activity recordings of PV and SOM interneuron subtypes. Note significantly better classification quality in the case of PV recordings, implying that PV interneurons are affected the most by the knock-out.	114
4-13	Accuracy (left panel) and AUC-ROC (right panel) scores in the WT vs. $\alpha 5$ SNP animal group classification task under control conditions and after nicotine application in the SNP group. Note the poor classification quality in the case of nicotine application, implying significant activity restoration in $\alpha 5$ SNP animals subject to nicotine application.	115
4-14	Accuracy score distribution over different train/test split in the task of classifying wild-type animals from the ones with $A\beta$ expression depending on the neuron type in the PFC. Note the low classification quality obtained with VIP activity data recordings suggesting the significant invariance of the VIP population to $A\beta$ expression.	116

4-15	Accuracy score distribution over different train/test splits in the task of classifying control nAChR knock-out animal groups from the knock-out groups expressing $A\beta$ . . . . .	117
4-16	AUC-ROC score distribution over different train/test splits in the task of classifying control $A\beta$ -expressing animal groups vs. $A\beta$ animals with vehicle and galantamine injections. . . . .	118
5-1	Top: Scatter plot of a two-dimensional PCA embedding of the motor cortex activity time-series encoded by top-30 tsfresh features (with features sorted by importance scores determined from a random forest classifier trained on the dataset). The class labels are color-coded with red points corresponding samples from control animals and blue points to samples from mutant ones. Bottom: similar scatter plot of the top-30 tsfresh features of neural activity time-series embedded with a t-SNE algorithm. . . . .	125
5-2	Top: Scatter plot of a two-dimensional PCA embedding of the animal speed signal time-series encoded by top-30 tsfresh features (with features sorted by importance scores determined from a random forest classifier trained on the dataset). The class labels are color-coded with red points corresponding samples from control animals and blue points to samples from mutant ones. Bottom: similar scatter plot of the top-30 tsfresh features of the animal speed signal time-series embedded with a t-SNE algorithm. . . . .	126
5-3	Top: Scatter plot of a two-dimensional PCA embedding of the top-30 feature vectors obtained by combining tsfresh features from both neural activity and speed signal time series (with features sorted by importance scores determined from a random forest classifier trained on the dataset). The class labels are color-coded with red points corresponding samples from control animals and blue points to samples from mutant ones. Bottom: similar scatter plot of the top-30 tsfresh features from both neural activity and speed signal time series embedded with a t-SNE algorithm. . . . .	127
5-4	Accuracy scores (average predicted labels) per animal over training trials of a Catch22 time-series classifier trained on (top) neural activity, (middle) movement data and (bottom) neural activity + movement data using the leave-one-out cross-validation scheme. . . . .	128
5-5	Top: Accuracy scores (average predicted labels) per animal over training trials of an FCN model trained on neural activity + movement data using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group). Thresholding the average labels with a value of 0.5 to get the animal-wide predictions leads to a single false positive error (animal ID 0) and two false negative errors (animal IDs 9 and 12). Bottom: same for per-sample average mutant class probabilities for each animal. . . . .	129

5-6	The dynamics of test set cross-entropy loss (top panels) and test set accuracy (bottom panels) during neural net training across different training trials for different animals. Animals 0 and 12 correspond to misclassified animals from WT and MUT groups, respectively, animals 2 and 13 are correctly classified animals from the two respective groups. Note that generally neural net training converges to a similar accuracy values regardless of random initialization. . . . .	131
5-7	Per-sample distributions of probabilities of test set samples belonging to the mutant group over training trials. Note the distributions for animal 2 highlighting the need for averaging the final predictions across training runs. . . . .	132
5-8	Accuracy score distributions (accuracy scores collected over different training runs) for the test set consisting of samples corresponding to the animal 2 (the rest of the samples in the training set) depending on the neural net architecture with (top panel) common convolutional architectures for time-series classification and (bottom panel) some other architectures widely used for sequence modeling such as recurrent networks and transformers. . . . .	135
5-9	Accuracy score distributions (accuracy scores collected over different training runs) for the test set consisting of samples corresponding to the animal 2 (the rest of the samples in the training set) for FCN models with different hyperparameter values – numbers of output channels, numbers of layers and inputs sizes. . . . .	136
5-10	Accuracy scores (average predicted labels) per animal over training trials of an FCN model with [32, 32, 32] output channel layout (other parameters same as in the default model) trained on neural activity + movement data using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group). . . . .	137
5-11	Accuracy scores (average predicted labels) per animal over training trials of a Transformer model with 64 self-attention heads trained on neural activity + movement data using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group). . . . .	138
5-12	Accuracy scores (average predicted labels) per animal over training trials of a base FCN model trained on (top panel) only neural activity time-series and (bottom panel) only speed signal time-series using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group). . . . .	141



5-13 Accuracy scores (average predicted labels) per animal over training trials of a base FCN model trained on time-series samples of both modalities corresponding to (top panel) high-mobility periods meaning that the average animal speed within the sample is higher than 0.08 m/s and (bottom panel) low-mobility periods meaning that the average animal speed within the sample is lower than 0.05 m/s. Evaluation is done using the leave-one-out cross-validation scheme. . . . 142

# List of Tables

2.1	Test set accuracy values for different spike train classification models over the three tasks proposed in the benchmark. . . . .	42
2.2	Test set accuracy and AUC-ROC scores in the retinal spike train stimulus classification task (moving bar vs. random checkerboard stimuli) achieved on quantized ISI sequences with a <i>tsfresh</i> + Random Forest model depending on the number of quantization bits. Median value $\pm$ standard deviation is shown. . . . .	46
2.3	Test set accuracy and AUC-ROC scores in the fcx1 WAKE/SLEEP state classification achieved on quantized ISI sequences with a <i>tsfresh</i> + Random Forest model depending on the number of quantization bits. Median value $\pm$ standard deviation is shown. . . . .	47
2.4	Test set accuracy values for the unsupervised temporal structure recognition tasks for different base spiking datasets and different transforms. . . . .	48
4.1	Test set accuracy in the WT vs. $\alpha$ 5KO classification task (with a fixed train/test split and class-balanced training and testing data) achieved with different time-series classification methods (and data pre-processing strategies). . . . .	111
4.2	Wasserstein distance values between time-series feature value distributions in WT and $\alpha$ 5KO animal groups. . . . .	113
5.1	Logistic loss score values computed over the animals in the data set for different models and training data preparation strategies evaluated.	140

# Chapter 1

## Introduction

### 1.1 Thesis summary

Machine learning tools could be very useful for neural data analysis not just merely for engineering applications but for discovering the state-dependent changes in the structure of neural activity. This is in particular important for the cases when one is aiming to build systems to automatically detect pathological activity patterns from neural circuit activity in animal models of nervous system disorders. Machine learning methods might help reveal hidden structure in the neural data and identify the features of activity predictive of disease. The question of how one could utilize machine learning methods to study what information is contained in the activity of single neurons and which of its features might serve as disease markers is a central question we attempt to tackle in this thesis. We first start with identifying the machine learning methods for time-series data that work well for neural activity signals from across different brain areas and states. To this end, we consider the task of spike train time-series classification and propose an evaluation benchmark consisting of a set of binary classification tasks based on open-access neural spiking data sets. The tasks include prediction of the animal's behavioural state, prediction of the input stimulus type and prediction of the neuronal sub-type given a single neuronal spike train. We establish a strong machine learning baseline for the benchmark based on massive time-series feature extraction coupled with effective classifiers such as decision tree based gradient boosting machines. This approach

is shown to outperform state-of-the-art deep learning approaches for neural decoding on our benchmark while also allowing to identify the time-series features of the spike trains useful for decoding across tasks and brain regions. We then analyze a data set of neuronal activity in the prefrontal cortex (PFC) in animal models of schizophrenia and early Alzheimer’s disease. We build a minimal computational model of interacting neuronal populations in the PFC and show that it is possible to capture experimental data by finding a suitable parametrization of the model. We use the model to predict restoration of activity in mice with the  $\alpha$ 5SNP mutation under nicotine treatment as well as the effect of galantamine in mice expressing the amyloid beta peptide. We further utilize the developed machine learning approach for these data. We show that it is possible to build predictive models that recognize patterns of pathological activity from single-neuron recordings and detect animals with mutations such as the  $\alpha$ 5SNP or animals expressing the amyloid beta peptide. We also use machine learning to assess the effects of neuromodulation and pharmacological manipulations on neural activity like in the case of activity level restoration from nicotine application in animals with the  $\alpha$ 5SNP mutation or galantamine treatment in animals expressing the amyloid beta peptide. Finally, we test different machine learning approaches on a multi-modal data set of neural activity in an animal model of early amyotrophic lateral sclerosis (ALS). We show that in order to design a system that is capable of accurately detecting the pathological activity inherent to early ALS, one has to train the model to extract information from the interaction between the modalities of cortical activity and animal movement. We demonstrate that deep learning models, specifically convolutional neural nets (CNNs), are particularly well-suited for this task and outperform the feature extraction approach by a large margin. Interestingly, we show that simple CNN architectures outperform more advanced convolutional and non-convolutional architectures like transformers. One central finding from the early ALS data set is that data that are the most predictive of the pathology are state-dependent correspond to the periods of high mobility of the animals. We hope that these findings could further help design paradigms of early diagnosis of the ALS in human patients.

## 1.2 Background

Understanding the underlying features and factors that are predictive of nervous system disorders is critical in medical applications. Machine learning has been employed to this end for neurodegenerative diseases like Alzheimer’s disease to estimate the relevance of neuroimaging characteristics (e.g. the importance scores of functional connectivity measures extracted from a machine learning model [Challis et al. \[2015\]](#)). In humans, deep learning approaches were also used to estimate the predictiveness of fMRI connectivity relationships in the context of schizophrenia [Kim et al. \[2016\]](#) and ADHD [Deshpande et al. \[2015\]](#). Highly interpretable deep learning methods have been designed to identify correlates of depression and other mental health disorders from noisy EEG data [Honke et al. \[2020\]](#). [Philips et al. \[2021\]](#) used time-series machine-learning methods cortical functional connectivity patterns of distinct subcortical regions. [Hultman et al. \[2016\]](#) identified variables of prefrontal cortex activity predictive of pathological behaviour in a mouse model of depression which helped them design a neural stimulation paradigm to restore normal behavior.

Deep learning is a common tool used to predict neural disorders from a variety of data modalities. The primary input data modalities for systems predicting Alzheimer’s disease are fMRI and PET data [Gautam and Sharma \[2020\]](#) or also speech data in some cases [Punjabi et al. \[2019\]](#). Convolutional neural networks are widely used for AD classification [Wen et al. \[2020\]](#). Several studies have addressed Parkinson’s disease classification using single-photon emission computed tomography [Choi et al. \[2017\]](#) or diffusion MRI [Zhang et al. \[2018\]](#).

The current state-of-the-art approaches in neural decoding in general are based on advanced machine learning methods such as deep neural networks and model ensembles. [Tampuu et al. \[2019\]](#) used deep recurrent neural network architectures such as LSTMs [Hochreiter and Schmidhuber \[1997\]](#) to decode self-location from hippocampal place cell activity. [Glaser et al. \[2020\]](#) benchmarked several common decoding methods in the tasks of predicting position of a rat chasing rewards on a platform from its hippocampal activity or predicting a position of a cursor controlled by a monkey via moving a manipulandum from its motor cortex activity. [Xu](#)

et al. [2019] have evaluated a set of different machine learning methods in the task of decoding from head direction cells. The common finding in these studies is that advanced techniques such as deep neural nets outperform more traditional decoding methods such as Kalman and Wiener filters. While some of the studies have been primarily aimed at improving the decoding accuracy, the machine learning model also helped better understand the data that they were trained on. For instance, Livezey et al. [2019] trained a deep neural network to predict produced speech syllables based on gamma cortical surface electric potentials recorded from human sensorimotor cortex and revealed hierarchical latent structure in the neural data using uncertainties in the network’s predictions as well as discovered high-gamma-to-beta activity coupling during speech production. There is a general trend in using interpretability techniques on trained neural decoders to analyze the underlying neural activity data Livezey and Glaser [2021]. Troullinou et al. [2020] have shown that is possible to infer neuron type from calcium imaging recordings with high accuracy using common deep learning architectures such as CNNs or RNNs. Chowdhury et al. [2020] investigated cell type classification using transcriptomic data with recurrent neural networks.

Notably, most of the studies applying deep learning to neural data rely on basic DNN architectures like simple VGG-style CNNs or LSTM RNNs. However, there have been many recent developments on improving machine learning methods for time series data Fawaz et al. [2019]. This includes specific deep architectures like XCM Fauvel et al. [2020], InceptionTime Fawaz et al. [2020] or the Time-Series Transformer Zerveas et al. [2020] as well as more traditional machine learning approaches like nearest-neighbor models with the Dynamic Time-Warping (DTW) measure Bagnall et al. [2017], time-series feature extraction approaches Christ et al. [2018], Fulcher and Jones [2017] and shapelet learning Ye and Keogh [2009]. A promising avenue of research is adopting these techniques that have been established as state-of-the-art approaches in time-series classification Bagnall et al. [2017] to neural decoding and in particular to the problems of detecting pathological activity from neural recordings, which is one of the central questions addressed in this thesis.

One way to model the pathological activity on the level of single neurons or neural populations is to build mechanistic models of neural function and fit them to available experimental data. This approach, based on population-level modeling of neural activity, has been successfully applied to study the cognitive deficits in schizophrenia [Wang \[2006\]](#) as well as the pathophysiology of the Parkinson's disease [Humphries et al. \[2018\]](#). A central idea here is based on modeling decision making or working memory processes via networks of connected excitatory neural populations modulated by a pool of interneurons [Barak and Tsodyks \[2014\]](#), [Albantakis and Deco \[2011\]](#). In the case of cortical activity, the inhibitory part of the network is modelled by a hierarchy of interneuron populations of different subtypes with a specific inter-population connectivity pattern [Litwin-Kumar et al. \[2016\]](#), [Hertäg and Sprekeler \[2019\]](#). Notably, these distinct interneuronal populations in the cortex are found to be differentially modulated by different types of nicotinic acetylcholine receptors [Poorthuis et al. \[2013\]](#). A computational model of nicotinic neuromodulation [Graupner and Gutkin \[2009\]](#) coupled with a population-level model of cortical activity then gives an effective tool to simulate cholinergic system dysfunctions typically associated with neural disorders such as schizophrenia, which is known to be linked to the mutation of the  $\alpha 5$  subunit of the nicotinic acetylcholine receptor [Besson et al. \[2018\]](#), [Koukouli et al. \[2016a\]](#).

## Chapter 2

# Neural activity classification with machine learning models trained on inter-spike interval time-series data

Ivan Lazarevich, Ilya Prokin, Boris Gutkin, Victor Kazantsev

Under review in *PLOS Computational Biology*

I.L., I.S. and B.G. conceived and designed research. I.L. designed and performed computational experiments. All authors wrote the manuscript.



# Abstract

Modern well-performing approaches to neural decoding are based on machine learning models such as decision tree ensembles and deep neural networks. The wide range of algorithms that can be utilized to learn from neural spike trains, which are essentially time-series data, results in the need for diverse and challenging benchmarks for neural decoding, similar to the ones in the fields of computer vision and natural language processing. In this work, we propose a spike train classification benchmark, based on open-access neural activity datasets and consisting of several learning tasks such as stimulus type classification, animal’s behavioral state prediction and neuron type identification. We demonstrate that an approach based on hand-crafted time-series feature engineering establishes a strong baseline performing on par with state-of-the-art deep learning based models for neural decoding. We release the allowing to reproduce the reported results.

## Chapter summary

Machine learning-based neural decoding has been shown to outperform the traditional approaches like Wiener and Kalman filters on certain key tasks [Glaser et al. \[2020\]](#). To further the advancement of neural decoding models, such as improvements in deep neural network architectures and better feature engineering for classical ML models, there need to exist common evaluation benchmarks similar to the ones in the fields of computer vision or natural language processing. In this work, we propose a benchmark consisting of several *individual neuron* spike train classification tasks based on open-access data from a range of animals and brain regions. We demonstrate that it is possible to achieve meaningful results in such a challenging benchmarks using the massive time-series feature extraction approach, which is found to be hard to beat using state-of-the-art deep learning approaches.

## 2.1 Introduction

The latest advances in multi-neuronal recording technologies such as two-photon calcium imaging [Pachitariu et al. \[2016\]](#), extracellular recordings with multi-electrode arrays [Tsai et al. \[2015\]](#), Neuropixels probes [Steinmetz et al. \[2018\]](#) allow producing large-scale single-neuron resolution brain activity data with remarkable magnitude and precision. Some of the neural spiking data recorded in animals has been released to the public in the scope of data repositories such as CRCNS.org [Teeters and Sommer \[2009\]](#). In addition to increasing experimental data access, various neural data analysis tools have been developed, in particular for the task of neural decoding, which is often posed as a supervised learning problem [Glaser et al. \[2020\]](#): given firing activity of a population of neurons at each time point, one has to predict the value of a certain quantity pertaining to animal’s behaviour such as its velocity at a given point in time.

Such a formulation of the neural decoding task implies that it is a multivariate time-series regression or classification problem. An array of supervised learning methods focused specifically on general time-series data has been developed over the years, ranging from classical approaches [Bagnall et al. \[2017\]](#) to deep neural networks for sequential data [Fawaz et al. \[2019\]](#). It is not fully clear, however, how useful these methods are for the specific tasks of learning from neural spiking data. In order to establish a sensible ranking of these algorithms for neural decoding, there is a need for a common spiking activity recognition benchmark. In this work, we propose a diverse and challenging spike train classification benchmark based on several open-access neuronal activity datasets. This benchmark incorporates firing activity from different brain regions of different animals (retina, prefrontal cortex, motor and visual cortices) and comprises distinct task types such as visual stimulus type classification, animal’s behavioral state prediction from individual spike trains and interneuron subtype recognition from firing patterns. All of these tasks are formulated as univariate time-series classification problems, that is, one needs to predict the target category based on an individual spike train recorded from a single neuron. The formulation of the classification problems implies that the predicted

category is stationary across the duration of the given spike train sample.

Our main contributions can be summarized as follows:

- We propose a diverse spike train classification benchmark based on open-access data.
- We show that global information such as the animal’s behavioral state or stimulus type can be decoded (with high accuracy) from *single-neuron* spike trains containing several tens of interspike intervals.
- We show that inter-spike interval encoding of spike trains in general contains more information predictive of neural circuit state and the cell properties compared to the spike count encoding and, consequently, the firing rate time-series
- We establish a strong baseline for spike train classification based on hand-crafted time-series feature engineering that performs on par state-of-the-art with deep learning models.
- We demonstrate that highly compressed representations of neuronal spike trains with as few as 2 bits per inter-spike interval could be used to decode relevant information about stimuli or animal state.

Well-established machine learning techniques such as gradient boosted decision tree ensembles and recurrent neural networks have been successfully applied both to neural activity decoding (predicting stimuli/action from spiking activity) [Glaser et al. \[2020\]](#) as well as neural encoding (predicting neural activity from stimuli) [Benjamin et al. \[2018\]](#). Neural decoding tasks are often formulated as regression problems, whereas binned spiking count time-series of a single fixed neural population are used to predict the animal’s position or velocity in time.

A number of previous studies on feature vector representations of spike trains also focused on defining a spike train distance metric [Tezuka \[2018\]](#) for identification of neuronal assemblies [Humphries \[2011\]](#). Several different definitions of the spike train distance exist such as van Rossum distance [Rossum \[2001\]](#), Victor-Purpura distance [Victor and Purpura \[1997\]](#), SPIKE- and ISI- synchronization distances [Mulansky](#)

and Kreuz [2016] (for a thorough list of existing spike train distance metrics see Tezuka [2018]). These distance metrics were used to perform spike train clustering and classification based on the k-Nearest-Neighbors approach Tezuka [2015]. Jouty et al. Jouty et al. [2018] employed ISI and SPIKE distance measures to perform clustering of retinal ganglion cells based on their firing responses to a given stimulus.

In addition to characterization with spike train distance metrics, some previous works relied on certain statistics of spike trains to differentiate between cell types. Charlesworth et al. Charlesworth et al. [2015] calculated basic statistics of multi-neuronal activity from cortical and hippocampal cultures and were able to perform clustering and classification of activity between these culture types. Li et al. Li et al. [2015] used two general features of the interspike interval (ISI) distribution to perform clustering analysis to identify neuron subtypes. Such approaches represent neural activity (single or multi-neuron spiking patterns) in a low-dimensional feature space where the hand-crafted features are defined to address specific problems and might not provide an optimal feature representation of spiking activity data for a general decoding problem. Finally, not only spike timing information can be used to characterize neurons in a supervised classification task. Jia et al. Jia et al. [2018] used waveform features of extracellularly recorded action potentials to classify them by brain region of origin.

The aforementioned works were aimed at, to some extent or another, trying to decode properties of neurons or stimuli given recorded spiking data. In some of the cases the datasets used were not released to be openly available, in some of the cases the predictive models used constituted quite simple baselines for the underlying decoding/cell identification tasks. In this work, we aim to propose a benchmark base on open-access datasets that is diverse and challenging enough to robustly demonstrate gains of advanced time-series machine learning approaches as compared to some of the simple baselines used in previous works. We release the code allowing to reproduce the reported results<sup>1</sup>.

---

<sup>1</sup><https://github.com/lzrvch/pyspikelib>

## 2.2 Materials and methods

### 2.2.1 Overview of time series classification methods

We applied general time series feature representation methods [Bagnall et al. \[2017\]](#) for classification of neuronal spike train data. Most approaches in time series classification are focused on transforming the raw time series data into an effective feature space representation before training and applying a machine learning classification model. Here we give a brief overview of state-of-the-art approaches one could utilize in order to transform time series data into a feature vector representation for efficient neural activity classification.

#### Neighbor-based models with time series distance measures

A strong baseline algorithm for time series classification is k-nearest-neighbors (kNN) with a suitable time series distance metric such as the Dynamic Time Warping (DTW) measure or the edit distance (ED) [Bagnall et al. \[2017\]](#). In this work, we evaluated performance of nearest-neighbor models for generic distance measures such as  $l_p$  and DTW distance, converting spike trains to the interspike-interval (ISI) time-series representation prior to calculating the spike-train distances. Some of the distance metrics we also used for evaluation are essentially distribution similarity measures (e.g. Kolmogorov-Smirnov distance, Earth mover’s distance) which allow comparing ISI value distributions within spike trains. Such a spike train distance definition would only use the information about the ISI distribution in the spike train, but not about its temporal structure. Alternatively, one can keep the original event-based representation of the spike train and compute the spike train similarity metrics such as van Rossum or Victor-Purpura distances or ISI/SPIKE distances [Tezuka \[2018\]](#).

The choice of the distance metric determines which features of the time series are considered as important. Instead of defining a complex distance metric, one can explicitly transform time series into a feature space by calculating various properties of the series that might be important (e.g. mean, variance). After assigning appropriate weights to each feature one can use kNN with any standard distance metric.

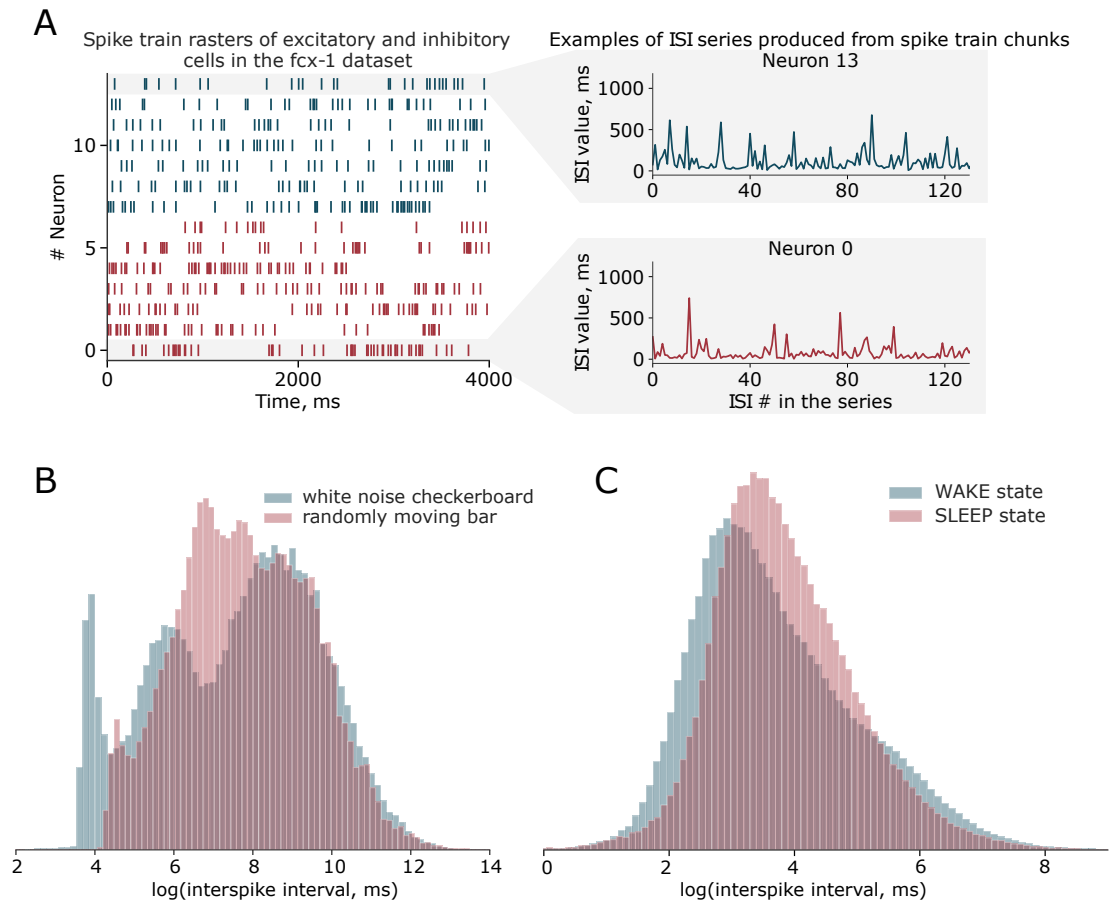


Figure 2-1 – (A) Examples of spiking activity recordings in the CRCNS fcx-1 dataset in the WAKE state. Left: spike train raster of a random subset of excitatory cells (red) and inhibitory cells (blue). Right: examples of ISI series produced from spike train chunks of inhibitory/excitatory cells in the fcx-1 dataset. (B) Interspike interval value distribution histograms generated from the aggregated spike trains of retinal ganglion cells in response to a "white noise checkerboard" visual stimulus (blue) and a "randomly moving bar" stimulus (red). (C) Interspike interval value distribution histograms generated from the aggregated PFC spike trains (fcx-1 dataset) corresponding to the WAKE (blue) or SLEEP (red) state of the rat.

Moreover, such a representation allows the application of any state-of-the-art machine learning classification algorithm beyond kNN to obtain better classification results. In the following, we discuss approaches using various feature space representations available for time series data.

### **Models using hand-crafted time series features**

One of the useful and intuitive approaches in time series classification is focused on manually calculating a set of descriptive features for each time series (e.g. their basic statistics, spectral properties, other measures used in signal processing and so on) and using these feature sets as vectors describing each sample series. There exist approaches which enable automated calculation of a large number of time series features which may be typically considered in different application domains. Such approaches include automated time series phenotyping implemented in the *hctsa* MATLAB package [Fulcher and Jones \[2017\]](#) and automated feature extraction in the *tsfresh* Python package [Christ et al. \[2018\]](#). Here we utilize the *tsfresh* package which enables calculation of 779 descriptive time series features for each spike train, ranging from Fourier and wavelet expansion coefficients to coefficients of a fitted autoregressive process.

Once each time series (spike train) is represented as a feature vector, the spiking activity dataset has the standard form of a matrix with size  $[n_{\text{samples}}, n_{\text{features}}]$  rather than the raw dataset with shape  $[n_{\text{samples}}, n_{\text{timestamps}}]$ . This standardized dataset can be then used as an input to any machine learning algorithm such as logistic regression or gradient boosted trees [Friedman \[2001\]](#). We found this approach to set a strong baseline for all of the classification tasks we considered.

### **Quantization/bag-of-patterns transforms**

Some state-of-the-art algorithms in general time series classification use text mining techniques and thus transform time series into bags-of-words (patterns). This is typically done the following way. First, a time series of real numbers is transformed into a sequence of letters. One of the methods to perform this transform is Symbolic Aggregate approXimation (SAX) [Lin et al. \[2007\]](#). In SAX, bins are computed for

each time series using gaussian or empirical quantiles. After that, each datapoint in the series is replaced by the bin it is in (a letter). Another algorithm commonly used for this task is Multiple Coefficient Binning (MCB). The idea is very similar to SAX and the difference is that the quantization is applied at each timestamp. The third algorithm for the series-letter transform is Symbolic Fourier Approximation (SFA) [Schäfer and Höggvist \[2012\]](#). It performs a discrete Fourier transform (DFT) and then applies MCB, i.e. MCB is applied to the selected Fourier coefficients of each time series. Once the time series is transformed into a sequence of letters, a sliding window of fixed size can be applied to define and detect words (letter patterns) in the sequence. After that, the bag-of-words (BOW) representation can be constructed whereby each "sentence" (time series) turns into a vector of word occurrence frequencies.

Several feature generation approaches were developed utilizing the BOW representation of time series data. One such method is Bag-of-SFA Symbols (BOSS) [Schäfer \[2015\]](#). According to the BOSS algorithm, each time series is first transformed into a bag of words using SFA and BOW. Features that are created after this transformation are determined by word occurrence frequencies.

Some classification algorithms which use this bag-of-patterns approach represent whole classes of samples with a set of features. One example of such a method is an algorithm called SAX-VSM [Senin and Malinchik \[2013\]](#). The outline of this algorithm is to first transform raw time series into bags of words using SAX and BOW, then merge, for each class label, all bags of words for this class label into a single class-wise bag of words, and finally compute term-frequency-inverse-document-frequency statistic (tf-idf) [Sparck Jones \[1972\]](#) for each bag of words. This leads to a tf-idf vector for each class label. To predict an unlabeled time series, this time series is first transformed into a term frequency vector, then the predicted label is the one giving the highest cosine similarity among the tf-idf vectors learned in the training phase (nearest neighbor classification with tf-idf features). A very similar approach is Bag-of-SFA Symbols in Vector Space (BOSSVS) [Schäfer \[2016\]](#) which is equivalent to SAX-VSM, but words are created using SFA rather than SAX. The choice of SAX/MCB or SFA representation of the time series depends on the task at hand –



in particular, SFA would work best if spectral characteristics of the time series are important for classification, while SAX/MCB would be efficient for describing the temporal structure (e.g. reoccurring patterns) of the series.

These time series representation methods are implemented in the *pyts* Python package [Faouzi \[2018\]](#), which was used in the present work. Whenever we apply BOSSVS and SAX-VSM algorithms to classify neural activity, we make use of the ISI representation of the corresponding spike trains.

### **Image representation of time series**

Several methods to represent time series as images (matrices with spatial structure) were developed and utilized for classification as well. One such image representation method is called the recurrence plot [Eckmann et al. \[1987\]](#). It transforms a time series into a matrix where each value corresponds to the distance between two trajectories (a trajectory is a sub time series, i.e. a subsequence of back-to-back values of a time series). The obtained matrix can then be binarized using some threshold value.

Another method of time series image representation is called Gramian Angular Field (GAF) [Wang and Oates \[2015\]](#). According to GAF, a time series is first represented as polar coordinates. Then the time series can be transformed into a Gramian Angular Summation Field (GASF) when the cosine of the sum of the angular coordinates is computed or a Gramian Angular Difference Field (GADF) when the sine of the difference of the angular coordinates is computed.

Yet another image representation method is the Markov Transition Field (MTF). The outline of the algorithm is to first quantize a time series using SAX, then to compute the Markov transition matrix (the quantized time series is treated as a Markov chain) and finally to compute the Markov transition field from the transition matrix. These image representations can be effectively used in junction with effective deep learning models for image classification (e.g. various available architectures of convolutional neural nets [Wang and Oates \[2015\]](#), [LeCun et al. \[2015\]](#)).

## Deep learning models

Lastly, there are deep learning based approaches working well for time-series classification Fawaz et al. [2019] such as deep recurrent networks like LSTMs and GRUs Karim et al. [2017] and 1D convolutional neural networks (1D-CNNs) Zhao et al. [2017], Fawaz et al. [2020]. While rather generic model architectures have been typically applied to neural decoding tasks Glaser et al. [2020], there exist models specifically designed for time-series classification and regression tasks, like InceptionTime Fawaz et al. [2020], achieving state-of-the-art results on benchmarks like the UCR Time Series Classification Archive Dau et al. [2019]. Recent developments in deep learning model for time series also include the Time Series Transformer Zerveas et al. [2020] and convolutional architectures like the Omniscale-CNN Tang et al. [2020]. Perhaps surprisingly, we found that deep learning models could not significantly outperform the baseline with hand-crafted time series features on the spike train classification tasks, oftentimes performing worse than the baseline.

### Ensembling: the best of all worlds

All model types listed in the above subsections (e.g. neighbor-based models, models based on hand-crafted features, bag-of-patterns classifiers) are different in their underlying feature representation of the time series and the kind of information they extract from these features. To attain better prediction performance, all these models may be effectively combined in an ensemble of models using model stacking/blending Džeroski and Ženko [2004] to improve classification results.

### 2.2.2 The proposed spike train classification benchmark

We propose a spike train classification benchmark comprising of several different open-access dataset and distinct classification tasks. The datasets used for the benchmark are as follows:

- **Retinal ganglion cell stimulus type classification** based on the published dataset Prentice et al. [2016], Loback [2016]: Spike time data from multi-electrode array recordings of salamander retinal ganglion cells under four

stimulus conditions: a white noise checkerboard, a repeated natural movie, a non-repeated natural movie, and a bar exhibiting random one-dimensional motion. We define the 4-class classification task to predict the stimulus type given the spike train chunk, also considering binary classification tasks for pairs of stimuli types (e.g. "white noise checkerboard" vs. "randomly moving bar").

- **WAKE/SLEEP classification** based on fcx-1 dataset [Watson et al. \[2016a,b\]](#) from [CRCNS.org](#) [Teeters and Sommer \[2009\]](#): Spiking activity and Local-Field Potential (LFP) signals recorded extracellularly from frontal cortices of male Long Evans rats during wake and sleep states without any particular behavior, task or stimulus. Around 1100 units (neurons) were recorded, 120 of which are putative inhibitory cells and the rest is putative excitatory cells. Figure 4-8 shows several examples of spiking activity recordings that can be extracted from the fcx-1 dataset. The authors classified cells into an inhibitory or excitatory class based on the action potential waveform (action potential width and peak time). Sleep states (SLEEP activity class) were labelled semi-automatically based on extracted LFP and electromyogram features, and the non-sleep state was labelled as the WAKE activity class. We define the binary classification task as the prediction of WAKE or SLEEP animal state given a spike train chunk recorded from a putative excitatory cell.
- **Interneuron subtype classification task** based on the Allen Cell Types dataset [All](#): Whole cell patch clamp recordings of membrane potential in neurons of different types. We selected the PV, VIP and SST interneurons from the whole dataset, as these interneuron groups comprise the majority of inhibitory cells in the prefrontal cortex [Rudy et al. \[2011\]](#). We selected the spike trains recorded under the naturalistic noise stimulation protocol (as a proxy for the *in vivo* spontaneous activity in these cells). The non-trivial prediction task is defined for VIP vs. SST spike train classification, since the PV interneuron spike trains can be easily distinguished from the other interneuron types. The latter is due to a significantly higher firing frequency in PV interneurons that we found in the Allen Cell Types dataset.

- **Unsupervised temporal structure recognition task.** We defined a set of spike train classification tasks constructed in a self-supervised manner [Jing and Tian \[2020\]](#). In such tasks, we take any set of (unlabelled) neuronal spike train recordings and generate a additional set of spike trains by applying a given transformation to the original data. The target classification task is to determine whether a given spike train chunk belongs to the original dataset or to the transformed one. Note that this task can be constructed for any spiking dataset without the need for the ground truth labels, i.e. in an unsupervised way. The spike train transformations we consider here are (i) adding spike timing jitter via, in particular, timing noise following a truncated normal distribution, (ii) random shuffling of the interspike intervals in the spike train, (iii) reversing the spike train. The models trained in such tasks learn to detect the temporal structure of the original spike trains, since the order/precise values of interspike intervals have been disrupted by the transformation (e.g. by ISI shuffling), while the ISI value distribution is preserved by some of the transformations (e.g. by the shuffling and reversal operations). The final trained model accuracy in a shuffled vs. non-shuffled spike train classification task can thus be thought of as a measure of temporal structure in the original spiking dataset (test set accuracy would be on the chance level if the ISI values in the original spike trains were independently sampled from a fixed value distribution, i.e. the exact ordering of the ISIs did not contain any predictive information). We consider the the fex-1 and retinal ganglion cell datasets described above to construct the temporal structure recognition tasks.

### 2.2.3 Cross-validation scheme and data preprocessing

Suppose we are given a dataset containing data from several mice each recorded multiple times with a large number of neurons captured in each recording. For each recorded neuron, we have a corresponding spike train captured over a certain period of time (assuming that the preprocessing steps like spike sorting or spiking time inference from fluorescence traces were performed beforehand). The number of

spikes within each spike-train is going to be variable. A natural way to standardize the length of spike-train sequences would be dividing the full spike train into chunks of  $N_{\text{size}}$  spike times, where  $N_{\text{size}}$  is fixed for each chunk. The chunks can be produced by moving a sliding window across the spike-timing-vector. Thus, each neuron would contribute a different number of spike-timing-chunks depending on its average firing rate. This sliding window procedure can be applied to both the sequences of inter-spike intervals (ISIs) and time-binned spike count time series. We empirically investigate the advantages of using either ISI-encoding or spike-count encoding of spike trains further in this work.

The cross-validation strategy we use in this work is based on group splits, whereby we determine the split into the training and the testing datasets based on neuron/animal identifiers available in the original data. The motivation is that, in cases recordings from several animals and corresponding animal identifiers are available, the set of animals used to construct the training dataset and the set of animals for the testing dataset should not overlap in order to test whether the trained decoding models could generalize across different animals. In case animal identifiers are not available, we split the dataset into training and testing based on non-overlapping neuron identifiers in train and test. We perform several random train/test splits for all datasets based on animal/neuron groups to evaluate the variance of classification metrics across splits.

The metrics we use to evaluate model performance in classification tasks is accuracy and AUC-ROC. The testing datasets are balanced by undersampling to mitigate the influence of class imbalance on the ranking properties of accuracy and AUC-ROC metrics. Since the class distribution in the testing datasets is balanced by construction, the values of these metrics reflect the real classification performance of the trained models relative to the chance level of 0.5.

The base task we consider for all benchmark datasets is classification given an individual spike train. However, prediction performance can be improved by aggregating predictions from several spike trains (in case such data is available) or from several sub-sequences of a single large spike train. If the final classification in such a setting is done by majority voting from all single spike-train predictions and we

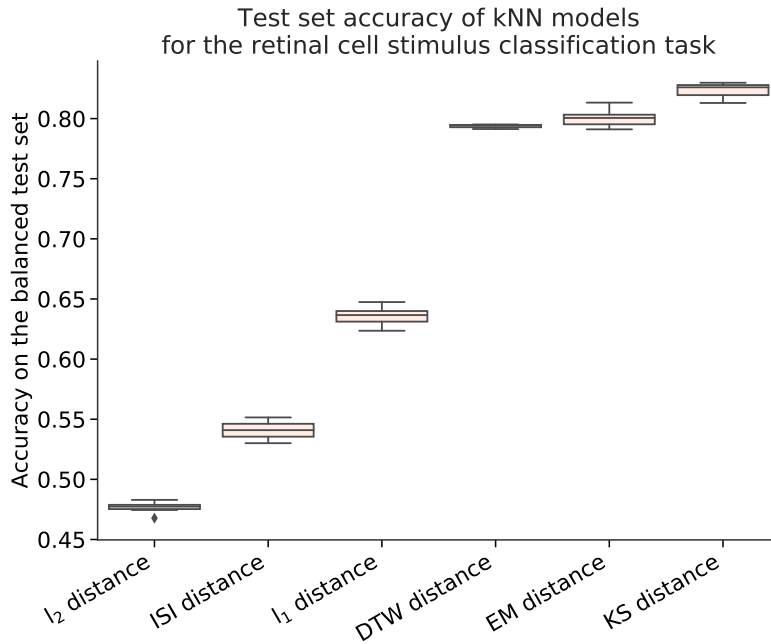


Figure 2-2 – Spike train classification accuracy values for the retinal neuron activity dataset for nearest-neighbor models with different distance metrics . The task is defined as binary classification of the stimulus type ("white noise checkerboard" or "randomly moving bar"), with the test set balanced in class distribution (that is, accuracy=0.5 corresponds to chance level).

assume that recorded spike-trains (neurons) are randomly sampled from the whole ensemble, the optimistic estimate for accuracy improvement with the number of spike trains (neurons)  $N_{\text{trains}}$  would be

$$\mu = \sum_{i=m}^{N_{\text{trains}}} C_{N_{\text{trains}}}^i p^i (1-p)^{N_{\text{trains}}-i} \quad (2.1)$$

where  $\mu$  is the probability that the majority vote prediction is correct,  $p$  is the probability of a single classifier prediction being correct (single spike train prediction accuracy),  $N_{\text{trains}}$  is the number of predictions made,  $m = \lfloor N_{\text{trains}}/2 \rfloor + 1$  is the minimal majority of votes. We found that the empirical values of accuracy improvement are close to the optimistic analytical estimate (2.1) in both cases when the spike train chunks are sampled from different neurons and from a large spike train of a single neuron (see Fig. 2-6).

## 2.3 Results

### 2.3.1 Visual stimulus type classification from retinal spike trains

We first start with looking the at the retinal ganglion cell spike train classification task. Recorded spike trains in the dataset are associated with one of the four categories corresponding to different visual stimulus types, labelled with "white noise checkerboard", "randomly moving bar", "repeated natural movie" and "unique natural movie". The classification task is, given a chunk of the spike train recording, predict the corresponding stimulus type category. The number of neurons in the dataset belonging to each category is 155, 140, 178, 152, respectively. The number of interspike intervals is quite variable among individual cells (due to firing rate variability) ranging from 100 ISIs per recording to as much as 60000 ISIs per recording.

We focus on the binary classification task aimed at predicting one of the two types of stimuli: "randomly moving bar" or "white noise checkerboard". We select recorded spike trains corresponding to those stimuli types and split 50% of recorded neurons for the training part of the dataset and the remaining 50% for the testing dataset. We encode spike trains using the ISI representation and apply a rolling window of size equal to 200 ISIs with a stride of 100 ISIs to each recorded neuron. We then perform undersampling to make the class distribution balanced and arrive at a dataset of 5188 training and 5272 testing examples (each containing 200 spikes), with an equal amount of spike-trains corresponding to both stimuli types in the training and testing datasets.

### 2.3.2 Nearest-neighbor models for spike train classification

We evaluated performance of nearest-neighbor models with different distance metrics on the retinal stimulus classification task, results are shown on Figure 2-2. We found that the nearest neighbor model with the DTW distance is amongst the best performing ones, but is still outperformed by nearest-neighbor models with Kolmogorov-Smirnov and Earth Mover's distances, suggesting that differences in ISI

distributions contain discriminative information helpful for the classification task.

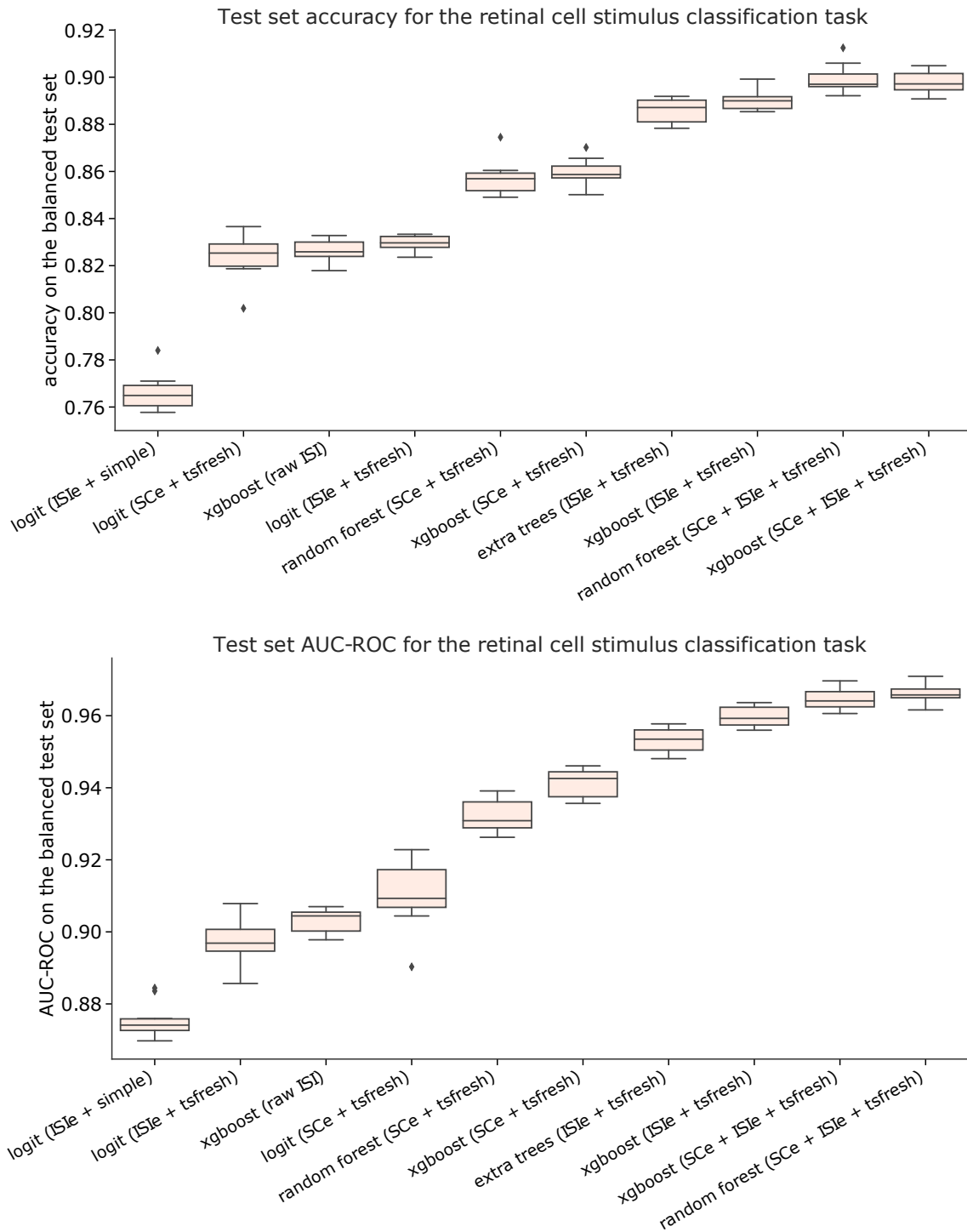


Figure 2-3 – (Caption next page.)



Figure 2-3 – (Previous page.) Spike train classification metric values for the retinal neuron activity dataset on a range of models. The task is defined as binary classification of the stimulus type ("white noise checkerboard" or "randomly moving bar"), with the test set balanced in class distribution (that is, accuracy=0.5 corresponds to chance level). Accuracy is shown on the left and AUC-ROC on the right for the same set of models and train/test dataset splits. The "simple" model tag corresponds to spike trains encoded with 6 basic distribution statistics (representing a simple baseline), the "raw ISI" tag implies that the model has been directly trained on ISI time-series data without encoding. The "tsfresh" tag corresponds to encoding with the full set of time-series features. "ISIE" stands for interspike-interval encoding of the spike train, "SCe" stands for spike-count encoding. "ISIE + SPE" means that feature vectors corresponding to both types of encoding are concatenated.

### 2.3.3 Hand-crafted feature extraction for time-series classification

The kNN results clearly suggest that characteristics of the interspike-interval distribution of the given spike train are predictive of the category label in our classification task. At the same time, one would expect the the temporal (sequential) information contained in the spike train also has certain predictive power. A straightforward way to incorporate both types of features in the model is to build a corresponding vector embedding of the spike train time-series. An efficient way to do so is to use a set of hand-crafted time-series features, like for example the set of 779 features provided in the *tsfresh* Python package. In order to compute vector embeddings for the spike trains in the training and testing datasets, one has to convert spike times into a time-series, which in principle could be done using either an interspike-interval encoding (the time-series is the sequence of ISIs) or a spike-count encoding (time is binned and spike counts in each time bin comprise the time series). The latter type of encoding depends on an additional hyperparameter which is the size of the time bin while ISI-encoding is parameter-free. We tested both types of spike-train encoding for our task and observed that on average models trained using the ISI-encoding of spikes perform better than the ones using binned spike counts. Furthermore, we found that combining features corresponding to both encoding types leads to better performance compared to using a single encoding scheme (see Figure 2-3).

For each spike-train encoding type, we computed the 779-dimensional *tsfresh*

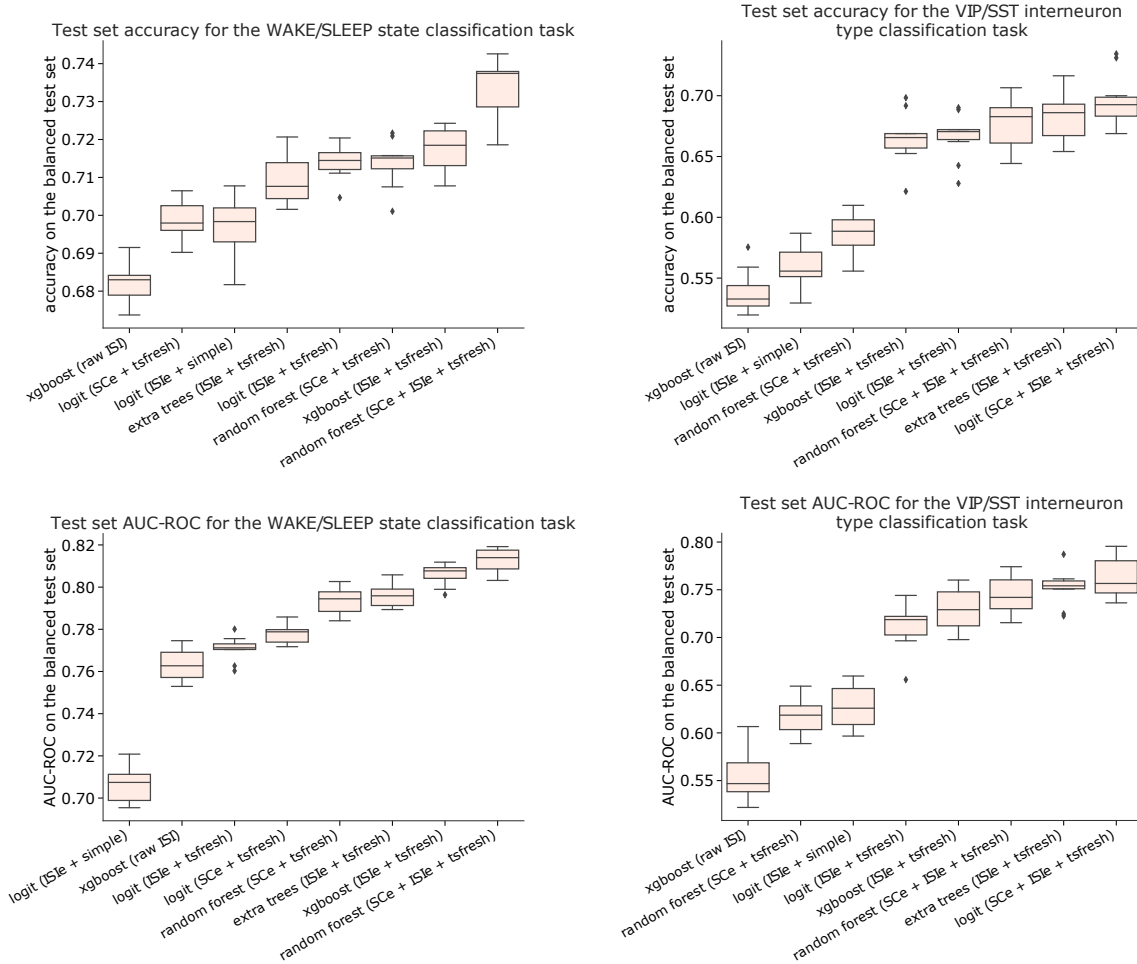


Figure 2-4 – Spike train classification metric values for the WAKE/SLEEP state prediction dataset (left) and VIP/SST interneuron type prediction dataset (right) on a range of models. The task is defined as binary classification, with the test set balanced in class distribution for both datasets (that is, accuracy=0.5 corresponds to chance level). Accuracy is shown on the top panes and AUC-ROC on the bottom panes for the same set of models and train/test dataset splits. The "simple" model tag corresponds to spike trains encoded with 6 basic distribution statistics (representing a simple baseline), the "raw ISI" tag implies that the model has been directly trained on ISI time-series data without encoding. The "tsfresh" tag corresponds to encoding with the full set of time-series features. "ISle" stands for interspike-interval encoding of the spike train, "SCe" stands for spike-count encoding. "ISle + SCe" means that feature vectors corresponding to both types of encoding are concatenated.

	Retinal stimulus (white noise vs. moving bar)	fex-1 WAKE/SLEEP state prediction	Allen Cell Types SST/VIP INs
Logistic regression on basic ISI statis- tics	76.49 $\pm$ 0.78	69.83 $\pm$ 0.82	55.57 $\pm$ 1.63
GBDT on unpro- cessed ("raw") ISI time-series	82.58 $\pm$ 0.47	68.30 $\pm$ 0.51	53.27 $\pm$ 1.74
Logistic regression on tsfresh-encoded ISI time-series	82.96 $\pm$ 0.32	71.44 $\pm$ 0.45	67.05 $\pm$ 1.89
Random forest on tsfresh-encoded ISI time-series	88.69 $\pm$ 0.48	72.31 $\pm$ 0.60	67.95 $\pm$ 2.21
GBDT on tsfresh- encoded ISI time- series	88.99 $\pm$ 0.47	71.84 $\pm$ 0.60	66.55 $\pm$ 2.11
Random forest on tsfresh-encoded ISI + spike count time-series	<b>89.70 <math>\pm</math> 0.62</b>	<b>73.74 <math>\pm</math> 0.75</b>	<b>68.27 <math>\pm</math> 2.29</b>
1D-CNN on un- processed ("raw") ISI time-series	61.57 $\pm$ 2.33	71.88 $\pm$ 1.58	57.13 $\pm$ 1.87
InceptionTime on unprocessed ("raw") ISI time- series	73.60 $\pm$ 1.34	72.35 $\pm$ 1.31	60.08 $\pm$ 2.10
Omniscale-CNN on unprocessed ("raw") ISI time- series	71.34 $\pm$ 2.91	68.10 $\pm$ 0.19	62.05 $\pm$ 2.23
LSTM-FCN on unprocessed ("raw") ISI time- series	65.12 $\pm$ 3.65	64.12 $\pm$ 2.42	54.14 $\pm$ 3.63

Table 2.1 – Test set accuracy values for different spike train classification models over the three tasks proposed in the benchmark.

time-series embeddings independently for each sample in the training and testing datasets (no statistic aggregation across samples is performed). We then performed simple pre-processing steps by (i) removing low-variance features from the embed-

ding (features  $f$  satisfying  $\text{std}(f)/(\text{mean}(f) + \varepsilon) < \theta$  with  $\theta = 0.2$  and  $\varepsilon = 10^{-9}$  were removed) and (ii) performing standard scaling for each feature using mean and variance statistics collected over the training dataset. Note that since the number of spikes in each data sample is fixed and interspike intervals are highly variable, the number of time bins also becomes variable from sample to sample. The *tsfresh* embeddings can nevertheless be computed since they are applicable to variable-length time series.

We then trained classification models on the resulting spike-train vector embeddings. We chose a representative set of models comprising (i) a linear model, namely logistic regression with an  $l_2$  penalty and (ii) various types of tree-based ensembles: a random forest classifier, randomized decision trees (extra trees), a gradient boosted decision tree ensemble. The classifier hyperparameter values used are specified in the Supplementary Materials.

Classification results obtained with the described approach are presented in Figure 2-3. We report classification accuracy and AUC-ROC values for each model type; note that the testing sets were constructed to be balanced such that the accuracy value of 0.5 corresponded to chance level. We generated 10 random balanced subsamples of the training and testing sets (80% of training/testing data randomly sampled) to estimate accuracy/AUC-ROC distributions presented via boxplots in Figure 2-3. We were able to reach significant performance levels ( $> 0.9$  accuracy,  $> 0.96$  AUC-ROC) with our best *tsfresh*-based models on the binary stimulus classification task ("randomly moving bar" vs. "white noise checkerboard"). To make better sense of these metric values, we compared our *tsfresh*-based models against two simple baselines: (a) a logistic regression model on ISI-encoded spike-trains represented by 6 basic statistical features – the mean, median, minimum and maximum ISI values, the standard deviation and the absolute energy of the ISI-sequence (the mean of squared ISI values) and (ii) a gradient-boosted decision tree ensemble trained directly on unprocessed ISI-encoded spike trains.

We found that the worst performing model was logistic regression trained on 6 basic statistical features of the ISI distribution, reaching the median test set accuracy of 0.7649 and AUC-ROC of 0.8741. We further demonstrated that one could

significantly improve upon those results by training a strong model directly on the unprocessed ISI-encoded samples, with a GBDT model reaching 82.63 median test set accuracy and 0.9030 median AUC-ROC (see Figure 2-3 and Table 2.1). We found that using *tsfresh* embeddings significantly improves upon that, with a GBDT model reaching lower metric values for spike-count-encoded spike trains (85.97 median accuracy; 94.13 median AUC-ROC) compared to ISI-encoded spike trains (89.05 median accuracy; 95.97 median AUC-ROC for the GBDT classifier). Furthermore, merging *tsfresh* features calculated for both encoding types and training a classifier on the extended set of features increased the test accuracy to 89.77 (with median AUC-ROC of 96.46) for the GBDT classifier.

We also evaluated the performance of state-of-the-art deep learning models in our retinal stimulus classification tasks, using implementation from *sktime-dl* Löning et al. [2019] and *tsai* Oguiza [2020]. Accuracy results for a range of deep learning models (1D-CNN, InceptionTime, LSTM-FCN and Omniscale-CNN) are shown in Table 2.1. It is clear that models using *tsfresh*-encodings typically perform better deep learning models in terms of classification accuracy on the retinal stimulus classification task as well as the other tasks we establish.

We performed the same preprocessing steps for the WAKE/SLEEP and VIP/SST datasets as for the retinal stimulus classification dataset. The rolling window of size equal to 200 ISIs and a stride of 100 ISIs together with undersampling for class balance produced a dataset of 19786 training examples and 5540 testing examples (according to a 70/30 train/test split) for the WAKE/SLEEP state dataset. A rolling window of size equal to 50 ISIs and a stride of 20 ISIs with undersampling was applied to the VIP/SST dataset, resulting in 1630 training and 872 testing examples.

We observed similar trends both for the WAKE/SLEEP state and VIP/SST interneuron classification tasks, shown in Figure 2-4 and Table 2.1. The best performing models were found to be *tsfresh*-based ones using the combined ISI and spike count encodings of the underlying spike trains.

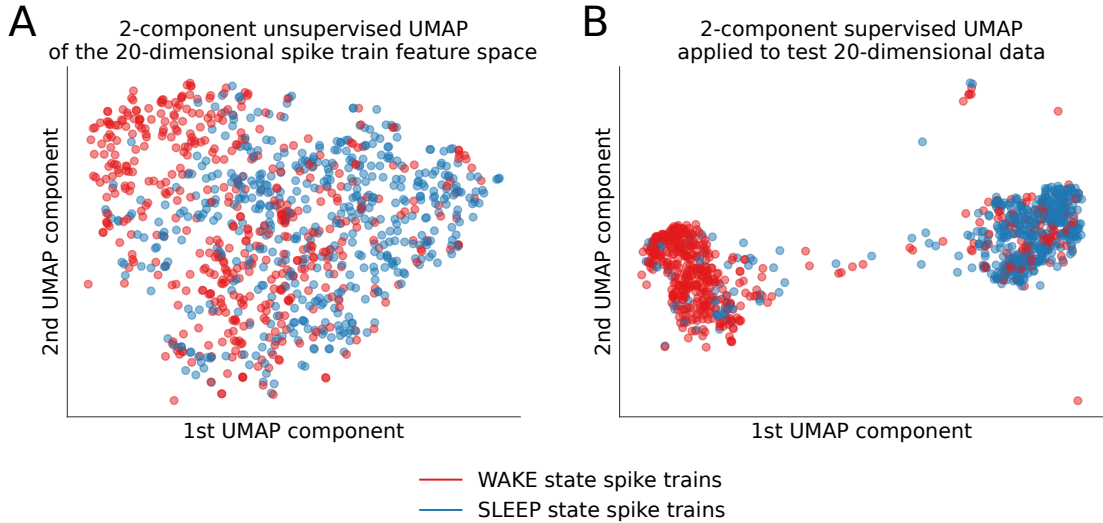


Figure 2-5 – Spike train feature embeddings for WAKE (points marked red) vs. SLEEP (points marked blue) activity states of the neural circuit. Two-dimensional embeddings of the (20-dimensional) selected-*tsfresh*-feature space using (A) unsupervised UMAP and (B) supervised UMAP embedding algorithms for spike trains corresponding to WAKE vs. SLEEP activity states.

### 2.3.4 Prediction accuracy from quantized spike trains

One might hypothesize that the information contained in the spike trains that is learned by the classifiers is not contained in the fine-grained timing of the spikes, but instead, the precise spike timing information might not be relevant for a classification task at hand, for instance for the stimulus type decoding. We tested this hypothesis by limiting the set of possible values in the ISI sequences by applying vector quantization to the spike trains. We first applied a log transformation the ISI representations of spike trains to get rid of the heavy tail in the value distributions, and then performed vector quantization using the k-means algorithm to determine the quantization centroids. After this, we applied the *tsfresh* feature extraction pipeline to the quantized ISI sequences and trained a Random Forest model on top of this vectorized feature-based representation. The results for the retinal spike train stimulus classification task (moving bar vs. random checkerboard stimuli) and the fcx-1 WAKE/SLEEP state classification task are shown in Tables 2.2 and 2.3, respectively. Interestingly, the accuracy and AUC-ROC scores do not significantly drop when the spike trains are quantized down to 2 bits, with the best performing models (with the same hyperparameter values) corresponding to 3-5 bits and even beating

the non-quantized baseline for both tasks. The accuracy scores drop significantly for 1 bit quantization even compared to the model trained on the 6 basic statistics of the non-quantized spike trains, which implies that the distortion introduced to the ISI value distribution by 1-bit quantization is too large compared for instance to quantization with 3 possible value levels. Overall, these results suggest that the information needed to decode the stimulus type or a global animal state contained in the spike trains (to the extent it is extracted by the *tsfresh*+RF model) is robust to extreme low-bitwidth quantization and can be extracted from symbolic representations of spike trains with as few as 3 or 4 possible characters (corresponding to 2 bit quantization).

Number of quantization bits	Test set accuracy	Test set AUC-ROC
No quantization	0.844404 $\pm$ 0.005098	0.931788 $\pm$ 0.003469
No quantization (basic stats.)	0.762265 $\pm$ 0.011024	0.858245 $\pm$ 0.010463
8	0.849228 $\pm$ 0.004748	0.959834 $\pm$ 0.002477
7	0.845782 $\pm$ 0.003613	0.958322 $\pm$ 0.002888
6	0.847712 $\pm$ 0.003883	0.958819 $\pm$ 0.002691
5	0.846747 $\pm$ 0.004296	<b>0.962793 <math>\pm</math> 0.001982</b>
4	0.843026 $\pm$ 0.005474	0.958753 $\pm$ 0.004400
3	<b>0.855016 <math>\pm</math> 0.004360</b>	0.952253 $\pm$ 0.003331
2	0.830485 $\pm$ 0.004602	0.943829 $\pm$ 0.003635
3 quantization levels	0.849641 $\pm$ 0.004377	0.924570 $\pm$ 0.001955
1	0.624586 $\pm$ 0.007284	0.713986 $\pm$ 0.007230

Table 2.2 – Test set accuracy and AUC-ROC scores in the retinal spike train stimulus classification task (moving bar vs. random checkerboard stimuli) achieved on quantized ISI sequences with a *tsfresh* + Random Forest model depending on the number of quantization bits. Median value  $\pm$  standard deviation is shown.

### 2.3.5 Unsupervised spike train temporal structure recognition

The spike train temporal structure recognition task is defined as follows: for a set of spike train activity data, we generate a binary classification task by producing an additional category of spiking data consisting of spike trains from the original dataset with a certain transformation applied to them. We consider the following spike train transformations: (i) ISI shuffling inside the spike train (random shuffling applied to

Number of quantization bits	Test set accuracy	Test set AUC-ROC
No quantization	0.722526 $\pm$ 0.004966	0.796485 $\pm$ 0.004133
No quantization (basic stats.)	0.687829 $\pm$ 0.002994	0.757096 $\pm$ 0.004226
8	0.752161 $\pm$ 0.004693	0.809916 $\pm$ 0.003708
7	0.753322 $\pm$ 0.003422	0.808470 $\pm$ 0.003674
6	0.751740 $\pm$ 0.003929	0.807201 $\pm$ 0.006228
5	0.751001 $\pm$ 0.005085	0.803855 $\pm$ 0.006503
4	<b>0.754798 <math>\pm</math> 0.006267</b>	<b>0.809998 <math>\pm</math> 0.007574</b>
3	0.741615 $\pm$ 0.005924	0.803788 $\pm$ 0.007489
2	0.750896 $\pm$ 0.004535	0.803850 $\pm$ 0.005467
3 quantization levels	0.730436 $\pm$ 0.006422	0.782436 $\pm$ 0.006388
1	0.540392 $\pm$ 0.005679	0.551334 $\pm$ 0.009612

Table 2.3 – Test set accuracy and AUC-ROC scores in the fcx1 WAKE/SLEEP state classification achieved on quantized ISI sequences with a *tsfresh* + Random Forest model depending on the number of quantization bits. Median value  $\pm$  standard deviation is shown.

the ISI time series), (ii) reversing the ISI time series and (iii) adding spike timing jitter sampled from the truncated normal distribution to the time series. Note that the first two transformation types do not change the value distribution of the time series, only its temporal structure (the exact ordering of the interspike intervals of the spike train). Hence, if it is possible to construct a classification model capable of distinguishing between the two activity classes (original spiking activity versus the transformed one), then one could say that the model has learned to detect the temporal structure in the time series. In the case of classifiers trained on *tsfresh* feature vectors, the classification accuracy metrics obtained can be thought of as measures of the amount of temporal structure contained in the spike trains (to the extent encoded in *tsfresh* features). In case of the ISI shuffling transformation, if the spiking is described by a Poisson model [Kass and Ventura \[2001\]](#), or ISIs are in general sampled from a stationary probability distribution, the classification accuracy in this task would be on the chance level, otherwise in case there is a dependence of a given inter-spike interval on the previous spiking history the classification accuracy value would be significantly higher compared to the chance level.

The classification results for different base spiking data and different transforms is shown in [Table 5.1](#). Notably, one could observe higher accuracy values for the



retinal ganglion cell spiking data in case of "structured" inputs (i.e. moving bar or natural movie) as compared to the white noise input for all of the three transforms considered. The same difference in accuracy values is observed for the fcx-1 dataset whereby classification accuracy is higher for all of the three transforms when the SLEEP state is used as the base spiking dataset as opposed to the WAKE state. The accuracy values that we observe are above the chance level in most cases except for, interestingly, the reverse transform for the WAKE spiking data. Time reversal appeared to be a strong invariant for the WAKE state spiking activity, since we could not achieve a significant level of accuracy in that classification tasks neither with *tsfresh*-based feature encodings, nor by training deep neural networks on this task.

	Reverse transform	Shuffling transform	Noise transform
Retinal ganglion cells (randomly moving bar input)	$0.7469471 \pm 0.002794$	$0.80608 \pm 0.00297$	$0.77501 \pm 0.00298$
Retinal ganglion cells (white noise checker-board input)	$0.696856 \pm 0.007963$	$0.77845 \pm 0.00423$	$0.77859 \pm 0.004575$
fcx-1 WAKE spike trains	$0.52777 \pm 0.00644$	$0.82476 \pm 0.00441$	$0.80936 \pm 0.00433$
fcx-1 SLEEP spike trains	$0.65619 \pm 0.002426$	$0.83085 \pm 0.00211$	$0.84983 \pm 0.00178$

Table 2.4 – Test set accuracy values for the unsupervised temporal structure recognition tasks for different base spiking datasets and different transforms.

### 2.3.6 The set of predictive spike train features

Being able to estimate feature importance ranks from trained decision tree ensembles allows us to detect the most discriminating features of ISI time-series for a particular classification problem. In order to select the important groups of discriminative features, we applied the following procedure to the full set of *tsfresh* features: first, we trained 10 logistic regression models with  $l_1$  regularization penalty (with different

random seeds,  $C = 0.01$ ) on the full feature set; we then selected the features which had non-zero values of corresponding weight coefficients in all trained models. After that, we identified highly correlated pairs of features ( $|R| > 0.98$ ), which represent almost equivalent quantities, and removed one randomly selected feature out of each such pair. Further, we trained several random forest classifier models (with different random seeds) and calculated the aggregated feature importance ranks across models to select groups of features relevant for the classification task. After merging important feature sets corresponding to all of the three classification tasks, the following groups of *tsfresh* features are selected (see also Fig. 2-7):

- *median, kurtosis, quantile\_q* – simple statistics of the ISI value distribution in the series like the median ISI value,  $q$  quantiles and kurtosis of the ISI value distribution
- *change\_quantiles* – this feature is calculated by fixing a corridor of the time series values (defined by lower and higher quantile bounds,  $q_l$  and  $q_h$ , which are hyperparameters), then calculating a set of consecutive change values in the series (differencing) and then applying an aggregation function (mean or variance). Another boolean hyperparameter *is\_abs* determines whether absolute change values should be taken or not.
- *fft\_coefficient* – absolute values of the fast Fourier transform coefficients (individual coefficient values and aggregates).
- *entropy* – values of the sample entropy, the approximate entropy and the binned entropy of the power spectral density of the time series.
- *agg\_linear\_trend* – features from linear least-squares regression (standard error in particular) for the values of the time series that were aggregated over chunks of a certain size (with different aggregation functions like min, max, mean and variance). Chunk sizes vary from 5 to 50 points in the series.

To visualize class separation for the WAKE vs. SLEEP state spike trains from the fex-1 dataset as point clouds in two dimensions, we took the top-20 importance

*tsfresh* features identified during the above feature selection procedure. We then used dimensionality reduction techniques on this reduced 20-dimensional dataset to visualize the structure of the data with respect to excitatory/inhibitory labels of the series. Results are shown in Fig. 2-5 for two Uniform Manifold Approximation and Projection, UMAP [McInnes et al. \[2018\]](#) low-dimension embedding algorithms. We also applied other methods such as PCA and t-SNE (t-distributed Stochastic Neighbor Embedding) which gave essentially the same results (not shown). In all cases, classes cannot be linearly separated in two-dimensional embedding spaces, however, there is a separation of large fraction of the points of the excitatory-cell and inhibitory-cell classes.

We conclude that the hand-crafted feature engineering approach combined with strong tree-based learning models sets a strong baseline for spike train classification for all of the tree studied tasks.

## **Classifier hyperparameter values.**

We did not perform dataset-specific hyperparameter search, rather just using sane settings for each of the models to see "out-of-the-box" performance for each classifier. Listed below are hyperparameter values and implementation references for all of the classifier types we used.

- Random Forest: [sklearn](#) implementation, 500 trees, maximal depth = 13
- Extra Trees Classifier: [sklearn](#) implementation, 500 trees, no limit on depth
- Logistic Regression: [sklearn](#) implementation,  $l_2$  penalty,  $C = 0.01$
- Decision Tree Gradient Boosting: [xgboost](#) implementation, `max_depth = 8`, `n_trees = 1000`, `learning_rate = 1e-1`, `gamma = 0`, `reg_lambda = 1`
- 1D-CNN, InceptionTime: [sktime-dl](#) implementation, batch size = 1024, number of epochs = 2000, data was normalized before inference
- Omniscale-CNN, LSTM-FCN: [tsai](#) implementation, batch size = 1024, maximal learning rate =  $1e-4$  with a flat cosine annealing schedule, 500 epochs, data was normalized before inference

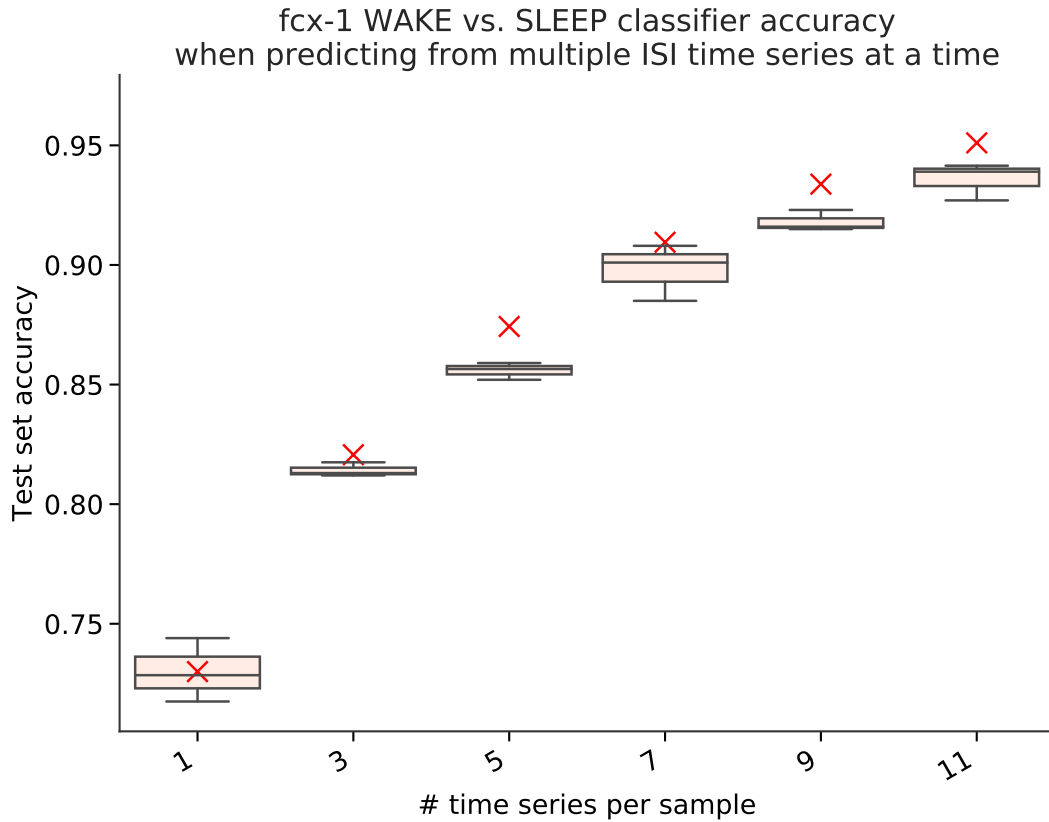


Figure 2-6 – Classification accuracy for the fcx-1 WAKE vs. SLEEP task in the case of multiple randomly sampled same-class spike train chunks per prediction (with prediction done via majority voting). The model trained in these trials is a random forest classifier on the full set of *tsfresh* features. The boxplots reflect the median accuracy and the variance between different train/test splits as done in the main text for the fcx-1 data set. The red crosses correspond to the theoretical estimate under the assumption of independently sampled spike train chunks.

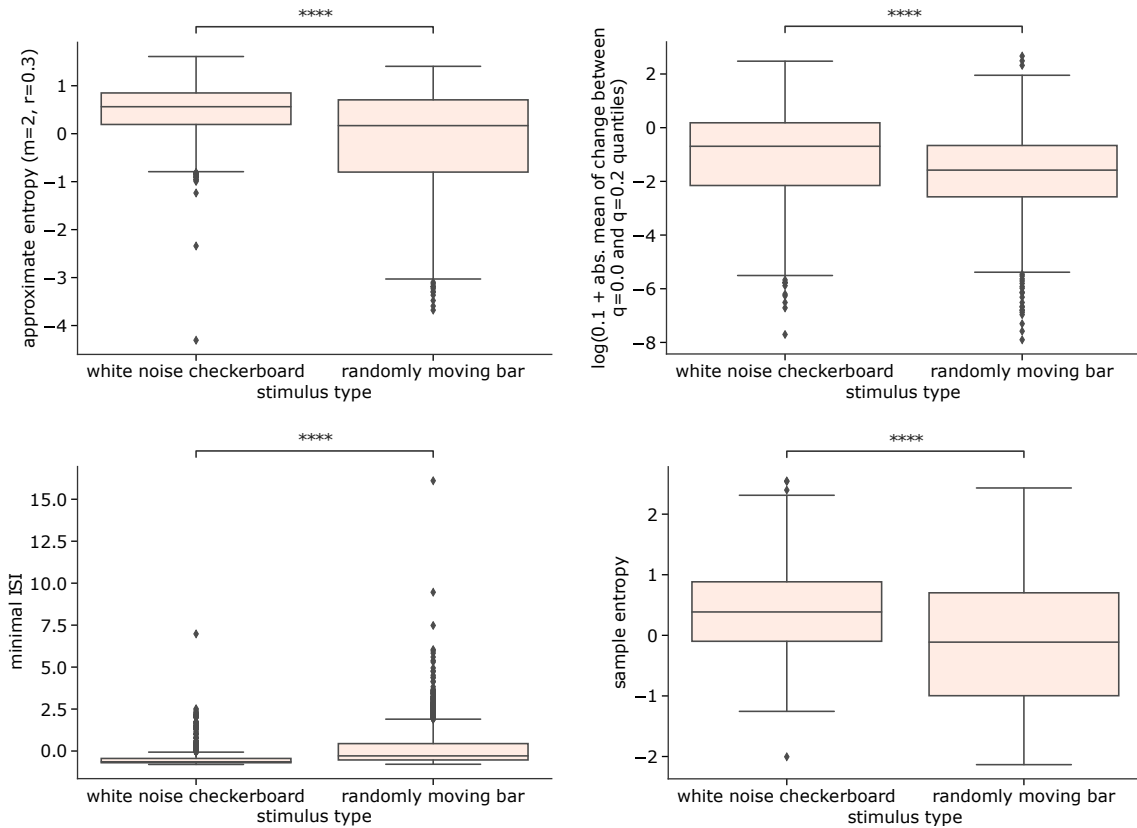


Figure 2-7 – Boxplots of *tsfresh*-extracted feature distributions for features with high discriminative power as detected by the trained decision tree ensemble classifiers in the retinal stimulus type prediction task. Two-sided Mann-Whitney-Wilcoxon test with Bonferroni correction is performed to assess statistical significance; \*\*\*\* denotes  $p < 1e - 4$ .

## 2.4 Discussion

In this work, we have introduced a diverse neuronal spike train classification benchmark to evaluate neural decoding algorithms. The benchmark consists of several single-neuron spike train prediction tasks spanning stimulus type prediction, neuron type identification and animal behavioural state prediction. We have demonstrated that individual neuronal spike trains contain information related to the global state of the underlying neural circuit and this information can be decoded if appropriate time-series learning models are used. Extensive experiments on several datasets that we have conducted imply that not only ISI value distribution is important for global state identification but also the temporal information contained in the spike trains, that is, features related to the exact sequences of interspike intervals in neural firing. We have identified groups of features highly informative for neural decoding tasks and established that this feature encoding combined with strong supervised learning algorithms such as gradient boosted tree ensembles establishes a strong baseline on the proposed benchmark that performs on par with state-of-the-art deep learning approaches. It was shown that significantly large accuracy values can be obtained on all of the proposed tasks using the hand-crafted feature encoding approach on single-neuron spike train chunks containing as low as 50 interpike intervals. We suggest that accuracy values can further be improved by model ensembling and test-time data augmentation. We propose that neural decoding models be evaluated on diverse and challenging tasks including the proposed benchmark in order to establish a sensible model performance ranking similarly to what is done for computer vision and natural language understanding problems. We believe that this would drive further development of highly accurate neural decoding/neural activity mining approaches enabling their application in precision-critical tasks such as identifying pathological disease-related firing activity patterns in the brain.

## 2.5 Conclusion

To summarize our contributions, we have proposed a challenging and diverse benchmark for *individual cell* spike train classification to evaluate neural decoding models.

We have shown that the deep learning models which tend to perform well on other decoding tasks could not outperform a classical machine learning baseline comprised of massive time-series feature extraction from different spike train encodings coupled with well-performing classification approaches such as gradient boosting. We have thus established a strong baseline for neuronal spike train classification and we hope that this would further drive the advancement of models/architectures for neural decoding. Furthermore, we have shown that the firing of individual neurons contains information about the global state of the organism as well as the information about the neuron type that can be decoded with machine learning approaches. This approach was further generalized to the unsupervised (self-supervised) setting, which helped reveal interesting structural properties of the spiking data we considered, in particular the WAKE-state time-reversal invariance of the cortical activity in the fcx-1 data set. The massive time-series feature engineering approach helped detect groups of time-series features that have discriminative power over a set of different tasks in our benchmark and might thus be useful in general neural decoding tasks.

## Chapter 3

# Local circuit modeling of schizophrenia and Alzheimer's disease related cholinergic system pathologies in the PFC

Published in part as: Rooy, M., Lazarevich, I., Koukouli, F., Maskos, U., Gutkin, B. (2021). Cholinergic modulation of hierarchical inhibitory control over cortical resting state dynamics: Local circuit modeling of schizophrenia-related hypofrontality. *Current Research in Neurobiology*, 2, 100018.

To be submitted to *Cell* in part as: Koukouli, F., Zhang, C.-L., Lazarevich, I., Rooy, M., Lamotte-d'Incamps, B., Peixoto, J., Thiberge, C., Pons, S., Changeux, J.-P., Bacci, A., Gutkin, B., Schmidt-Hieber, C., Maskos, U. (2021). Nicotinic receptors mediate network dysfunction in early Alzheimer's disease.

M. R., I. L. and B. G. conceived and designed research. F.K. and U.M. designed experiments. F.K. performed experiments. M.R. and I.L. analyzed experimental data. M. R. and I. L. performed computational experiments.



Nicotinic acetylcholine receptors (nAChRs) modulate the cholinergic drive to a hierarchy of inhibitory neurons in the superficial layers of the PFC, critical to cognitive processes. It has been shown that genetic deletions of the various types of nAChRs impact the properties of neuronal activity in the PFC in mice during quiet wakefulness. The impact characteristics depend on specific interneuron populations expressing the manipulated receptor subtype. In addition, recent data indicate that a genetic mutation of the  $\alpha 5$  nAChR subunit, located on vasoactive intestinal polypeptide (VIP) inhibitory neurons, the rs16969968 single nucleotide polymorphism ( $\alpha 5$  SNP), plays a key role in the hypofrontality observed in schizophrenia patients carrying the SNP. Data also indicate that chronic nicotine application to  $\alpha 5$  SNP mice reverses the hypofrontality. In this Chapter, we use a computational model of PFC activity to show that the activity patterns recorded in the genetically modified mice can be explained by changes in the dynamics of the local PFC circuit. Notably, the model shows that these altered PFC circuit dynamics are due to changes in the stability structure of the activity states. We identify how this stability structure is differentially modulated by cholinergic inputs to the parvalbumin (PV), somatostatin (SOM) or the VIP inhibitory populations. We demonstrate how nicotine-induced desensitization and upregulation of the  $\beta 2$  nAChRs located on the SOM interneurons, as opposed to the activation of  $\alpha 5$  nAChRs located on VIP interneurons, is sufficient to explain the nicotine-induced activity normalization in  $\alpha 5$  SNP mice. We demonstrate that different parametrizations of the rate-based model fitting the observed experimental data result in similar predictions with regard to activity restoration in  $\alpha 5$ SNP animals under nicotine application. The model further implies that subsequent nicotine withdrawal may exacerbate the hypofrontality over and beyond one caused by the SNP mutation. We also use the developed rate-based modeling approach to find parametrizations that fit the PFC activity data in different wild-type and nAChR knock-out mice expressing the amyloid beta ( $A\beta$ ) peptide, linked to the pathogenesis of the Alzheimer's disease. The model predicted a reduction in the activity of the principal neurons of the PFC in case of enhanced  $\alpha 5$  nAChR currents, a prediction confirmed by the experimental data on galantamine application in  $A\beta$ -expressing mice.

### 3.1 Introduction

Alteration in the resting activity of the prefrontal cortex (PFC) occurs at the very onset of schizophrenia [Barch et al. \[2001\]](#). Cortical acetylcholine (ACh) release exerts strong modulation of the PFC via the metabotropic muscarinic as well as the ionotropic nicotinic acetylcholine receptors. In this work, we focus on modelling the role of nicotinic acetylcholine receptors (nAChRs) in structuring the PFC resting state activity. These receptors are ligand-gated ion channels that mediate depolarizing current in response to ACh and nicotine. Within the layer II/III of the PFC nAChRs are specifically expressed on a hierarchically organized circuit of inhibitory neurons [Bloem et al. \[2014\]](#). The subunit composition of these receptors determines their properties and their specific expression targets among the interneuronal populations. Individuals with nAChR gene variants appear to be susceptible to mental disorders and cognitive deficits [Sinkus et al. \[2015\]](#), [Koukouli and Changeux \[2020\]](#). Notably, a mutation of the  $\alpha 5$  nAChR subunit, the  $\alpha 5$  SNP, has been observed in a subpopulation of patients with schizophrenia [Maskos \[2020\]](#). This mutation has been specifically linked to nicotine addiction and to a functional cortical deficit, hypofrontality, a characteristic of schizophrenia patients [Koukouli et al. \[2017\]](#), [Hong et al. \[2010\]](#). Experiments in mice  $\alpha 5$  SNP mutation show that the resting-state PFC neural activity exhibits a reduction that is qualitatively similar to the hypofrontality seen in humans. This tell-tale hypofrontality is reversed after 7 days of chronic nicotine application [Koukouli et al. \[2017\]](#). In this work we build upon these findings and use computational modelling to support a specific hypothesis for the generative mechanism of the hypofrontality – that such hypofrontality is generated by a nicotinic receptor pathology in a local cortical circuit in the superficial layers of the PFC. More specifically, our model tests the hypothesis that the  $\alpha 5$  SNP alters the hierarchical interneuronal inhibitory/disinhibitory layer II/III subcircuits within this local PFC circuit.

During quiet wakefulness, neural activity in layers II/III of the mouse PFC is characterized by synchronous ultra-slow fluctuations, with alternating periods of high and low activity [Koukouli et al. \[2016a\]](#). Genetic knock-outs (KO) of spe-

cific nAChRs subunits were shown to disrupt these ultra-slow fluctuations, leading to changes in duration of high and low activity states (H-states and L-states, respectively). These patterns of activity are functionally significant since they may optimize information transmission in the context of lowered metabolism, such as quiet wakefulness, and because they could play an important role in memory consolidation processes [Droste and Lindner \[2017\]](#). Furthermore, multi-stable dynamics in recurrent networks have been suggested to play a crucial role in working memory and decision-making processes [Durstewitz et al. \[2000\]](#), [Wong and Wang \[2006\]](#), as well as in disease [Koukouli et al. \[2016b\]](#). Interestingly, patterns of hypofrontality in schizophrenia are associated with working memory deficits [Carter et al. \[1998\]](#), hypothesized to be a core feature of the disease [Lee and Park \[2005\]](#).

Using the designed and validated local circuit computational model, we studied the modulatory role of the cholinergic inputs to the layer II/III GABAergic interneurons, mediated by the different nAChR subtypes. Specifically we focused on the nicotinic influence on the PFC ultra-slow fluctuations and the changes of the neural firing rate dynamics seen in mice with altered nAChR gene function. We used our model to examine the chronic nicotine impact on specific nAChR subunit-types to pinpoint the principal target of nicotine-dependent restoration of neural activity in  $\alpha 5$  SNP mice. These modelling results may lend support to the self-medication hypothesis for smoking in schizophrenia patients as previously suggested in [Koukouli et al. \[2017\]](#). Furthermore, we used our model to predict the consequence of nicotine withdrawal following chronic nicotine applications in  $\alpha 5$  SNP mice. Our model showed a significant reduction in the PFC activity for this phenotype during nicotine withdrawal.

We also applied our modeling framework to the imaging activity data in mice expressing the amyloid beta peptide, which is known to be linked to the pathogenesis of the the Alzheimer’s disease. The AD-like deficits in mice were elicited using an adeno-associated viral vector expressing the human mutated amyloid precursor protein (AAV-hAPP) [Koukouli et al. \[2016b\]](#). We have demonstrated experimentally that the differential disruption of nAChR subtypes results in substantial deficits in network activity, and found parametrizations of the rate-based neural population

model that reproduced activity levels in different knock-out mice groups, including the ones expressing APP. We then use different fitted parametrizations of the model to predict the changes in pyramidal neuron population activity under different simulated nAChR manipulations. We found that a block of  $\beta$ 2-containing nAChRs and enhancement of  $\alpha$ 5-containing nAChRs leads to activity reduction relative to wild-type APP animals, an effect confirmed experimentally.

## 3.2 Methods

### 3.2.1 Neural population rate model

The model describes the dynamics of the firing rates of the various neuronal populations ( $r_e$ ,  $r_p$ ,  $r_s$ , and  $r_v$ , for PYR, PV, SOM and VIP neurons, respectively) in a local PFC circuit using a generalization of the Wilson-Cowan model [Papasavvas et al. \[2015\]](#). We follow the theoretical framework for dominant subtractive vs. divisive inhibition by the SOM and PV interneuron populations, respectively [Chance and Abbott \[2000\]](#). Following [Chance and Abbott \[2000\]](#), we can heuristically justify this modelling choice, based on the relative location of the SOM and PV synapses on the layer II/III pyramidal neurons. SOM synapses are located largely on the distal part of the dendritic tree. They are electrotonically distant from the soma, hence their impact on the cell's activity is mostly subtractive (a de facto hyperpolarizing current). The PV interneurons tend to furnish synapses peri-somatically and onto the proximal dendritic segments, hence the synaptic conductance effects would yield a de facto divisive inhibition (synaptic-activation dependent reduction of the input resistance and change in the input-output pyramidal gain). See below for further discussion.

We further implemented structured subtractive inhibitory-inhibitory connections between SOM, VIP and PV interneuronal populations [Papasavvas et al. \[2015\]](#). The full set of equations describing the firing rate dynamics of all the populations can be written as

$$\begin{cases} \tau_s \frac{dr_e}{dt} = -r_e + (A_e k_e (\omega_{pe} r_p) - r_e) F_e \left( \frac{\omega_{ee} r_e / A_e - (1 - k_d) \omega_{pe} r_p / A_p - \omega_{se} r_s / A_s + I_{ext-e}}{1 + k_d \omega_{pe} r_p / A_p} \right) + \sigma_s \xi(t) \\ \tau_s \frac{dr_p}{dt} = -r_p + (A_p k_p - r_p) F_p \left( \frac{\omega_{ep} r_e}{A_e} - \frac{\omega_{pp} r_p}{A_p} - \frac{\omega_{vp} r_v}{A_v} + I_{ext-p} \right) + \sigma_s \xi(t) \\ \tau_s \frac{dr_s}{dt} = -r_s + (A_s k_s - r_s) F_s \left( \frac{\omega_{es} r_e}{A_e} - \frac{\omega_{vs} r_v}{A_v} + I_{ext-s} \right) + \sigma_s \xi(t) \\ \tau_s \frac{dr_v}{dt} = -r_v + (A_v k_v - r_v) F_v \left( \frac{\omega_{ev} r_e}{A_e} - \frac{\omega_{sv} r_s}{A_s} + I_{ext-v} \right) + \sigma_s \xi(t) \end{cases} \quad (3.1)$$

with  $I_{ext-p} = I_{0-p} + I_{\alpha 7-p} + I_{adapt}$ ,  $I_{ext-s} = I_{0-s} + I_{\alpha 7-s} + I_{\beta 2-s}$  and  $I_{ext-v} = I_{0-v} + I_{\alpha 5-v}$ .  $I_{\alpha 7-p}$ ,  $I_{\alpha 7-s}$ ,  $I_{\beta 2-s}$  and  $I_{\alpha 5-v}$  are cholinergic external currents regulated by nAChRs.  $I_{0-p}$ ,  $I_{0-s}$  and  $I_{0-v}$  are non-specific constant external currents.  $I_{adapt}$  is the adaptation current determined by the equation 2 below.

We set  $\tau_s = 20$  ms, close to each population type's membrane time constant Pfeiffer et al. [2013].  $F_x$  is a sigmoid response function characteristic of an excitatory (inhibitory) population, which gives a nonlinear relationship between input currents to a population, and its output firing rate.  $k_x$  modulate the amplitude of the firing rate response to an input current, dependent on PV activity for PYR neurons. See Pappasavvas et al. [2015] for the details of  $F_x$  and  $k_x$  functions, which are set by two constants,  $\theta_x$  (minimum displacement) and  $\alpha_x$  (maximum slope).

We note that most of the PV interneurons are Basket cells, and thus have a lower input resistance compared to PYR neurons Beierlein et al. [2003]. On the other hand, SOM interneurons, mostly Martinotti cells, and VIP interneurons have higher input resistance compared to PYR neurons Beierlein et al. [2003]. The highest input resistances are recorded in VIP interneurons Beierlein et al. [2003]. Hence, we changed the maximal slopes  $\alpha_x$  according to those experimental findings.

As mentioned above, we modelled PV modulation of PYR activity as a combination of both divisive and subtractive inhibition, which can be thought as more biologically realistic Jadi et al. [2012], Wilson et al. [2012]. PV interneurons were specifically found to exert a divisive inhibition effect, which is assumed to be caused by powerful somatic, rather than dendritic, inhibition Wilson et al. [2012]. An additional constant parameter  $k_d = 0.8$  is introduced in the model in order to express the fraction of divisive modulation that is delivered to the excitatory population.

The rest of the modulation,  $1 - k_d$ , is delivered as subtractive.

The  $\omega_{xx}$  (with  $x \in \{e, p\}$ ) are the self-excitatory (or self-inhibitory) synaptic coupling weights of the excitatory (inhibitory) neural populations. The  $\omega_{xy}$  (with  $x \neq y, x \in \{e, p, s, v\}$  and  $y \in \{e, p, s, v\}$ ) are the excitatory (or inhibitory) synaptic coupling weights from one population to another. We did not consider self-inhibition in the SOM and VIP interneuron populations, since inhibitory chemical synapses between those neurons are rarely observed [Pfeffer et al. \[2013\]](#), [Tremblay et al. \[2016\]](#), and did not consider the direct SOM-PV connections in the main body of the paper, since regional discrepancies have been reported [Pfeffer et al. \[2013\]](#), [Gibson et al. \[1999\]](#), [Ma et al. \[2012\]](#), [Hu et al. \[2011\]](#). However, we looked at the impact of these connections parametrically (see Fig. 3-5) and found that for a range of connectivity strengths our results were not significantly altered. The parameter  $\sigma_s$  controls the strength of fluctuations of neural populations' firing rate, and  $\xi(t)$  is a Gaussian white noise with mean 0 and variance 1.

As in [Papasavvas et al. \[2015\]](#), activities of the various neural populations were normalized in the range 0 to 0.5. We added factors ( $A_e, A_p, A_s, A_v$ ) in order to fit the range of firing rates in the model to experimental values.

In order to model spike frequency adaptation in the PYR neuron population of the network model, we used the following equation, adapted from [Hayut et al. \[2011\]](#):

$$\tau_{adapt} \frac{dI_{adapt}}{dt} = -I_{adapt} + r_e J_{adapt} \quad (3.2)$$

where  $\tau_{adapt}$  and  $J_{adapt}$  are the adaptation time constant and the adaptation strength of the PYR population. We chose  $\tau_{adapt} = 600$  ms according to [Destexhe \[2009\]](#). The resulting adaptation current  $I_{adapt}$  was added to the right hand side of the equation for the activity of the pyramidal population in eq. 1 above.

The modelling framework for the input terms dependent on the nicotinic acetylcholine receptors is described below.

### 3.2.2 Model parameter search procedure

The random search algorithm sampled  $10^6$  random parameter sets (points in parameter space), in a fixed value range for each parameter (values 1.0-55.0 for synaptic connection strengths, and 0.1 to 0.55 for external input currents). A single constraint was added to the parameter values of the model at this stage: VIP to PV connection strength value was set to be lower (by 50%) than VIP to SOM connection strength value [Pi et al. \[2013\]](#). After that, the parameter sets sampled by the random search algorithm were filtered out through a set of consecutive constraints:

- First, we performed a basic sanity check for each parameter set to determine if it could produce the bistable firing rate dynamics for the WT (baseline) parameter values. To do so, we computed the roots of the non-linear set of equations  $dr_e/dt = 0$ ,  $dr_p/dt = 0$ ,  $dr_s/dt = 0$ ,  $dr_v/dt = 0$ , using MINPACK's `hybrd` and `hybrj` algorithms. We then selected the parameter sets that corresponded to the three roots of the above equations, hence two stable steady states of firing rate activity. These parameter sets represent  $\sim 6\%$  of the total number of tested value sets.
- Second, we selected connection sets for which the normalized H-state is far from saturation ( $< 0.45$ ), and with a higher activity in H-state than L-state for all neuron types, so that the transitions between H-states and L-states are simultaneous across cell types. 93%, 61% and 96% of networks exhibited a higher PV, SOM and VIP activity all together in PYR high activity state (H-state) compared to the low activity state (L-state), according to experimental data.
- During the third step, parameters were selected based on residual error values for H- to L-state firing rate ratios for the simulated values versus the experimental ones. We fitted the scaling factors  $A_e, A_p, A_s, A_v$  so that the H-state for each simulated population type corresponded to the one found experimentally (to the median firing rate during the H-state in the experiments). We selected the networks for which the absolute error of the L-state rate level between the

model and experiments did not exceed 5 spikes/min for each neural population type.

- From those selected networks, we simulated the firing rate time evolution of all neural types, in order to compute the distribution of H-state and L-state durations (with at least 500 state transitions in each simulation) for varying levels of noise  $\sigma_s$  (values between 0.001 and 0.02). For each given parameter set, we selected the noise level that best reproduced the mean and mode of H-state and L-state duration distributions in terms of the summed mean squared error. Then, for the matched noise level value, we selected parameter sets with low mean absolute percentage error (MAPE) values for the properties of ultraslow fluctuations extracted from the experimental data in WT animals. Parameter sets corresponding to MAPE values below 100% were taken.

At this step of the parameter selection pipeline, we were left with  $\sim 50$  parameter set candidates out of the initially sampled  $10^6$  parameter sets. This "training" stage of the parameter selection procedure allowed us to effectively downselect the large initial collection of models to several plausible ones using only the WT experimental data to formulate constraints.

In order to select one final of the 50 last parameter set candidates, we introduced an additional "validation" selection stage, which determined the model with the best predictive power outside the previously used WT data. Instead of pure WT data, for the validation stage we looked at the change of H-state firing rate levels between WT and  $\alpha 5$  KO states. We chose the network in which the change of SOM H-state firing rate was the closest to the experimental value in terms of absolute error. Further on, we tested the selected parameter set against experimental data for all neural populations across different KO states. As we demonstrate, the selected model performs quite well during testing on KO states and predicts the key features of experimental data. Note that the other candidate parameter sets found during our search procedure also have relatively low values of the WT fitting error and decent performance when predicting activity levels in different knockout variants we considered (see Fig. 3-3). These parameter sets were selected based on



the above procedure as leading to small values of the WT fitting error and well predicting the change in the SOM H-state firing rate in the  $\alpha 5$  knockout state relative to the WT state. The set of candidate model parameters providing a good fit was found to span a relatively wide area in parameter space, however we found that all the major predictions inferred from the parameter set we ended up selecting were well reproduced qualitatively by the other found parameter sets. In particular, predictions concerning pyramidal population activity changes under nicotine treatment in wild-type and  $\alpha 5$  SNP animals were found to be similar for different candidate parameter sets (see Fig. 3-3E).

### 3.2.3 Modeling nAChRs

We used a minimal model of subtype-specific activation and sensitization of nAChRs Graupner et al. [2013], from which we determined the amplitude of each cholinergic current:

$$I_{x-n} = w_{x-n} a_x s_x, w_n = 0.01 N_{x-n}, x \in \{\alpha 4\beta 2, \alpha 5\alpha 4\beta 2, \alpha 7\}, n \in \{s, p, v\} \quad (3.3)$$

where  $a_x$  is the activation variable, while  $s_x$  is the sensitization variable. They both take values between 0 and 1. The receptor is fully activated for  $a_x = 1$  and fully sensitized for  $s_x = 1$ , while it is closed for  $a_x = 0$  and fully desensitized for  $s_x = 0$ . We did not take into account desensitization of nAChRs by physiological levels of ACh because of its rapid breakdown through acetylcholinesterase Graupner et al. [2013], Dani et al. [2001], such that  $s_x = 1$ . So the desensitization plays a role only for the nicotine simulations.

Using formulas from Graupner et al. [2013], with  $ACh = 1.77 \mu M$ , relevant for in vivo simulations, we found  $a_{\alpha 4\beta 2} = a_{\alpha 5\alpha 4\beta 2} = 0.0487$ , and  $a_{\alpha 7} = 0.0014$ .  $N_{x-n}$  is the number of nAChRs of each type that best reproduced the changes of activity patterns recorded in experimental data (see Fig. 3-7A1-A2-A3). For the nicotine application simulations, we used a physiologically relevant blood concentration for smokers  $Nic = 1 \mu M$  and computed the change in sensitization and activation of

different nAChR types. For  $\alpha 5\alpha 4\beta 2$  nAChRs, more resistant to desensitization, we shifted  $DC_{50}$  from 61 nM to 610 nM [Vyazovskiy and Harris \[2013\]](#).

In the main body of the paper, we modeled the different KO mice by setting the corresponding cholinergic currents to zero. We also studied how the results depend on this assumption and found that the same trends in activity changes can be observed for small non-zero amplitudes of the nAChR currents, see Fig. 2. In particular, for the modelled  $\alpha 5$  SNP mice, expressing a human polymorphism of the  $\alpha 5$  nAChRs associated with both schizophrenia and heavy smoking, decreasing the VIP-expressed  $\alpha 5$  receptor activation by 30% was sufficient to give results compatible with the experiments.

## 3.3 Results

### 3.3.1 Summary of the experimental approach and results

We studied experimentally the spontaneous activity of neurons in the prelimbic cortex of PFC in awake mice by two-photon calcium imaging [Koukouli et al. \[2017\]](#). Male  $\alpha 7$ KO,  $\beta 2$ KO,  $\alpha 5$ KO and WT C57BL/6J mice were used and experiments were performed at 3 months of age. Mice engineered to harbor the  $\alpha 5$  D398N variant ( $\alpha 5$ SNP mice) were obtained via homologous recombination. Briefly, a chronic cranial window was prepared and 200 nl of AAV1.syn.GCaMP6f.WPRE.SV40 were injected bilaterally in the prelimbic cortex (PrL) (coordinates: AP, +2.8 mm from the bregma; L,  $\pm 0.5$  mm; DV, -0.3 to -0.1 mm from the skull) for recordings of pyramidal neurons. To record the activity of interneurons we used Cre mouse lines (VIP-Cre, SOM-Cre and PV-Cre mouse lines) and 200 nl of AAV1.syn.Flex.GCaMP6f.WPRE.SV40 were injected as before. A sterile small stainless steel bar was embedded over the cerebellum in order to head-fix the mouse for imaging. Mice were trained for awake imaging by gentle handling for 4 days and were habituated to rest in a support tube on the mouse stage during the recordings, as previously described [Koukouli et al. \[2017\]](#). *In vivo* imaging was performed using an Ultima IV two-photon laser-scanning microscope system (Bruker). For pyramidal neurons expressing GCaMP6f, time-series movies were acquired at the frame rate of almost 7

Hz, with a movie duration of approximately 215 seconds. The frame rate used for the interneuron recordings was about 30 Hz and the duration of each focal plane movie was approximately 165 seconds. Detection of individual neuron  $\text{Ca}^{2+}$  transients was performed automatically using a preprocessing and deconvolution pipeline written in MATLAB (Mathworks) Koukouli et al. [2017].

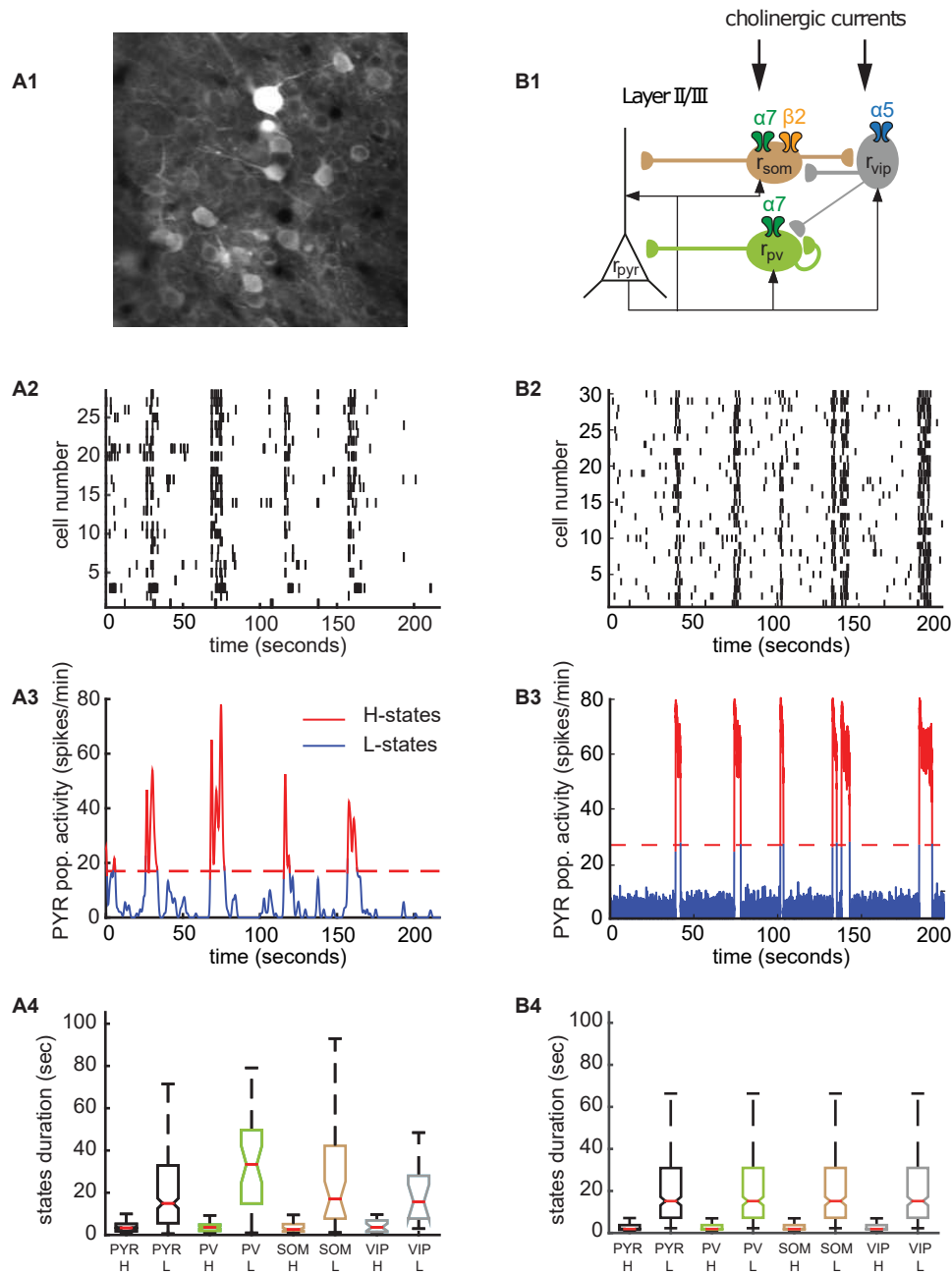


Figure 3-1 – (Caption next page.)

Figure 3-1 – (Previous page.) Bistable firing rate dynamics of interconnected neural populations replicates ultra-slow fluctuations recorded in the PFC of WT mice. (A1) Two-photon image of GCaMP6f expressing neurons, from Koukouli et al. [2017]. (Scale bar: 20  $\mu m$ ). (A2) Inferred events of a population of simultaneously recorded cells in a WT mouse, obtained by deconvolving the spontaneous Ca<sup>2+</sup>-transients. 80% of the recorded cells are PYR neurons Koukouli et al. [2017]. (A3) Time varying population mean activity of the neurons shown in A2. The dashed red line delineates the threshold between high and low activity states (H-states and L-states, respectively). Red periods correspond to H-states and blue periods to L-states. See Koukouli et al. [2016a] for more info on the methods. (A4) H-state and L-state durations recorded in the different neuron types. H-state durations are similar between neuron types (mean of 3 seconds, no statistical differences), as well as L-state durations (mean of 20 seconds, no statistical differences). (B1) Schematic of the studied circuitry. (B2) A simulated rastergram of neuronal activity, generated for illustration purposes from the population rate model, using a Poisson process with  $\lambda(t)=mean\_population\_rate(t)$ . (B3) Time varying mean population activity of pyramidal neurons, computed from the network model. We use the same method as Koukouli et al. [2016a] to delineate H-states (in red) and L-states (in blue). (B4) The model reproduces the similar H-state and L-state durations across neural types. Distribution of a total of 500 state durations collected are plotted.

Briefly, we identified the temporal structure of the spontaneous activity in the layer II/III as alternating between low and high activity states (see Fig. 3-1A2-3 for an example of transition in a WT mouse). The typical time-scale of the high activity states (H-states) and low activity states (L-states) was on the order of several to tens of seconds. We then studied how the average activity and the temporal structure is altered by the experimentally induced genetic mutations of the nAChRs. Notably, we showed that mice expressing the human  $\alpha 5$ SNP exhibit reduced pyramidal cell activity. Our results also showed that the different nAChR subunits control the spontaneous PFC activity through a hierarchical inhibitory circuit. Specifically, in  $\alpha 5$ SNP mice and  $\alpha 5$  KO mice, lower activity of VIP-expressing interneurons appeared to result in an increased SOM-interneuron inhibitory drive over the pyramidal neurons. Chronic nicotine administration reversed the hypofrontality observed in  $\alpha 5$ SNP mice through possible desensitization of  $\beta 2$ -containing nAChRs in SOM interneurons. Specific experimental data on the SNP as well as other KO mice will be shown below alongside the simulations for comparison and discussion of the circuit mechanisms.

### 3.3.2 Modelling formalism and strategy

For our local circuit model we used a neural population approach first pioneered by Wilson and Cowan [Destexhe and Sejnowski \[2009\]](#) and subsequently widely used in modelling studies. In this approach, the firing rates of cell ensembles are taken as the dynamical variables of the model (as opposed to modelling the biophysics of individual cells). Each variable in the model represents the activity of the specific cell-type population. The circuit is constructed by weighted connections between the cell populations so that the inputs to a given population represent the summarized synaptic connectivity between the neural populations. External inputs to the circuit can be included in a similar way. These inputs are then put through a non-linear input-output function specific for each neuronal population.

Central to our work was to explicitly include the influence of the nicotinic cholinergic modulation exerted on specific cellular targets of the layer II/III PFC circuitry. We have previously shown how acetylcholine-dependent currents mediated by the nAChRs can be incorporated in the population rate models using simple kinetic schemes modified from [Katz and Thesleff \[1957\]](#) (see Methods and [Koukoulis et al. \[2017\]](#) for more details). The key point is that these currents can be parameterized to reflect the pharmacological and electrophysiological properties of specific receptor subtypes (temporal scales and ligand affinities for the activation and the desensitization) and incorporated into the specific neural population dynamics, reflecting their expression targets. We can then use the model to perform analysis, parameter fitting, validation and *in silico* genetic manipulation to understand how the local circuit dynamic mechanisms could explain the observed data. In other words, the strategy is to use a highly reduced model that distills away much biological complexity.

The logic for using such a highly reduced modelling approach is two-fold. First of all, the calcium imaging data analysis focused mainly on the dynamics and alterations of the mean activity of the recorded cell populations. Second, after incorporating the proposed structure of the local circuit and the relevant receptor-mediated currents, we were able to arrive at a model that was sufficient to explain key aspects of the data and yet was still tractable and understandable.

Overall, the circuit model was structured as shown in Figure 3-1B. It reflects the hierarchical structure of the local inhibitory circuitry in the PFC layer II/III . Inspired by [Chance and Abbott \[2000\]](#) we used an extended Wilson-Cowan formalism to account for two kinds of inhibition impinging on the pyramidal neuron population: divisive inhibition due the PV interneurons and subtractive inhibition due to the SOM interneurons. Please note that the effect of PV interneurons was modelled as a mix of subtractive and divisive inhibition, reflecting that these neurons target the pyramidal cells peri-somatically and hence exert a shunting effect [Jadi et al. \[2012\]](#). The ratio of divisive to subtractive inhibition in the model is controlled by the parameter  $k_d$ , Fig. 3-4 shows the robustness of the obtained results to the value of this parameter. The divisive inhibition acts to modulate the gain of PYR response, while the subtractive inhibition shifts this response. In fact one can heuristically think of two recurrent PYR-interneuronal pathways in this circuit: the PYR-PV divisive inhibitory one and the net additive disinhibitory one through the PYR-VIP-SOM neurons. We further included the inhibitory interactions between these sub-circuits. The specific neuronal population variables were also coupled to the nAChR models as indicated by the circuit scheme. As we will see below, the interplay of the recurrent excitation with the multiple interneuronal sub-circuits can lead to non-trivial dynamical outcomes in the model. The equations are described in full detail in the Methods.

Our strategy was to identify a model (i.e. the set of parameter values) that was able to account for the experimental data by first fitting the model parameters to match the control data of ongoing spontaneous activity, and then subjecting it to the simulated variations reflecting the genetic manipulations of the nAChRs. To do so we made an ansatz that the spontaneous high-low activity state alterations are due to a bistability in the local cortical circuit and the switching is controlled by random noise and firing rate adaptation in the PYR population. In order to identify the model regimes that could account for the experimental data, we used a multi-step model selection (fitting and validation) procedure (see Methods). First, we performed a semi-analytic analysis of the model dynamics as a function of the various inhibition strengths (see below). These results were then used as a guide for

a global search procedure to identify a subset of model parameters that exhibited bistable dynamics. We then selected the model parameters so as to minimize the quantitative discrepancy (mean error) between the observed control experimental activity data and the data obtained by simulating the model. We found that the connectivity parameters that minimize this error were generally widely distributed. We thus added further constraints on the model parameters. Since we were ultimately interested in accounting for the effects of the  $\alpha 5$ -SNP on the population firing patterns, we further selected the parameters so that we could focus on those models that could produce the mean activity changes when the ACh input to the VIP neurons was turned off. See Fig.3-3 for the distribution of the fitting (control activity) and validation (knockout activity) errors across different parameters sets.

We then performed parametric manipulations of the receptor models to reflect the genetic alterations in the  $\alpha 7$ KO and  $\beta 2$ KO animals (to validate the model on obtained experimental data), as well as to measure the influence of nicotine in those phenotypes. Below we present the model predictions for average firing activity level alterations in manipulated animals.

The details of model fitting and validation procedures are given in the Methods section; details of the calcium imaging analysis are given in Koukouli et al. [2017]. Further simulation results are given in the methods and the Supplement. Our model strategy and model selection procedures allowed us to potentially identify the local circuit pathways linking the genetic alterations and the *in vivo* PFC activity observations. Furthermore, we could use the model to profile future experiments and make predictions on the effects of nicotine withdrawal in the WT and  $\alpha 5$ SNP animal phenotypes.

Furthermore, we relaxed our model selection and examined multiple parametric sets of the model, to understand how the impact of the nicotinic receptor manipulations are distributed across multiple parameter sets (see Fig. 3-3). Indeed we found that there exist several different parameter sets that could predict the activity changes during simulated nAChR knockouts while giving a slightly worse fit in terms of the control wild-type data. Moreover, we found that the predictions regarding activity shifts during nicotine treatment were consistent across different parameter

sets (see Fig. 3-3). Hence, the parameter set that we selected to obtain the main model predictions (marked as a red dot in Fig. 3-3) is representative of the whole model population.

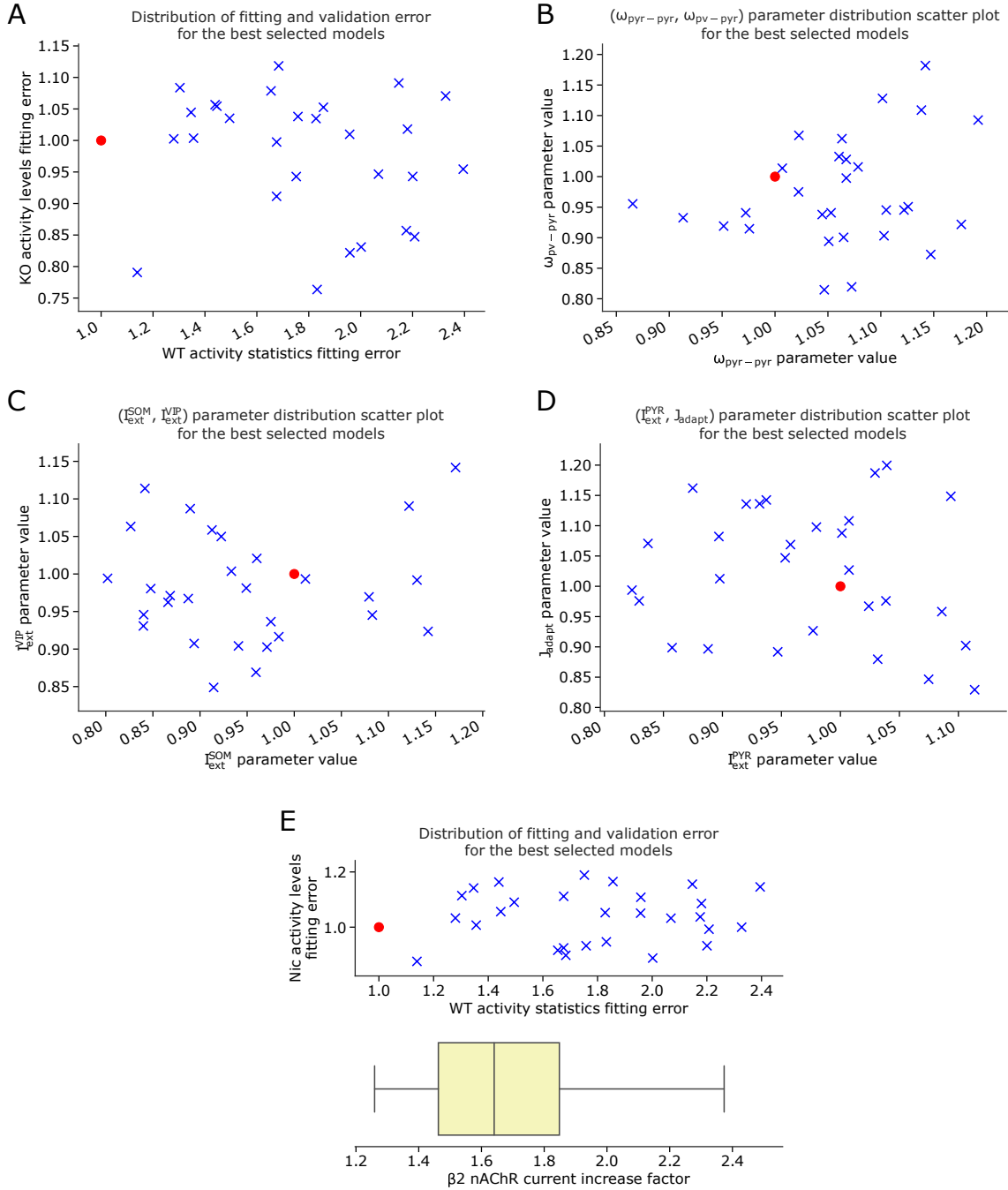


Figure 3-2 – (Caption next page.)



Figure 3-3 – (Previous page). The distribution of fitting (WT) and validation (KO) errors along with some of the parameter values for the set of candidate models found during parameter optimization. The main model candidate chosen in the paper is denoted with the red color. (A) Scatter plot for a range of candidate model parameter sets whereby every parameter set is represented by its WT activity statistics fitting error (MAPE of WT high-low activity statistics) and the KO activity prediction error (MAPE of PYR firing rates in  $\alpha 5$ ,  $\alpha 7$  and  $\beta 2$  knockout states, see Fig. 4A1), normalized to the error values of the selected parameter set (denoted with red color), (B) scatter plot of  $\omega_{ee}$ ,  $\omega_{pe}$  parameter values for the candidate parameter sets normalized to the selected values, (C) same for  $I_{ext-s}$ ,  $I_{ext-v}$  parameters, (D) same for  $I_{ext-e}$ ,  $J_{adapt}$  parameters, (E) distributions of the WT fitting errors and nicotine treatment activity level prediction errors (error in predicted change in WT and  $\alpha 5$ SNP PYR firing rates after nicotine treatment), along with the  $\beta 2$  nAChR enhancement factors required for the candidate models to reproduce the normalization of  $\alpha 5$ SNP activity to WT levels under nicotine treatment.

### 3.3.3 Bistable layer II/ III local PFC circuit firing rate dynamics replicate ultraslow fluctuations in WT mice

The basis for this modelling study are our experimental studies, briefly described above, of the spontaneous activity in the layers II/III of the prelimbic cortex in awake mice by two-photon calcium imaging [Koukoulis et al. \[2017\]](#). As shown by an example in Fig. 3-1A2-3, a population of simultaneously recorded cells in a WT mouse exhibits transitions between high activity states (H-states) and low activity states (L-states) lasting several to tens of seconds. We then set out to model the local circuitry that may produce this activity pattern. The circuit model sketched out in Fig. 3-1B1 simulates the firing rate evolution of populations of pyramidal (PYR) neurons intercoupled with a hierarchy of interneurons. Parvalbumin (PV) interneurons, expressing  $\alpha 7$  nAChRs subunits [Bloem et al. \[2014\]](#), target PYR cells axosomatically, with strong reciprocal connections [Holmgren et al. \[2003\]](#). Somatostatin (SOM) interneurons, expressing both  $\alpha 7$  and  $\alpha 4\beta 2$  nAChRs subunits [Bloem et al. \[2014\]](#), target the dendrites of the PYR cells. The  $\alpha 5\alpha 4\beta 2$  nAChRs subunits are expressed only by vasoactive intestinal polypeptide (VIP) interneurons [Porter et al. \[1999\]](#), that preferentially inhibit the SOM cells, and to a lesser extent PV cells [Pi et al. \[2013\]](#). Both the SOM and the VIP interneurons receive excitatory feedback from the PYR neurons [Silberberg and Markram \[2007\]](#), [Porter et al. \[1998\]](#). The

model is able to reproduce the ultra-slow fluctuations of PYR population activity recorded in WT mice (Fig. 3-1B2-3) by assuming that two stable states of activity arise from the connectivity between neural populations (see Methods for more information on the model and fitting procedure). Simultaneous network transitions between activity states, for all neural types, drive the activity fluctuations. This prerequisite is consistent with our experimental findings, showing that the various neuron types have similar H-state and L-state durations (Fig. 3-1A4, Fig. 3-1B4). Establishing that the local circuit model can replicate central properties of the WT data provided us with a computational model platform to turn to the data from the genetically modified animals.

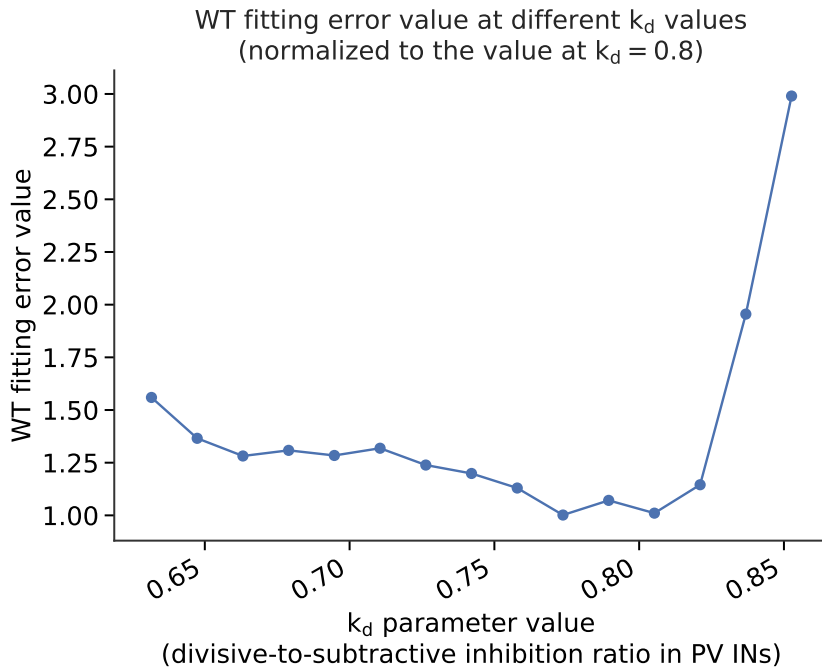


Figure 3-4 – Evolution of WT activity statistics fitting error (normalized to the error value at  $k_d=0.8$ ) for different values of the divisive-to-subtractive inhibition ratio  $k_d$  while the other parameter values in the model are fixed (corresponding to the main chosen parameter set). The default value of  $k_d$  is 0.8 for the fitted model.

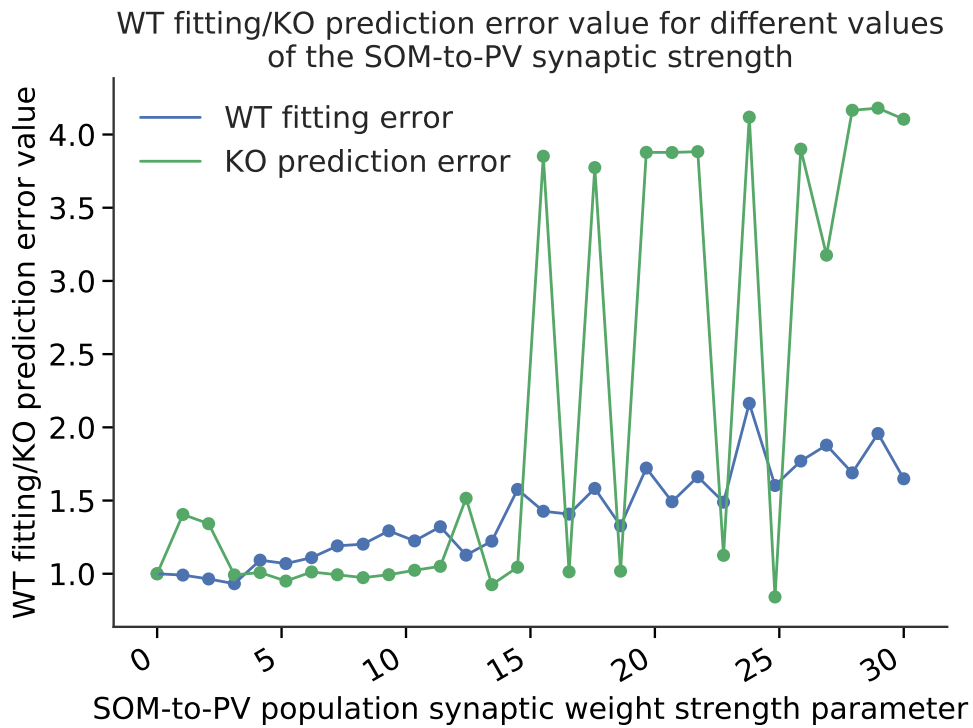


Figure 3-5 – Evolution of the normalized WT statistics fitting and KO activity level prediction errors as a function of the SOM-PV synaptic connection strength added to the model with the main chosen parameter set. Note that model outputs are not significantly affected in terms of the fitting and validation errors for low values of the SOM-PV connection strength. The maximal value of the synaptic strength parameters that we considered in the model was equal to about 50 (in arbitrary units).

### 3.3.4 Heuristic analysis of impact of inhibitory population activity variation on network state stability.

Before proceeding to modelling the genetic nAChR manipulations, we wanted to develop an intuition about the influence of the different inhibitory inputs on the dynamics of the pyramidal neuron population using a reduced feed-forward inhibitory circuit (see Supplementary Materials). Here we explicitly differentiated the subtractive (SOM) versus the divisive (PV) inhibition. The PV-divisive inhibition controls the gain of the PYR input-output function. Our analysis showed that reducing the PV-divisive inhibition of the PYR population, increases this gain and makes both the high and the low states more stable. Therefore, decreases in the divisive inhibition should lead to increases in the state-durations during the spontaneous activity. Reducing the subtractive inhibition of the PYR population shifts the PYR activation function (without changing its shape). This leads to the low activity state becoming less stable and the high activity state gaining in stability. In other words, decreasing the SOM-dependent subtractive inhibition increases the duration of the H-states and decreases the duration of the L-states. Since the VIP neurons project to the SOMs and inhibit them, decreases in VIP activity lead to L-states increasing their durations and H-states becoming shorter on average. In summary, the preliminary analysis points out that SOM and VIP activity decrements should have an opposing effect on the PYR activity: former increasing it and latter decreasing it.

Taking the above into account, we now turn to the fully connected network, where we take into account the excitatory feedback from pyramidal neurons to the various interneurons subtypes, as well as the inhibitory inputs from VIP to PV neurons. We found that generally the simplified network intuition (see Supplementary Materials) holds for the full network (Fig. 3-6). The full model shows the expected increase of both H-state and L-state duration for decreased external inputs to PV population (Fig. 3-6B1). When the external inputs to the SOM population is decreased, we saw the predicted increase of H-state- and a decrease of L-state-duration for (Fig. 3-6B2). A decrease of H-state duration was seen for decreased external inputs to VIP population (Fig. 3-6C2). Note the shape of the bifurcation

diagram as a function of the external input to PV population (Fig. 3-6A1). Due to the strong excitatory feedback from PYR to PV population Lee and Park [2005], for external inputs higher than a critical value, the network loses its bistability to a single H-state. Second, we observe a slight decrease of L-state duration for decreased external inputs to the VIP population (Fig. 3-6C2). This is where the VIP-PV connection plays the defining role: the decreased VIP activity increases the PV activity in both the H- and L-states, which in turn decreases both H-state and L-state duration, due to the increased divisive inhibition impinging onto the PYR population. The results of this analysis indicate that if the knockout does not fully abolish the nAChR-mediated current, we would still observe the same qualitative behaviour in the modelled circuit, due to the smoothness and monotonicity of the activity level curves in Fig. 3-6. Interestingly, we observed the same trends in terms of increases/decreases of L- and H-state durations and firing rates for the different sets of parameters that we found during the model search procedure (the parameter set distribution is reflected in Fig. 3-3).

### 3.3.5 Impact of the nAChR genetic manipulations on the temporal structure of the ultraslow activity fluctuations

In order to analyze the nicotinic modulation of the resting state temporal structure through the  $\beta 2$ - and  $\alpha 7$ -nAChR-mediated currents on their target inhibitory neurons, we explicitly modeled nAChR activation levels following a computational framework developed in Graupner et al. [2013] (see also Methods). Our model reflected that the various receptor subclasses are expressed on specific neuronal targets (here exclusively on the different interneuronal subtypes). The model also took into account the ligand-gated electrophysiological properties of the modelled nAChR subclasses. According to the computational framework,  $\beta 2$  nAChRs, which have high affinity to acetylcholine (ACh), activate such that their cholinergically evoked input to the target cell population is  $\sim 35$  fold the amplitude of the cholinergic current due to the  $\alpha 7$  nAChRs. Note that due to the rapid ACh breakdown by the acetylcholinesterase Dani et al. [2001], Giniatullin et al. [2005] and following our

previously developed nAChR/neural circuit modelling framework, we chose not to take into account desensitization of nAChRs by physiological levels of ACh.

We then simulated our model with the *in silico* receptor knock-outs and tracked the duration of the high and low activity states as well as the proportion of neural populations (out of 100 simulated samples of our models generated from the pre-determined model parameter set as explained above) that showed the bistable dynamics. You can see in Fig. 3-7A1 that a knock-out of  $\alpha 7$  nAChRs, located on both PV and SOM interneurons, induces an increase of H-state mean duration as compared to the WT animals,  $4.2 \pm 0.3$  to  $6.1 \pm 0.5$  sec, with  $P < 0.05$ . Knock-outs of SOM-localized  $\beta 2$  nAChRs induce increased H-state durations ( $8.7 \pm 0.5$  sec) that are much larger than those in the  $\alpha 7$  KO animals. We modeled the KO case by setting the relevant receptor-mediated currents to zero. In our model we observed that setting the  $\alpha 7$ -mediated current to zero induces an increase of mean H-states duration from  $4.0 \pm 0.2$  sec, for simulated WT animals, to  $5.6 \pm 0.3$  sec, for simulated  $\alpha 7$  KO animals ( $P < 0.001$ , see Fig. 3-7B1). The model also was able to account for the strong effect of the high affinity  $\beta 2$  nAChRs manipulation, with a more than 2-fold increase of H-state mean duration to  $9.7 \pm 0.8$  sec. The model further showed reduced H-state mean duration for the  $\alpha 5$  KOs ( $2.0 \pm 0.1$ ,  $P < 0.001$ ) compared to WT mice, similar to experimental findings ( $2.6 \pm 0.2$  sec,  $P < 0.001$ ). These results are consistent with  $\alpha 5$ -containing nAChRs having a modulatory effect on VIP activity, which in turn inhibit both PV and SOM interneurons. We found no significant changes in the H-state mean duration for  $\alpha 5$  SNP compared to WT mice, both experimentally and through modeling ( $4.0 \pm 0.3$  sec of H-state duration for experiments, and  $3.3 \pm 0.2$  sec for simulations). Our analysis above lead us to expect a significant decrease of mean L-state duration for  $\beta 2$  KO animals. Indeed we observed a drop from  $22.6 \pm 1.3$  sec to  $13.9 \pm 0.8$  sec,  $P < 0.01$  (see Fig 3-7B2), which we identify in the model as a combined effect of the decreased cholinergic input to the SOM and VIP populations, and which reproduces experimental findings ( $21.7 \pm 1.3$  sec to  $15.3 \pm 1.0$  sec,  $P < 0.001$ , Fig 3-7A2, modified from Koukouli et al. [2017]). No significant change of L-state durations was found for  $\alpha 7$ ,  $\alpha 5$  and  $\alpha 5$ SNP compared to WT, both experimentally and through modeling. This is consistent

with our prediction that VIP change of input to PV population, exerting divisive type of inhibition over PYR activity, would counterbalance VIP effects on L-state stability through SOM population (Fig. 3-6B3).

Having determined that our model can account for the mean statistics of the high-low spontaneous dynamics under various genetic manipulations of the nAChRs in the hierarchical inhibitory sub circuit of the PFC, we then examined the proportion of neural populations that would show bistable high-low dynamics in the spontaneous activity and how these are modified by the  $\beta 2$ -receptor modulation. To do so, we constructed distribution histograms of the firing rates observed for a given phenotype (see also Methods and Koukouli et al. [2017] for further analysis details) and identified what proportion of the populations showed multi-modal activity distributions. In our data (also see Koukouli et al. [2017] for fuller discussion), we found that  $65.9 \pm 5.7\%$  of populations exhibited high and low activity states transitions dynamics in  $\beta 2$  KO mice, a significant decrease compared to WT animals ( $90.3 \pm 5.0\%$  of populations,  $P < 0.05$ , Fig. 3-7A3, modified from Koukouli et al. 2017 Koukouli et al. [2017]). We simulated the  $\beta 2$  KO in our model by setting the cholinergic current strength terms to zero for the SOM neuron population. According to our model analysis, a decrease of cholinergic currents in the SOM population should induce an increase of H-state stability while a decrease of L-state stability (Fig. 3-6B2). Our simulations indeed reproduce the decreased proportion of populations exhibiting L-state/H-state transitions from  $88.6 \pm 3.6\%$  in simulated WT animals to  $64.3 \pm 4.9\%$  for simulated  $\beta 2$  KO mice (Fig. 3-7B3). Furthermore, the model predicted that knocking out either the  $\alpha 7$  or the  $\beta 2$  nAChRs should decrease the H-state mean rate amplitude, and induce no changes in the L-state amplitudes, analysis of experimental data was in fact consistent with this prediction (see Fig. 3-7C).

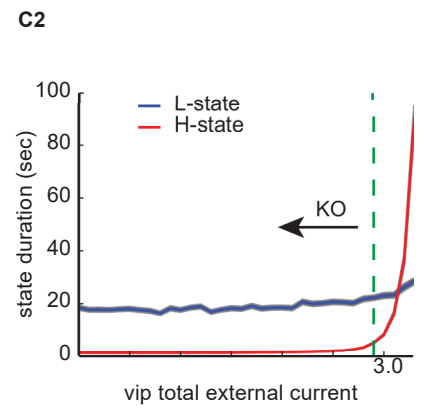
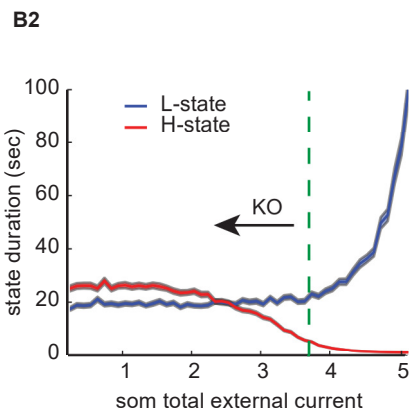
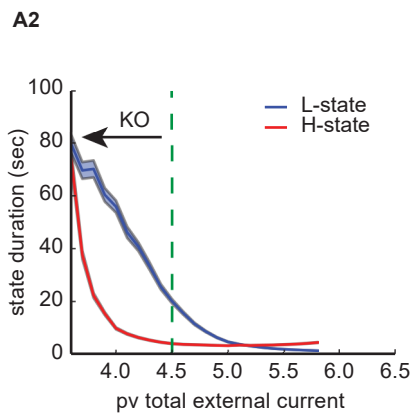
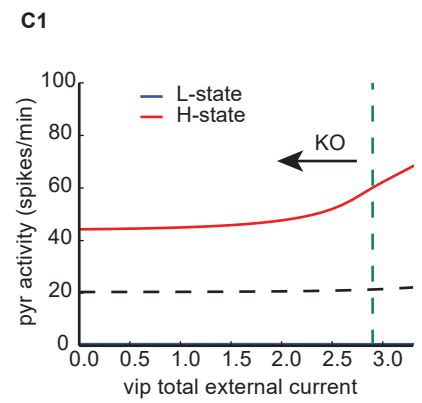
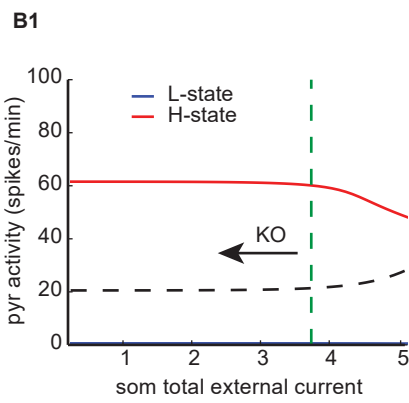
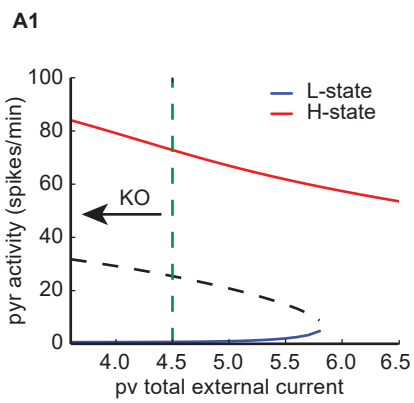
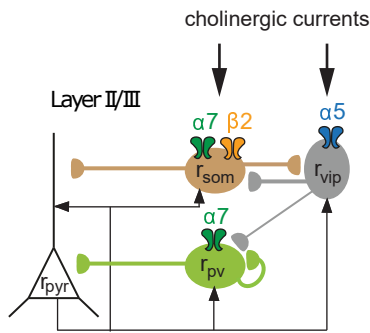


Figure 3-6 – (Caption next page.)



Figure 3-6 – (Previous page.) Effects of changing external inputs to inhibitory populations on network state stability in the fully connected network (A1) H-state (red line) and L-state (blue line) PYR activities as a function of the external input current to VIP interneurons population. The dashed green line shows the selected parameter value to reproduce WT mice neural dynamics. KO of nAChRs is associated with a decrease of external currents (black arrow). (A2) H-state (red line) and L-state (blue line) durations as a function of the external input current to PV interneurons population. The shaded areas delineate  $\pm$ sem. The dashed green line shows the selected parameter value to reproduce WT mice neural dynamics. (B1) Same as (A1), but for the external input current to SOM INs population. (B2) Same as (A2), but for the external input current to SOM INs population. (C1) Same as (A1), but for the external input current to VIP INs population. (C2) Same as (A2), but for the external input current to VIP INs population.

### 3.3.6 Layer II/III circuit model accounts for the VIP and SOM neuron firing rate changes under schizophrenia-associated $\alpha 5$ pathology.

To further confirm our hypothesis that the decreased disinhibition in  $\alpha 5$  KO and  $\alpha 5$  SNP mice accounts for hypofrontality, we compared the changes of VIP, PV, and SOM interneurons under  $\alpha 5$  knock-down in experiments and simulations. In the experiments, clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated endonuclease (Cas)9 technology was used to knock-down the  $\alpha 5$  subunits in vivo, as shown in Koukouli et al. [2017]. We implemented this manipulation in the model by decreasing the activation of the  $\alpha 5$ -associated input to the VIP interneurons. Experimentally it was found that VIP neuron median activity decreased drastically under the CRISPR technology from  $21.5 \pm 2.17$  spikes/min to  $3.5 \pm 0.7$  spikes/min ( $P < 0.001$ , see Fig. 3f in Koukouli et al. [2017]). The model accounting for these results yielded decreased H-state network durations and decreased VIP H-state activity (see Fig. 3-7A1, Fig. 3-7B1). As a result, the simulated spike frequency of VIP interneurons in VIP  $\alpha 5$  knock down mice ( $5.7 \pm 0.2$  spikes/min) was significantly lower than VIP neural activity in simulated WT animals ( $21.4 \pm 1.2$  spikes/min,  $P < 0.001$ , see Fig. 3-8A2,B2). The model predicted that the decreased VIP levels of activity should result in increased levels of SOM activity through disinhibition. Experimental findings endorsed this prediction, through a robust in-

crease in SOM interneuron spontaneous activity ( $39.1 \pm 3.1$  spikes/min) compared to control mice ( $5.6 \pm 1.3$  spikes/min,  $P < 0.001$ , see Fig. 3n in Koukoulis et al. [2017]). The model could reproduce this increase of SOM activity quantitatively (see Fig. Fig. 3-8C3) despite the significant decrease of network H-state duration (Fig. 3-7A1). Please note that the H-state durations are determined with all neuronal populations without being differentiated. In the model the over-all H-state duration decrease was associated with an unexpected increase in SOM H-state level of activity specifically. SOM WT mice simulated median activity increased from  $13.9 \pm 0.7$  spikes/min to  $32.0 \pm 1.2$  spikes/min ( $P < 0.001$ , see Fig. 3-8C3). Experimentally, the decrease in PV interneuron activity was slight and not statistically significant ( $6.1 \pm 0.6$  spikes/min) compared to WT mice ( $5.6 \pm 0.4$  spikes/min, see Fig 3j in Koukoulis et al. [2017]). We confirmed that in our model with parameters optimized to quantitatively reproduce the  $\alpha 5$  SNP affect as reviewed above, we saw only a slight decrease of PV activity in CRSPR mice, as compared to WT simulated mice, from  $7.9 \pm 0.3$  spikes/min to  $5.0 \pm 0.2$  spikes/min ( $P < 0.001$ ). Because PV interneurons receive high levels of excitatory input from PYR neurons, we would expect a decrease of excitatory input to this population in  $\alpha 5$  KO and CRISPR mice, due to decreased PYR neurons firing rate. This decreased excitatory input may be compensated by a decreased inhibition from the less active VIP interneurons, yet this connection is rather weak.

### **3.3.7 Nicotine re-normalizes $\alpha 5$ SNP PFC network activity through desensitization and upregulation of SOM $\beta 2$ nAChRs**

Experimentally it was observed that nicotine administration to  $\alpha 5$  SNP mice by mini-pump infusion increased their PFC PYR neuron activity to WT levels Koukoulis et al. [2017], implying that it could reduce some of the cognitive deficits linked to schizophrenia, we used our model to pinpoint the specific nAChRs responsible for this normalization. We know that  $\beta 2$ -dependent nAChR currents, but not  $\alpha 7$ , inner-vating interneurons in layer II/III of PFC, completely desensitize after exposure to

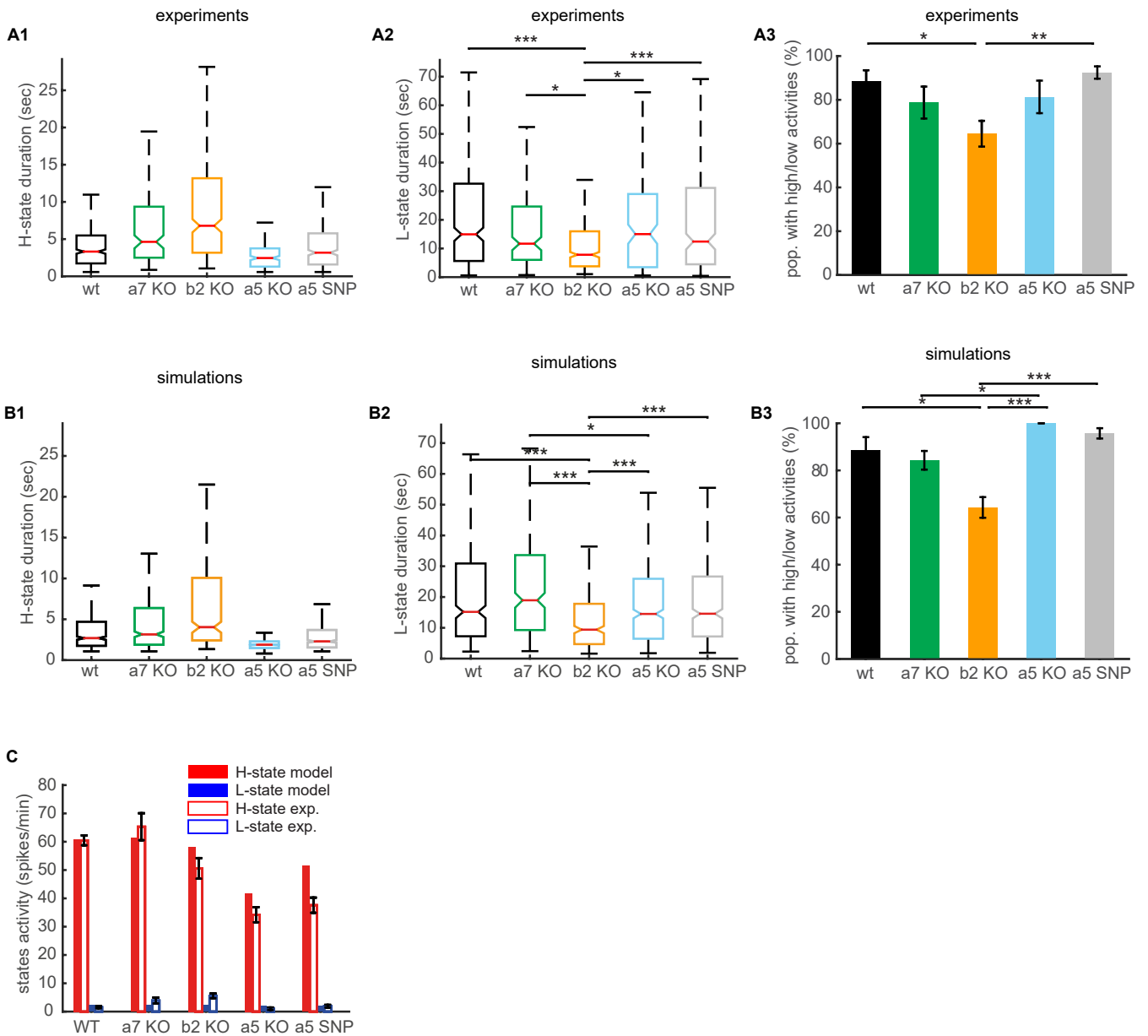


Figure 3-7 – (Caption next page.)

Figure 3-7 – (Previous page.) Accounting for the nAChR KO and mutation impact on ultraslow fluctuations (A1) Mean of H-state durations, for WT and mutant mice, modified from Fig 3C in Koukouli et al. [2016a] for  $\alpha 7$  KO and  $\beta 2$  KO mice. For  $\alpha 5$  KO and  $\alpha 5$  SNP mice, we use the same method as in Koukouli et al. [2016a]. All mutant mice distributions are significantly different from WT (Kruskal-Wallis,  $P < 0.001$ ), except for  $\alpha 5$ SNP mice. The error bars are  $\pm$ sem. (A2) Mean of L-state durations, for WT and mutant mice, modified from Fig. 3B in Koukouli et al. [2016a] for  $\alpha 7$  KO and  $\beta 2$  KO mice.  $\beta 2$  KO mice L-state durations are significantly lower compared to WT (Kruskal-Wallis,  $P < 0.001$ ). The error bars are  $\pm$ sem. (A3) Mean % of populations with H-states and L-states transitions. The error bars are  $\pm$ sem. The circle shows the proportions computed for single mice.  $\beta 2$  KO mice exhibit significantly lower % of populations with H-states and L-states transitions (ANOVA,  $P < 0.05$ ), compared to WT mice. Modified from Fig. 3D in Koukouli et al. [2016a] for  $\alpha 7$  KO and  $\beta 2$  KO mice. (B1-2-3) Same as (A1-2-3), computed from simulations. (C) H-states (red bars) and L-states (blue bars) activity levels from simulations (filled bars) and experiments (empty bars). Error bars are  $\pm$ sem. Experimental data described in Koukouli et al. [2016a]. A total of 200 simulation repetitions with 500 state transitions each were carried out to produce the simulated data.

smoking concentrations of nicotine in slice preparation Poorthuis et al. [2013], with an exception for  $\alpha 5\alpha 4\beta 2$  nAChRs, that are more resistant to desensitization Grady et al. [2012]. Modeling nAChRs levels of activation and desensitization Graupner et al. [2013] in contact of physiologically realistic levels of nicotine during smoking permits to predict the exact change of cholinergic currents amplitude, for each specific interneuron subtype. The model predicts activations of  $\alpha 7$  and  $\alpha 5$ -containing nAChRs in contact of  $1 \mu M$  of nicotine and a strong desensitization of  $\beta 2$ -containing nAChRs (see Fig. 3-14A1). As a result, you can see in Fig. 3-14A1 and Fig. 3-14A2 our predictions for PYR activity variations in WT and  $\alpha 5$  SNP mice when nicotine targets selectively each type of nicotinic receptor. Desensitization of  $\beta 2$  nAChRs decreases cholinergic inputs to SOM interneurons, and should increase the H-state network duration, leading to higher PYR firing rates in both WT and  $\alpha 5$  SNP animals. An increase of cholinergic inputs to both SOM and PV interneurons, through the activation of  $\alpha 7$  nAChRs, is assumed to induce a decrease of H-state durations, reducing PYR activity in WT and  $\alpha 5$  SNP mice. Activation of  $\alpha 5$  nAChRs, increasing cholinergic inputs to VIP interneurons, should disinhibit PYR neurons, causing higher PYR activities in both WT and  $\alpha 5$  SNP mice. However,  $\alpha 5$  SNP mice with  $\alpha 5$

nAChRs activated by nicotine application still has lower PYR activities compared to WT mice without treatment.

According to our work,  $\alpha 5$  nAChRs activation is not high enough to overcome  $\alpha 5$  SNP receptor malfunction. As a consequence, in  $\alpha 5$  SNP mice under treatment, higher PYR activities compared to WT control mice can only be replicated by nicotine interaction with  $\beta 2$  nAChRs located on SOM interneurons (Fig. 3-14A2). Experimental results (Fig. 4c in Koukouli et al. [2017]) show that in both WT mice and  $\alpha 5$  SNP mice, nicotine induces high increase of PYR activity after two days of nicotine administration, consistent with simulations predictions (see Fig. 3-14B). We know from our preliminary analysis in Fig. 3-14A2 and Fig. 3-14A3 that those effects are almost entirely due to the desensitization of  $\beta 2$  nAChRs on SOM interneurons.

We can notice in Fig. 4d, in Koukouli et al. [2017], that the increase in PYR activity after 7 days of nicotine administration is reduced in both WT mice and  $\alpha 5$  SNP mice compared to the 2 days treatment, such that PYR neurons firing rate in  $\alpha 5$  SNP mice treated with nicotine are at the level of WT mice. Previous studies have indicated that long-term nicotine exposure over days increases or up-regulates the number of high-affinity nicotine binding sites on  $\alpha 4\beta 2$  nAChRs Govind et al. [2009]. In addition to this, no nicotine-induced upregulation was observed for  $\alpha 5\alpha 4\beta 2$  nAChRs Mao et al. [2008]. Hence, we tested through modeling the effect of the upregulation of  $\beta 2$  nAChRs, located on SOM interneurons, on the PYR neuron firing rates. We were able to reproduce the normalization of PYR activity to WT levels in  $\alpha 5$  SNP mice after 7 days of administration by considering a 1.8-fold increase of the number of  $\beta 2$  nAChRs located on SOM interneurons (see Fig. 3-14C). The factor by which one needs to increase the  $\beta 2$  nAChR current in the SOM population in order to reproduce this activity normalization was found to be similar for a range of different parameter sets found during model search (see Fig.3-3), with a mean of about 1.7. You can see that the increase of PYR activity in WT animals after 7 days of nicotine treatment is accompanied by an increase of H-state durations, that is reproduced in simulations. It is also accompanied by a significant decrease of SOM activity (Fig. 4h in Koukouli et al. [2017]), consistent with the

model predictions (Fig. 3-14B).

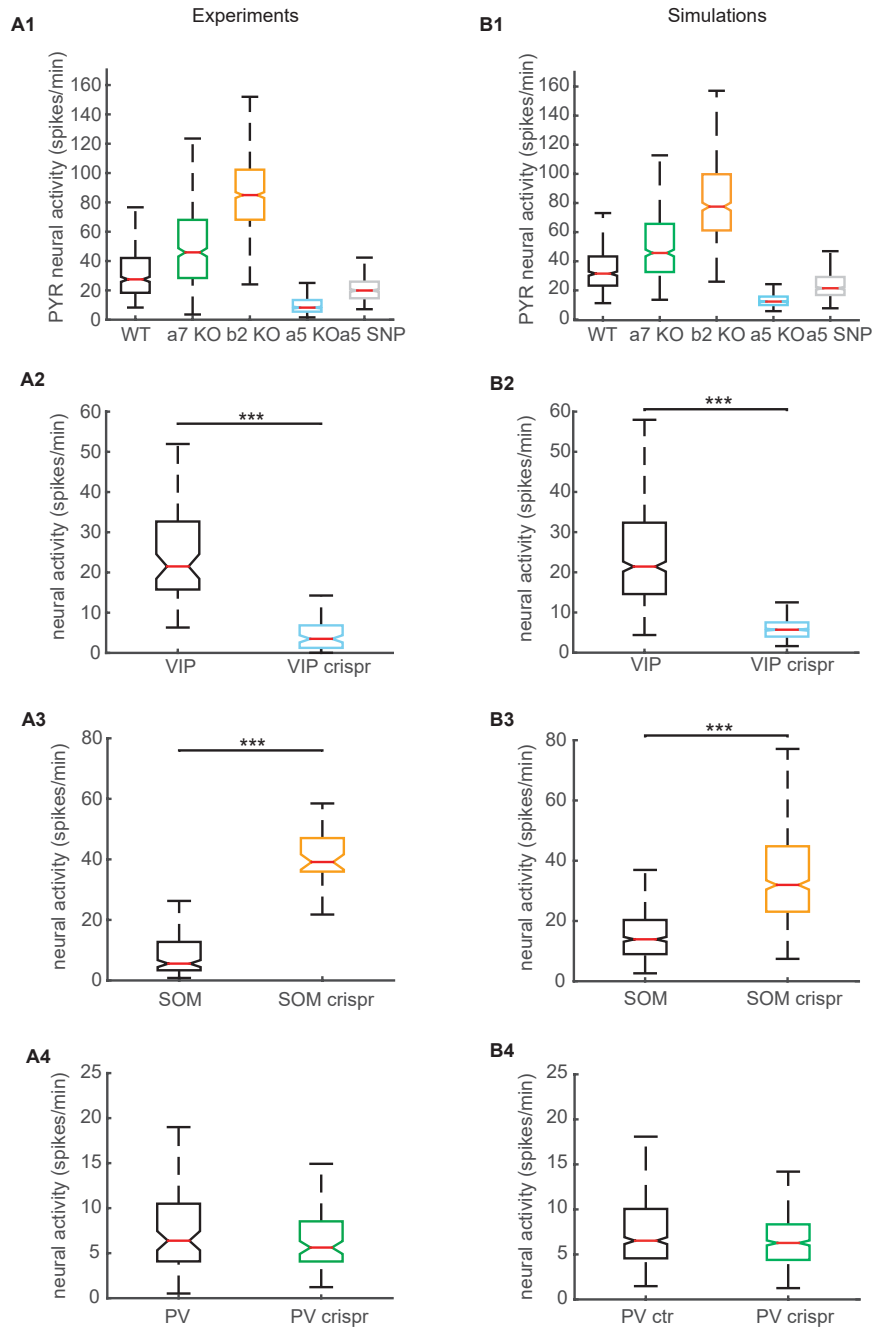


Figure 3-8 – (Caption next page.)

Having validated our modelling framework on available experimental data, we set out to test the effects of nicotine withdrawal. We considered the neural circuit activity alterations that follow a 7-day nicotine application. We hypothesized that

Figure 3-8 – (Previous page.) Reproduction of the effects of KO and mutations of nAChRs on neural firing rates (A) Boxplots of PYR neurons firing rates for WT and mutant mice, computed from simulations. All distributions are significantly different (Kruskal-Wallis,  $P < 0.001$ ). (B) Boxplots of VIP interneurons baseline activities for WT and CRISPR mice. CRISPR mice exhibit lower neural activities compared to WT mice (Kruskal-Wallis,  $P < 0.001$ ). (C) Boxplots of SOM interneurons baseline activities for WT and CRISPR simulated mice. CRISPR mice exhibit higher neural activities compared to WT mice (Kruskal-Wallis,  $P < 0.001$ ). (D) Boxplots of PV interneurons baseline activities for WT and mutant mice affected by the CRISPR technology for the deletion of the  $\alpha 5$  subunits, according to simulations. CRISPR mice exhibit similar levels of neural activity compared to WT mice. A total of 200 simulation repetitions with 500 state transitions each were carried out to produce the simulated data.

nicotine withdrawal would rapidly resensitize  $\beta 2$ -containing nAChRs. On other hand, the renormalization of the number of receptors, which had been increased through upregulation, would occur on much slower time scales. As a result, the SOM interneurons would end up with higher levels of cholinergic innervation in the post- compared to the pre-treatment condition. Model simulations predict that WT and  $\alpha 5$  SNP mice in the withdrawal condition to show a significant suppression of PYR activity as compared to the initial (pre-nicotine) state (Fig. 3-14D). Therefore our modelling results predict that post-nicotine withdrawal may exacerbate the  $\alpha 5$  SNP-associated hypofrontality.

### **3.3.8 Population model of the PFC fitted to data in APP-expressing mice predicts PYR hyper-activity reduction by galantamine**

In order to make predictions regarding activity changes in mice with amyloid beta expression as a result of its interaction with nAChRs, we have used the same modeling framework as described above to fit model parameters to replicate mean activity levels recorded in a variety of knock-out and APP-expressing animal groups (see Figs. 3-9 and 3-10) in pyramidal as well as interneuronal populations. We have relaxed the bistability assumption on population activity, as the low-high state transitions were not consistently observed across all experimental activity recordings.

Starting with the parameters best fitted on the nAChR activity data before as an initial guess point for optimization, we ran a global optimization procedure based on differential evolution with consequent local non-gradient optimization based on the Nelder-Mead algorithm. All of the knock-out and APP groups were used to formulate mean activity level targets for the fitting procedure. The nAChR inactivation levels due to interaction with the amyloid beta peptide were treated as free parameters for all three receptor subtypes and were fitted along with other model parameters such as synaptic weights between populations. We found that it is possible to find parameter sets that closely replicate the activity trends between different animal groups using the population model (Figure 3-12).

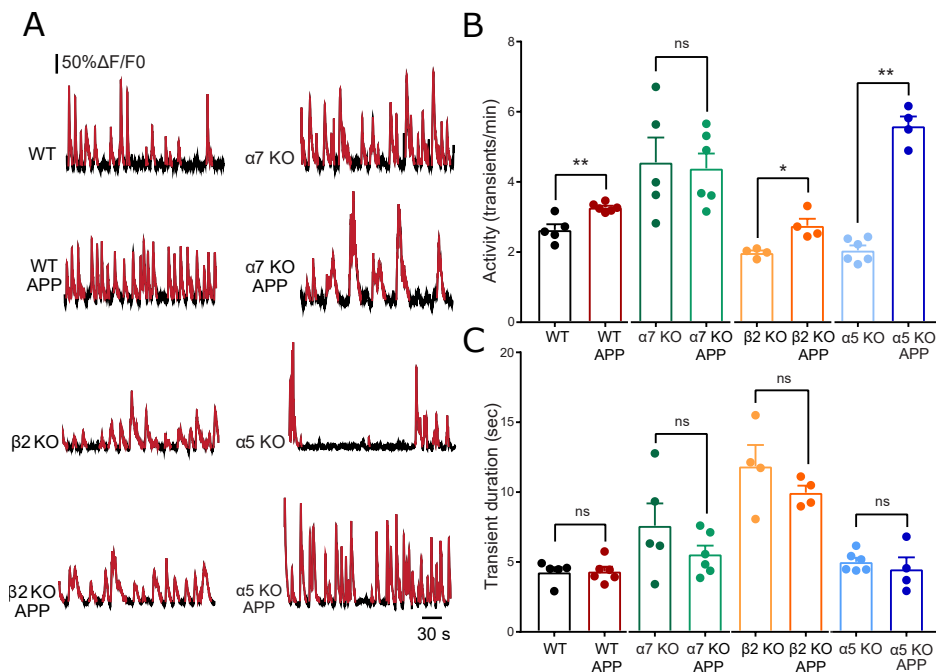


Figure 3-9 – Summary of activity levels in different groups of APP-expressing mice vs. control groups. (A) Representative recordings of spontaneous Ca<sup>2+</sup> transients in WT, WT APP, α7KO, α7KO APP, β2KO, β2KO APP, α5KO and α5KO APP mice. The detected calcium transients are indicated in red. (B) Median frequency of spontaneous Ca<sup>2+</sup> transients/min of WT, WT APP, α7KO, α7KO APP, β2KO, β2KO APP, α5KO and α5KO APP mice. (C) Same, but for mean transient duration in seconds.

We then used the fitted parameters that reproduce the activity in different animal groups to make predictions regarding pharmacological manipulations of different nicotinic receptor subtypes. In particular, we used the computational model to



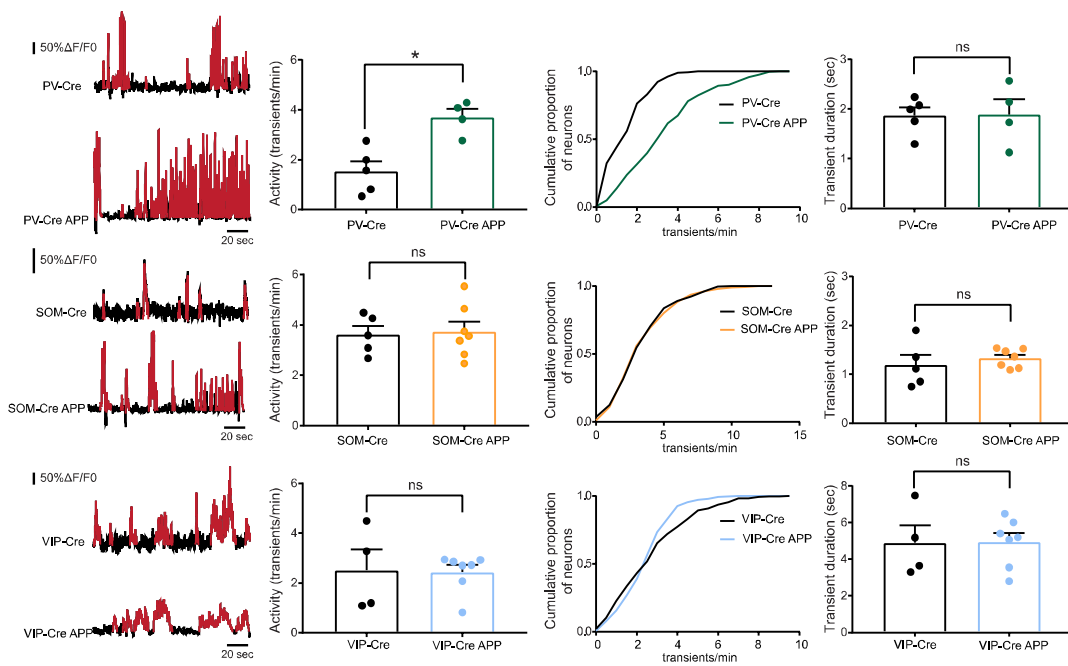


Figure 3-10 – Summary of interneuron activity levels in different groups of APP-expressing mice. (Left column) Representative recordings of spontaneous  $\text{Ca}^{2+}$  transients in different interneuron populations in control and APP mice. (Middle left column) Median frequency of spontaneous  $\text{Ca}^{2+}$  transients/min in different interneuron populations in control and APP mice. (Middle right column) Cumulative distribution of firing frequency (transient/min) in different interneuron populations in control and APP mice. (Right column) Median transient durations in different interneuron populations in control and APP mice.

estimate the effects of  $\beta 2$  nAChR block (by setting the corresponding current value to zero) as well as the effect of enhanced  $\alpha 5$  nAChR activity (by setting the value of the corresponding current to its maximal value) (see Figure 3-13). We have found a reduction in pyramidal cell activity in both cases relative to the wild-type APP activity level. This observation was then confirmed experimentally by treating the wild-type APP mice with galantamine, which at a specific concentration used acts as a positive allosteric modulator (PAM) of the  $\alpha 5$ -containing nAChRs. A reduction in pyramidal cell activity was observed in mice treated with galantamine as opposed to APP mice treated with vehicle (Figure 3-11), in correspondence with model findings.

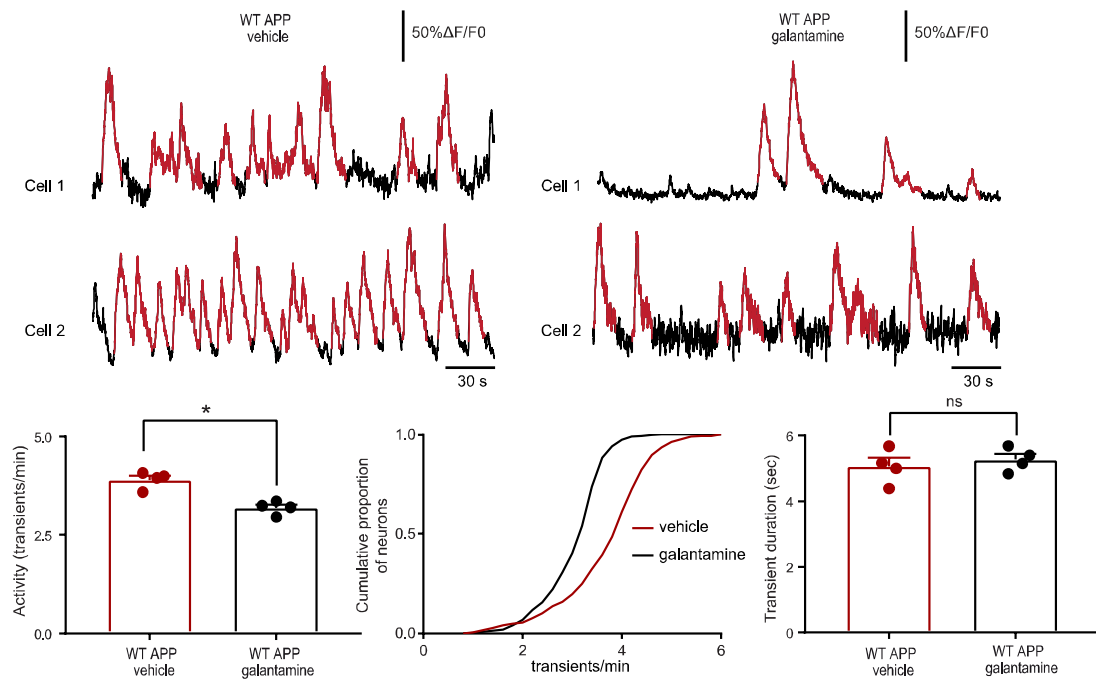


Figure 3-11 – Galantamine restores PYR neuron hyperactivity early in AD. Top: Representative recordings of spontaneous Ca<sup>2+</sup> transients in WT APP vehicle and WT APP galantamine treated mice. Bottom left: Median frequency of spontaneous Ca<sup>2+</sup> transients/min of WT APP vehicle ( $3.969 \pm 0.11$  transients/min;  $n = 4$  mice) and WT APP galantamine treated mice ( $3.22 \pm 0.09$  transients/min;  $n = 4$  mice).  $p = 0.0286$ , Mann-Whitney test. Bottom center: Cumulative distribution of firing frequency (transients/min) of WT APP vehicle and WT APP galantamine treated mice. Bottom right: Median transient durations of WT APP vehicle ( $5.078 \pm 0.27$  secs) and WT APP galantamine treated mice ( $5.269 \pm 0.18$  secs).  $p = 0.6857$ , Mann-Whitney test.

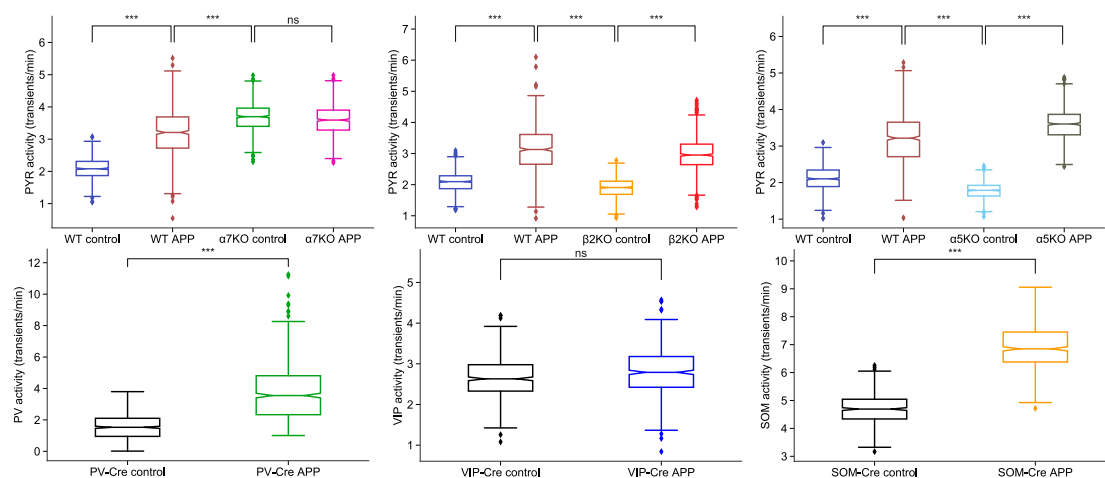


Figure 3-12 – Median simulated activity levels for the fitted population model in pyramidal neurons and interneurons corresponding to different animal groups with nAChR knock-outs and APP expression.

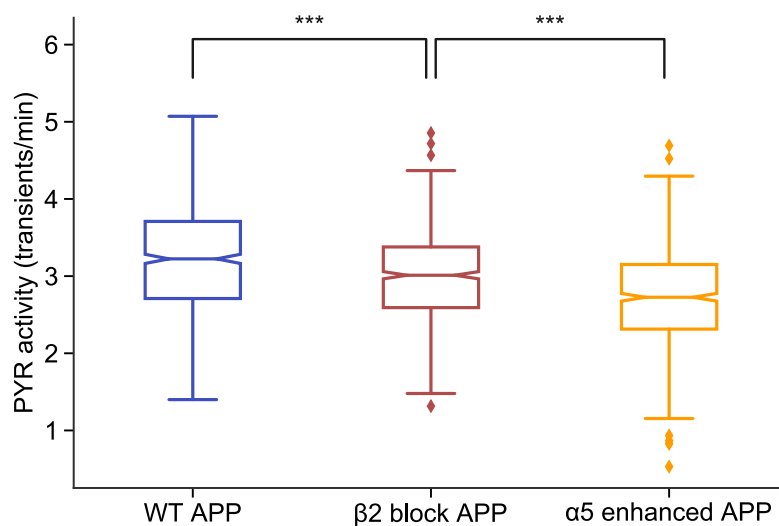


Figure 3-13 – Predictions on activity level change in the model fitted on APP data upon a block of the  $\beta 2$ -containing nAChRs or enhanced activity of the  $\alpha 5$ -containing nAChRs.

## 3.4 Discussion

In this work we developed a data-driven framework to model the influence exerted by nicotinic cholinergic neuromodulation of the hierarchical inhibitory local circuitry in the prefrontal cortex. We also showed how this inhibitory sub-circuit in turn controls the spontaneous resting state activity. We used the resulting computational model to account for effects of alterations in nAChR function by genetic manipulations and mutations associated with schizophrenia. We specifically applied to model to cellular imaging data obtained from the superficial layers of the prefrontal cortex of genetically modified mice in quiet wakefulness.

### 3.4.1 Summary of the results.

What we learned from this modeling framework is two-fold.

First, the change of cholinergic input to the various GABAergic neurons, due to the knockout of various types of nAChRs, could fully account for the change of activity patterns recorded in the various mouse lines. The KO of the high-affinity  $\beta 2$  nAChRs, decreased the cholinergic input to the SOM population, thereby decreased the subtractive inhibition of PYRs. Decreasing this specific type of inhibition increases the stability of the network high activity states (H-states), increasing their durations. In addition reduction of SOM-mediated additive inhibition decreases the stability of the network low activity states (L-states), decreasing their durations. Simulated KO of  $\alpha 5$  nAChRs decreases the cholinergic inputs to the VIP interneuronal population, decreasing synaptic inhibition of SOM interneurons. This in turn increases the SOM-mediated inhibition of PYRs and leads to a significant decrease of H-state stability, decreasing their durations. However, this SOM-inhibition increase does not appear to lead to a significant increase of L-states durations. This is because the VIP inhibition of the PV neurons is also decreased, boosting the divisive inhibition of PYRs and having an opposite effect on L-state stability. Hence the impact of VIP alterations on the SOM-PYR inhibition and the PV-PYR inhibition balance out. The  $\alpha 7$  nAChR have a lower affinity to ACh. Hence KO of the  $\alpha 7$  nAChR leads to a relatively weaker decrease of cholinergic inputs to SOM and PV

interneurons, which both, through subtractive and divisive inhibition respectively, increase the stability, thus durations, of the H-states.

Second, simulated 7-day nicotine application could restore the PYR activity to WT levels as compared to the  $\alpha 5$ SNP case because of a mixture of desensitization and upregulation of  $\beta 2$  nAChRs. The model showed that activation of  $\alpha 5$  nAChRs by nicotine was not sufficient to compensate the  $\alpha 5$ SNP activity deficits. Upregulation of  $\beta 2$  nAChRs after 7 days of treatment should lead to activity depression in both WT and  $\alpha 5$ SNP cases when nicotine is removed, which could represent a highly critical situation in schizophrenia patients. Hence, this work provides experimentally testable predictions that the severity of symptoms in schizophrenia linked to decreased neural activity might increase upon nicotine withdrawal.

### 3.4.2 Predictions of the model.

We summarize the predictions obtained with the proposed modelling framework as follows:

- Based on the constructed model, we hypothesize that the ultra-slow firing rate fluctuations in the prefrontal cortex are due to the internal bistable dynamics of the connected neuronal populations in the layer II/III PFC local circuit, with low-high activity state transitions triggered by inherent noise.
- The stability of the low-high states, and thus their life-times, are differentially controlled by a hierarchy of interneuronal populations. In particular, PV interneurons are key in supporting the activity balance in the PFC circuit, with the synaptic VIP-PV connection pathway being dominant in controlling network bistability.
- Reduced pyramidal neuron activity in  $\alpha 5$  nAChR dysfunctions associated with schizophrenia (e.g. in  $\alpha 5$  SNP animals) can be normalized by enhancing the  $\beta 2$  nAChR activity, in particular under acute nicotine treatment.
- The model predicts that the existing  $\alpha 5$  SNP-associated hypofrontality can be significantly worsened by nicotine withdrawal.

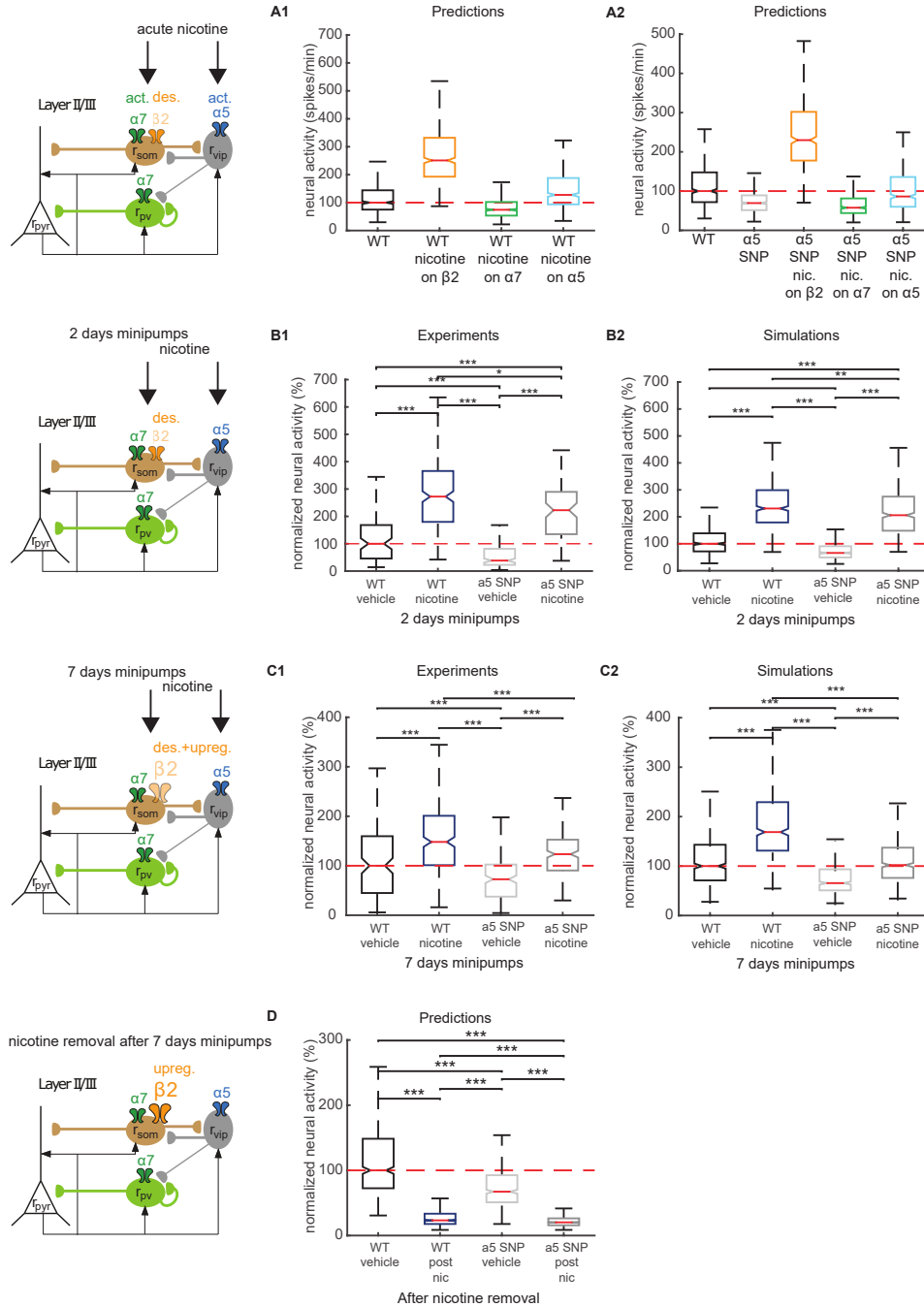


Figure 3-14 – (Caption next page.)

Figure 3-14 – (Previous page.) Desensitization and upregulation of  $\beta 2$  nAChRs normalizes  $\alpha 5$  SNP mice network activity to WT levels after chronic nicotine application (A1) Distribution of PYR neurons firing rates for WT mice, computed from simulations of nicotine effects on  $\alpha 7$ ,  $\beta 2$ , and  $\alpha 5$  nAChRs. All distributions are significantly different (Kruskal-Wallis,  $P < 0.001$ ). (A2) Distribution of PYR neurons firing rates for WT and  $\alpha 5$ SNP mice, computed from simulations of nicotine effects on  $\alpha 7$ ,  $\beta 2$ , and  $\alpha 5$  nAChRs in  $\alpha 5$ SNP mice. All distributions are significantly different (Kruskal-Wallis,  $P < 0.001$ ). (B) Distribution of PYR neurons firing rates for WT and  $\alpha 5$ SNP mice, control and treated with 2 days of chronic nicotine application, obtained from simulations. All distributions are significantly different (Kruskal-Wallis,  $P < 0.001$ ). (C) Distribution of PYR neurons firing rates for WT and  $\alpha 5$ SNP simulated mice, control and treated with 7 days of chronic nicotine application. All distributions are significantly different (Kruskal-Wallis,  $P < 0.001$ ). The model predicts an upregulation of  $\beta 2$  nAChRs. (D) Distribution of PYR neurons firing rates for WT and  $\alpha 5$ SNP mice, control and after nicotine removal following 7 days of chronic nicotine administration, predicted from simulations. All distributions are significantly different (Kruskal-Wallis,  $P < 0.001$ ). A total of 200 simulation repetitions with 500 state transitions each were carried out to produce the simulated data.

- The population model fitted on activity data from APP-expressing mice predicted reduction in PYR population activity upon enhanced  $\alpha 5$  nAChR currents, experimentally validated by galantamine treatment experiments in APP mice.

### 3.4.3 Limitations of the model and future direction.

The prefrontal cortex is a well-connected brain region receiving inputs from multiple brain regions. The complexity of the PFC circuitry, function and the multiplicity of the potential mechanisms that could influence its activity dynamics suggest that one may need to consider a range of compensatory, secondary and off-target effects when analysing experimental conditions like receptor knockout or/and nicotine treatment in animals [Bernard \[2020\]](#), [Grashow et al. \[2010\]](#), [O’Leary et al. \[2014\]](#), [Wolff \[2018\]](#). These multiple effects could be potentially reflected in the constructed computational models. However, in the scope of this work, we primarily focused on the cholinergic signaling components of the prefrontal cortex local circuitry as the main source triggering activity changes in the PFC. Within this reductionist approach to modelling, we were aiming to arrive at a minimal model that would be sufficient to

explain our experimental data, assuming that the dynamics of the PFC circuitry is controlled by a hierarchy of interneurons expressing distinct nicotinic receptor types. While acknowledging that this is a strong hypothesis, there is experimental evidence supporting our local activity modulation proposal. Our previous experimental work [Koukouli et al. \[2016a\]](#) demonstrated that targeted nicotinic acetylcholine receptor re-expression in layers II/III of the prelimbic cortex of the knockout mice completely restores pyramidal neuron activity to the levels of the wild-type mice, justifying our assumption that the activity change effects are due to receptor knockouts and are not significantly influenced by the developmental issues of the studied animals. This local re-expression also suggests that the effects we observed are unlikely to be simply due to changes in the inputs to our circuit. Furthermore, knockouts of  $\beta 2$ -containing nAChRs were shown to completely shut down the activity of these receptors, with its complete restoration by local re-expression [Guillem et al. \[2011\]](#), [Maskos et al. \[2005\]](#), [Avale et al. \[2008\]](#). Knockouts of  $\alpha 5$ -containing nAChRs were found to cause dramatic shifts in dose-response curves for nicotine [Morel et al. \[2014\]](#), [Besson et al. \[2019, 2018\]](#).

#### **3.4.4 Implications for nicotine withdrawal in schizophrenia**

Based on our work we may further speculate on the neurobiological explanation for the high prevalence of smoking and low smoking cessation rate observed among individuals with schizophrenia [Anthenelli et al. \[2016\]](#). Our modelling framework gives predictions that nicotine cessation should decrease the prefrontal activity in both WT and the  $\alpha 5$ SNP PFC, with the most drastic hypofrontality seen in  $\alpha 5$ SNP mice under nicotine removal. This predicted exacerbated decrease of pyramidal activity is due to the upregulation of  $\beta 2$  nAChRs located on SOM interneurons, induced by several days of chronic nicotine application. These model results beg the question: might not the lower cessation rates seen in schizophrenia patients be caused by a pronounced hypofrontality, induced by a combination of mutated  $\alpha 5$  nAChRs located on VIP interneurons and the upregulation of  $\beta 2$  nAChRs located on SOM interneurons. In fact, previous work showed that negative affect, one aspect of the negative schizophrenia symptoms associated with hypofrontality, is a key



contributor to the low quitting rate seen in smoker schizophrenia patients [Tidey et al. \[2008\]](#). At the same time pharmacotherapy, through the use of varenicline, having similar nAChRs interaction mechanisms to nicotine, increases the abstinence rate in smokers and even more drastically in schizophrenia patient who are smokers (from 4.1% to 23.2%) [Anthenelli et al. \[2016\]](#). We may suggest that these studies lend support to our hypothetical conjecture.

### 3.5 Conclusions

Schizophrenia is a severe mental disorder implicating a large variety of symptoms, among which apathy, abolition or social withdrawal, grouped as negative symptoms. Nowadays, no specific treatment can be recommended to treat negative aspects of schizophrenia pathology. Yet, it has been suggested recently that a mutation of a specific type of nicotinic receptor was implicated in the reduced neural activity levels recorded in the prefrontal cortex (PFC) of schizophrenia patients. Chronic nicotine injections in mice expressing this mutation ( $\alpha 5$  SNP mice) permits to restore neural activities to control levels, consistent with the idea that schizophrenia patients smoke to self-medicate. Using computational modeling, we showed that nicotinic receptors located on a hierarchy of inhibitory neurons were able to control ultra-slow neural activity fluctuations recorded in the PFC of mice. Furthermore, our modelling framework suggests that  $\beta 2$  receptors are the nicotine's main target in restoring neural activity to control levels in  $\alpha 5$  SNP mice. We provide a testable model prediction that nicotine withdrawal in schizophrenia patients with the  $\alpha 5$  SNP mutation should lead to a progressively severe hypofrontality. Lastly, we apply our modeling framework to fit activity data in APP-expressing mice and make predictions for activity restoration upon pharmacological manipulations targeting  $\beta 2$ - and  $\alpha 5$ -containing nicotinic acetylcholine receptors.

To cast a wider perspective to our circuit-based dynamic modelling approach opens a number of further avenues to both study specific disease-related alteration of nicotinic modulation in cortical circuits and to identify potential points-of-entry for therapeutic interventions.

## Chapter 4

# Discovery of the cholinergic system pathologies in the PFC from cortical activity using machine learning

Ivan Lazarevich, Fani Koukouli, Uwe Maskos, Boris Gutkin

Unpublished results

I.L. and B.G. conceived and designed research. F.K. and U.M. designed experiments. F.K. performed experiments. I. L. analyzed the data and performed computational experiments.

In Chapter 2 of the thesis we have established that for a wide range of neural spiking activity data sets one could effectively utilize approaches from machine learning for time-series data to build predictive models and study the structure of recorded neuronal activity and what information about the outside world is contained therein. Machine learning approaches to neural decoding have been successfully applied to practical tasks with potential engineering applications such as predicting position of a rat chasing rewards on a platform from its hippocampal activity or predicting a position of a cursor controlled by a monkey via moving a manipulandum from its motor cortex activity [Glaser et al. \[2020\]](#). What, however, remains relatively unexplored is the application of machine learning methods to detect patterns of pathological activity in neural circuits in animal models of neural disorders. In this chapter, we consider a data set introduced in the previous chapter consisting of recordings of pyramidal cell activity in the prefrontal cortex obtained *in vivo* in genetically-modified animals with dysfunctions of specific nAChR subtypes associated with disorders such as schizophrenia and Alzheimer’s disease. We show that it is possible to construct accurate predictive models that classify between healthy control animals and animals with mutations specific for neural disorders.

## 4.1 Analysing prefrontal cortex activity in mice with nAChR dysfunctions

We start by looking at the two-photon calcium imaging data from [Koukouli et al. \[2017\]](#) introduced in the previous chapter comprised of recordings from pyramidal cell from the layers II/III of the PFC in mice with genetic alterations of nAChR expression. Alterations of resting dynamics in the prefrontal cortex have been hypothesised to serve as biomarkers of schizophrenia [Barch et al. \[2001\]](#). A mutation of the  $\alpha 5$  nAChR subunit, the rs16969968 single nucleotide polymorphism ( $\alpha 5$ SNP), is linked to both nicotine addiction and a functional cortical deficit associated with reduced neural activity (hypofrontality) and is characteristic to schizophrenia patients [Hong et al. \[2010\]](#), [Koukouli et al. \[2017\]](#). Here we look neural activity data recorded *in vivo* in prefrontal cortices of mice with this specific nAChR mutation.

A reduction of pyramidal cell firing is generally observed in these mice, similarly to hypofrontality seen in humans, but an important question is whether there exist activity patterns specific to this mutation that could help separate single-neuron and ensemble activity in such animals from activity recorded in healthy, control mice. We investigate this question in depth in the rest of the chapter and identify the key variables predictive of the  $\alpha 5$  nAChR subunit dysfunction.

#### 4.1.1 Two-photon imaging data pre-processing

In case of an activity classification task, the choice of the particular data pre-processing pipeline could be solely driven by its contribution to the final attainable classification accuracy metric. To that end, the data pre-processing procedure should remove as much irrelevant information and noise from the signal as possible to maximize final prediction accuracy. One might assume that if the  $\text{Ca}^{2+}$  fluorescence traces are given as the input, it could be beneficial to not only perform denoising of the extracted traces, but also to apply a deconvolution transformation to produce a proxy signal of the firing activity of the given neuron [Friedrich et al. \[2017\]](#). In our work, we first perform trace normalization via baseline subtraction regardless of the other transformations in the pipeline. This step is performed by calculating the baseline resting fluorescence signal as the 8<sup>th</sup> percentile of the values within a sliding window (20 seconds in size) applied to the fluorescence trace. The baseline signals are then subtracted from each of the traces in the dataset and the traces are also divided by the baseline values to standardize the variance. We then optionally apply the deconvolution procedure to empirically validate whether the spike extraction transformation could boost the attainable classification scores. Being aware of the issues encountered when inferring the spiking activity from fluorescence traces [Stringer and Pachitariu \[2019\]](#), we are not interested in obtaining a perfect proxy for the spiking activity of the neurons in question. Rather, we aim to transform the signals in way that maximizes the classification accuracy. One might hypothesize that using firing proxy signals instead of raw fluorescence traces might help classify pathological activity on the single-neuron level. We test this hypothesis by comparing the baseline classification accuracy obtained with the raw traces to

the values obtained with trace deconvolved using the OASIS algorithm [Friedrich et al. \[2017\]](#). An example set of activity traces prior to and after the pre-processing procedure (without deconvolution) is shown in [Figure 4-1](#).

### 4.1.2 PFC activity structure revealed with time-series features

In terms of building a classification model from time-series data, we apply the same approach to the pre-processed Ca<sup>2+</sup> fluorescence signals recorded from the PFC as the one we defined for ISI-encoded spike trains in Chapter 1. We first construct sets of binary classification tasks by picking pairs of states from the data set and measure the classification accuracies attainable by first encoding the time-series with *tsfresh* features and then building an efficient nonlinear classifier on that embedding (e.g. a random forest classifier). The cross-validation scheme that we employ for this task is based on individual animal identifiers so that the train and test subsets do not contain recordings from the same animals. Each sample in the data set corresponds to a full recording from a single neuron in a certain experiment containing 1500 timestamps. We first consider two main models: (i) a simple baseline comprised of a random forest classifier built on a set of 6 basic statistical features of the time series and (ii) an RF classifier trained on top of a full *tsfresh* embedding of the data. The difference in median accuracies between these models reveals the amount of discriminative information that is contained in time-series characteristics beyond the simple statistics like the mean and variance of the activity trace. It is common in experimental literature to measure the significance of the difference between functional states and/or animal lines by running statistical tests on the aggregates of the activity like the median firing rate between experiments. Much information remains hidden, however, in other activity covariates and the states that are deemed to be insignificantly different might appear to not be so if the appropriate activity measure is chosen for the analysis. This is our main motivation for probing the activity data in different animal lines for differences by building discriminative machine learning models in our constructed classification tasks instead of just looking

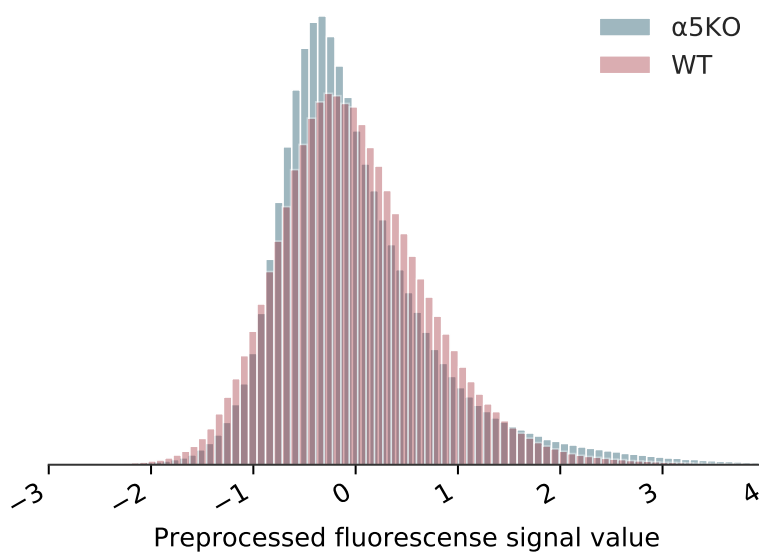
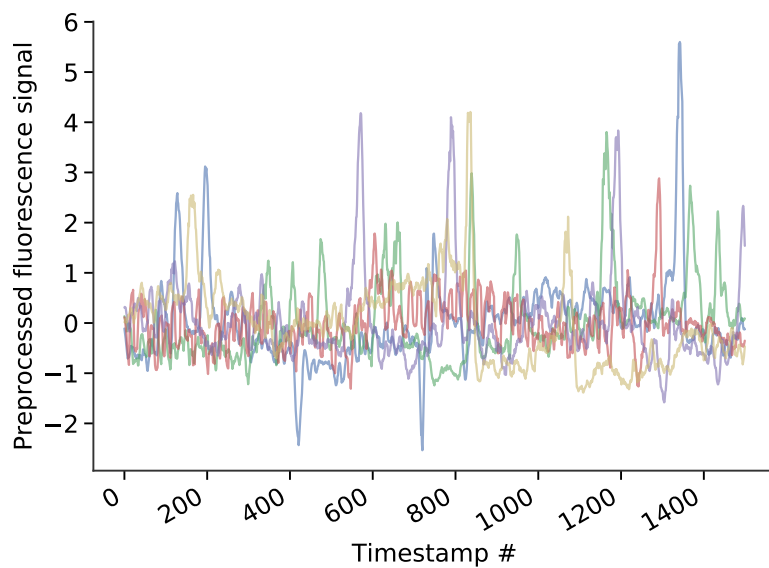
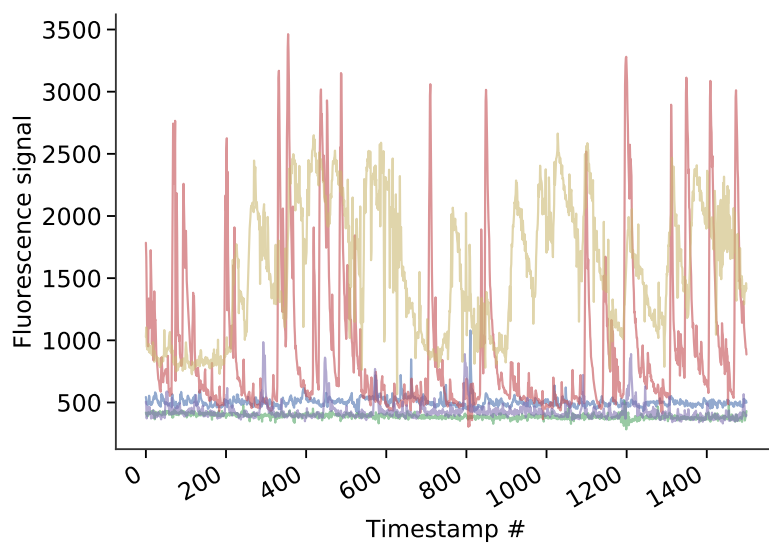


Figure 4-1 – (Caption next page.)

Figure 4-1 – Top: activity traces obtained by two-photon imaging of GCaMP6f expressing neurons (data from Koukouli et al. [2017]), middle: a set of neural activity traces from the same data set with pre-processing applied to the extracted signals, bottom: value distribution histograms obtained from the pre-processed time-series for two separate groups of animals – the wild-type control group and the group with the  $\alpha 5$ -containing nAChR knockout mutation (aggregated over neurons over animals).

at arbitrary activity aggregates. The power of the massive time-series feature engineering approach here is then in the ability to automatically detect the activity characteristics that allow distinguishing between activity states and then proceed to decide whether the found differences are significant from a functional point of view.

In terms of our classification models, if the median accuracy attainable on the full *tsfresh* embeddings does not significantly exceed that of the simple baseline, it would mean that not much of discriminative information is contained in the data (e.g. in the structure of the time-series) beyond the difference in basic statistics like mean and variance. If the accuracies of the both types of models we consider are on the chance level, it means that the activity series are indeed indistinguishable, at least to the extent of the *tsfresh* representation. This is much stronger evidence of similarity between activity states than the mere similarity between the mean or the median activity levels. We have observed, perhaps quite surprisingly, such a similarity between *tsfresh* embeddings of the PFC activity states that were previously identified to be similar by just looking at the average firing rates of neural ensembles inferred from Ca2+ fluorescence traces.

In the case when the simple baseline RF model shows a near-chance-level accuracy but the full-*tsfresh*-embedding model shows a significantly higher accuracy level, that means that the differences between activity states perhaps would not have been observed by doing a classical statistical test on the average firing rates of the neurons. Rather a more advanced approach such as building a machine learning classifier is required in this case.

On the other hand, the absolute accuracy values obtainable by the full *tsfresh* models are representative of what accuracy could be achieved in general in tasks of predicting patterns of pathological neural activity (patterns induced by modeled

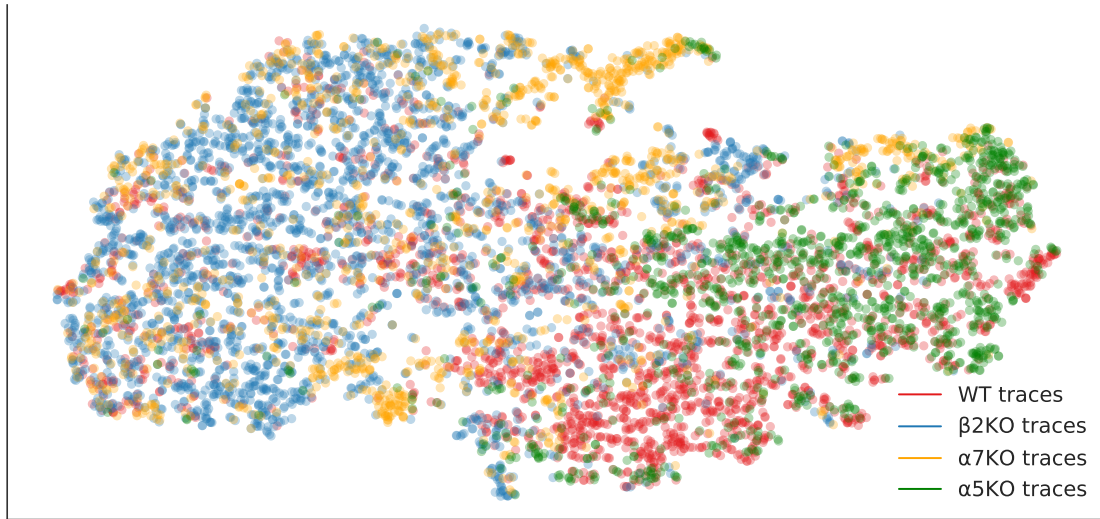


Figure 4-2 – Two-dimensional tSNE projection of the top-25 discriminative *tsfresh* features for recorded neuronal activity traces in the imaging experiments for normal and nAChR knockout states. Each point in the scatter plot corresponds to a trace from a single neuron; activity states are color-coded.

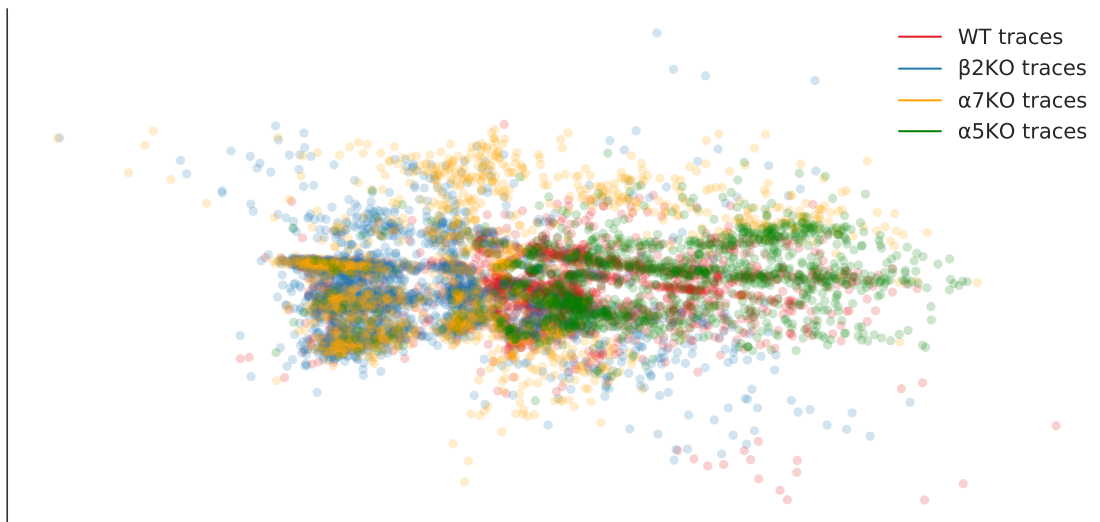


Figure 4-3 – Two-dimensional projection by a VAE model trained on vectors of the top-25 discriminative *tsfresh* features for recorded neuronal activity traces in the imaging experiments for normal and nAChR knockout states. Each point in the scatter plot corresponds to a trace from a single neuron; activity states are color-coded.



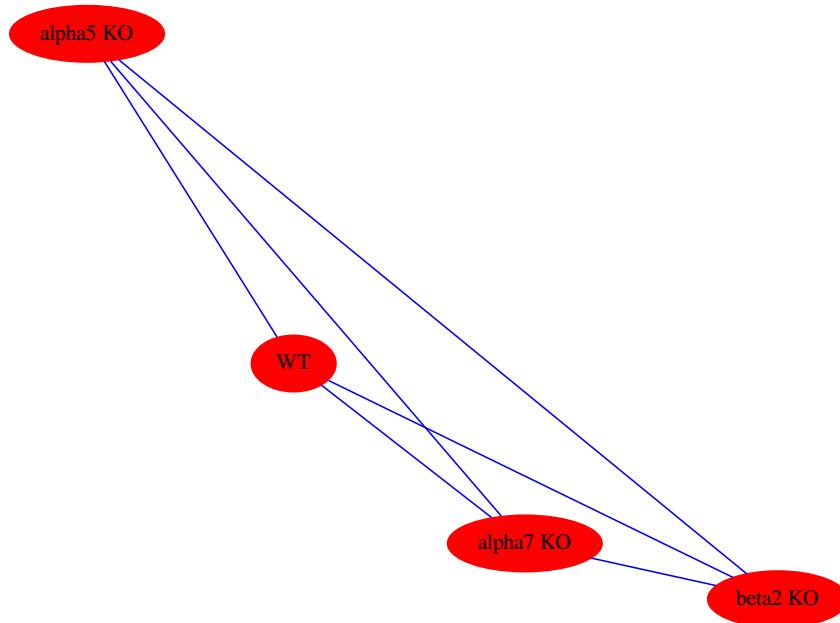


Figure 4-4 – Graph representation of the Wasserstein distance matrix between the distributions of the first tSNE mapping component in different knockout states. Note the proximity of  $\alpha 7$ KO and  $\beta 2$ KO states.

dysfunctions of the cholinergic system in the PFC, in particular).

Figures 4-2 and 4-3 show two-dimensional scatter plots as a result of two dimensionality reduction techniques applied to *tsfresh*-encoded time-series in all nAChR knock-out animal groups as well as the wild-type group. There is an apparent structural difference between the distributions of embedded time-series samples in different animal groups, also reflected as a graph representation of the Wasserstein distance matrix of the first t-SNE mapping component of the data 4-4. Activity patterns observed in  $\alpha 7$  and  $\beta 2$  nAChR knock-out groups appear to be close, with the  $\alpha 5$  knock-out state appearing to be the farthest from all other states.

### 4.1.3 Detecting nAChR dysfunction from single-neuron activity data with machine learning

As mentioned previously, we are aiming to build a predictive model for the animal group (wild-type animals vs. animals with specific nAChR mutations) based on

single-cell firing patterns, rather than activity data from a whole ensemble of cells. The latter potentially contains more predictive information compared to the single-cell case, given that the observed synchronicity patterns are found to change in animals with nAChR knock-outs compared to wild-type animals [Koukoulis et al. \[2016a\]](#). However, in part motivated by the small size of the available data set, we found that it is indeed possible to build sufficiently accurate predictive models based on single-cell activity traces as inputs. These results are promising, since they imply that it would also be possible to build accurate predictive models using scarce aggregate signals of neural activity such as local field potential recordings from just several electrode locations. The problem of data scarcity, as we demonstrate further on, could also be tackled with the use of data augmentation techniques for time-series data.

We begin by defining a single binary classification problem: given activity recordings from wild-type animals and animals with  $\alpha 5$ -subunit containing nAChR knock-out on the level of single neurons in the PFC, build a model that would predict if the activity recording corresponds to the WT group or the mutant group. We extracted signal chunks of 500 timestamps (corresponding to a recording of about 70 seconds) with a rolling window with a step of 500 timestamps (hence 3 chunks per full neuron recording) from across different experiments on different animals belonging to the two groups. We first started with a *tsfresh*-based time-series feature extraction approach whereby we encoded the pre-processed signal chunks as feature vectors and trained a random forest classifier on that feature representation. First of all, we compared the performance of the RF model on the full *tsfresh* vector representation against the RF model trained on vectors of 6 basic statistics of the input time-series (the mean, median, standard deviation values, maximal and minimal value across the time-series and its absolute energy value). We found that the accuracy values of the full *tsfresh* models is significantly higher than that of the baseline trained on simple statistics (see [Figure 4-5](#)). This means that there is information to be decoded that would help classify the animal groups contained in the time-series beyond the simple statistics. We have also found ([Fig. 4-5](#)) that denoising the input signals via convolving with a small Gaussian kernel further boosted the accuracy scores of the

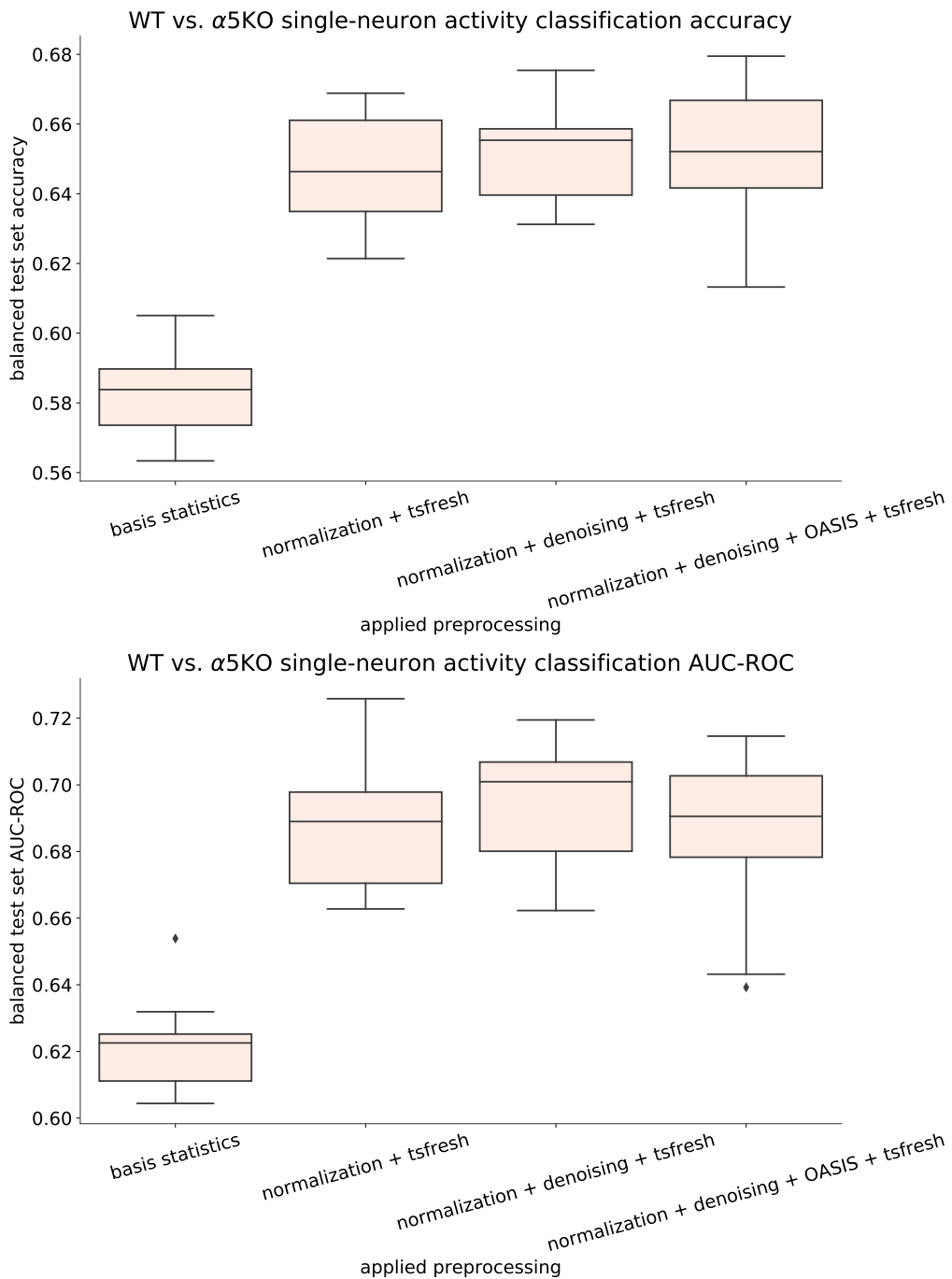


Figure 4-5 – Test set accuracy and AUC-ROC score distributions in the  $\alpha$ 5 nAChR KO detection task depending on the feature representation of the time-series and its pre-processing. Distributions are computed over different train/test splits of the data set and undersampling is performed to keep the class balance in the training and testing data sets.

*tsfresh*+RF models. On the other hand, applying the OASIS algorithm to get an estimate of neuronal firing rate to be used as an input to the model did not result in higher classification accuracy scores, but in similar accuracy values. This means that most of the information needed for predictions is contained in the time-series of neuronal firing rate, but the OASIS transformation itself leads to a distortion of the signal that slightly decreases the accuracy scores.

We extracted feature importance scores from the random forest classifiers trained on *tsfresh*-encoded data and used the top-30 most important features to produce two-dimensional embeddings of the data using PCA and t-SNE algorithms, as shown in Figure 4-6. There is a clear separation between the two-color coded clusters corresponding to the two different animals groups. Figure 4-7 shows the ranking of time-series features in the task of classifying the WT vs.  $\alpha 5$ KO animals. Note that the most discriminative features revealed by the classifier are the features related to the temporal structure of the time-series, rather than to the statistics of the value distribution (like e.g. the mean or the median activity value). These results show that rather than looking at basic statistics of recorded activity, one has to consider the local and structure of neural activity to distinguish well between the healthy control animals and the mutant group. Figure 4-11 shows the increase in accuracy when several samples are available to make a single prediction, with the accuracy value saturating at  $\sim 80\%$  if the base accuracy level (for a single sample) is around 67%, as we typically found for feature extraction based models in the WT vs.  $\alpha 5$ KO classification task.

We further looked into what temporal patterns are most predictive of the nAChR dysfunction. We employed a time-series classification algorithm named shapelet learning [Ye and Keogh \[2009\]](#), whereby a representation of the time-series is learned as a vector of distances to a set of fixed-size time-series named shapelets that are defined as subsets of a time series, that is a set of values from consecutive time points. The distance between a shapelet and a time series in the data set is defined as the minimum of the distances between the shapelet and all the shapelets of same length extracted from this time series. The most discriminative shapelets for the classification task are learned via gradient descent. The number and sizes of

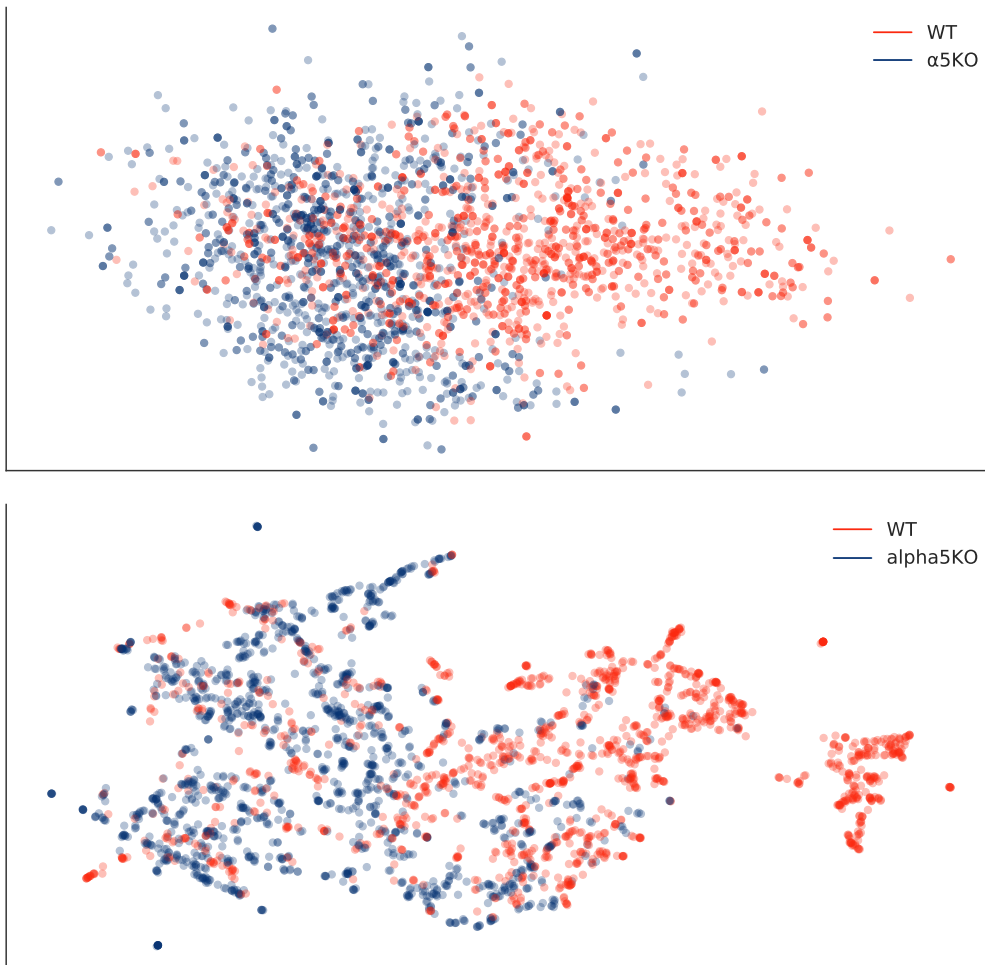


Figure 4-6 – Two-dimensional PCA (top) and t-SNE (bottom) embeddings of the WT vs.  $\alpha$ 5KO neuronal signals data set encoded with top-30 most important *tsfresh* features. Animal group is color-coded.

shapelets are hyperparameters of the algorithm and were chosen based on a heuristic from Grabocka et al. [2014].

The results obtained with the shapelet learning algorithm in the WT vs.  $\alpha$ 5KO animals classification task were interestingly on par with time-series extraction approaches (see Table 4.1) with only 6 shapelets and hence a 6-dimensional representation of the time-series. Learned shapelet waveforms and their corresponding spectra are shown in Figure 4-9. Note the dominance of oscillatory components in the delta range (specifically, the 0.75-1.25 Hz band) in the shapelets, suggesting that these are the frequencies most predictive of the nAChR dysfunction.

To further look into the contribution of the different frequency band components

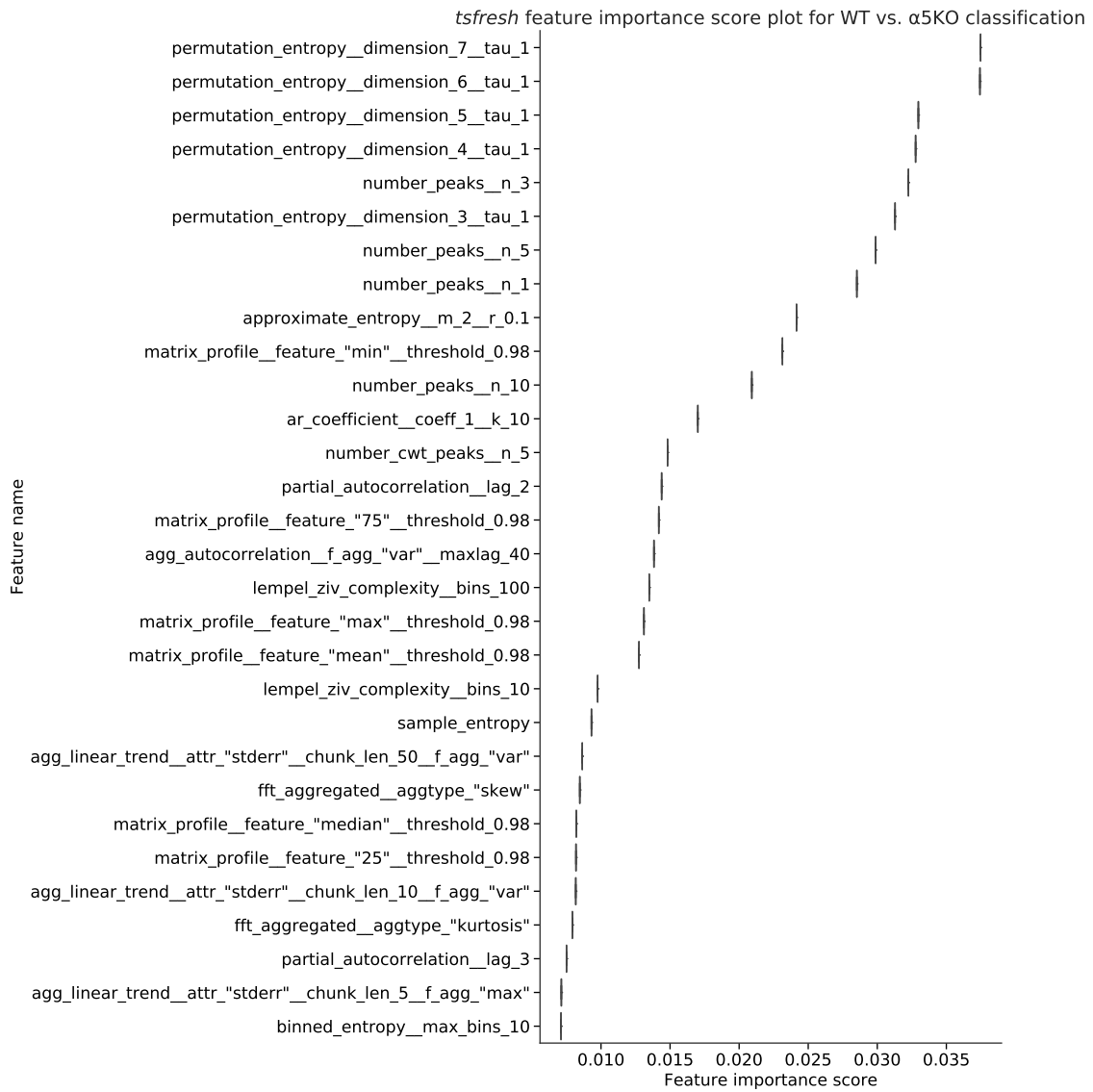


Figure 4-7 – Feature importance scores extracted from a random forest classifier trained on the *tsfresh* embeddings in the WT vs.  $\alpha$ 5KO animal classification task.

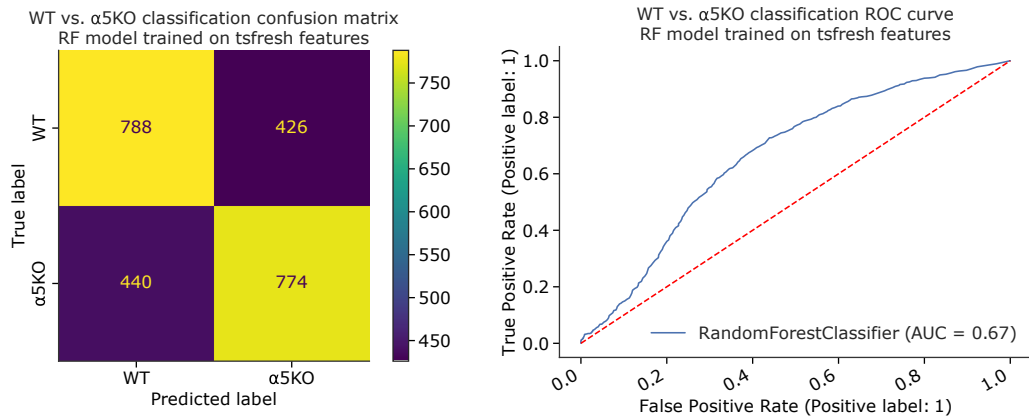


Figure 4-8 – Confusion matrix (left panel) and ROC curve (right panel) of a random forest classifier trained on *tsfresh* embeddings in the WT vs.  $\alpha 5KO$  animal classification task.

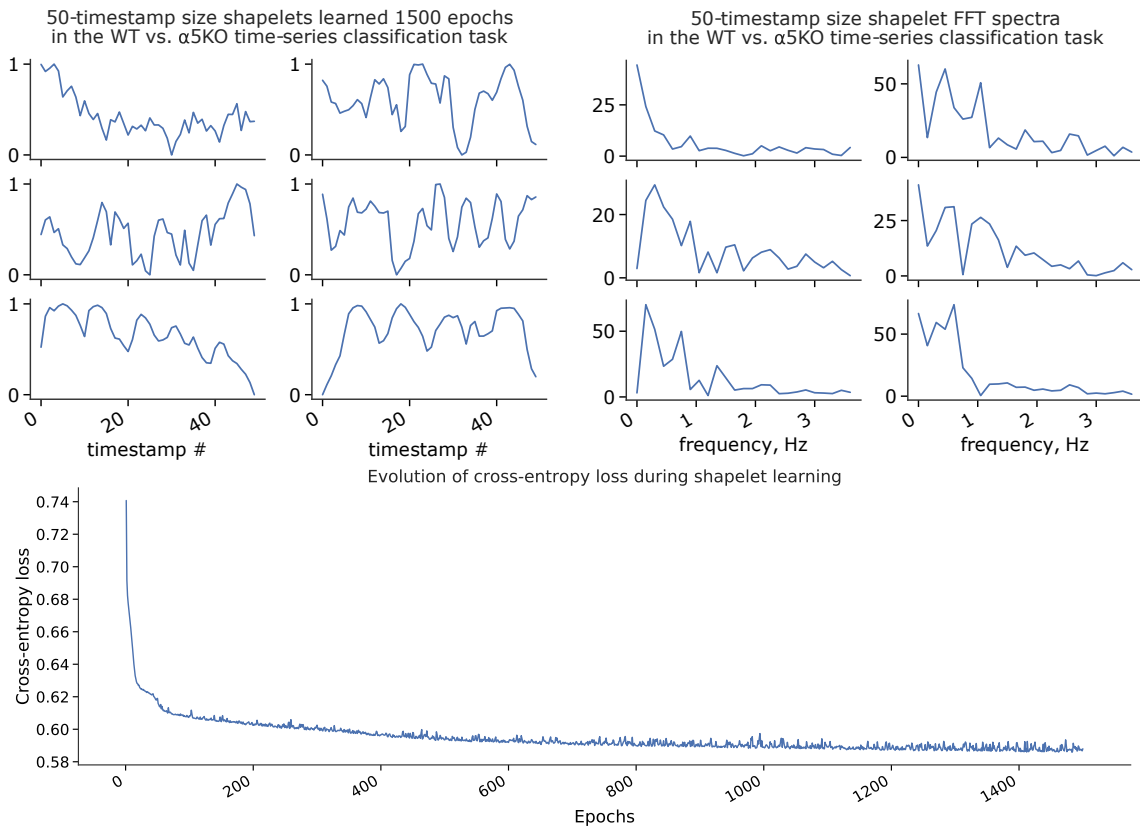


Figure 4-9 – Training results from the shapelet learning algorithm applied to the WT vs.  $\alpha 5KO$  single-neuron activity classification task. Bottom: evolution of cross-entropy loss value during shapelet learning. Top left: Learned shapelet waveforms after the end of the training. Top right: FFT spectra of the shapelet waveforms. Note the peaks in the 0.75-1.25 Hz frequency band (the delta band) in many shapelets.

Model (training data)	Balanced test set classification accuracy
catch22 + Random Forest	0.64251
tsfresh + Random Forest	0.64815
Shapelets (1500 epochs, Adam with lr=1e-2, $\lambda = 0.01$ )	0.64734
Shapelets (1500 epochs, Adam with lr=1e-2, $\lambda = 0.001$ )	0.65942
Vanilla FCN, 20 epochs with lr=1e-3	<b>0.68518</b>
catch22 + Random Forest (band-pass filtered, 0.25-0.75 Hz)	0.570048
catch22 + Random Forest (band-pass filtered, 0.75-1.25 Hz)	0.644122
catch22 + Random Forest (band-pass filtered, 1.25-1.75 Hz)	0.568438

Table 4.1 – Test set accuracy in the WT vs.  $\alpha$ 5KO classification task (with a fixed train/test split and class-balanced training and testing data) achieved with different time-series classification methods (and data pre-processing strategies).

in the input signals to the classification accuracy, we carried out training experiments on band-pass filtered input signals for a range of 0.5 Hz wide frequency bands in the delta range. Figure 4-10 shows the classification accuracies of the random forest classifiers trained on catch22-encoded band-pass filtered data depending on the frequency range (also see 4.1). Note a peak around the 0.75-1 Hz range for both the accuracy and the AUC-ROC metrics, suggesting that this range contains the most discriminative information useful for the considered classification task.

Table 4.2 shows the Wasserstein distances between feature value distributions for the wild-type and the 5KO animal groups for basic statistics and *tsfresh* features with the highest importance scores extracted from a trained random forest model. Note that the separation between the two animal groups is higher when the *tsfresh* features that reflect the temporal structure of the samples are considered rather than simple statistics. This is in particular important when considering the effect of pharmacological interventions in both animal groups on the activity levels.



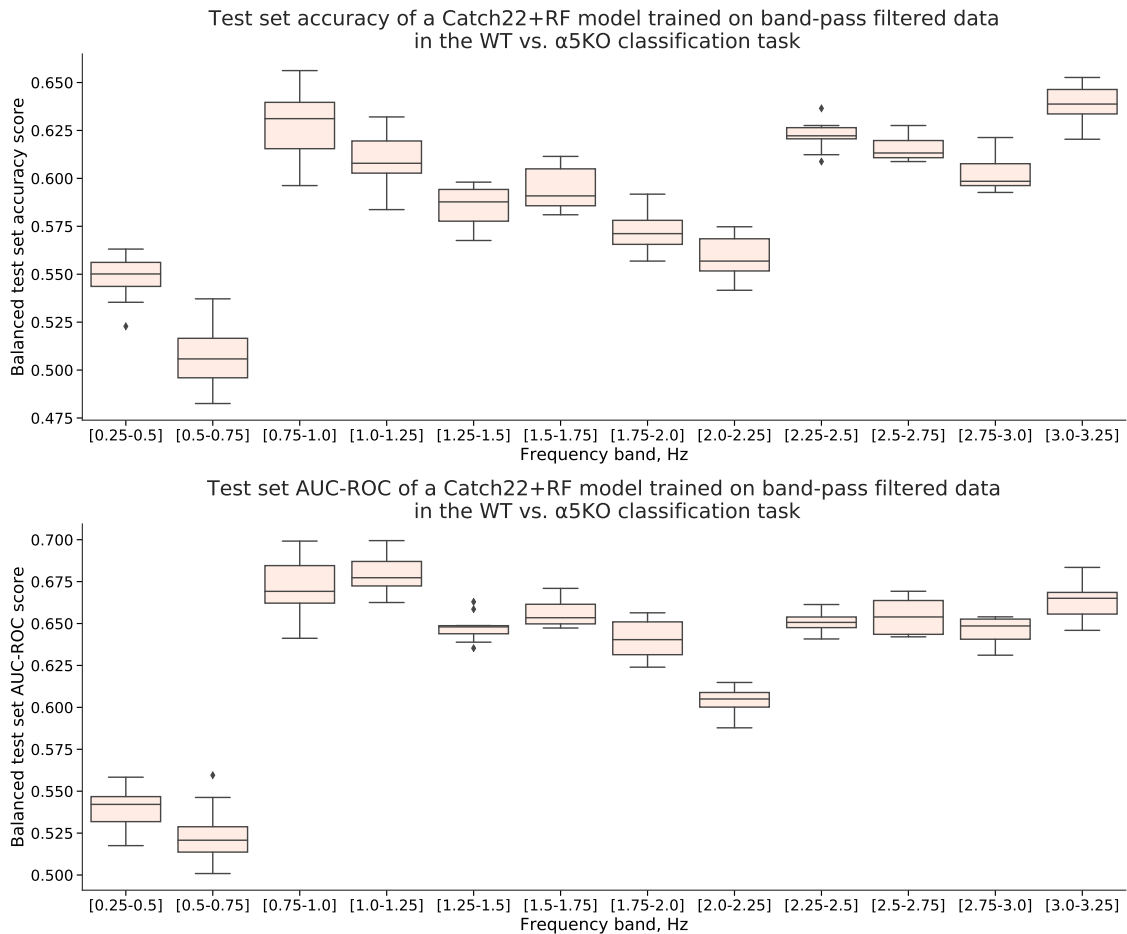


Figure 4-10 – Classification accuracy (top panel) and AUC-ROC scores (bottom panel) in a WT vs.  $\alpha$ 5KO neural activity classification task with Catch22 + Random Forest classification models trained on band-pass filtered signals depending on the frequency range in the band.

Time-series feature	Wasserstein distance between WT and $\alpha 5$ KO groups
Mean value	0.194852
Median value	0.250456
Standard deviation value	0.429985
Permutation entropy (dimension=7, $\tau = 1$ )	<b>0.871484</b>
1st AR coefficient (k=10)	0.751611
Matrix profile (feature="75", threshold=0.98)	0.614185

Table 4.2 – Wasserstein distance values between time-series feature value distributions in WT and  $\alpha 5$ KO animal groups.

## 4.2 Detection of $\alpha 5$ subunit containing nAChR dysfunction from interneuronal activity and the effect of nicotine application

We have further investigated whether the pathological activity could also be inferred from the activity of interneurons in the PFC. Previously, we have shown that the imaged PYR neuron activity is predictive of the  $\alpha 5$ -containing nAChR mutations. Figure 4-12 shows classification accuracy and AUC-ROC scores when using activity of PV and SOM interneurons as input data to the model. Significantly higher classification scores for PV interneurons suggest that the activity changes caused by the  $\alpha 5$  nAChR knockout are most prominent in PV interneurons compared to SOM interneurons and pyramidal cells as well. Given that PV interneurons could be accurately classified from other interneuron types by their activity trace features Troullinou et al. [2020], one could first predict the neuron type in case it is not known and then weigh predictions from different neurons based on the predicted type, with predictions from PV interneuron traces having the largest weights.

Figure 4-13 shows the accuracy and AUC-ROC scores for two binary classification tasks: (i) WT vs.  $\alpha 5$ SNP animal classification and (ii) classification of WT vs.  $\alpha 5$ SNP animals after nicotine application. One could observe the reduction of classification scores to almost chance level when trying to train a model to predict the presence of the  $\alpha 5$ SNP mutation after nicotine application, with above-chance-level accuracy for the  $\alpha 5$ SNP animals without nicotine application. This result strongly

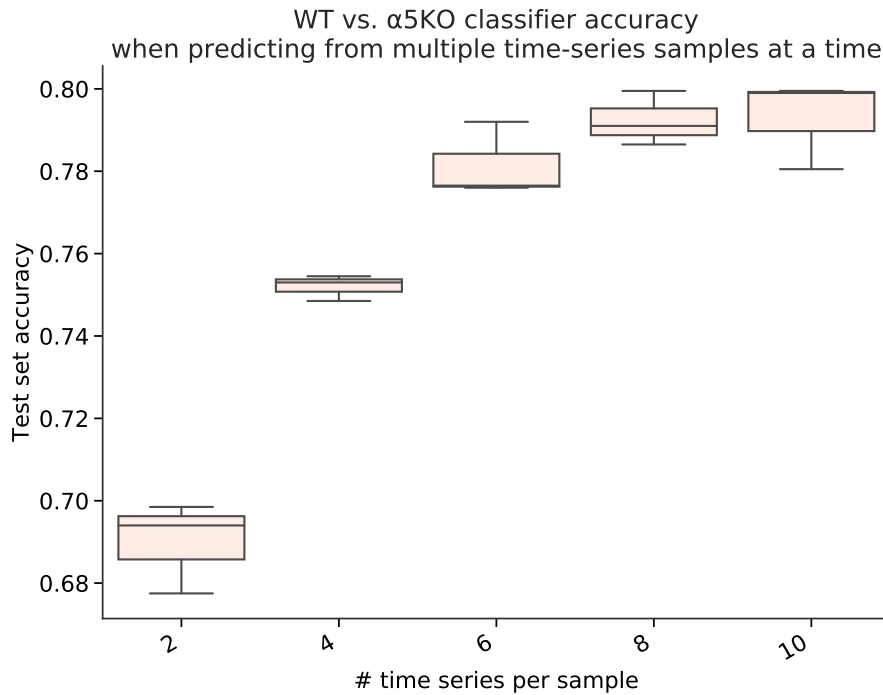


Figure 4-11 – Accuracy scores of a random forest classifier model trained on *tsfresh*-encoded data in the WT vs.  $\alpha$ 5KO classification task when predictions are made for several samples at a time and averaged depending on the number of samples. Samples are independently taken from the training set.

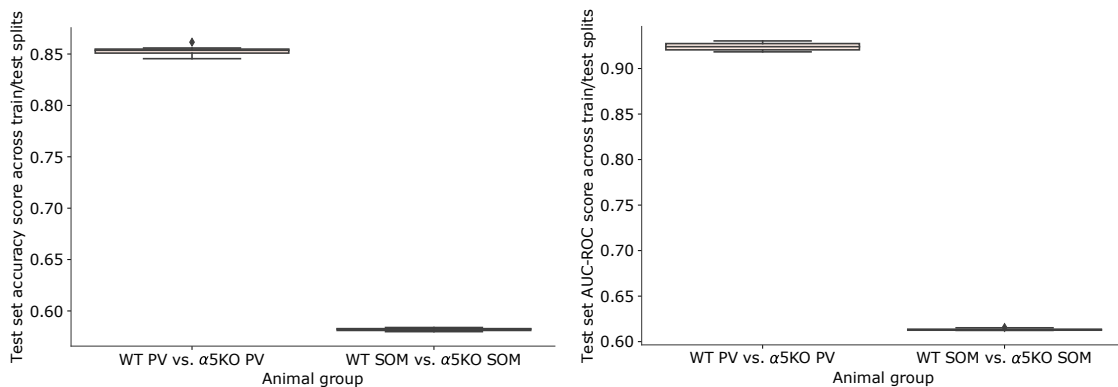


Figure 4-12 – Accuracy (left panel) and AUC-ROC (right panel) scores in the WT vs.  $\alpha$ 5KO animal group classification task from the imaging activity recordings of PV and SOM interneuron subtypes. Note significantly better classification quality in the case of PV recordings, implying that PV interneurons are affected the most by the knock-out.

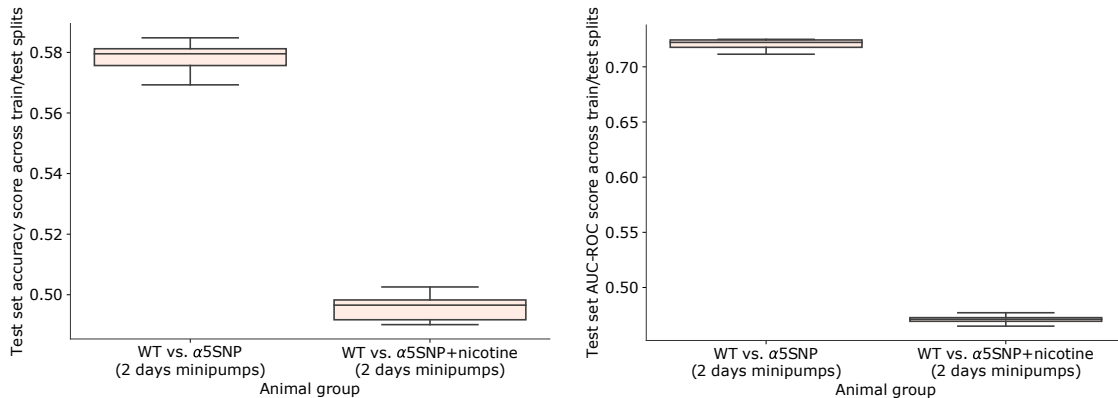


Figure 4-13 – Accuracy (left panel) and AUC-ROC (right panel) scores in the WT vs.  $\alpha$ 5SNP animal group classification task under control conditions and after nicotine application in the SNP group. Note the poor classification quality in the case of nicotine application, implying significant activity restoration in  $\alpha$ 5SNP animals subject to nicotine application.

supports the activity re-normalization hypothesis due to nicotine in  $\alpha$ 5SNP animals. It is important to note that the previously observed nicotine normalization effect [Koukouli et al. \[2016a\]](#) was measured the by lack of difference in estimated average per-animal firing rates. We have, however, demonstrated that it is not possible to train an accurate machine learning model to differentiate between WT  $\alpha$ 5SNP-Nic animal groups based on a range of different time-series features computed from the imaged activity traces. This means that the activity is restored after nicotine application as measured not only by the mean firing rate or other simple statistics of the activity traces, but also by comprehensive vector encodings of the corresponding time-series.

### 4.3 Detecting the effect of beta-amyloid expression on the activity of PFC neurons

We next turn to the imaging data in mice with AD-like activity deficits, elicited using an adeno-associated viral vector expressing the human mutated amyloid precursor protein (AAV-hAPP) [Koukouli et al. \[2016b\]](#). Intracranial injection of AAV-hAPP causes  $A\beta$  production as early as one month post-injection, leading to increased pyramidal cell activity when injected in the prefrontal cortex. We collected data

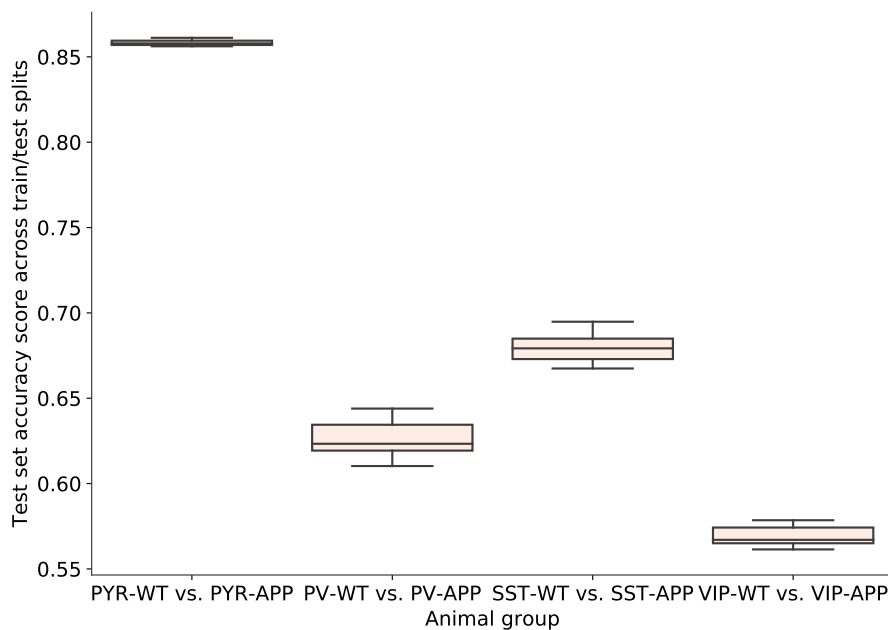


Figure 4-14 – Accuracy score distribution over different train/test split in the task of classifying wild-type animals from the ones with  $A\beta$  expression depending on the neuron type in the PFC. Note the low classification quality obtained with VIP activity data recordings suggesting the significant invariance of the VIP population to  $A\beta$  expression.

from wild-type APP expressing mice as well as nAChR knock-out mice expressing APP and formulated a set of different neural activity classification tasks between animal groups.

Figure 4-14 shows the accuracy scores in the tasks of predicting animals with the APP expression depending on the neuron type that the activity traces are taken from. The most significant change in activity, and hence highest classification scores, are seen in the pyramidal neuron traces, with classification scores almost as low as the chance level for the VIP interneuron traces. The latter observation hints at the apparent invariance of VIP activity to APP expression, not as apparently present in PV and SST interneuron groups.

Figure 4-15 shows obtained classification scores for the three classification tasks of detecting the animals with expressed  $A\beta$  versus the control group in animals with genetic knock-outs of three different receptor groups –  $\alpha 7$ ,  $\beta 2$ ,  $\alpha 5$ -subunit containing nAChRs, respectively. The accuracy score level allows to evaluate the strength of  $A\beta$  expression in disrupting neural activity in different knock-out animal groups.

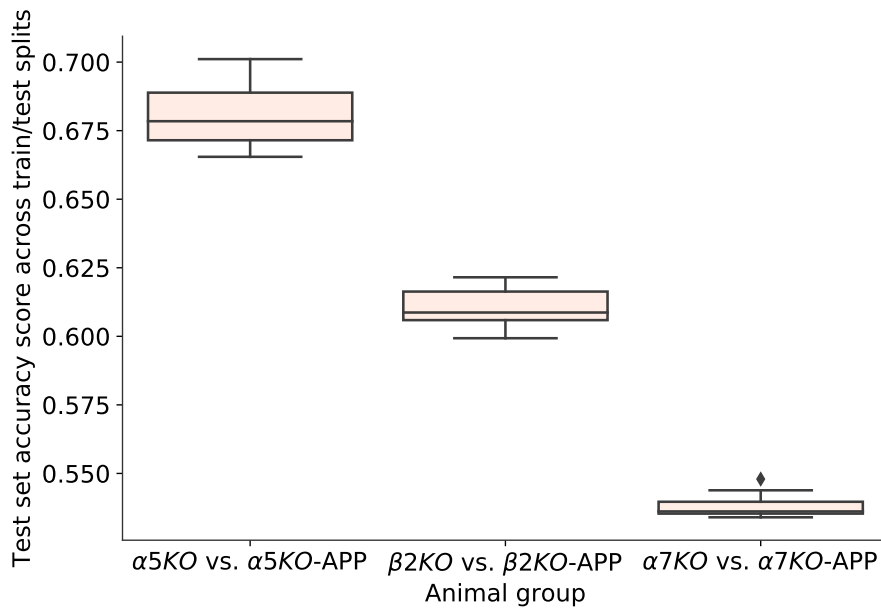


Figure 4-15 – Accuracy score distribution over different train/test splits in the task of classifying control nAChR knock-out animal groups from the knock-out groups expressing  $A\beta$ .

Figure 4-16 shows AUC-ROC classification scores when classifying between APP-expressing mice versus two different groups with pharmacological manipulations – one control one (with vehicle injection) and another with galantamine injection. We have demonstrated in Chapter 3 that galantamine application leads to reduced activity levels relative to the wild-type APP-expressing animal group. This is directly reflected in the classification score values, which could be interpreted as inter-group activity difference scores as reflected by the trained machine learning models. The classification scores in the APP vs. APP with galantamine application classification task are found to be significantly higher than scores in the control task with vehicle-injected animals.

## 4.4 Conclusions

In this Chapter, we have shown that the nicotinic acetylcholine receptor dysfunctions in the prefrontal cortex associated with common nervous system disorders lead to changes in activity of all neuron types which can be accurately detected with machine learning approaches on the single-neuron scale. Applying a range of different

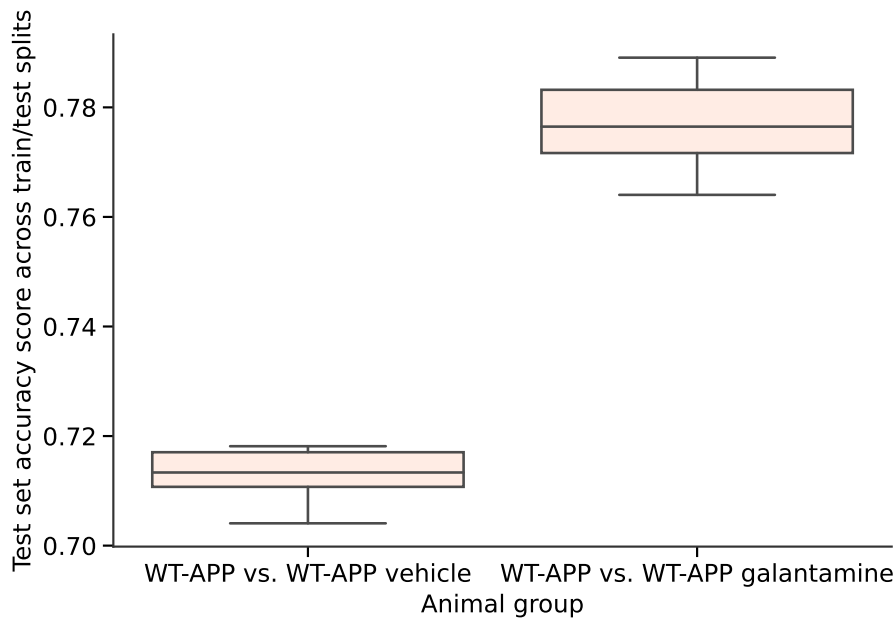


Figure 4-16 – AUC-ROC score distribution over different train/test splits in the task of classifying control  $A\beta$ -expressing animal groups vs.  $A\beta$  animals with vehicle and galantamine injections.

machine learning algorithms to the dataset of imaged neural activity in the PFC in different animal groups allowed us to reveal the time-series features most predictive of dysfunction as well as the primary frequency bands associated with patterns of pathological activity. We have also used the obtained classification scores in tasks involving different animal groups to support the hypothesis of activity restoration in animals with the  $\alpha 5$ SNP mutation after nicotine application and the hypothesis of VIP interneuron activity invariance in animals with beta-amyloid expression relative to the control group. In summary, we have demonstrated the effectiveness of machine learning as a tool to extract the features of neural activity most predictive of disease as well as a tool to evaluate effects of genetic and pharmacological manipulations on neural circuit activity.

## Chapter 5

# Deep learning models detect markers of early amyotrophic lateral sclerosis from neural activity and movement data

Ivan Lazarevich, Suhel Tamboli, Lisa Topolnik

To be submitted to *Cerebral Cortex* as a part of: Amalyan, S., Tamboli, S., Lazarevich, I., Topolnik, D., Bouman, L. H., Topolnik, L. (2021). Impaired motor cortex function in female mice with a C9orf72 genetic expansion.

I.L. and L.T. conceived and designed computational experiments. I.L. performed computational experiments. S.T. and L.T. performed all *in vivo* experimental work.



In this Chapter, we apply the approach developed in the previous chapters to a data set of cortical activity in animal models of early amyotrophic lateral sclerosis. The task is to detect the markers of the disease from the recordings of ensemble activity in the motor cortex in freely-behaving animals. We demonstrate that the time-series feature extraction approach does not work well for this task and propose an alternative based on deep convolutional networks. All experimental data used in this chapter were acquired at the Topolnik lab, Universite Laval, Quebec City, Canada.

## 5.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a devastating incurable disease, in which progressive loss of upper and lower motor neurons causes muscle weakness and loss [Rowland and Shneider \[2001\]](#). Most of the cases of ALS are sporadic, whereas only 10% have a family inherited history [Gros-Louis et al. \[2006\]](#), and until now, no definitive cause is known for ALS. The most common genetic abnormality identified in familial and sporadic ALS cases is a GGGGCC (G<sub>4</sub>C<sub>2</sub>) hexanucleotide repeat expansion in the C9orf72 gene [DeJesus-Hernandez et al. \[2011\]](#). Recently, the first mouse model has been developed, which reproduces the human form of C9orf72 repeat expansion with  $\sim 500$  G<sub>4</sub>C<sub>2</sub> repeats (C9-500; [Liu et al. \[2016\]](#)), and, depending on genetic and environmental factors [Mordes et al. \[2020\]](#), demonstrates some behavioral and neuropathological features of ALS/FTLD [Liu et al. \[2016\]](#). We used the C9-500 mouse model to directly explore the impact of an ALS/FTLD risk gene mutation on the activity of neuronal circuits involved in the control of motor functions. We utilized calcium imaging data recorded in these mice and machine learning approaches to link motor cortex activity and movement in healthy animals and in ALS pathology.

## 5.2 Neural activity data in the model of the early amyotrophic lateral sclerosis

The data comprises simultaneous recordings of neural activity in the motor cortex captured by calcium imaging and movement data (speed signals) in healthy (control) mice and mice carrying the most common ALS mutation – c9orf72. Calcium signals in axons of pyramidal neurons in the motor cortex were imaged using fiber-photometry on awake mice. During experiments, mice were freely moving with a wireless miniature fiberscope attached, the animal speed was recorded simultaneously with the calcium signals. The sampling rate of the calcium signal was close to 100 Hz, with recording duration of around 300 seconds per animal (per experiment). The video camera used for speed recording captured 10 frames per second, so we upsampled the raw recorded speed signals using standard linear one-dimensional interpolation with an upsampling factor of about 10 to match the number of timestamps in the imaged calcium signal. Then, the dataset consists of recordings from 7 wild-type (control) animals and 7 mutant animals, with a single signal of about 30000 timestamps per animal (in both neural activity and speed modalities). The neural activity signal is an aggregate proxy for the ensemble activity of the motor cortex, rather than single-neuron activity. The amount of data is hence limited with a single two-channel signal per animal and 14 animals recorded in total.

We applied standard pre-processing to signals of both modalities: a linear detrending procedure and standard scaling along the time axis (subtracting the mean and dividing by the standard deviation). We further applied a denoising procedure to both channels in the signal by convolving them with a Gaussian kernel with  $\sigma = 10$  timepoints. We then applied a rolling window of size equal to 500 timestamps with a step of 100 timestamps to produce a dataset of fixed-length time series of both activity classes (control mice and mutant mice). This procedure resulted in a dataset of 3609 samples with 2 channels (modalities) and 500 timestamps each. The objective is to train a machine learning model to predict the class label (0 for control, 1 for mutant) which essentially boils down to a multivariate binary time-series classification problem. The average target variable value in the dataset is 0.4989,

the dataset is thus fairly balanced. The presence of two distinct modalities in the data allows us further to evaluate the amount of predictive information contained in the neural activity data and in the movement data alone and in their joint dynamics (i.e. features of the interaction between motor cortex activity and movement) and to also estimate how much the recorded data contributes to prediction quality depending on the animal state (e.g. during low-mobility and high-mobility periods).

### 5.3 Model validation scheme

Due to the small size of the full dataset, we did not perform a single train/test split, but rather utilized a leave-one-out cross-validation scheme, whereby we performed 14 training experiments for each model corresponding to every one of the animals being consecutively put into the test set while the rest of the data comprised the train set. Therefore, for each of the 14 experiments, the test set is comprised of the 500-timestamp time-series chunks corresponding to a particular animal ID, with that animal being in either the control or the mutant group. The trained model is then used to output the probability of the time-series sample belonging to the mutant group independently of each sample and either the average probability or the average predicted label (thresholded probability) is taken as the output probability of the animal belonging to the mutant group. We did not perform class balancing of the training datasets in any of the reported training results, as we found empirically that the inherent class imbalance caused by the described train/test dataset construction did not significantly impact the obtainable accuracy scores. Example visualization of model predictions distributed across training trials of an FCN model can be seen in Fig. 5-5. We base the ranking of different approaches on the cross-entropy loss function for median model predictions averaged across animals.

## 5.4 Visualizing data set structure by hand-crafted time-series feature encoding

We first began exploring the dataset using the massive time-series feature extraction approach whereby we encoded each of the time-series samples as vectors of features computed as pre-determined manually-engineered functions of the time-series. We started with the neural activity modality only and encoded the samples using the *tsfresh* package [Christ et al. \[2018\]](#). After sample-wise normalization of every feature column and discarding of the low-variance features, we trained a random forest classifier model (with 1000 decision tree estimators with a maximal depth of 5) to predict whether a sample came from a WT or a mutant animal. We used the feature importance scores from the random forest model to condense the feature set to 30 most discriminative ones (the ones with highest importance scores). We then performed a two-dimensional embedding of the computed 30-dimensional feature vectors using a linear (PCA) and a non-linear (t-SNE) embedding techniques to visualize the dataset structure (Figure 5-1). One can note the lack of clear visual separation between points corresponding to either WT or mutant animals (points being color-coded). We observed a similar picture for *tsfresh*-encoded speed signal samples (Figure 5-2) as well as feature vectors comprised of features extracted from both neural activity and speed time-series (Figure 5-3). This lack of clear separation in manually-constructed time-series feature space is also reflected in poor classification quality obtained with models trained on time-series feature representation of the data. Figure 5-4 shows the model predictions across time-series samples for each animal separately in the test set obtained with random forest classifier trained on a Catch22 encoding of the dataset [Lubba et al. \[2019\]](#) for (i) only the neural activity modality, (ii) only the speed modality and (iii) both modalities (with features from modalities concatenated to form larger feature vectors). What we observed was poor performance in predicting the wild-type animals (particularly for animals with IDs 0, 1, 2) and overall high variance (hence uncertainty) including the predictions for the mutant animal group. This feature extraction approach is inherently designed for univariate time-series and the final predictions could only be obtained as a function

of aggregate features each extracted from a single given modality. As we demonstrate further, the interaction between cortical activity and movement time-series is essential to obtain accurate model predictions in this task.

## 5.5 Pathological activity detection with deep convolutional neural networks

We next turn to deep learning models for time-series to solve our classification problem. Deep neural nets can tackle multi-variate time-series classification by design, with each modality incorporated as a separate channel in the input tensor. We found that a simple one-dimensional convolutional architecture named FCN could provide sufficiently good predictions when trained on signals of both modalities represented as different input (one-dimensional) image channels (Figure 5-5). The results we obtained with this baseline model in terms of per-animal predictions correspond to 3 falsely classified animals (animal 0 is a false positive – a WT animal classified as a mutant and animals 9 and 12 are false negatives – mutant classified as control animals). The predictions are done by comparing the median predicted label over training trials with a pre-defined confidence threshold, in our case equal to 0.5 and specified with a dashed horizontal line in Fig. 5-5. In other words, after each training trial the model is used to predict a discrete label (0 or 1) for each of the samples in the testing set (all of these samples corresponding to a single animal), the per-sample mean of predicted labels is aggregated over several training runs (7 in our case) and the median value is given as a final output probability of the animal belonging to the mutant group. The decision whether the animal belongs to the mutant group is made whenever the output probability for the animal exceeds the 0.5 threshold. Figure 5-5 also shows the distribution of mean per-sample probabilities of being in the MUT group over separate training runs in the bottom panel. The decision to aggregate results from different training runs stems from the variability in neural net training observed depending on the random initialization of its weights. This variability is demonstrated in Figure 5-6 where the dynamics of the test set loss (the cross-entropy loss on the test set) and the test set accuracy during



Figure 5-1 – Top: Scatter plot of a two-dimensional PCA embedding of the motor cortex activity time-series encoded by top-30 tsfresh features (with features sorted by importance scores determined from a random forest classifier trained on the dataset). The class labels are color-coded with red points corresponding samples from control animals and blue points to samples from mutant ones. Bottom: similar scatter plot of the top-30 tsfresh features of neural activity time-series embedded with a t-SNE algorithm.

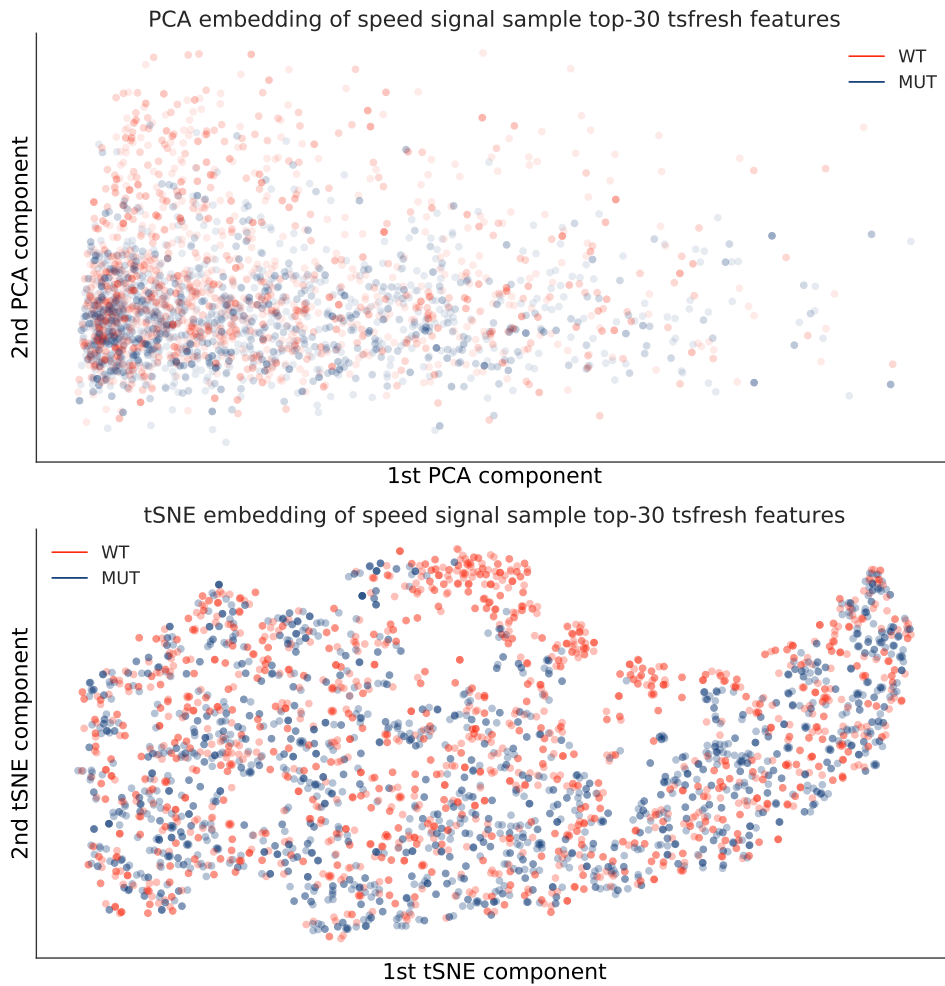


Figure 5-2 – Top: Scatter plot of a two-dimensional PCA embedding of the animal speed signal time-series encoded by top-30 tsfresh features (with features sorted by importance scores determined from a random forest classifier trained on the dataset). The class labels are color-coded with red points corresponding samples from control animals and blue points to samples from mutant ones. Bottom: similar scatter plot of the top-30 tsfresh features of the animal speed signal time-series embedded with a t-SNE algorithm.

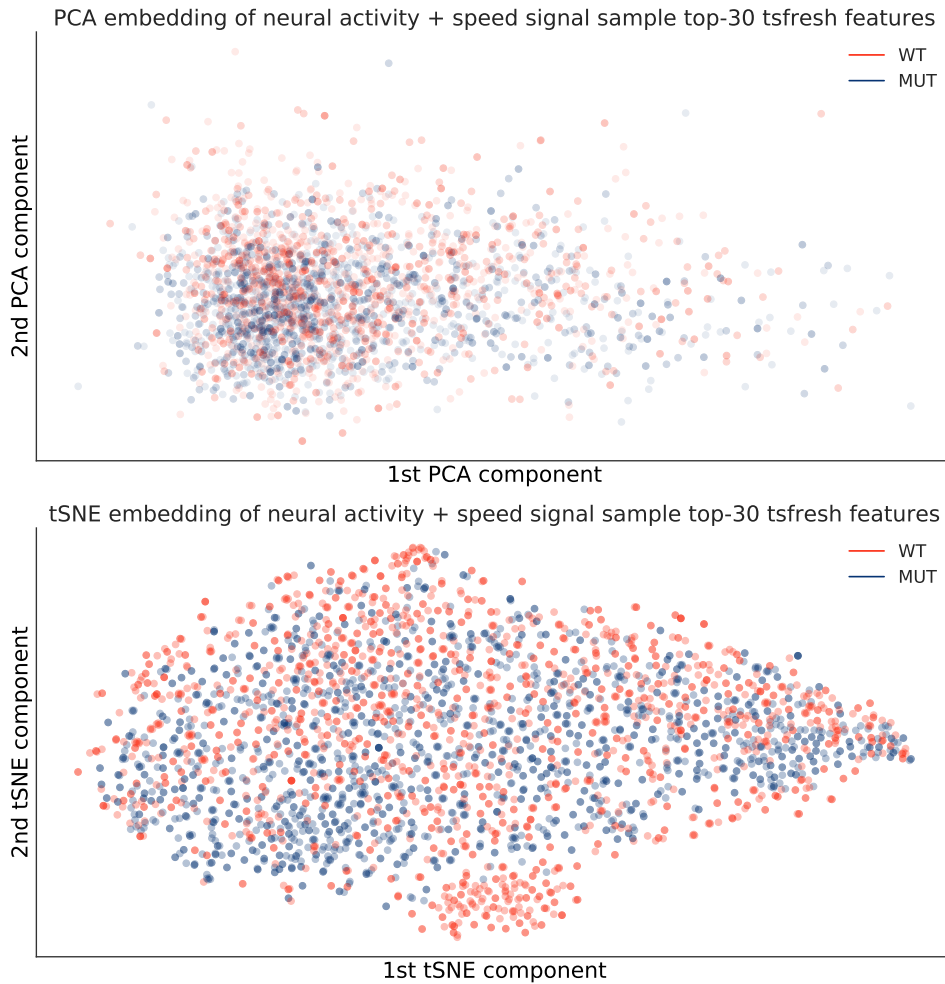


Figure 5-3 – Top: Scatter plot of a two-dimensional PCA embedding of the top-30 feature vectors obtained by combining tsfresh features from both neural activity and speed signal time series (with features sorted by importance scores determined from a random forest classifier trained on the dataset). The class labels are color-coded with red points corresponding samples from control animals and blue points to samples from mutant ones. Bottom: similar scatter plot of the top-30 tsfresh features from both neural activity and speed signal time series embedded with a t-SNE algorithm.



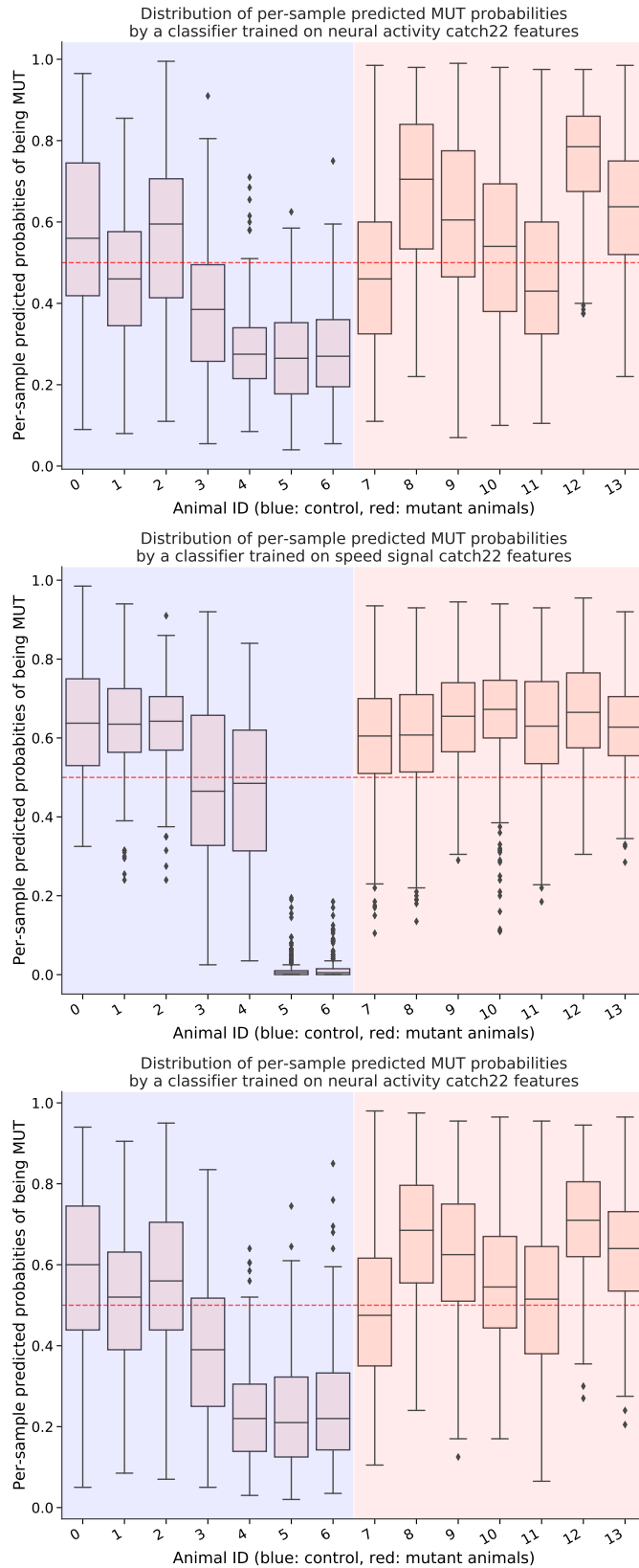


Figure 5-4 – Accuracy scores (average predicted labels) per animal over training trials of a Catch22 time-series classifier trained on (top) neural activity, (middle) movement data and (bottom) neural activity + movement data using the leave-one-out cross-validation scheme.

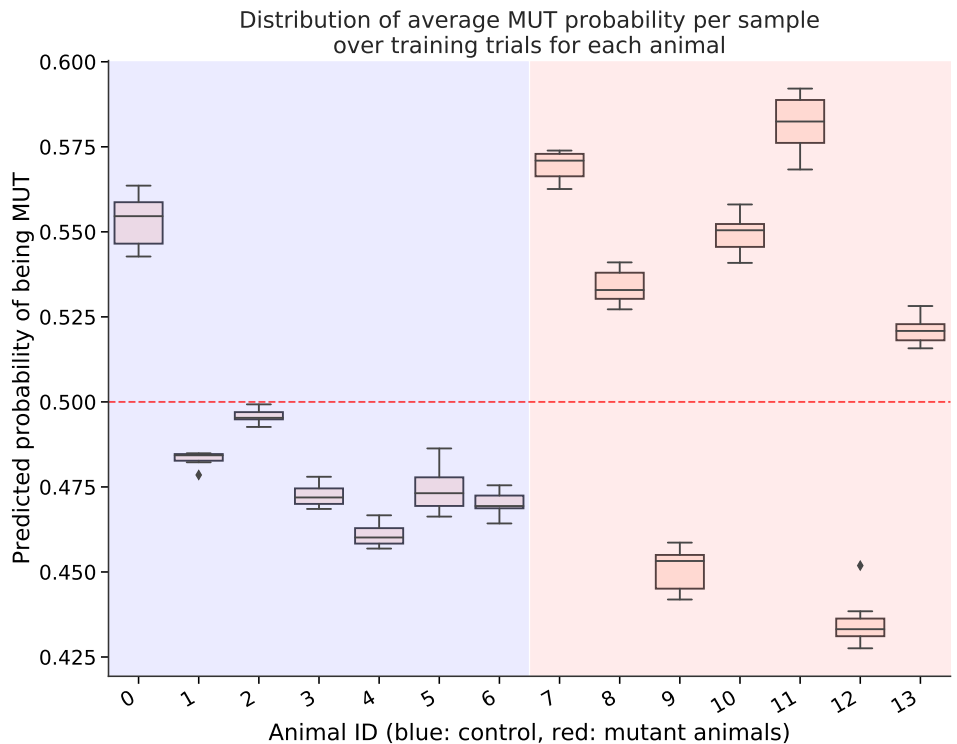
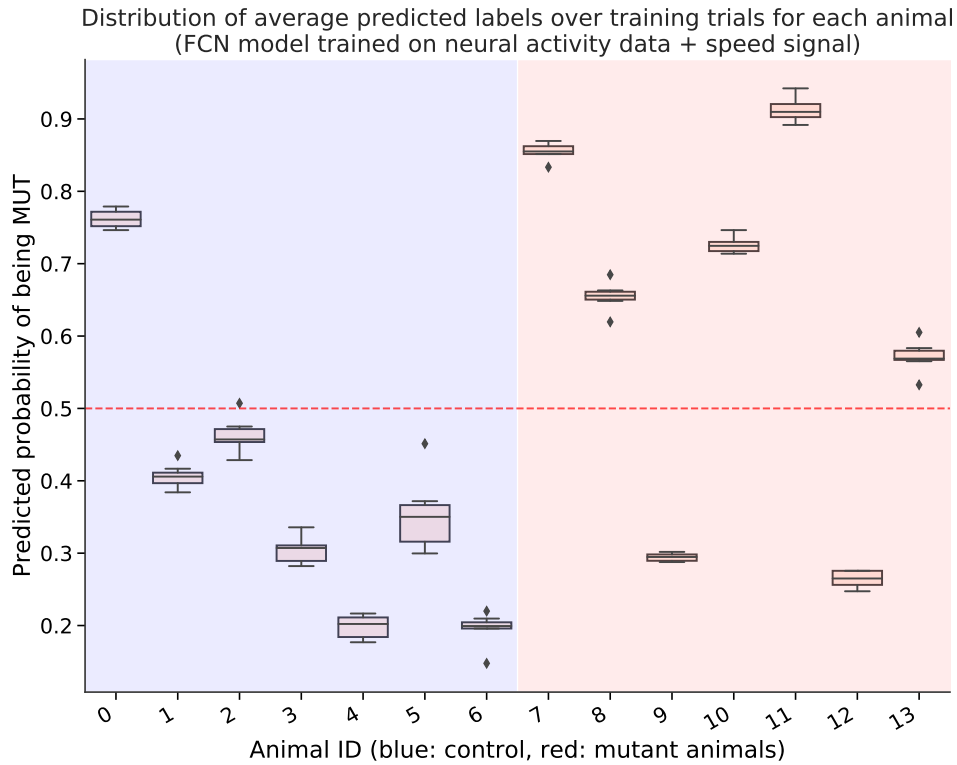


Figure 5-5 – Top: Accuracy scores (average predicted labels) per animal over training trials of an FCN model trained on neural activity + movement data using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group). Thresholding the average labels with a value of 0.5 to get the animal-wide predictions leads to a single false positive error (animal ID 0) and two false negative errors (animal IDs 9 and 12). Bottom: same for per-sample average mutant class probabilities for each animal.

training is visualized for different training runs and different animals. The animals 2 and 13 are correctly classified, but the accuracy of a single run could be close to 0.5 meaning that approximately half of the samples in the test set are misclassified. This is resolved by averaging predictions over different training runs, which consistently converge to an accuracy value higher than 0.5, also demonstrated in Figure 5-7 depicting probability distributions over samples in the training set across training runs. Figure 5-6 also demonstrates that the default amount of iterations of the FCN model (50 training epochs) is sufficient for the model to converge to certain local optimum.

The model used to generate predictions for the referred figures is a model named FCN (fully-convolutional network) which is a VGG-like model (FCN baseline from Fawaz et al. [2020]). The default architecture consists of three convolutional layers (with BatchNorm layers and ReLU activations following the convolutions) with kernel sizes [7, 5, 3] and output channel counts [128, 256, 128], respectively, followed by a global average pooling layer and a fully-connected classification head. We found this simple architecture hard to beat with more advanced deep learning approaches, as detailed in the following section.

## 5.6 Choice of DNN architecture: an ablation study

We compared performance of a range of different neural net architectures commonly used for sequence modeling in general and time-series classification in particular such as various one-dimensional convolutional architectures such as the FCN Wang et al. [2017], XCM Fauvel et al. [2020] and InceptionTime Fawaz et al. [2020] as well as non-convolutional architectures such as the fully-connected MLP Fawaz et al. [2019], Transformers and more recent Time-Series Transformers (TSTs) Zerveas et al. [2020] as well as an array of different recurrent architectures like LSTMs Hochreiter and Schmidhuber [1997] and GRUs Chung et al. [2014]. Surprisingly, we found that the simple FCN baseline appears quite hard to beat with more advanced approaches. The default hyper-parameters for the training procedure used throughout the work are a constant learning rate schedule with a learning rate of 1e-3 and 50 training

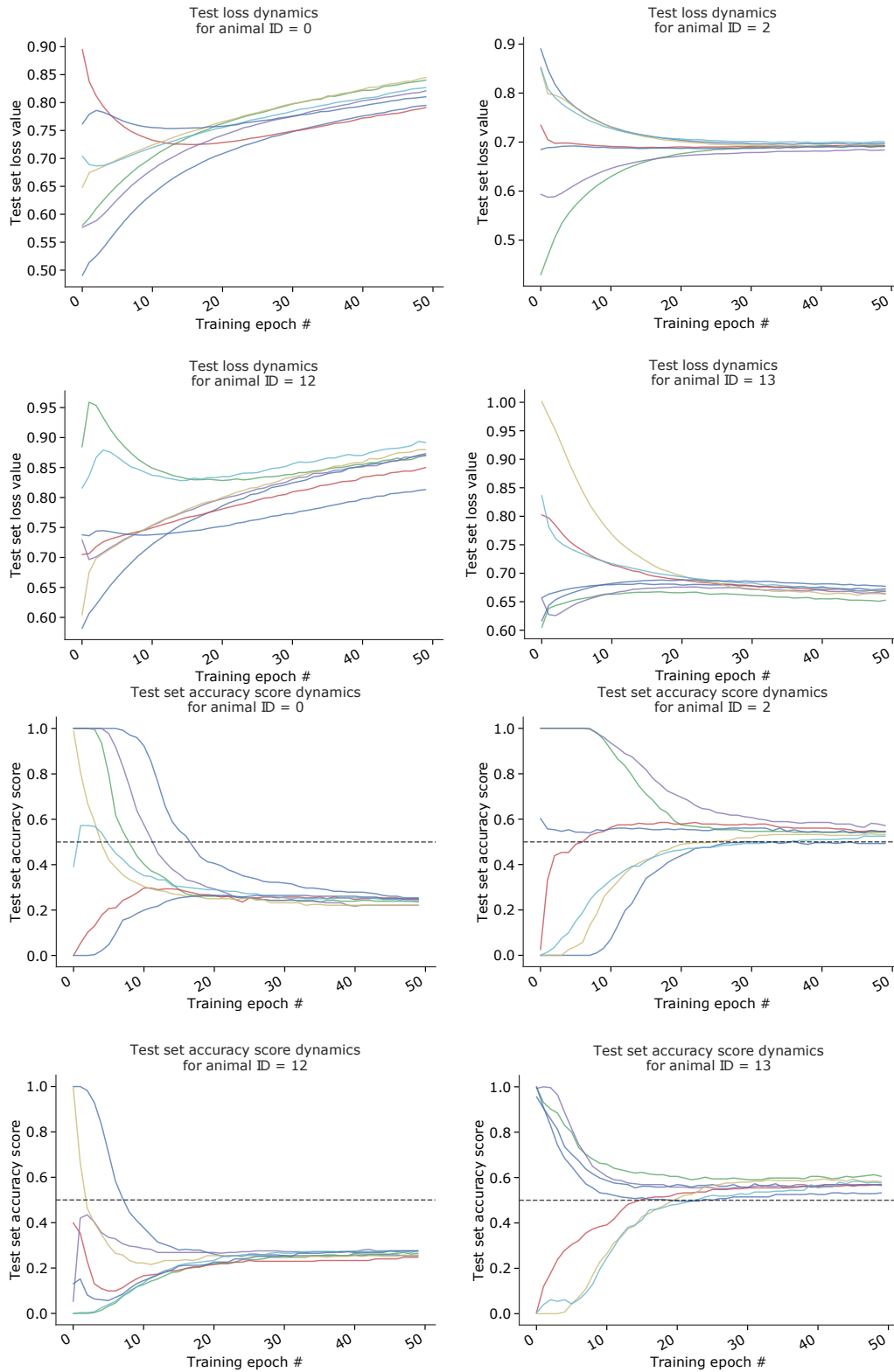


Figure 5-6 – The dynamics of test set cross-entropy loss (top panels) and test set accuracy (bottom panels) during neural net training across different training trials for different animals. Animals 0 and 12 correspond to misclassified animals from WT and MUT groups, respectively, animals 2 and 13 are correctly classified animals from the two respective groups. Note that generally neural net training converges to a similar accuracy values regardless of random initialization.

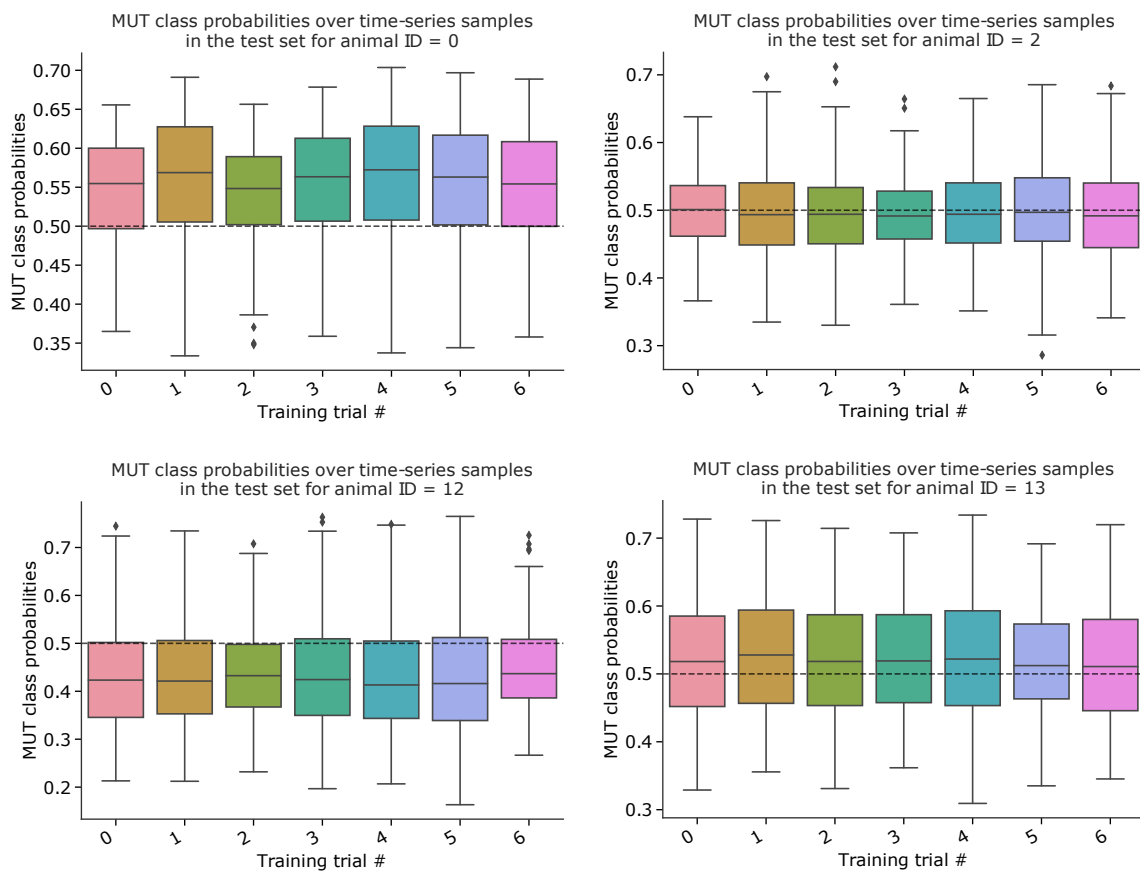


Figure 5-7 – Per-sample distributions of probabilities of test set samples belonging to the mutant group over training trials. Note the distributions for animal 2 highlighting the need for averaging the final predictions across training runs.

epochs and an SGD optimizer, with a PyTorch-based implementation of the training procedure provided in the *tsai* package Oguiza [2020]. We did not use data augmentations unless specified (augmentations were used in the state-dependent sample selection experiments to speed up the training). First, we selected a single animal to first evaluate a whole zoo of architectures available in the *tsai* package on. We observed that the animals typically misclassified by the FCN models after training (such as animal 0 and animal 9) were also incorrectly classified by the other architectures we tested. We thus focused on an animal that was correctly classified in most cases but had a high uncertainty in neural net predictions, in particular animal 2. Data recorded from that animal was used to construct a test set of samples and the corresponding performance results of the different DNN architectures are shown in Figure 5-8. The top panel of the figure 5-8 demonstrates scores obtained with different convolutional architectures, with the FCN model being an apparent leader, both in terms of median accuracy and in terms of having low variability of accuracy scores across training runs. The bottom panel of Figure 5-8 shows performance across primarily non-convolutional architectures. LSTM has significantly better accuracy scores compared to vanilla RNNs and GRUs (which results in a misclassification for animal 2), but we found that LSTM performance was not consistent across other animals (ones from the mutant group in particular). Transformer models were found to be among the better performing ones, in particular when the learning rate was reduced by an order of magnitude (with a fixed value of  $1e-4$  throughout training), however this class of models was observed to have large variance in final accuracy scores. This effect of large variance in predictions was further confirmed during evaluation on every one of the 14 animals (see Figure 5-11). Recognizing the FCN as one of the best performing architectures on our task, we have further looked into how the hyperparameters of the model affect the obtained classification accuracy. Figure 5-9 shows how the accuracy obtained on the test set comprised of the animal 2 varies with different output channel count layouts, number of layers and the input size. Reduction of the per-layer output channel counts by factors of 4 or 8 results in higher accuracy variance, while the median accuracy remains on a sufficiently good level even for the [16, 32, 16] channel layout. Accuracy variance is somewhat reduced

with a higher number of channels in the first layer of the network. We have also found that the vanilla layer layout can benefit from a lower input resolution, which we obtained by average pooling the input time-series with a factor of 4 (hence, a resolution of 125 timestamps instead of original 500). Classification performance did not significantly improve from the addition of the squeeze-and-excitation blocks to convolutional layers, nor from increasing the depth of the network to 4 or 5 layers (see the bottom panel of Figure 5-9). We further checked whether increased accuracy variance due to reduced channel count was present across all animals during the leave-one-out validation, which turned out to be true (see Figure 5-10), with an increased variance in predictions in spite of the same number of misclassified animals as for the vanilla FCN model.

## 5.7 Prediction accuracy as a function of input data modality and animal state

An important question to tackle when analyzing the predictions obtained in the given task with a deep convolutional neural net is whether predictive information that the network learns to extract from input signals is contained in the interaction between the two input modalities (cortical activity and movement) or it is contained in each modality separately. We have previously seen that nonlinear combinations of aggregate time-series features from both modalities separately provide poor predictive quality. We hence designed ablation experiments for our neural network training whereby we trained our best performing architecture (vanilla FCN) on input tensors with only a single input channel corresponding either to cortical activity time-series or speed signal time-series. The results obtained with FCN models trained on single-modality data are shown in Figure 5-12. Interestingly, we have found that relatively poor prediction quality is observed in both cases, with the cortical activity data apparently containing more predictive information compared to movement data in general. These results, however, strongly suggest that it is the interaction between the two modalities that is learned by the neural networks that allows accurate predictions, since this level of classification performance cannot be reached considering

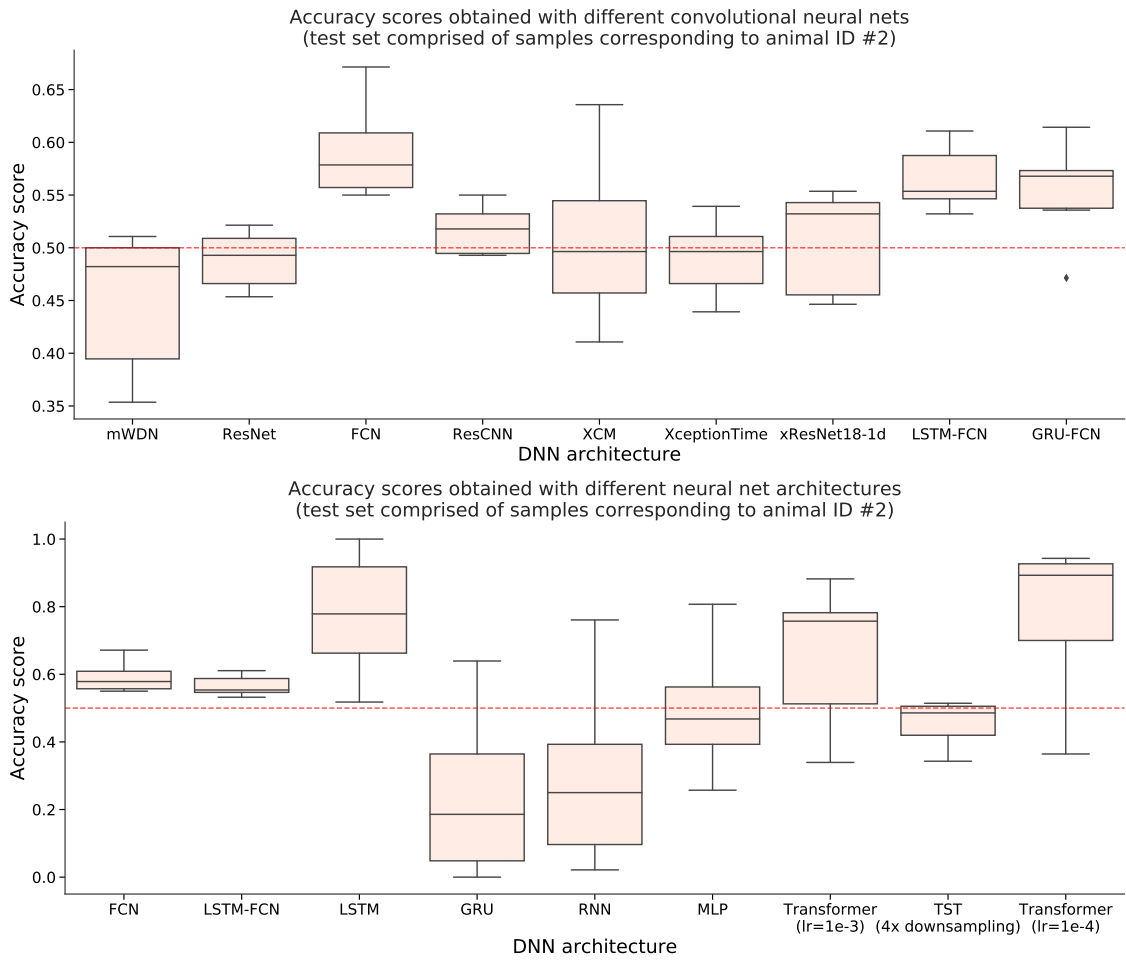


Figure 5-8 – Accuracy score distributions (accuracy scores collected over different training runs) for the test set consisting of samples corresponding to the animal 2 (the rest of the samples in the training set) depending on the neural net architecture with (top panel) common convolutional architectures for time-series classification and (bottom panel) some other architectures widely used for sequence modeling such as recurrent networks and transformers.



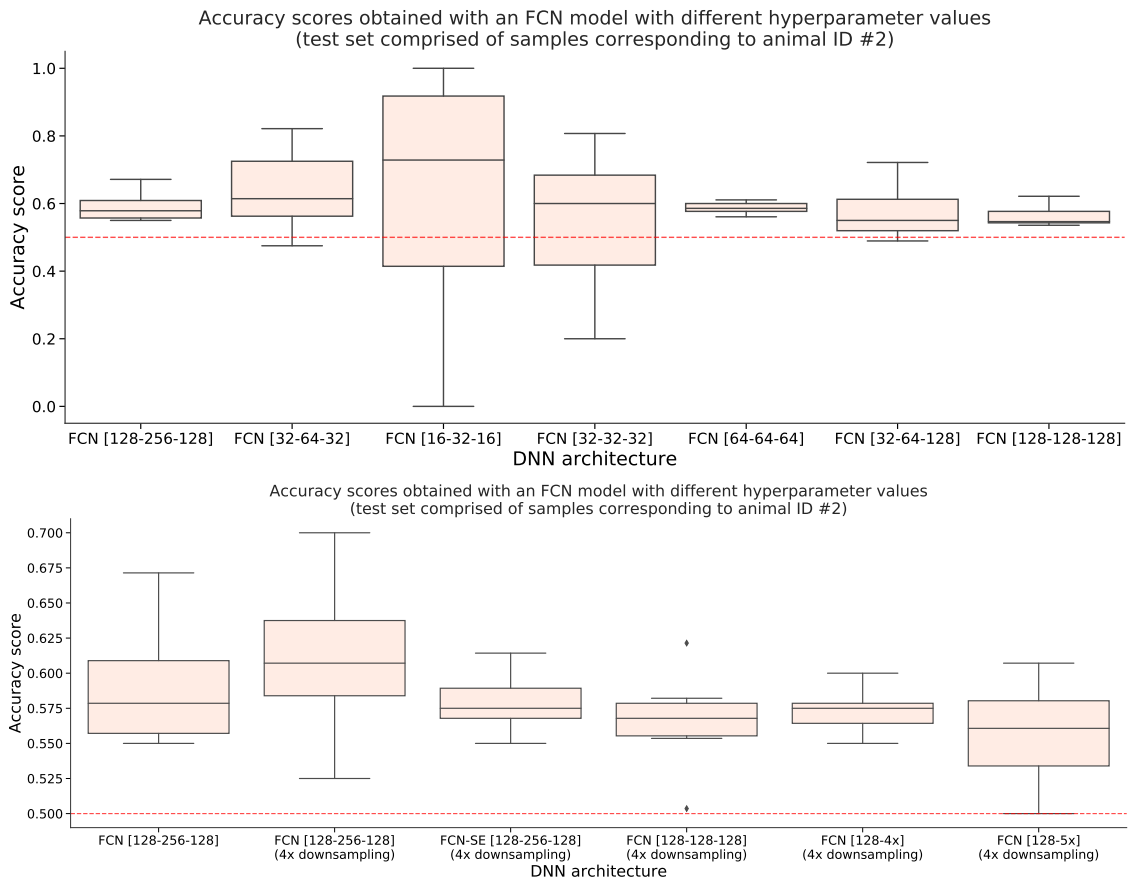


Figure 5-9 – Accuracy score distributions (accuracy scores collected over different training runs) for the test set consisting of samples corresponding to the animal 2 (the rest of the samples in the training set) for FCN models with different hyperparameter values – numbers of output channels, numbers of layers and inputs sizes.

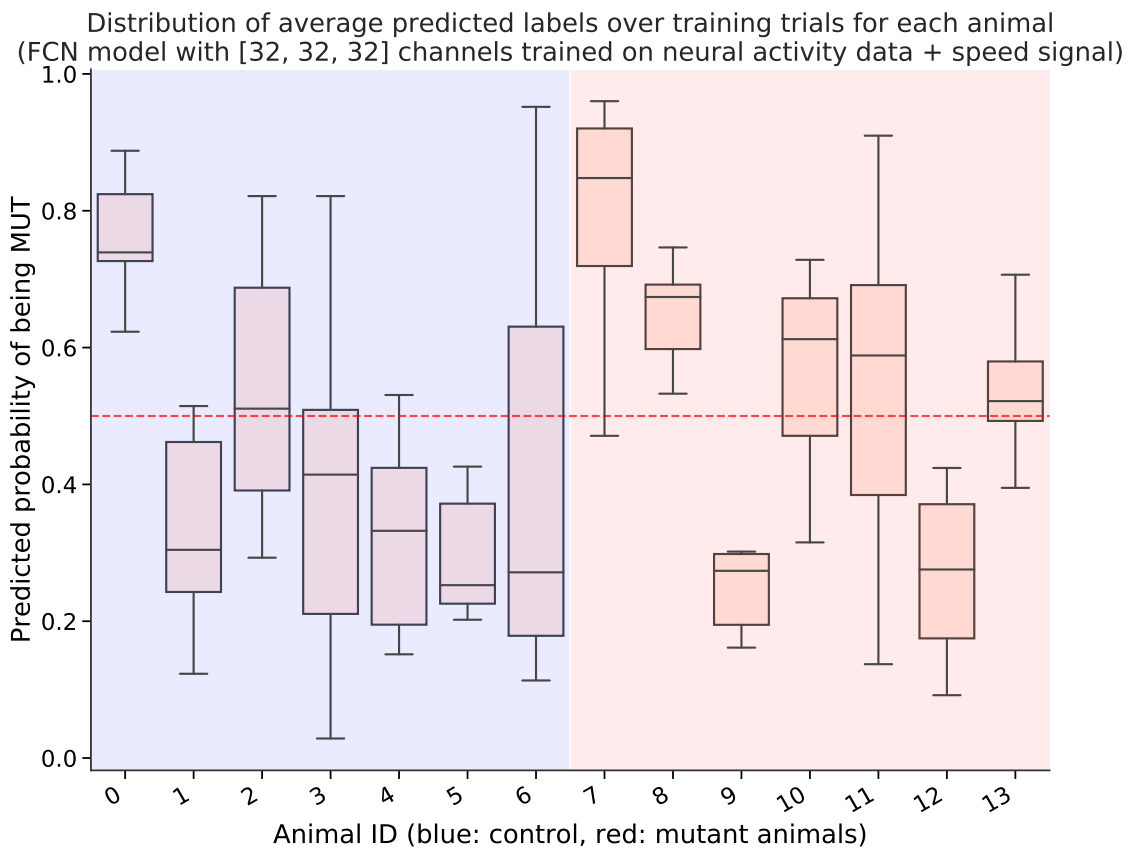


Figure 5-10 – Accuracy scores (average predicted labels) per animal over training trials of an FCN model with [32, 32, 32] output channel layout (other parameters same as in the default model) trained on neural activity + movement data using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group).

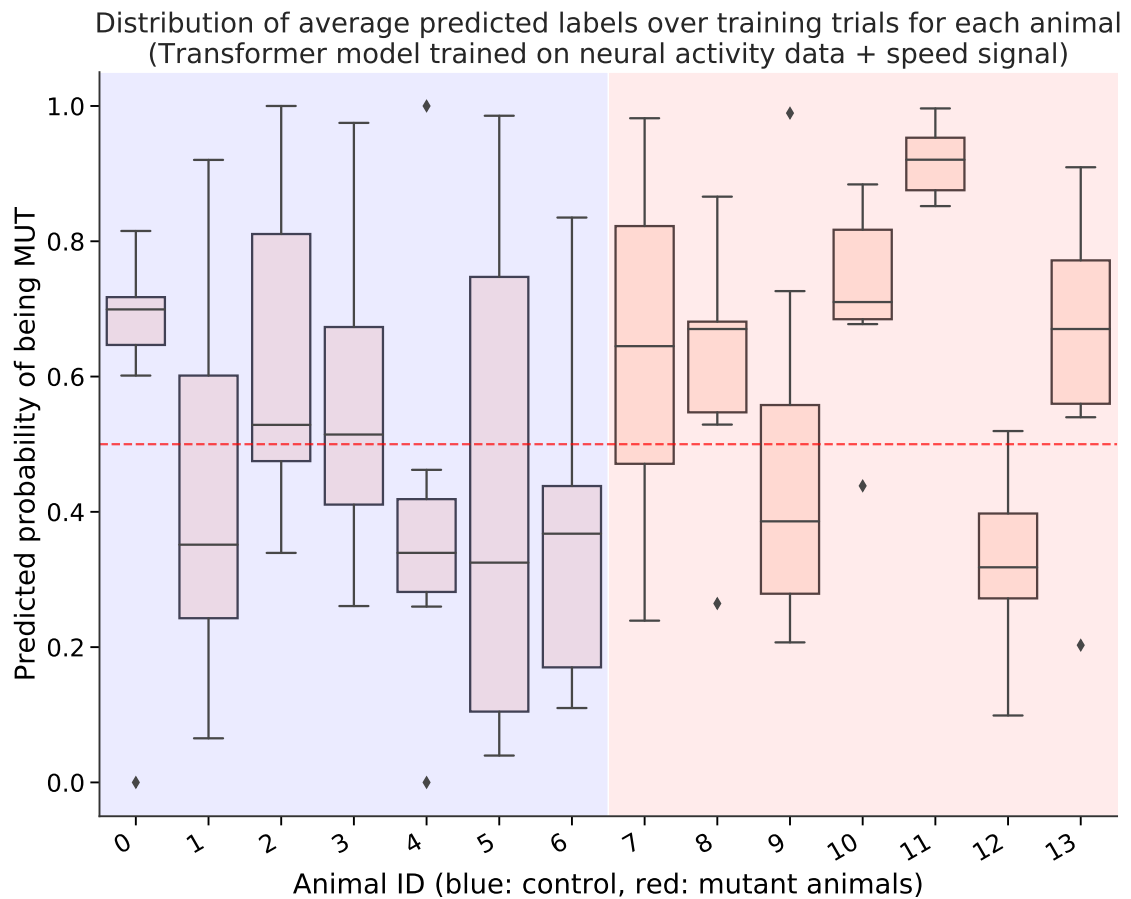


Figure 5-11 – Accuracy scores (average predicted labels) per animal over training trials of a Transformer model with 64 self-attention heads trained on neural activity + movement data using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group).

single modalities separately.

Another important question to address here is whether all the recorded data evenly contributes to neural net training, in other words, whether some of the time-series samples contain significantly more predictive information than others and whether there exist a simple heuristic to filter out the unimportant samples. This issue can be addressed by considering that the ALS-inducing mutation leads to movement impairment in general, so it would be sensible to hypothesize that the periods of high mobility (reflected in the speed data) could more strongly reveal the presence of the pathology rather than the resting low-mobility periods also present in the recordings. We tested this hypothesis by simply filtering the samples in the dataset based on the average speed value in every sample. We defined the samples corresponding to high-mobility periods as the ones with the mean speed value in the sample exceeding a certain threshold (which an adjustable parameter of the dataset filtering procedure) and, conversely, the low-mobility period samples as the ones with the mean speed value below a certain threshold. Based on the speed distribution and in order to keep around 1500-2000 samples in the training set after the filtering, we selected a threshold value for high-mobility periods as 0.08 m/s and a threshold for the low-mobility periods as 0.05 m/s. Sample filtering based on the average speed value was applied to both the training set and the testing set to preserve the similarity in the training and testing data distributions. This, in particular, means that the test set in all of the evaluation runs is reduced in terms of the amount of samples. Since the size of the training set is reduced by the filtering procedure, we found that a larger amount of epochs is required to reach convergence or, alternatively, data augmentation could be used to increase the size of the dataset. We found that both approaches work equally well in terms of final classification quality. To augment our time-series dataset, we use simple transformations such as adding Gaussian noise, time warp (random changes of the timeline) and drift (random drift of a set of consecutive values in the time-series).

The classification results obtained using the FCN model trained on samples corresponding to low-mobility and high-mobility periods in the recording are shown in Figure 5-13. Interestingly, the classification quality significantly drops when the

Model (training data)	Log-loss value
FCN (neural activity + speed)	0.56551
FCN (neural activity + speed; high mobility periods)	<b>0.54462</b>
FCN (neural activity + speed; low mobility periods)	8.02311
FCN-[32,32,32] (neural activity + speed)	0.62741
Transformer (lr=1e-3) (neural activity + speed)	0.58114
FCN (neural activity)	0.56623
FCN (speed data)	0.60688

Table 5.1 – Logistic loss score values computed over the animals in the data set for different models and training data preparation strategies evaluated.

network is trained on samples extracted from low-activity periods in the recording. On the other hand, the classification performance slightly increases overall when only high-mobility samples are used for training. Using predictions from the best performing model we obtained on high-mobility period samples (see the top panel of Figure 5-13) we only ended up with two misclassified animals (one from the WT and one from the mutant group). These results imply that the predictive information about whether the animal demonstrates pathological activity is mostly contained in signals recorded during high-mobility periods and thus isolating time-series samples corresponding solely to these periods could potentially improve predictive power of the models trained on these data.

## 5.8 Conclusions

In summary, we have demonstrated that one could detect patterns of pathological activity from the signals recorded in the motor cortices of C9orf72 mice. The discriminative features to be used for accurate classification of mutant animals have to be learned by a deep neural network, and the amount of predictive information contained in the cortical activity alone does not lead to the best classification results alone. We have shown that it is the interaction between the movement signal and the cortical activity that is highly predictive of pathology, with the amount of predictive information in the data being state-dependent, found to be highest during high-mobility periods of the animal. We hope that these observations could further help design paradigms of early diagnosis of the ALS in human patients.

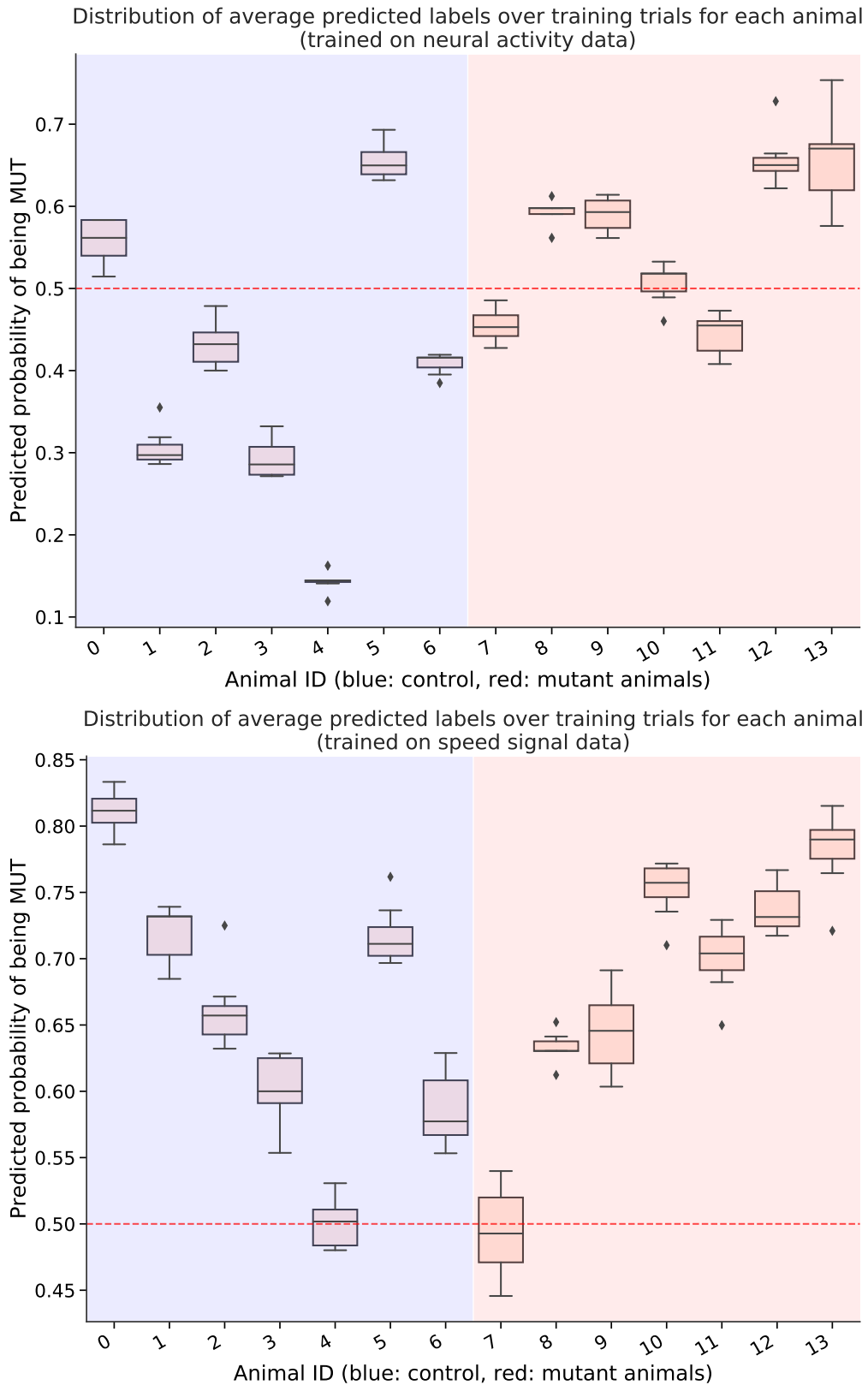
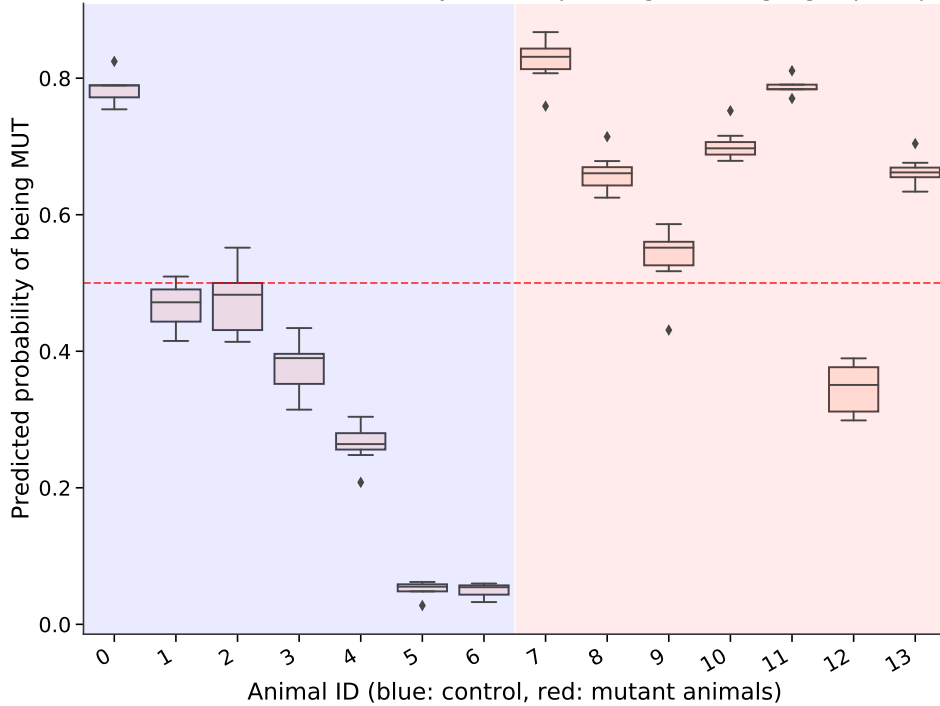


Figure 5-12 – Accuracy scores (average predicted labels) per animal over training trials of a base FCN model trained on (top panel) only neural activity time-series and (bottom panel) only speed signal time-series using the leave-one-out cross-validation scheme. Blue part of the chart (animal IDs from 0 to 6 correspond to WT animals, the rest to the mutant group).

Distribution of average predicted labels over training trials for each animal  
(FCN model trained on neural activity data + speed signal during high-speed periods)



Distribution of average predicted labels over training trials for each animal  
(FCN model trained on neural activity data + speed signal during low-speed periods)

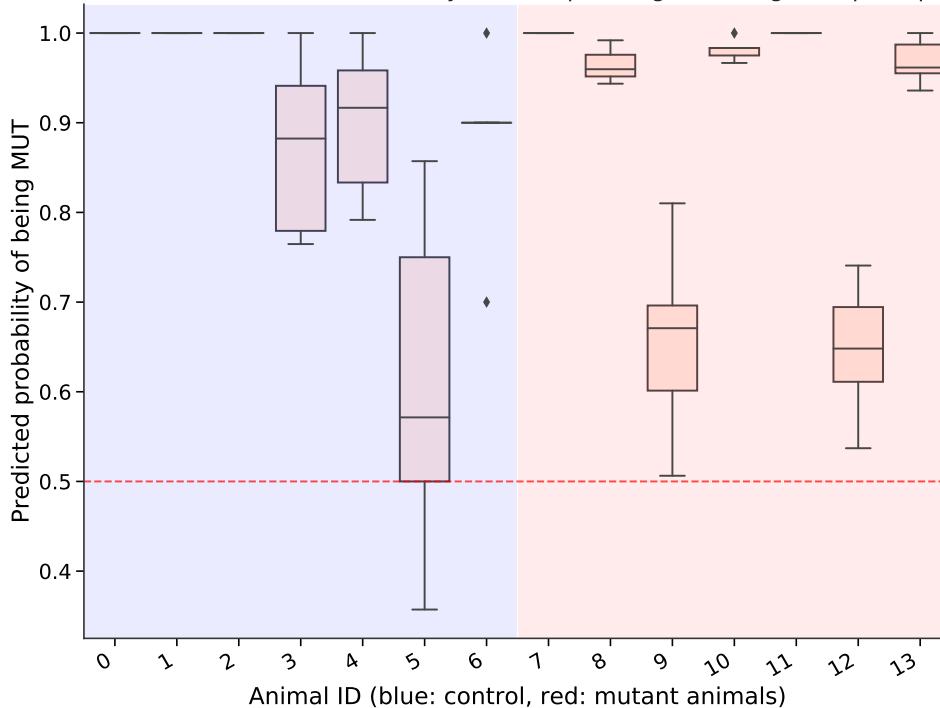


Figure 5-13 – Accuracy scores (average predicted labels) per animal over training trials of a base FCN model trained on time-series samples of both modalities corresponding to (top panel) high-mobility periods meaning that the average animal speed within the sample is higher than 0.08 m/s and (bottom panel) low-mobility periods meaning that the average animal speed within the sample is lower than 0.05 m/s. Evaluation is done using the leave-one-out cross-validation scheme.

# Chapter 6

## Discussion

Across the chapters of this thesis, we aimed to better understand how to apply machine learning and computational modelling to create and improve systems capable of detecting signs of pathological activity from neural circuit level data. We started by defining a common evaluation benchmark for single-neuron activity classification models in Chapter 2, keeping in mind that a similar kind of problem setup would further be extended to pathological activity detection problems. The evaluation benchmark for spike train classification is based on several open-access data sets, and we hope that this benchmark would facilitate the advancement of machine learning approaches (and, in particular, deep learning approaches) for spike train data and for neural decoding in general. We found that a strong baseline for spike train classification is an approach based on time-series feature extraction, which also has an advantage of being interpretable. We then went on to apply this approach to calcium imaging data of pathological neural activity in the PFC in Chapter 4 and have shown that it is indeed possible to detect dysfunctions of cholinergic signalling associated with common neurodegenerative disorders in the brain from single-neuron or neural ensemble activity data (like e. g. the local-field potential time-series).

In Chapter 5, we have shown that for the particular task of detecting pathologies associated with early-stage ALS, one has to resort to multivariate time-series analysis methods, and that deep learning is particularly well suited to capture the interaction between different data modalities to make predictions, rather than trying to extract predictive information from distinct data modalities separately. One



central finding in this study is that data that are the most predictive of the pathology correspond to the periods of high mobility of the animals. These findings could further help design paradigms of early behaviour-dependent diagnosis of the ALS in human patients.

# Bibliography

- Allen cell types dataset. URL <https://celltypes.brain-map.org/>.
- Larissa Albantakis and Gustavo Deco. Changes of mind in an attractor network of decision-making. *PLoS computational biology*, 7(6):e1002086, 2011.
- Robert M. Anthenelli, Neal L. Benowitz, Robert West, Lisa St Aubin, Thomas McRae, David Lawrence, John Ascher, Cristina Russ, Alok Krishen, and A. Eden Evins. Neuropsychiatric safety and efficacy of varenicline, bupropion, and nicotine patch in smokers with and without psychiatric disorders (eagles): a double-blind, randomised, placebo-controlled clinical trial. *The Lancet*, 387(10037):2507–2520, 2016.
- Maria Elena Avale, Philippe Faure, Stéphanie Pons, Patricia Robledo, Thierry Deltheil, Denis J. David, Alain M. Gardier, Rafael Maldonado, Sylvie Granon, and Jean-Pierre Changeux. and others interplay of  $\beta 2^*$  nicotinic receptors and dopamine pathways in the control of spontaneous locomotion. *Proceedings of the National Academy of Sciences*, 105(41):15991–15996, 2008.
- Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3):606–660, 2017.
- Omri Barak and Misha Tsodyks. Working models of working memory. *Current opinion in neurobiology*, 25:20–24, 2014.
- Deanna M. Barch, Cameron S. Carter, Todd S. Braver, Fred W. Sabb, Angus MacDonald, Douglas C. Noll, and Jonathan D. Cohen. Selective deficits in prefrontal cortex function in medication-naive patients with schizophrenia. *Archives of general psychiatry*, 58(3):280–288, 2001.
- Michael Beierlein, Jay R. Gibson, and Barry W. Connors. Two dynamically distinct inhibitory networks in layer 4 of the neocortex. *Journal of neurophysiology*, 90(5):2987–3000, 2003.
- Ari S Benjamin, Hugo L Fernandes, Tucker Tomlinson, Pavan Ramkumar, Chris VerSteeg, Raed H Chowdhury, Lee E Miller, and Konrad P Kording. Modern machine learning as a benchmark for fitting neural responses. *Frontiers in computational neuroscience*, 12, 2018.

- Christophe Bernard. On fallacies in neuroscience. *Eneuro*, 7:6, 2020.
- Morgane Besson, Benoît Forget, Caroline Correia, Rodolphe Blanco, and Uwe A. Maskos. human polymorphism in chrna5 is linked to relapse to nicotine seeking in transgenic rats. *Current Biology*, 28(20):3244–3253, 2018.
- Morgane Besson, Benoît Forget, Caroline Correia, Rodolphe Blanco, and Uwe Maskos. Profound alteration in reward processing due to a human polymorphism in chrna5: A role in alcohol dependence and feeding behavior. *Neuropsychopharmacology*, 44(11):1906–1916, 2019.
- Bernard Bloem, Rogier Bernard Poorthuis, and Huibert Daniel Mansvelder. Cholinergic modulation of the medial prefrontal cortex: the role of nicotinic receptors in attention and regulation of neuronal activity. *Frontiers in neural circuits*, 8:17, 2014.
- Cameron S. Carter, William Perlstein, Rohan Ganguli, Jaspreet Brar, Mark Mintun, and Jonathan D. Cohen. Functional hypofrontality and working memory dysfunction in schizophrenia. *American Journal of Psychiatry*, 155(9):1285–1287, 1998.
- Edward Challis, Peter Hurley, Laura Serra, Marco Bozzali, Seb Oliver, and Mara Cercignani. Gaussian process classification of alzheimer’s disease and mild cognitive impairment from resting-state fmri. *NeuroImage*, 112:232–243, 2015.
- Frances S. Chance and L. F. Abbott. Divisive inhibition in recurrent networks. *Network: Computation in Neural Systems*, 11(2):119–129, 2000.
- Paul Charlesworth, Ellese Cotterill, Andrew Morton, Seth GN Grant, and Stephen J Eglén. Quantitative differences in developmental profiles of spontaneous activity in cortical and hippocampal cultures. *Neural development*, 10(1):1, 2015.
- Hongyoon Choi, Seunggyun Ha, Hyung Jun Im, Sun Ha Paek, and Dong Soo Lee. Refining diagnosis of parkinson’s disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical*, 16:586–594, 2017.
- Shanta Chowdhury, Xishuang Dong, Oscar A Solis, Lijun Qian, and Xiangfang Li. Cell type identification from single-cell transcriptomic data via gene embedding. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 258–263. IEEE, 2020.
- Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package). *Neurocomputing*, 2018.
- Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- John A. Dani, Daoyun Ji, and Fu-Ming Zhou. Synaptic plasticity and nicotine addiction. *Neuron*, 31(3):349–352, 2001.

- Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.
- Mariely DeJesus-Hernandez, Ian R Mackenzie, Bradley F Boeve, Adam L Boxer, Matt Baker, Nicola J Rutherford, Alexandra M Nicholson, NiCole A Finch, Heather Flynn, Jennifer Adamson, et al. Expanded ggggcc hexanucleotide repeat in noncoding region of c9orf72 causes chromosome 9p-linked ftd and als. *Neuron*, 72(2):245–256, 2011.
- Gopikrishna Deshpande, Peng Wang, D Rangaprakash, and Bogdan Wilamowski. Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE transactions on cybernetics*, 45(12):2668–2679, 2015.
- Alain Destexhe. Self-sustained asynchronous irregular states and up–down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *Journal of computational neuroscience*, 27:3, 2009.
- Alain Destexhe and Terrence J. Sejnowski. The wilson–cowan model, 36 years later. *Biological cybernetics*, 101(1):1–2, 2009.
- Felix Droste and Benjamin Lindner. Up-down-like background spiking can enhance neural information transmission. *Eneuro*, 4:6, 2017.
- Daniel Durstewitz, Jeremy K. Seamans, and Terrence J. Sejnowski. Neurocomputational models of working memory. *Nature neuroscience*, 3(11):1184–1191, 2000.
- Saso Džeroski and Bernard Ženko. Is combining classifiers with stacking better than selecting the best one? *Machine learning*, 54(3):255–273, 2004.
- J-P Eckmann, S Oliffson Kamphorst, and David Ruelle. Recurrence plots of dynamical systems. *EPL (Europhysics Letters)*, 4(9):973, 1987.
- Johann Faouzi. pyts: a Python package for time series transformation and classification, May 2018. URL <https://doi.org/10.5281/zenodo.1244152>.
- Kevin Fauvel, Tao Lin, Véronique Masson, Élisabeth Fromont, and Alexandre Termier. Xcm: An explainable convolutional neural network for multivariate time series classification, 2020.
- Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.
- Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F Schmidt, Jonathan Weber, Geoffrey I Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020.

- Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- Johannes Friedrich, Pengcheng Zhou, and Liam Paninski. Fast online deconvolution of calcium imaging data. *PLoS computational biology*, 13(3):e1005423, 2017.
- Ben D Fulcher and Nick S Jones. hctsa: A computational framework for automated time-series phenotyping using massive feature extraction. *Cell systems*, 5(5):527–531, 2017.
- Ritu Gautam and Manik Sharma. Prevalence and diagnosis of neurological disorders using different deep learning techniques: a meta-analysis. *Journal of medical systems*, 44(2):1–24, 2020.
- Jay R. Gibson, Michael Beierlein, and Barry W. Connors. Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, 402(6757):75–79, 1999.
- Rashid Giniatullin, Andrea Nistri, and Jerrel L. Yakel. Desensitization of nicotinic ach receptors: shaping cholinergic signaling. *Trends in neurosciences*, 28(7):371–378, 2005.
- Joshua I Glaser, Ari S Benjamin, Raaed H Chowdhury, Matthew G Perich, Lee E Miller, and Konrad P Kording. Machine learning for neural decoding. *Eneuro*, 7(4), 2020.
- Anitha P. Govind, Paul Vezina, and William N. Green. Nicotine-induced upregulation of nicotinic receptors: underlying mechanisms and relevance to nicotine addiction. *Biochemical pharmacology*, 78(7):756–765, 2009.
- Josif Grabocka, Nicolas Schilling, Martin Wistuba, and Lars Schmidt-Thieme. Learning time-series shapelets. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 392–401, 2014.
- Sharon R. Grady, Charles R. Wageman, Natalie E. Patzlaff, and Michael J. Marks. Low concentrations of nicotine differentially desensitize nicotinic acetylcholine receptors that include  $\alpha 5$  or  $\alpha 6$  subunits and that mediate synaptosomal neurotransmitter release. *Neuropharmacology*, 62(5-6):1935–1943, 2012.
- Rachel Grashow, Ted Brookings, and Eve Marder. Compensation for variable intrinsic neuronal excitability by circuit-synaptic interactions. *Journal of Neuroscience*, 30(27):9145–9156, 2010.
- Michael Graupner and Boris Gutkin. Modeling nicotinic neuromodulation from global functional and network levels to nachr based mechanisms. *Acta Pharmacologica Sinica*, 30(6):681–693, 2009.
- Michael Graupner, Reinoud Maex, and Boris Gutkin. Endogenous cholinergic inputs and local circuit mechanisms govern the phasic mesolimbic dopamine response to nicotine. *PLoS Comput Biol*, 9:8, 2013.

- Francois Gros-Louis, Claudia Gaspar, and Guy A Rouleau. Genetics of familial and sporadic amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1762(11-12):956–972, 2006.
- Karine Guillem, Bernard Bloem, Rogier B. Poorthuis, Maarten Loos, August B. Smit, Uwe Maskos, Sabine Spijker, and Huibert D. Mansvelder. Nicotinic acetylcholine receptor  $\beta 2$  subunits in the medial prefrontal cortex control attention. *Science*, 333(6044):888–891, 2011.
- Itai Hayut, Erika E. Fanselow, Barry W. Connors, and David Golomb. Lts and fs inhibitory interneurons, short-term synaptic plasticity, and cortical circuit dynamics. *PLoS Comput Biol*, 7:10, 2011.
- Loreen Hertäg and Henning Sprekeler. Amplifying the redistribution of somatodendritic inhibition by the interplay of three interneuron types. *PLoS computational biology*, 15(5):e1006999, 2019.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Carl Holmgren, Tibor Harkany, Björn Svennenfors, and Yuri Zilberter. Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of physiology*, 551(1):139–153, 2003.
- L. Elliot Hong, Colin A. Hodgkinson, Yihong Yang, Hemalatha Sampath, Thomas J. Ross, Brittany Buchholz, Betty Jo Salmeron, Vibhuti Srivastava, Gunvant K. Thaker, David Goldman, et al. A genetically modulated, intrinsic cingulate circuit supports human nicotine addiction. *Proceedings of the National Academy of Sciences*, 107(30):13509–13514, 2010.
- Garrett Honke, Irina Higgins, Nina Thigpen, Vladimir Miskovic, Katie Link, Sunny Duan, Pramod Gupta, Julia Klawohn, and Greg Hajcak. Representation learning for improved interpretability and classification accuracy of clinical factors from eeg. *arXiv preprint arXiv:2010.15274*, 2020.
- Hang Hu, Yunyong Ma, and Ariel Agmon. Submillisecond firing synchrony between different subtypes of cortical interneurons connected chemically but not electrically. *Journal of Neuroscience*, 31(9):3351–3361, 2011.
- Rainbo Hultman, Stephen D Mague, Qiang Li, Brittany M Katz, Nadine Michel, Lizhen Lin, Joyce Wang, Lisa K David, Cameron Blount, Rithi Chandy, et al. Dysregulation of prefrontal cortex-mediated slow-evolving limbic dynamics drives stress-induced emotional pathology. *Neuron*, 91(2):439–452, 2016.
- Mark D Humphries. Spike-train communities: finding groups of similar spike trains. *Journal of Neuroscience*, 31(6):2321–2336, 2011.
- Mark D Humphries, Jose Angel Obeso, and Jakob Kisbye Dreyer. Insights into parkinson’s disease from computational models of the basal ganglia. *Journal of Neurology, Neurosurgery & Psychiatry*, 89(11):1181–1188, 2018.

- Monika Jadi, Alon Polsky, Jackie Schiller, and Bartlett W. Mel. Location-dependent effects of inhibition on local spiking in pyramidal neuron dendrites. *PLoS Comput Biol*, 8:6, 2012.
- Xiaoxuan Jia, Josh Siegle, Corbett Bennett, Sam Gale, Daniel Denman, Christof Koch, and Shawn Olsen. High-density extracellular probes reveal dendritic back-propagation and facilitate neuron classification. *bioRxiv*, page 376863, 2018.
- Longlong Jing and Yingli Tian. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- Jonathan Jouty, Gerrit Hilgen, Evelyne Sernagor, and Matthias Hennig. Non-parametric physiological classification of retinal ganglion cells. *bioRxiv*, page 407635, 2018.
- Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. Lstm fully convolutional networks for time series classification. *IEEE access*, 6:1662–1669, 2017.
- Robert E Kass and Valérie Ventura. A spike-train probability model. *Neural computation*, 13(8):1713–1720, 2001.
- Bernard Katz and S. Thesleff. A study of the desensitization produced by acetylcholine at the motor end-plate. *The Journal of physiology*, 138:1, 1957.
- Junghoe Kim, Vince D Calhoun, Eunsoo Shim, and Jong-Hwan Lee. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage*, 124:127–146, 2016.
- Fani Koukouli and Jean-Pierre Changeux. Do nicotinic receptors modulate high-order cognitive processing? *Trends in Neurosciences*, 43:550–564, 2020.
- Fani Koukouli, Marie Rooy, Jean-Pierre Changeux, and Uwe Maskos. Nicotinic receptors in mouse prefrontal cortex modulate ultraslow fluctuations related to conscious processing. *Proceedings of the National Academy of Sciences*, 113(51):14823–14828, 2016a.
- Fani Koukouli, Marie Rooy, and Uwe Maskos. Early and progressive deficit of neuronal activity patterns in a model of local amyloid pathology in mouse prefrontal cortex. *Aging (Albany NY)*, 8:12, 2016b.
- Fani Koukouli, Marie Rooy, Dimitrios Tziotis, Kurt A Sailor, Heidi C O’Neill, Josien Levenga, Mirko Witte, Michael Nilges, Jean-Pierre Changeux, Charles A Hoefler, et al. Nicotine reverses hypofrontality in animal models of addiction and schizophrenia. *Nature Medicine*, 23(3):347–354, 2017.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

- Junghee Lee and Sohee Park. Working memory impairments in schizophrenia: a meta-analysis. *Journal of abnormal psychology*, 114:4, 2005.
- Meng Li, Fang Zhao, Jason Lee, Dong Wang, Hui Kuang, and Joe Z Tsien. Computational classification approach to profile neuron subtypes from brain activity mapping data. *Scientific reports*, 5:12474, 2015.
- Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.
- Ashok Litwin-Kumar, Robert Rosenbaum, and Brent Doiron. Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *Journal of neurophysiology*, 115(3):1399–1409, 2016.
- Yuanjing Liu, Amrutha Pattamatta, Tao Zu, Tammy Reid, Olgert Bardhi, David R Borchelt, Anthony T Yachnis, and Laura PW Ranum. C9orf72 bac mouse model with motor deficits and neurodegenerative features of als/ftd. *Neuron*, 90(3):521–534, 2016.
- Jesse A Livezey and Joshua I Glaser. Deep learning approaches for neural decoding across architectures and recording modalities. *Briefings in bioinformatics*, 22(2):1577–1591, 2021.
- Jesse A Livezey, Kristofer E Bouchard, and Edward F Chang. Deep learning as a tool for neural data analysis: speech classification and cross-frequency coupling in human sensorimotor cortex. *PLoS computational biology*, 15(9):e1007091, 2019.
- Gašper; Prentice Jason S.; Ioffe Mark L.; Berry II Michael J.; Marre Olivier Loback, Adrianna R.; Tkačik. Multi-electrode retinal ganglion cell population spiking data. *Dryad. Dataset.*, 2016.
- Markus Löning, Anthony Bagnall, Sajaysurya Ganesh, Viktor Kazakov, Jason Lines, and Franz J Király. sktime: A unified interface for machine learning with time series. *arXiv preprint arXiv:1909.07872*, 2019.
- Carl H Lubba, Sarab S Sethi, Philip Knaute, Simon R Schultz, Ben D Fulcher, and Nick S Jones. catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, 33(6):1821–1852, 2019.
- Yunyong Ma, Hang Hu, and Ariel Agmon. Short-term plasticity of unitary inhibitory-to-inhibitory synapses depends on the presynaptic interneuron subtype. *Journal of Neuroscience*, 32(3):983–988, 2012.
- Danyan Mao, David C. Perry, Robert P. Yasuda, Barry B. Wolfe, and Kenneth J. Kellar. The  $\alpha 4\beta 2\alpha 5$  nicotinic cholinergic receptor in rat brain is resistant to up-regulation by nicotine in vivo. *Journal of neurochemistry*, 104(2):446–456, 2008.
- Uwe Maskos. *The nicotinic receptor alpha5 coding polymorphism rs16969968 as a major target in disease: Functional dissection and remaining challenges*. Journal of Neurochemistry, 2020.



- Uwe Maskos, B. E. Molles, S. Pons, M. Besson, B. P. Guiard, J. P. Guilloux, A. Evrard, P. Cazala, A. Cormier, and M. Mameli-Engvall. and others nicotine reinforcement and cognition restored by targeted expression of nicotinic receptors. *Nature*, 436(7047):103–107, 2005.
- Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861, 2018.
- Daniel A Mordes, Brett M Morrison, Xanthe H Ament, Christopher Cantrell, Joanie Mok, Pierce Eggan, Carolyn Xue, Jin-Yuan Wang, Kevin Eggan, and Jeffrey D Rothstein. Absence of survival and motor deficits in 500 repeat c9orf72 bac mice. *Neuron*, 108(4):775–783, 2020.
- C. Morel, L. Fattore, Stéphanie Pons, Y. A. Hay, F. Marti, B. Lambolez, M. De Biasi, M. Lathrop, Walter Fratta, and Uwe Maskos. and others nicotine consumption is regulated by a human polymorphism in dopamine neurons. *Molecular psychiatry*, 19(8):930–936, 2014.
- Mario Mulansky and Thomas Kreuz. Pyspike—a python library for analyzing spike train synchrony. *SoftwareX*, 5:183–189, 2016.
- Ignacio Oguiza. tsai - a state-of-the-art deep learning library for time series and sequential data. Github, 2020. URL <https://github.com/timeseriesAI/tsai>.
- Timothy O’Leary, Alex H. Williams, Alessio Franci, and Eve Cell types Marder. network homeostasis, and pathological compensation from a biologically plausible ion channel expression model. *Neuron*, 82(4):809–821, 2014.
- Marius Pachitariu, Carsen Stringer, Sylvia Schröder, Mario Dipoppa, L Federico Rossi, Matteo Carandini, and Kenneth D Harris. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *Biorxiv*, page 061507, 2016.
- Christoforos A. Papasavvas, Yujiang Wang, Andrew J. Trevelyan, and Marcus Kaiser. Gain control through divisive inhibition prevents abrupt transition to chaos in a neural mass model. *Physical Review E*, 92:3, 2015.
- Carsten K. Pfeffer, Mingshan Xue, Miao He, Z. Josh Huang, and Massimo Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature neuroscience*, 16(8):1068–1076, 2013.
- Ryan T Philips, Salvatore J Torrisi, Adam X Gorka, Christian Grillon, and Monique Ernst. Dynamic time warping identifies functionally distinct fmri resting state cortical networks specific to vta and snc: A proof of concept. *Cerebral Cortex*, 2021.
- Hyun-Jae Pi, Balázs Hangya, Duda Kvitsiani, Joshua I. Sanders, Z. Josh Huang, and Adam Kepecs. Cortical interneurons that specialize in disinhibitory control. *Nature*, 503(7477):521–524, 2013.

- Rogier B. Poorthuis, Bernard Bloem, Benita Schak, and Jordi Wester. Christiaan pj de kock, and huibert d mansvelder. *Layer-specific modulation of the prefrontal cortex by nicotinic acetylcholine receptors*, 23(1):148–161, 2013.
- James T. Porter, Bruno Cauli, Jochen F. Staiger, Bertrand Lambollez, Jean Rossier, and Etienne Audinat. Properties of bipolar vipergic interneurons and their excitation by pyramidal neurons in the rat neocortex. *European Journal of Neuroscience*, 10(12):3617–3628, 1998.
- James T. Porter, Bruno Cauli, Keisuke Tsuzuki, Bertrand Lambollez, Jean Rossier, and Etienne Audinat. Selective excitation of subtypes of neocortical interneurons by nicotinic receptors. *Journal of Neuroscience*, 19(13):5228–5235, 1999.
- Jason S Prentice, Olivier Marre, Mark L Ioffe, Adrianna R Loback, Gašper Tkačik, and Michael J Berry. Error-robust modes of the retinal population code. *PLoS computational biology*, 12(11):e1005148, 2016.
- Arjun Punjabi, Adam Martersteck, Yanran Wang, Todd B Parrish, Aggelos K Katsaggelos, and Alzheimer’s Disease Neuroimaging Initiative. Neuroimaging modality fusion in alzheimer’s classification using convolutional neural networks. *PloS one*, 14(12):e0225759, 2019.
- MCW van Rossum. A novel spike distance. *Neural computation*, 13(4):751–763, 2001.
- Lewis P Rowland and Neil A Shneider. Amyotrophic lateral sclerosis. *New England Journal of Medicine*, 344(22):1688–1700, 2001.
- Bernardo Rudy, Gordon Fishell, SooHyun Lee, and Jens Hjerling-Leffler. Three groups of interneurons account for nearly 100% of neocortical gabaergic neurons. *Developmental neurobiology*, 71(1):45–61, 2011.
- Patrick Schäfer. The boss is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 29(6):1505–1530, 2015.
- Patrick Schäfer. Scalable time series classification. *Data Mining and Knowledge Discovery*, 30(5):1273–1298, 2016.
- Patrick Schäfer and Mikael Höggqvist. Sfa: a symbolic fourier approximation and index for similarity search in high dimensional datasets. In *Proceedings of the 15th International Conference on Extending Database Technology*, pages 516–527. ACM, 2012.
- Pavel Senin and Sergey Malinchik. Sax-vsm: Interpretable time series classification using sax and vector space model. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 1175–1180. IEEE, 2013.
- Gilad Silberberg and Henry Markram. Disynaptic inhibition between neocortical pyramidal cells mediated by martinotti cells. *Neuron*, 53(5):735–746, 2007.

- Melissa L. Sinkus, Sharon Graw, Robert Freedman, Randal G. Ross, Henry A. Lester, and Sherry Leonard. The human *chrna7* and *chrfa7a* genes: A review of the genetics, regulation, and function. *Neuropharmacology*, 96:274–288, 2015.
- Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.
- Nicholas A Steinmetz, Christof Koch, Kenneth D Harris, and Matteo Carandini. Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology*, 50:92–100, 2018.
- Carsen Stringer and Marius Pachitariu. Computational processing of neural recordings from calcium imaging data. *Current opinion in neurobiology*, 55:22–31, 2019.
- Ardi Tampuu, Tambet Matiisen, H Freyja Ólafsdóttir, Caswell Barry, and Raul Vicente. Efficient neural decoding of self-location with a deep recurrent network. *PLoS computational biology*, 15(2):e1006822, 2019.
- Wensi Tang, Guodong Long, Lu Liu, Tianyi Zhou, Jing Jiang, and Michael Blumenstein. Rethinking 1d-cnn for time series classification: A stronger baseline. *arXiv preprint arXiv:2002.10061*, 2020.
- Jeff L Teeters and Friedrich T Sommer. Crcns. org: a repository of high-quality data sets and tools for computational neuroscience. *BMC Neuroscience*, 10(S1):S6, 2009.
- Taro Tezuka. Spike train pattern discovery using interval structure alignment. In *International Conference on Neural Information Processing*, pages 241–249. Springer, 2015.
- Taro Tezuka. Multineuron spike train analysis with r-convolution linear combination kernel. *Neural Networks*, 102:67–77, 2018.
- Jennifer W. Tidey, Damaris J. Rohsenow, Gary B. Kaplan, Robert M. Swift, and Amy B. Adolfo. Effects of smoking abstinence, smoking cues and nicotine replacement in smokers with schizophrenia and controls. *Nicotine & Tobacco Research*, 10(6):1047–1056, 2008.
- Robin Tremblay, Soohyun Lee, and Bernardo Rudy. Gabaergic interneurons in the neocortex: from cellular properties to circuits. *Neuron*, 91(2):260–292, 2016.
- Eirini Troullinou, Grigorios Tsagkatakis, Spyridon Chavlis, Gergely F Turi, Wenke Li, Attila Losonczy, Panagiotis Tsakalides, and Panayiota Poirazi. Artificial neural networks in action for an automated cell-type classification of biological neural networks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2020.
- David Tsai, Esha John, Tarun Chari, Rafael Yuste, and Kenneth Shepard. High-channel-count, high-density microelectrode array for closed-loop investigation of neuronal networks. In *Engineering in Medicine and Biology Society (EMBC)*,

- 2015 37th Annual International Conference of the IEEE, pages 7510–7513. IEEE, 2015.
- Jonathan D Victor and Keith P Purpura. Metric-space analysis of spike trains: theory, algorithms and application. *Network: computation in neural systems*, 8(2):127–164, 1997.
- Vladyslav V. Vyazovskiy and Kenneth D. Harris. Sleep and the single neuron: the role of global slow oscillations in individual cell rest. *Nature Reviews Neuroscience*, 14(6):443–451, 2013.
- X-J Wang. Toward a prefrontal microcircuit model for cognitive deficits in schizophrenia. *Pharmacopsychiatry*, 39(S 1):80–87, 2006.
- Zhiguang Wang and Tim Oates. Imaging time-series to improve classification and imputation. *arXiv preprint arXiv:1506.00327*, 2015.
- Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017.
- BO Watson, D Levenstein, JP Greene, JN Gelinas, and G. Buzsaki. Multi-unit spiking activity recorded from rat frontal cortex (brain regions mpfc, ofc, acc, and m2) during wake-sleep episode wherein at least 7 minutes of wake are followed by 20 minutes of sleep. *crns.org*. 2016a. doi:<http://dx.doi.org/10.6080/K02N506Q>.
- Brendon O Watson, Daniel Levenstein, J Palmer Greene, Jennifer N Gelinas, and György Buzsáki. Network homeostasis and state dynamics of neocortical sleep. *Neuron*, 90(4):839–852, 2016b.
- Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, Olivier Colliot, et al. Convolutional neural networks for classification of alzheimer’s disease: Overview and reproducible evaluation. *Medical image analysis*, 63:101694, 2020.
- Nathan R. Wilson, Caroline A. Runyan, Forea L. Wang, and Mriganka Sur. Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, 488(7411):343–348, 2012.
- Steffen B. E. and Wolff. and ölveczky, bence p the promise and perils of causal circuit manipulations. *Current opinion in neurobiology*, 49:84–94, 2018.
- Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.
- Zishen Xu, Wei Wu, Shawn S Winter, Max L Mehlman, William N Butler, Christine M Simmons, Ryan E Harvey, Laura E Berkowitz, Yang Chen, Jeffrey S Taube, et al. A comparison of neural decoding methods and population coding

across thalamo-cortical head direction cells. *Frontiers in neural circuits*, 13:75, 2019.

Lexiang Ye and Eamonn Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 947–956, 2009.

George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. *arXiv preprint arXiv:2010.02803*, 2020.

Xi Zhang, Lifang He, Kun Chen, Yuan Luo, Jiayu Zhou, and Fei Wang. Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson’s disease. In *AMIA Annual Symposium Proceedings*, volume 2018, page 1147. American Medical Informatics Association, 2018.

Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, 2017.