

BIO-482 Neuroscience: Cellular Circuit Mechanisms.

Cortical Neuron Classification: Leveraging Interpretable Machine Learning Techniques.

Wesley Monteith-Finas, SCIPER: 324745

In collaboration with Marianne Scoglio & Sander Miesen

December 23, 2023

Abstract—This study explores the use of interpretable Machine Learning tools in order to try and identify what are the most distinct characteristics of known cortical neuron types. As such we analyzed the contribution and importance of different features, on supervised learning classification algorithms. In addition, dimensionality reduction and unsupervised learning techniques were employed to enrich our analysis. Section II is dedicated to answering the demanded questions.

I. INTRODUCTION

Neural dynamics and cellular behavior within the brain presents an ongoing challenge for understanding its complexities. In this pursuit, the present study delves into the classification of distinct neuronal cell types based on a set of carefully selected features derived from electrophysiological data. Focusing on “free whisking” sweeps, this study employs a range of machine learning techniques, both supervised and unsupervised, to unravel patterns and relationships within the dataset. The chosen predictors, spanning from firing rates to spectral domain information, aim to capture nuanced aspects of neuronal activity. Through rigorous model testing, hyperparameter optimization, and dimensionality reduction, the study seeks not only to classify cell accurately types but also to uncover the underlying features pivotal in this classification.

II. ANSWERS TO QUESTIONS

All relevant figures to answer questions, can be visualized in Annex.

1) *Impact of AP Threshold, Mean V_m, and SD of V_m on the Mean Firing Rate*: The mean firing rate of neurons is intricately linked to properties such as the action potential (AP) threshold, mean membrane potential (V_m), and standard deviation (SD) of V_m . The firing rate exhibits a significant correlation with the mean V_m (Fig. 9 bottom left), and to a lesser extent, with the standard deviation of V_m (Fig. 9 bottom right). A higher mean V_m generally corresponds to a lower firing rate, which is consistent with established principles. Intuitively, the lower the AP threshold, the higher the firing rate, however, this relationship does not hold uniformly across all cell classes as no clear correlation was observed with the AP threshold (Fig. 9 top right).

2) *Specificities of each cortical neuron class allowing to distinguish excitatory vs inhibitory neurons (Fig. 10)*:

- Excitatory Neurons: Excitatory neurons exhibit overall low action potential firing rates and hyperpolarized membrane potential.

- PV-expressing GABAergic Neurons: PV neurons fire at higher rates than other classes and have short AP durations.
- VIP-expressing GABAergic Neurons: VIP neurons, have longer AP durations than other inhibitory neurons.
- SST-expressing GABAergic Neurons: SST neurons, relatively low FR compared to other inhibitory neurons, a much smaller FFT amplitude and a much smaller SD in their membrane potential.

3) *Whisking versus Active-Contact Onset Time*:

Whisking Onset Time:

- Excitatory Neurons: Excitatory neurons depolarize at whisking onset, with a brief transient hyperpolarization shortly after. Superficial neurons exhibit more depolarization compared to deeper neurons.
- PV-expressing GABAergic Neurons: PV neurons depolarize at whisking onset, with a more prominent increase in firing rate in deeper neurons.
- VIP-expressing GABAergic Neurons: VIP neurons depolarize at whisking onset, with late excitation during whisking. They are more depolarized compared to quiet periods, influenced by nicotinic input and glutamatergic input from other cortical regions.
- SST-expressing GABAergic Neurons: On average, SST neurons hyperpolarize at whisking onset. Superficial SST neurons may show a prominent hyperpolarization.

Active-Contact Onset Time:

- Excitatory Neurons: Excitatory neurons show robust depolarization with a small increase in firing rate during active touch. There's a frequency-dependent suppression of touch-evoked responses at short intercontact intervals.
- PV-expressing GABAergic Neurons: PV neurons exhibit large and fast depolarization during active touch. Touch-evoked responses are suppressed at short intercontact intervals, but with a reliable peak V_m .
- VIP-expressing GABAergic Neurons: VIP neurons depolarize and increase firing rate during active touch. They show late excitation during active touch, consistent with a delayed response.
- SST-expressing GABAergic Neurons: SST neurons display relatively weak active touch responses, with a unique response to intercontact intervals, showing a small hyperpolarization at long intervals and excitatory PSP response at short intervals.

Now we transition to the actual study.

III. METHODS

A. Data Preprocessing

1) *Predictors*: We decided to focus only on “free whisking” sweeps, which amounted to 497 different samples. From these sweeps we decided to use as predictors the calculated firing rate (*firing_rate*), action potential (AP) threshold and duration (*ap_threshold* and *ap_duration*), the mean and the standard deviation of the membrane potential (*mean_vm* and *std_vm*), and spectral domain information in the form of the lowest and the highest Fourier amplitudes (*fft_low* and *fft_high*). Whenever a sweep did not have an AP, we replaced both AP duration and threshold by a value of 0. In hindsight this is biologically nonsensical. These seven features created a data matrix $X_1 \in \mathbb{R}^{497 \times 7}$ to be paired with the response variable $y \in \mathbb{R}^{497 \times 1}$ corresponding to each cell class.

2) *Additional Extracted features*: In addition to the 7 features mentioned above, the 13 features described in Table-I were extracted and added from each sweep, widening our data size $X_2 \in \mathbb{R}^{497 \times 20}$. One can observe the correlation between each variable present in X_2 , in Annex (Fig. 17).

3) *Standardization*: For supervised learning, the $X_2 \in \mathbb{R}^{497 \times 20}$ data was used. When splitting, training data was standardized, and the same transformation was applied to the test data. For unsupervised learning the $X_1 \in \mathbb{R}^{497 \times 7}$ data was used. It was standardized in its entirety prior to dimensionality reduction.

B. Supervised Learning

1) *Models*: In this study, a range of established machine learning algorithms available in the *Sci-kit Learn* library were tested. These are Logistic Regression, Random Forest, Gradient Boosting, AdaBoost, Decision Tree, Support Vector Machines, and a Gaussian Naive Bayes. Despite their effectiveness, deep learning models were not implemented as they would be hard to interpret.

2) *Evaluation Metric*: In order to get an estimate of the performance of each model on unseen data, we opted for a stratified 5-fold Cross Validation (CV) -which retains the original distribution when splitting the data- paired with the multi-class F1 score as an evaluation metric. It takes into account multi-class imbalance which was prevalent in our data (much more EXC or SST cells than VIP cells).

3) *Hyperparameter Optimization (HPO)*: Hyperparameters (HP) of a model were tuned using a Bayesian optimization algorithm available in the *Scikit-Optimize* library, in order to maximise the stratified 5-fold CV F1 score.

4) *Feature Importance*: The (HPO) models that achieved the best test score were trained on the entire data, and each individual feature importance was assessed using a metric specific to the type of classifier used. For tree-based ensemble methods (Gradient Boosting or Random Forest), the default feature importance attribute was used. It is computed as the mean and standard deviation of accumulation of the impurity decrease within each tree. For linear models utilizing weights (Linear SVC and Logistic Regression), the mean absolute weights across each class was used.

5) *Ensembling*: After having found the top models, we wanted to see whether combining their predictions would yield a higher result. As such, a final “ensemble” voting classifier with a majority voting rule was created and evaluated.

Variable	Description
<i>ap_amp_mean</i>	Average action potential amplitude for each sweep; any missing values or NaNs were replaced with 0.
<i>ap_amp_cv</i>	Coefficient of variation (= std / mean) of the action potential amplitudes for each sweep; any missing values or NaNs were replaced with 0.
<i>mean_ap_upstroke</i>	Average maximal rate of increase of the membrane potential during the action potential upstroke; any missing values or NaNs were replaced with 0.
<i>ap_upstroke_cv</i>	Coefficient of variation of the maximal rate of increase of the membrane potential during the action potential upstroke; any missing values or NaNs were replaced with 0.
<i>mean_ap_downstroke</i>	Average maximal rate of decrease of the membrane potential during the action potential downstroke; any missing values or NaNs were replaced with 0.
<i>ap_downstroke_cv</i>	Coefficient of variation of the maximal rate of decrease of the membrane potential during the action potential downstroke; any missing values or NaNs were replaced with 0.
<i>isi_mean</i>	Average duration between consecutive action potentials (inter-spike intervals); any missing values or NaNs were replaced with 50.
<i>isi_cv</i>	Coefficient of variation of the duration between consecutive action potentials; any missing values or NaNs were replaced with 0.
<i>irregularity_index</i>	Index indicating how regular the inter-spike intervals are, i.e., the mean of the difference between the durations of consecutive ISIs; any missing values or NaNs were replaced with 0.
<i>adaptation_index</i>	Another index indicating how regular the inter-spike intervals are; equal to the sum of the durations of consecutive ISIs divided by their difference. It is zero for a constant firing rate, negative for an increasing firing rate, and positive for a decreasing firing rate; any missing values or NaNs were replaced with 0, and very large values were capped to 10^{10} .
<i>nb_bursts</i>	Number of bursts per sweep. Two consecutive action potentials are considered part of the same burst if the duration between them is less than 10ms; any missing values or NaNs were replaced with 0.
<i>mean_burst_dur</i>	Mean duration of bursts; any missing values or NaNs were replaced with 0.
<i>burst_dur_cv</i>	Coefficient of variation of the duration of bursts; any missing values or NaNs were replaced with 0.

TABLE I: Description of Added Variables

C. Unsupervised Learning

1) *Dimensionality Reduction*: Dimensionality reduction techniques were employed in order to visualize the data in lower dimensional sub-spaces (2D and 3D). Hence, we would be able to visualize whether there is a spatial arrangement between each cell class. The three employed techniques were Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP).

2) *K-Means Clustering*: K-Means clustering algorithm was used in the lower dimensional subspace projections of our data. We fixed K to be equal to the number of cell classes present ($K = 4$), in order to determine whether each cluster could represent a specific cell class.

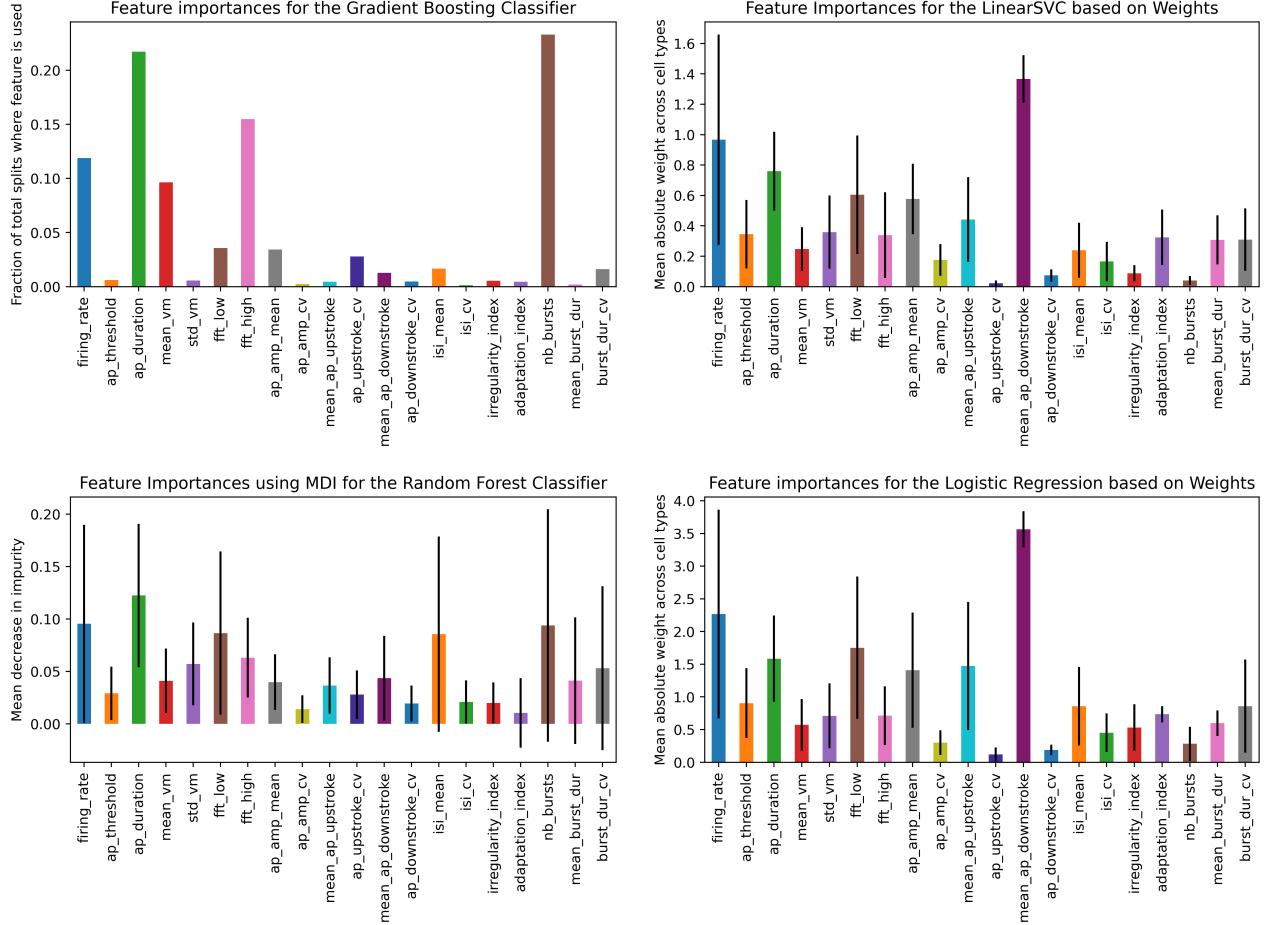


Fig. 1: Feature importance for the top 4 best classifiers. Left column are tree-based ensembles and right column are linear models

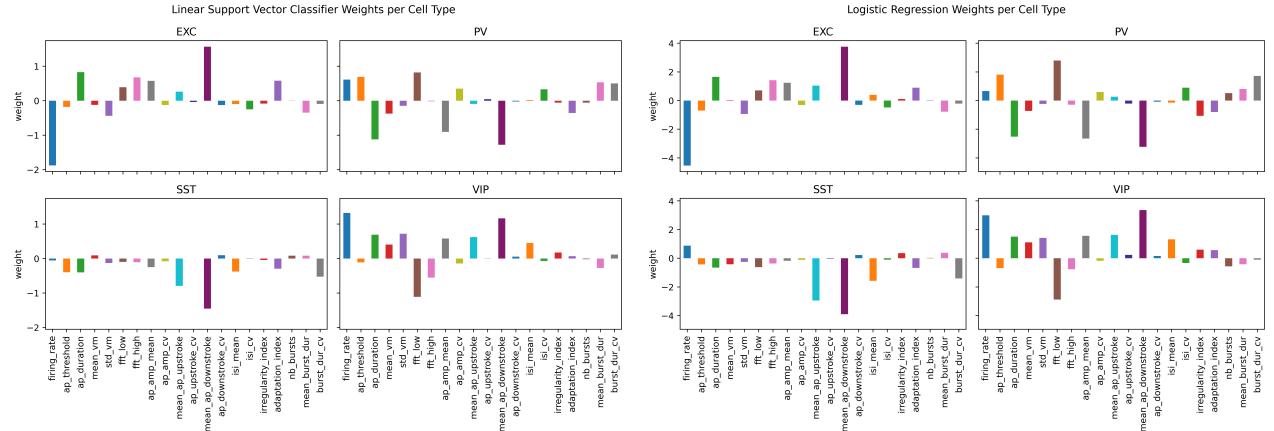


Fig. 2: Linear model weights across cell classes. Top four plots, are the weights of the Linear SVC (2nd best) and bottom four plots are the weights of the Logistic Regression (4th best).

IV. RESULTS

A. Best Models

From Fig. 3 we can see that the four best unoptimized models, based on test performance, are: 1st GradientBoosting with a score of 0.877, 2nd Logistic Regression, 3rd Random Forest and 4th Linear SVC. After having done HPO (Table-II), these models improve by about 1 to 2 %, every model is now above 0.8756, and Linear SVC and Logistic Regression swap places. Furthermore, by ensembling these four best models, it is possible to increase

test performance by an additional 1.77 % (from 0.8888 ± 0.0366 to 0.9065 ± 0.0206). It is interesting to notice that tree-based methods (Gradient Boosting and Random Forest) overfit to a greater extent, and have a higher standard deviation than linear models.

In Fig. 4 we can see that VIP cells cannot be classified as well as the others. This is consistent throughout each model, with the VIP F_1 score being 0.841, 0.774, 0.808 and 0.788, respectively from best to worst ranked classifier. On the other hand PV cells are by far the best classified cell class, followed by EXC and then

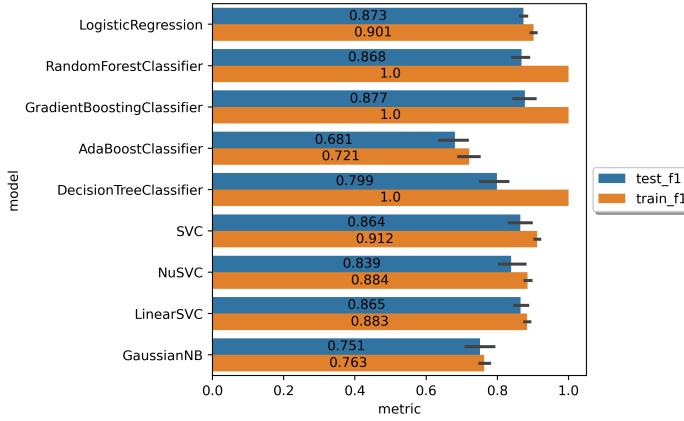


Fig. 3: F1 score performances, on a stratified 5-Fold CV, for a range of models tested.

Model	Inference	Mean	Std
Gradient Boosting	Test	0.8888	0.0366
	Train	1.0	0.0
Linear Support Vector Machine	Test	0.8832	0.0090
	Train	0.9094	0.0125
Random Forest	Test	0.8789	0.0334
	Train	1.0	0.0
Logistic Regression	Test	0.8756	0.0188
	Train	0.9166	0.0136
Ensemble	Test	0.9065	0.0206
	Train	0.9631	0.0072

TABLE II: Stratified 5-fold CV F1 score performance of top 4 best (HPO) models, plus the ensemble combination of them.

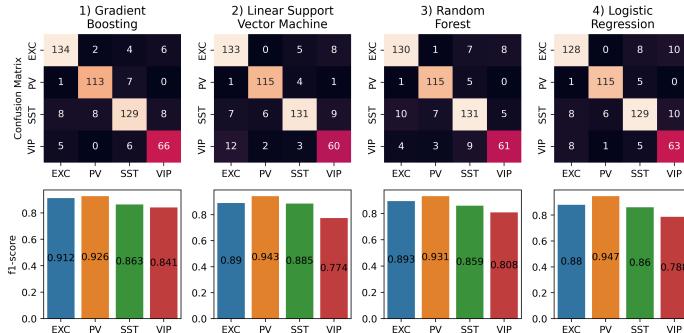


Fig. 4: Top 4 best model performances. Each column correspond to a model. Top row are confusion matrices based on a 5-Fold Stratified CV prediction. For each one of them the x-axis corresponds to predictions and the y-axis to the ground truth. Bottom row are the F1 scores for each cell type, calculated with the same prediction.

by SST cells. When we look at the miss-classified VIP cells 3 out of 4 times, none are classified as PV. Only one EXC cell is miss-classified as a PV cell.

B. Feature Importance

1) *Gradient Boosting*: For the Gradient Boosting Classifier the 5 most significant features are: the number of bursts, the AP duration, the highest amplitude of the Fourier transform, the firing rate and the mean of the membrane potential (Fig. 1).

2) *Linear SVC*: For the Linear SVC model, the 5 most significant features are: the mean AP downstroke, the firing rate,

AP duration, the lowest amplitude of the Fourier transform, and the AP amplitude mean (Fig. 1). More specifically if we look at the weight of each feature to determine the cell type (Fig. 2), there is a positive correlation with AP downstroke for EXC and VIP cells, versus negative for PV and SST cells. The firing rate is highly negatively correlated with EXC, positively for PV and VIP, and close to no correlation for SST cells. AP duration is positively correlated with EXC and VIP whereas negative for PV and SST.

3) *Random Forest*: For the Random Forest Classifier the 5 most significant features are: the AP duration, the number of bursts, the firing rate, the mean inter-spike intervals and the highest amplitude of the Fourier transform, (Fig. 1).

4) *Logistic Regression*: For the Linear SVC model, the 5 most significant features are the mean AP downstroke, the firing rate, the lowest amplitude of the Fourier transform, AP duration, and the mean AP upstroke. (Fig. 1). More specifically if we look at the weight of each feature to determine the cell type (Fig. 2), the same observations as the linear SVC can be made.

C. Dimensionality Reduction

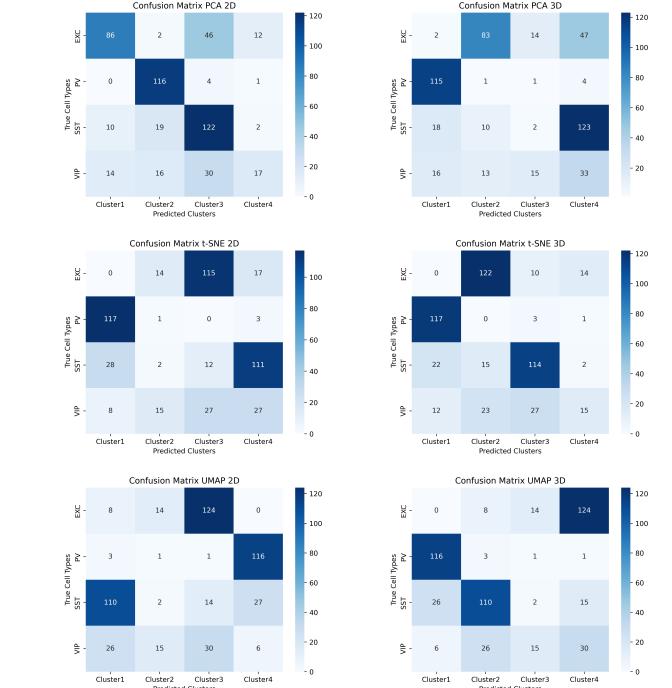


Fig. 5: K-Means Clustering Confusion Matrices. Top row is PCA. Middle row is t-SNE. Bottom row is UMAP. Left column in 2D. Right column in 3D.

From both PCA, t-SNE and UMAP (Fig. 6, 7, 8), there emerges similar patterns in the way cell classes are predisposed. PV cells seem to be more separated, and less mixed with other cell types. VIP cells are scattered and are not captured within a region of space, unlike the other cell classes. SST and EXC cells are depicted close to each other. Finally, a small, mixed class, grouping of points is always present in the bottom left corner of 2D projections.

D. Unsupervised Performance

K-Means clustering was a majority of times able to attribute a cluster to the region with the highest density of EXC, SST and

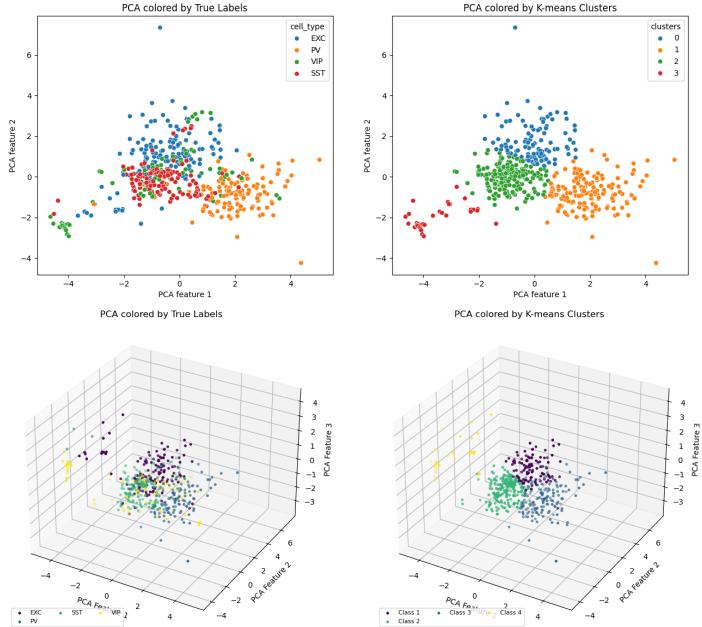


Fig. 6: 2D (top row) and 3D (bottom row) PCA projections of the data. Left column is colored by true labels, and right column is colored by K-Means clustering.

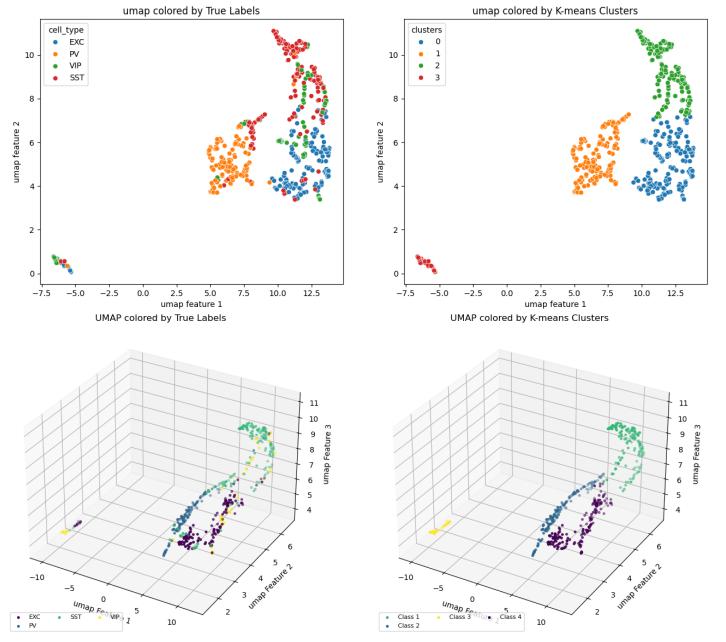


Fig. 8: 2D (top row) and 3D (bottom row) UMAP projections of the data. Left column is colored by true labels, and right column is colored by K-Means clustering.

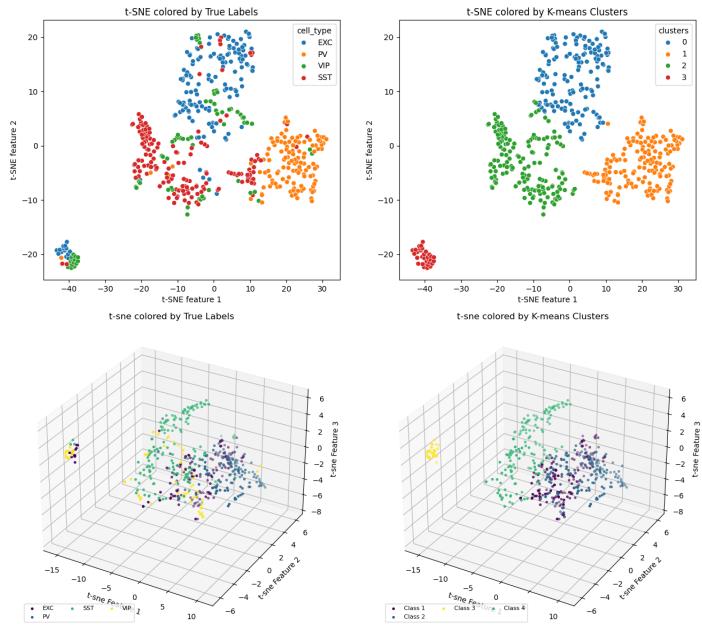


Fig. 7: 2D (top row) and 3D (bottom row) t-SNE projections of the data. Left column is colored by true labels, and right column is colored by K-Means clustering.

PV cells (Fig. 6, 7, 8). Due to the presence of the separated mixed grouping in the bottom left corner of each different dimensional space, it is always predicted as the final cluster. When looking at the confusion matrix from Fig. 5 one can observe the same phenomenon as described above. A cluster is always attributed to PV cells (cluster 2 and 1, for PCA, 1 and 1 for t-SNE, 4 and 1 for UMAP, respectively for 2D and 3D data). Although less pronounced, the same can be said for EXC and SST cells (except for PCA data where EXC cells are more sparse). One cluster captures the small mixed grouping consisting of roughly

14 EXC, 1 PV, 1 SST and 15 VIP cells. Also, VIP cells are found in similar ratios within each cluster.

V. DISCUSSION

A. Performance

It is important to note that for a 4-class classification like this one, a “chance level” score is $F1_{chance} = 0.25$. In that regard, the four best models achieve above 0.87 performance and the ensemble method is able to reach a score above 0.9. This is quite impressive considering the relatively low sample size ($n = 497$) and the moderate class imbalance present in our data. This means that there are definitely characteristics that can help to differentiate and discriminate between the different cell classes. Due to this high performance, we can also be more confident in our feature importance analysis.

B. Cell Class Specificity

Despite the fact that each supervised learning model is built differently, they all agree on certain aspects. It is hard to predict VIP cells but easy to predict PV cells.

This is also supported by unsupervised learning. K-Means was able to recognize PV cells quite accurately by grouping all of them into a cluster, but it did not attribute group VIP cells into a specific cluster.

The small mixed grouping of points found bottom left of dimensionally reduced spaces (Fig. 6, 7, 8), is probably due to the biologically nonsensical preprocessing step of replacing nonexistent AP durations and thresholds by a value of zero. This could mean that the 14 EXC, 1 PV, 1 SST and 15 VIP cells, mostly present in that grouping would have had no AP in their sweep.

C. General and Cell Specific Important Features

While being relatively simple, an advantage of linear model becomes clear due to the ease of interpretability. The important features for each supervised method repeat themselves. They are very similar in between types of models i.e. tree-based methods and linear methods, which is expected. The shared importance of features highlights the robustness of these factors in classifying cell types. This consistency suggests a biological relevance to these features. AP duration, firing rate and spectral characteristics were already identified as being good discriminants, but new features such as the number of bursts, the downward or upward phase of AP downstroke, influenced by ion channel dynamics in repolarization, seem to vary across EXC, PV, SST and VIP cell types. This generally reflects the overall balance of excitatory and inhibitory inputs.

VI. CONCLUSION

This study has undertaken a comprehensive exploration of machine learning methodologies applied to electrophysiological data. The supervised learning models, ranging from Gradient Boosting to Support Vector Machines, showcase impressive performance in differentiating cell types, with an ensemble model further elevating classification accuracy. Feature importance analyses reveal recurring themes across models, that were previously reported, emphasizing once again the significance of action potential duration, firing rates, and spectral characteristics, but also newly explored features such as the number of bursts or the action potential downstroke. In the realm of unsupervised learning, dimensionality reduction techniques provide insightful visualizations, hinting at spatial arrangements and inter-class relationships. While challenges persist, such as the elusive predictability of VIP cells, the study offers valuable insights into the discriminative features governing neuronal classifications, contributing to the broader understanding of neural dynamics and function.

VII. ANNEX

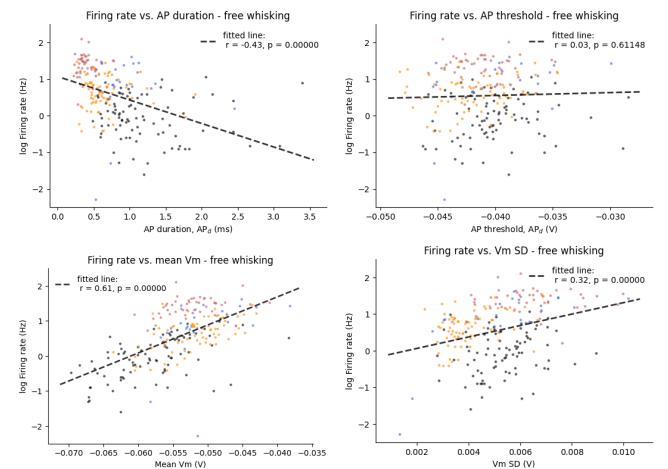


Fig. 9: 1. Mean Firing Rate vs. Action Potential Duration. 2. Mean Firing Rate Correlations with Action Potential Threshold. 3. Mean Firing Rate Correlations with Mean Membrane Potential. 4. Mean Firing Rate Correlations with Membrane Potential Standard Deviation.

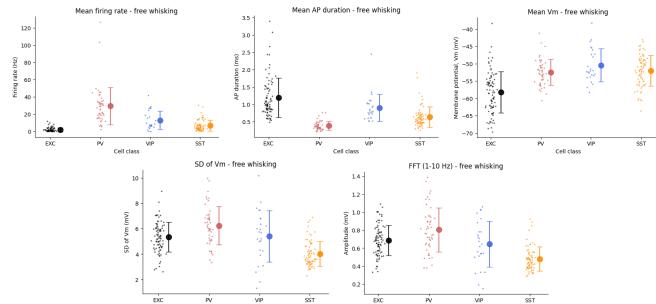


Fig. 10: 1. Mean Firing Rate. 2. Mean Action Potential Duration. 3. Mean Membrane Potential. 4. Mean Membrane Potential Standard Deviation. 5. Mean Fourier Transform Low Frequency.

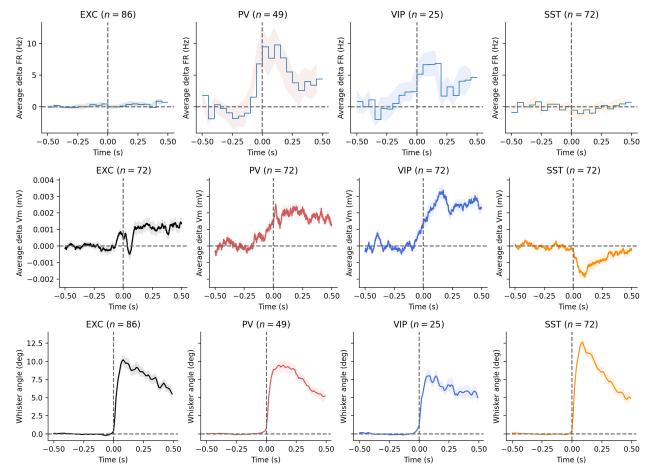


Fig. 11: 1. Whisker Onset Gradient-Averaged Firing Rate. 2. Whisker Onset Gradient-Averaged Membrane Potential. 3. Whisker Onset Gradient-Averaged Whisker Angle.

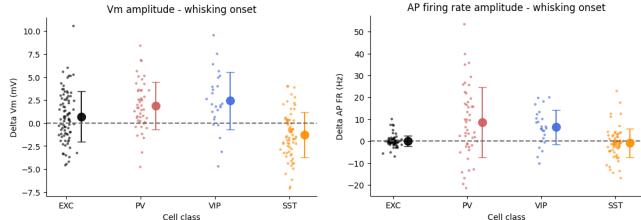


Fig. 12: 1. Delta Membrane Potential at Whisking Onset. 2. Delta Action Potential Firing Rate at Whisking Onset.

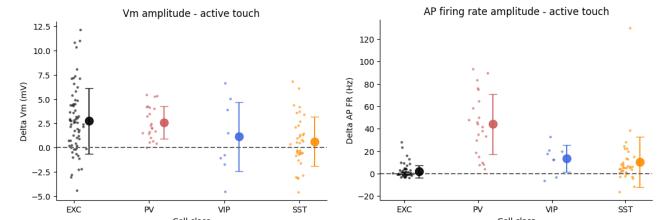


Fig. 15: 1. Delta Membrane Potential during Active Touch. 2. Delta Firing Rate during Active Touch.

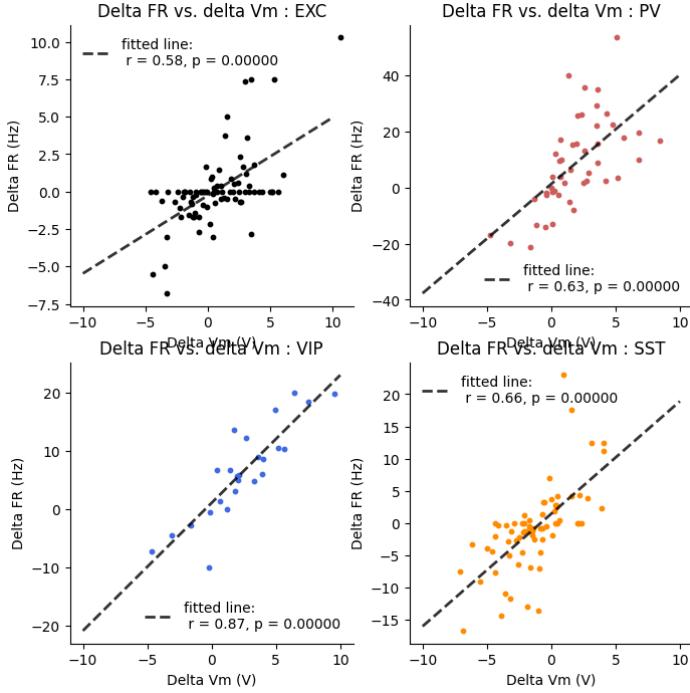


Fig. 13: Delta Firing Rate vs. Delta Membrane Potential.

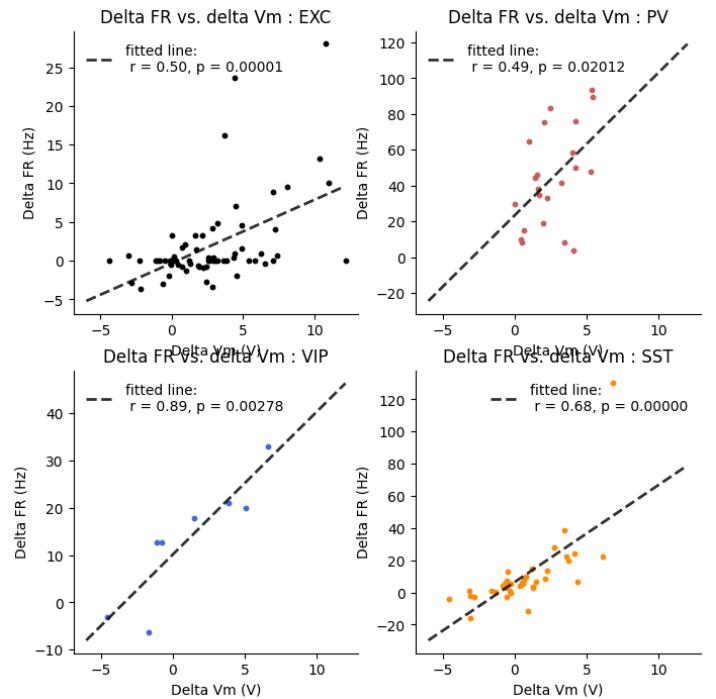


Fig. 16: Delta Firing Rate vs. Delta Membrane Potential during Active Touch.

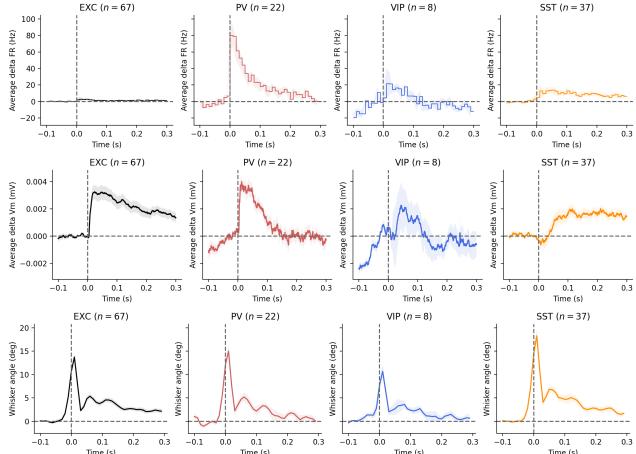


Fig. 14: 1. Touch Onset Gradient-Averaged Firing Rate. 2. Touch Onset Gradient-Averaged Membrane Potential. 3. Touch Onset Gradient-Averaged Whisker Angle.

Model	Hyperparameters
Gradient Boosting	Learning Rate: 1.0, Max Depth: 2, N Estimators: 176
Linear SVC	C: 0.8428, Multi-class: Crammer-Singer
Random Forest	Max Depth: 9, N Estimators: 174, Min Samples Leaf: 1
Logistic Regression	C: 5.241

TABLE III: Hyperparameters of Different Models. Any parameter not mentioned is equal to the default parameter in Sci-kit Learn.

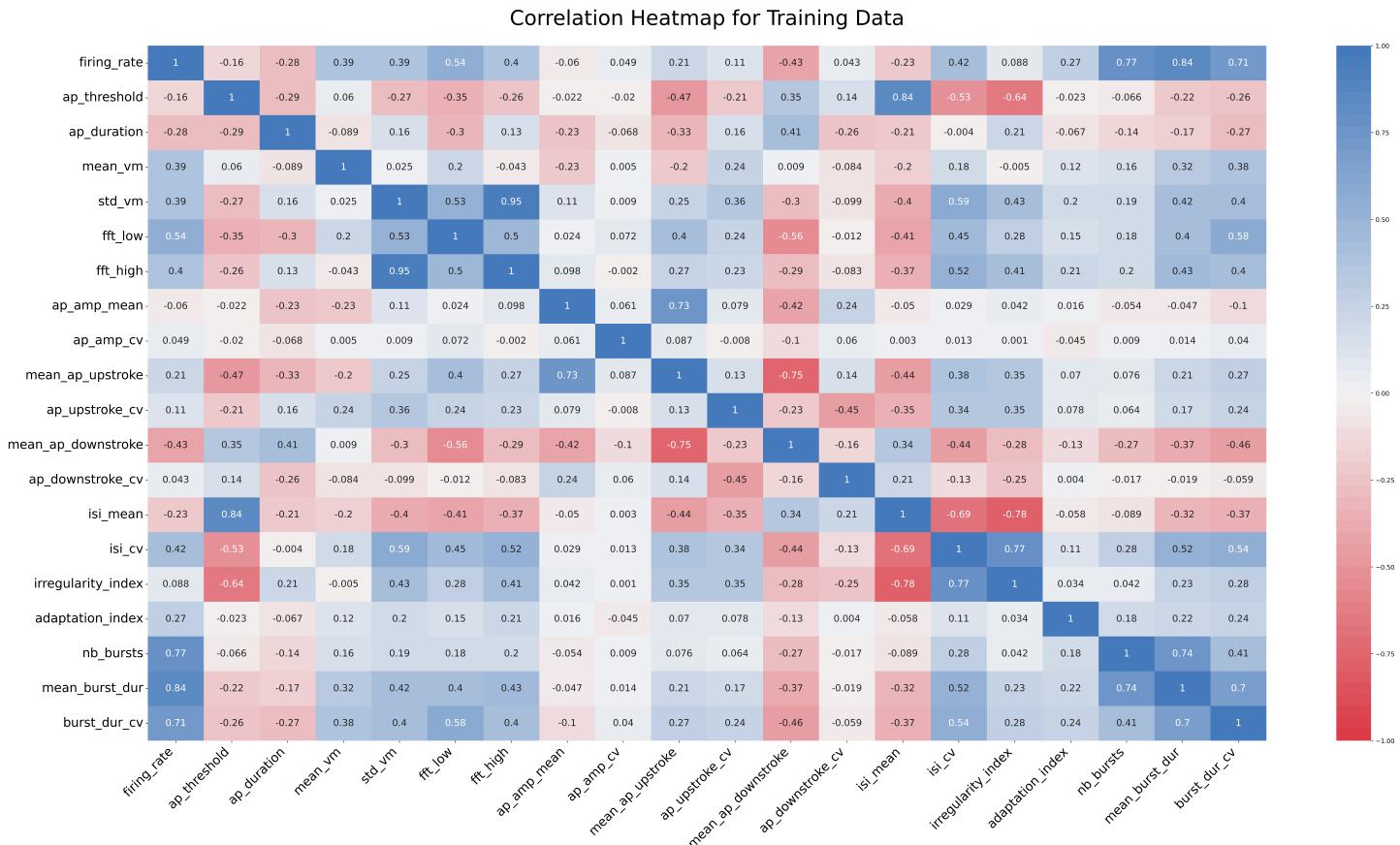


Fig. 17: Correlation of each predictor with one another within the training data