School of Informatics

Informatics Project Proposal
Machine Learning analysis of neural activity

B209251
May 2022

Abstract

This project proposes to use Machine Learning / Deep Learning techniques to analyze neural activity recorded from visual cortex of mice while performing learning based tasks. In our approach we plan to use these recordings as time series data. One of the problems in using complex models is that it is hard to decipher information regarding decisions made by the model. In this proposal we try to solve this using frameworks such as SHAP to better interpret results from the trained models. This analysis helps us better understand how interactions between different cell types in cortical regions change before and after learning tasks.

Date: Saturday 14$^{\text{th}}$ May, 2022

Tutor: Cristina Alexandru
Supervisor: Angus Chadwick

1 Motivation

Computational neuroscience is an interdisciplinary field of neuroscience, cognitive science, and mathematics that focuses on understanding the structure, development and cognitive abilities of the brain using mathematical models[1]. The main focus of this field is to explain how information is processed in brain and understand the underlying mechanisms. In the recent years due to advancements in the field of deep learning it had a huge influence in this field.

In order to better understand the brain it is essential to understand the role of evolution. From studies it is shown that evolution lead to enlargment of cortical areas of brain in mammals[2]. This lead to improvements in neo-cortical regions of the brain which is responsible for higher cognitive functions[3]. It is also a known fact that the size and shape of the neocortex differ based on the species. But some of the key features of neocortex remain same in all of them. This is one of the main reasons of using animals as test subjects to draw general inferences regarding the brain.

It's always been a mystery regarding how learning takes place in brain and this is hard to understand because of the complex structure and computations in neurons. The visual cortex region which is a part of neo-cortex deals with receiving and processing of visual data from the retina. This is further divided into 5 regions V1-V5. Visual information is first received by primary visual cortex region(V1) and then it is sent to ventral stream and dorsal stream[4]. In our project we focus on V1 region of the visual cortex which further contains 6 layers.

At the microscopic level connections between the neurons are driven by synapses(exchange of chemical also called as neurotransmitters). The neurons are classified as excitatory and inhibitory neurons based on how they effect other neurons. The excitatory neurons produces neurotransmitters that help the post synaptic cell to fire whereas the inhibitory neurons releases neurotransmitters that reduce the likeliness of connected cell to fire. In this project, data is taken from Layer 2/3 of V1 region in mice brain which consists of richly interconnected excitatory neurons (Pyramidal cells or PYR cells) and inhibitory interneurons(Paravalbium PV, Somatostatin SOM, Vasoactive Intestinal Peptide VIP cells)[5].

In the following parts of section 1 we discuss about the problem statement and objectives of the project. Followed by its novelty, importance and applications in the field. Then in section 2 details regarding the related works and background materials are discussed. Following this in section 3 further details regarding the potential solutions are discussed. In sections 4,5 discussing is done regarding the evaluation process and expected outcomes of the project. Section 6 contains the brief plan and research goals.

1.1 Problem Statement

From research it is known that the inhibitory neurons play a key role in suppressing neural activity so that only selective neurons play a key role during processing of information and stabilizing the network dynamics[6]. During learning process the sensory information plays a huge role on circuit dynamics. The representations in the network become more selective and easily discriminate stimulus during the process of learning[7][8]. But it is still poorly understood how these changes in networks take place based on the interactions between the classes of excitatory and inhibitory neurons.

In this project we try to analyze the interactions between the PYR,PV,SOM,VIP cells from the data which is recorded from primary visual cortex of mice brain during learning process.

1.2 Research Hypothesis and Objectives

The main research hypothesis that this project aims to analyze are as follows:

- In the original research [9] the authors use linear model(MVAR mode) for analysis of interactions between various classes. It would be interesting to see if there are any non linear correlations between the neurons that are not captured by the linear model.
- If the above mentioned research hypothesis turns out to be true it would be very useful in comparing different variations of non linear models and understand which model helps us best capture the data.
- It is hard to understand a non linear model because of the complexity of computations but with the recent advancements in interpretability of the non linear models we can better understand what the models capture[10]. This can be one of the objectives that help us in answering our hypothesis.

1.3 Timeliness and Novelty

It is a well known fact that the latest advancements of Deep learning techniques had a huge impact on many fields. Especially in the field of biomedical AI, but it had fair share of drawbacks when it came to analysis of these models. This is because of the fact that more complex models help capture the data well but these are black box models (i.e) we cannot infer logic regarding models behaviour. To counter these problems there is lot of research taking place in explainability of models called XAI(Explainable AI) or IML(Interpretible Machine Learning). These research already finds its applications in the field of financial data analysis and would be very useful in the field of Computational neuroscience as well[11].

1.4 Significance

This project/proposal aims to use non linear ML/DL techniques to analyze the data collected from various cell types in the primary visual cortex. The original research done includes using linear model with different variations but not non linear models. It would be a unique proposal because most of the research regarding neural activity does not use non linear models because of its problem of interpretibility. In this project we use the tools developed to solve these problems whose applications are prevalent in various other fields. This could help further research in the field to use more complex models to understand neural activity with an added advantage of having an interpretible model.

1.5 Feasibility

Since the data is available from past experiments and was used to train a linear model it is feasible to build a non-linear ML/DL model and use tools/frameworks such as SHAP(described in section 4) to evaluate the results of the models. Since the data is not very huge it is computationally feasible to train the proposed models. As for the implementation of these models it is based on the data (i.e) experiments regarding architecture can be done based on the data and can be done in the proposed time.

1.6 Beneficiaries

This research can help guide future research in the field to use non linear ML/DL techniques for drawing insights from more complex data. By use of existing frameworks such as SHAP the created models can be more trusted. This way it will improve use of ongoing research in AI to develop better models in computational neuroscience. All the results from this project will be published online after completing it so that future studies can use this as a reference.

2 Background and Related Work

2.1 Dataset

The neural data collected from mice brain is using calcium imaging technique[12]. With this method we can record the data of important neurons using calcium as dye and then identify them with use of fluorescent molecules that react to the binding of Ca^{+2} ions by fluorescence properties. From the fig 1 we can observe that the different type of neuron cells which we are interested in are color coded and their rate of firing. The data is converted from image to signals i.e improved brightness in image is recorded as firing of the neuron this way all the data is to signals which can be considered as preprocessing step. It is to be noted that data from 8 mice are taken for the experiment where the mice were trained to perform a learning task [9] and hence the data is collected before the learning task and after the learning task to understand how learning affected the groups of different cell types. It is to be noted that we do not focus too much on how it is created since it is beyond the scope of this project.

2.2 Time series Analysis

Time series analysis is defined as analysis of data which has correlation with its previous values[13]. It can be classified into univariate and multivariate analysis based on the parameters used for analysis. In univariate analysis only one variable is present and it is dependent on its previous values. Whereas in multivariate analysis there are more than 2 features and they are interdependent among themselves and also their previous values. This time series analysis has huge applications in fields of finance like stock prediction, risk estimation etc. In this project we use the multi variate time series analysis as there are different type of cells correlated among themselves and with respect to time.

2.3 Machine Learning and Deep Learning for time series analysis

With the developments unfolding in the field of machine learning and deep learning it had a huge influence on many fields. One of them include time series analysis which has applications in various sectors such as finance[14], medical domain[15], astronomy[16], etc. Some of the main machine learning and deep learning techniques which we focus in this project are discussed as follows:

XGBoost: It was first introduced by authors of [17] as a tree boosting technique which has given state of the art performance at that time for various machine learning challenges. It had its applications in various fields, [18] uses XGBoost for stock prediction volatility. [19] uses XGBoost to perform short term time series analysis etc. In our project we plan to use it as a time series prediction task similar to that used by baseline model(MVAR model)[9].

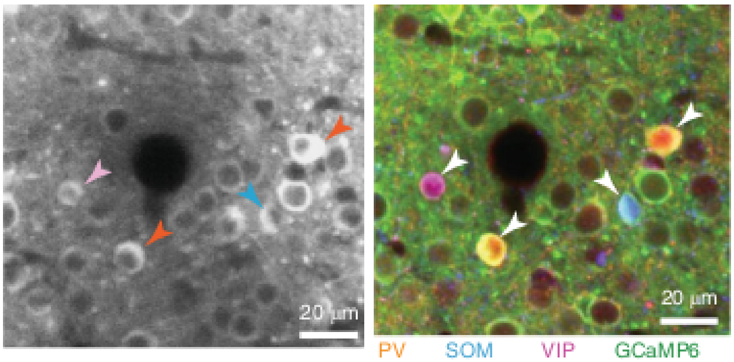

Figure 1: Images showing identification of cell types using calcium imaging[9]

LSTM(Long Short Term Memory): It is a form of recurrent neural network first introduced by [20] which was a novel approach that improved performance of recurrent neural networks. Since then it has found its advantages in various fields such as NLP(Natural Language Processing), computer vision and many other deep learning tasks. An LSTM can process sequence of data whereas the traditional networks can only process single data point(like images). An LSTM contains an input, output,forget gate which are learnable parameters and the network tries to understand which information is important or not by handling these gates that help in flow of information. These LSTM networks can be used for both time series prediction [21] and classification tasks [22]. It is often found that the LSTM networks needs lot of data for training and it is partially based on the architecture as well.

GRU is a form of LSTM network introduced to solve this problem [23]. In GRU networks the input and forget gate are combined to form one gate called as update gate. This reduces the no of parameters. GRU's are very useful when the datasets are small as compared to LSTM which requires large datasets to train. These explained forms of recurrent neural networks provide state of the art results for time series problems and in our project we try to utilize these proposed networks to analyze dynamics in cortical circuits.

2.4 Studies regarding interactions between different cell types

Is is important to understand the neuroscience perspective of the problem before looking for solutions using machine learning techniques. In the past research it is proven that learning has a huge impact on cortical networks which process sensory information. They tend to discriminate stimuli better after the process of learning[7],[24]. Due to the complexity of these networks it was still poorly understood how the interactions actually help distinguish the sensory information. The authors of the paper [25] observed that sensory perception is dependent on the context. When they tested the interactions in the auditory cortex of the mice they found that excitatory neurons were suppressed during behaviour task and some of the inhibitory neurons were still active. This study did not try to find exactly how the interactions uncovered which inspired other research in this field. Some research also suggests that specific cell types contribute to plasticity of the networks during learning[26],[27]. The gap in the knowledge was each type of interneuron were seperately analyzed but there were no studies that tried to understand interactions between more than two groups of cells. This was achieved by the authors [9] which acts as a baseline study for this proposal to further build on it.

3 Programme and Methodology

In this project we propose a standard protocol to achieve the specified milestones (table 1). There would be some initial time allocated for reading and research regarding best suitable algorithms for the dataset. So excluding the reading time and the dissertation writing the whole process can be broken down into two work packages(fig 2).

As the initial part of the plan we try to explore different non-linear machine learning algorithms as mentioned in 2. This is achieved by firstly implementing the machine learning algorithm according to the provided dataset, after which verification of the implementation and training process is done. Once this is complete we try to analyze the models performance by gathering information regarding feature importance at particular time interval and infer the interactions between different cell types in that time interval. The same protocol is followed for different machine learning algorithms for time series data the only change would be the algorithm used such as XGBoost, Prophet.

Once the analysis of these models and done they are compared with the baseline model [9](MVAR model). This helps us answer our hypothesis of whether non linear analysis provide same results or different results. In the second part of the work plan we try to understand the affect of non-linearity in the results so we use some deep learning techniques which give promising results for time series analysis2 like LSTM, GRU. For these experiments we start with a basic architecture and try to fit the data without the problem of overfitting[28]. This can be achieved by trying architectures with varying number of hidden layers and neurons based on the data. After implementing proposed networks we can use the mentioned framework SHAP to understand the feature importance in non linear models(described in detail in sec 4).

The unique aspect of this project would be testing various models on a unique dataset where it has not been tested before. It is important to note that both the work packages proposed are independent from each other since the methods are different but the important aspect of the project is the analysis part of the models. It is fair to say that both the packages should be dealt independently and given the time constraint it should be feasible.

3.1 Risk Assessment

For this project the dataset is already present in the original published paper[9] and the preprocessed version is available as well so there is no risk of finding a dataset for this task. Regarding the proposed machine learning and deep learning techniques implementation is feasible as there is open source tools and by the use of packages such as Tensorflow, scikit learn, scipy all machine learning and deep learning models can be created. It is expected that the project does not require very high computation power and the available resources should be sufficient to finish the project.

As for the analysis part of the problem statement this proposal tries to use open source framework known as SHAP(further details are discussed in sec 4). A slight risk involves in use of deep learning techniques as there is no perfect way to analyze them. The proposed methods show good performance in analysis of time series data in financial field so it is expected to provide basic information regarding patterns in the data.

3.2 Ethics

This project does not require any ethical approval since the data is already available through past experiments and has been used in other papers[9],[29]. The data is used confidentially and not available publicly. If the proposal is successful it would be a good contribution to the field and there is no expected ethical problem if the results are to be published.

4 Evaluation

The main aspect of the project is to analyze the results obtained from the model(2) since it helps us solve our problem statement. As mentioned in previous sections one of the toughest problem to solve is interpreting the results from complex models and for this project we propose to use SHAP(SHapely Additive exPlanaltions) framework [30]. It works on the priciples of Shapely values which is derived from game theory. In simple terms if we have n actors to finish a job and interrelation between them exists (for example some combination of actors work good together and some do not)the shapely values explains the contribution of each actor to finish the job taking the correlation into context. This basic game theory principle is used in most of the games for fair distribution of rewards or costs based on the contribution.

It is to be noted that there are other frameworks that exist for explainable AI solutions but it is always not very clear regarding which one of them can give valid explanation. To remedy this problem the authors of the paper [30] used Shapley values as basic method and added 6 other existing methods to create a unified approach that gives the feature importance while making a decision. The important thing to know about this framework is that it explains the final prediction of the model from the average prediction of the model i.e let's say a model gave prediction(in range of 0-25) of 20 for a particular input and on average the model gives 5 when tried on many inputs. The SHAP framework explains how the model reaches from 5 to 20 rather than how the model reaches from 0 to 20.

This is relevant to our scenario because we can now understand the contributions of each neuron at desired interval of time. So it is possible to use data before the learning process and after the learning process to compare how the contributions changed with respect to time.

5 Expected Outcomes

The original paper [9] uses linear model for analysis and found out the following interactions:

- After the learning process it is observed that the PYR and PV cells reorganized their interactions which lead to formation of subnetworks which are stimulus specific.
- The SOM cells were de-correlated from the network after the learning process suggesting that these cells are responsible for gating response plasticity.

Since our hypothesis is to see whether non linear models give rise to same interactions or different interactions to better understand the changes in cortical networks during the process of learning. An added advantage to this project is that if we prove that we get similar results using non linear models then it would add extra proof to the original paper. Another advantage would be end to end non linear model for analyzing further experiments which require more complex models without the disadvantage of black box modeling. If the results turn out to be different then the original results then it would fill the gap in the knowledge regarding the network organisation of cortical areas in the primary visual cortex regions.

6 Research Plan, Milestones and Deliverables

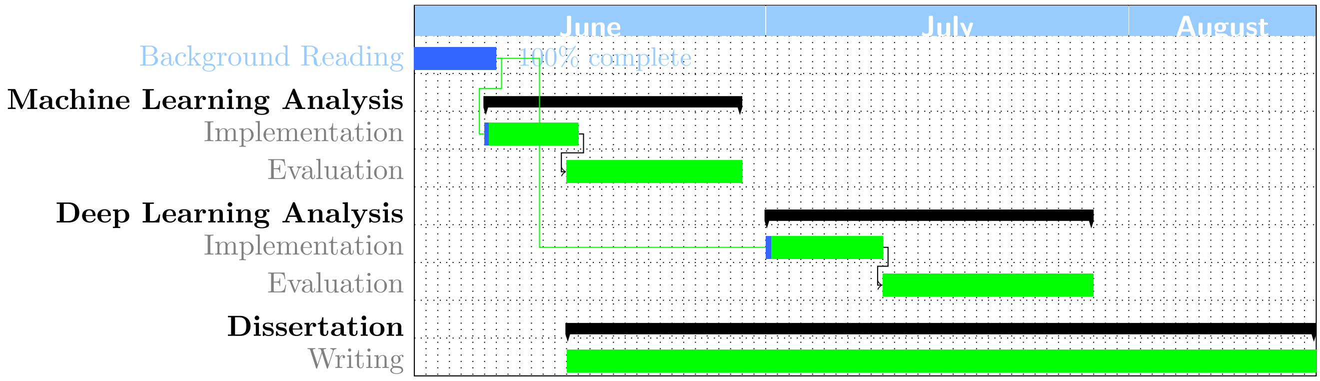

Figure 2: Gantt Chart of the activities defined for this project.

Milestone	Week	Description
M_1	1	Feasibility study to choose algorithms
M_2	2	Implementing machine learning algorithms
M_3	6	Implementing deep learning methods
M_4	10	Submission of dissertation

Table 1: Milestones defined in this project.

Deliverable	Week	Description
D_1	5	Analysis of ML models
D_2	8	Evaluation of deep learning methods
D_3	10	Dissertation

Table 2: List of deliverables defined in this project.

References

[1] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[2] Robert A Barton and Paul H Harvey. Mosaic evolution of brain structure in mammals. *Nature*, 405(6790):1055–1058, 2000.

[3] Jan H Lui, David V Hansen, and Arnold R Kriegstein. Development and evolution of the human neocortex. *Cell*, 146(1):18–36, 2011.

[4] Kalanit Grill-Spector and Rafael Malach. The human visual cortex. *Annu. Rev. Neurosci.*, 27:649–677, 2004.

[5] Zixiu Xiang, John R Huguenard, and David A Prince. Gabaa receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex. *the Journal of Physiology*, 506(3):715–730, 1998.

[6] Mark M Churchland, M Yu Byron, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378, 2010.

[7] Gregg H Recanzone, Christoph E Schreiner, and Michael M Merzenich. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience*, 13(1):87–103, 1993.

[8] Tianming Yang and John HR Maunsell. The effect of perceptual learning on neuronal responses in monkey visual area v4. *Journal of Neuroscience*, 24(7):1617–1626, 2004.

[9] Adil G Khan, Jasper Poort, Angus Chadwick, Antonin Blot, Maneesh Sahani, Thomas D Mrsic-Flogel, and Sonja B Hofer. Distinct learning-induced changes in stimulus selectivity and interactions of gabaergic interneuron classes in visual cortex. *Nature neuroscience*, 21(6):851–859, 2018.

[10] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10):749, 2018.

[11] Sherin Mary Mathews. Explainable artificial intelligence applications in nlp, biomedical, and malware classification: a literature review. In *Intelligent computing-proceedings of the computing conference*, pages 1269–1292. Springer, 2019.

[12] Christoph Stosiek, Olga Garaschuk, Knut Holthoff, and Arthur Konnerth. In vivo two-photon calcium imaging of neuronal networks. *Proceedings of the National Academy of Sciences*, 100(12):7319–7324, 2003.

[13] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer, 2000.

[14] Ruey S Tsay. *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons, 2013.

[15] Jaydip Sen and Tamal Chaudhuri. A time series analysis-based forecasting framework for the indian healthcare sector. *Journal of Insurance and Financial Management*, 3(1), 2017.

[16] T Subba Rao, MB Priestly, and O Lessi. Applications of time series analysis in astronomy and meteorology, 1998.

[17] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[18] Yan Wang and Yuankai Guo. Forecasting method of stock market volatility in time series data based on mixed model of arima and xgboost. *China Communications*, 17(3):205–221, 2020.

[19] Raza Abid Abbasi, Nadeem Javaid, Muhammad Nauman Javid Ghuman, Zahoor Ali Khan, Shujat Ur Rehman, et al. Short term load forecasting using xgboost. In *Workshops of the International Conference on Advanced Information Networking and Applications*, pages 1120–1131. Springer, 2019.

[20] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[21] Felix A Gers, Douglas Eck, and Jürgen Schmidhuber. Applying lstm to time series predictable through time-window approaches. In *Neural Nets WIRN Vietri-01*, pages 193–200. Springer, 2002.

[22] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. Lstm fully convolutional networks for time series classification. *IEEE access*, 6:1662–1669, 2017.

[23] 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.

[24] Aniek Schoups, Rufin Vogels, Ning Qian, and Guy Orban. Practising orientation identification improves orientation coding in v1 neurons. *Nature*, 412(6846):549–553, 2001.

[25] Kishore V Kuchibhotla, Jonathan V Gill, Grace W Lindsay, Eleni S Papadoyannis, Rachel E Field, Tom A Hindmarsh Sten, Kenneth D Miller, and Robert C Froemke. Parallel processing by cortical inhibition enables context-dependent behavior. *Nature neuroscience*, 20(1):62–71, 2017.

[26] Johannes J Letzkus, Steffen BE Wolff, Elisabeth MM Meyer, Philip Tovote, Julien Courtin, Cyril Herry, and Andreas Lüthi. A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature*, 480(7377):331–335, 2011.

[27] Simon X Chen, An Na Kim, Andrew J Peters, and Takaki Komiyama. Subtype-specific plasticity of inhibitory circuits in motor cortex during motor learning. *Nature neuroscience*, 18(8):1109–1115, 2015.

[28] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

[29] Jasper Poort, Katharina A Wilmes, Antonin Blot, Angus Chadwick, Maneesh Sahani, Claudia Clopath, Thomas D Mrsic-Flogel, Sonja B Hofer, and Adil G Khan. Learning and attention increase visual response selectivity through distinct mechanisms. *Neuron*, 110(4):686–697, 2022.

[30] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.