Adapting Supervised Machine Learning for Analysis of Neurobiological Signals

Andrei Ciuparu
Transylvanian Institute of Neuroscience
Technical University of Cluj-Napoca
Email: ciuparu@tins.ro

Ann Christin Garvert
The University of Oslo (UiO)
Email: a.c.garvert@medisin.uio.no

Koen Gerard Alois Vervaeke
The University of Oslo (UiO)
Email: k.g.a.vervaeke@medisin.uio.no

Raul C. Muresan
Transylvanian Institute of Neuroscience
STAR-UBB Institute, Babes-Bolyai University
Email: muresan@tins.ro

***Abstract*—Machine learning algorithms are uniquely suited to decipher high-dimensional neural data, which often contains multiple overlapping signals carried by the same set of cells, leading to a low signal-to-noise ratio for individual signals. In this study, we explored the role of the retrosplenial cortex (RSC) in navigation, specifically investigating whether landmarks are encoded in this area and if they serve as an error signal enabling the RSC to adjust its internal position representation. We gathered data from hundreds of neurons while mice ran along a dark corridor with textured patches on the ground. This data was then processed using a multi-layer perceptron (MLP) to classify patch identity, patch encounter number, or all events. Furthermore, we performed a position regression analysis to test the error signal hypothesis, which failed to confirm it, and highlighted potential interpretability issues arising from the use of machine learning. Nevertheless, our findings demonstrated that even with very small and shallow networks, it is feasible to classify and regress information from raw neural data. Our research underscores the continued need for comprehensive testing and cautious interpretation in machine learning applications, especially when dealing with multifaceted and complex data such as neural signals.**

***Index Terms*—GCaMP, supervised machine learning, retrosplenial cortex, navigation**

I. Introduction

The incorporation of supervised machine learning algorithms has become a staple in modern engineering practices and data analysis contexts, influencing a broad spectrum of disciplines. Despite the pervasive application, conceptualizing the utilization of supervised machine learning for the purpose of data analysis often presents an inherent challenge, primarily attributed to the 'black box' nature of these algorithms. In essence, the internal working mechanics of machine learning models, whilst producing efficient results, tend to obscure transparency, making it difficult for researchers to interpret or understand the direct correlations between input variables and the output predictions.

NO (Norway) Grants 2014-2021 (Project contract number 20/2020 (RO-NO-2019-0504)), RO (Romanian) Grants CNCS-UEFISCDI (ERA-NET-FLAG-ERA-ModelDXConsciousness, and ERANET-NEURON-Unscrambly), and a H2020 grant funded by the European Commission (grant agreement 952096, NEUROTWIN)

This black box problem stems from the complex nature of supervised machine learning algorithms, which are designed to autonomously learn patterns and relationships within vast amounts of data. These algorithms employ intricate mathematical calculations and optimization techniques to adjust model parameters and minimize prediction errors. As a result, the models become highly sophisticated and intricate, often comprising numerous layers and millions of interconnected parameters. While this complexity allows for impressive accuracy and predictive power, it also poses a significant challenge when it comes to interpretability. The models capture patterns and correlations that are difficult for humans to discern, especially as the number of input features and complexity of the models increase. Unlike traditional statistical models, where the relationships between variables can be explicitly expressed through equations, machine learning models operate in a high-dimensional space, making it challenging to trace how specific input variables influence the final predictions.

However, one particularly compelling property of supervised learning algorithms, specifically Multilayer Perceptrons (MLPs), counters this challenge: the concept of the Universal Approximation Theorem [1]. The theorem essentially proclaims that an MLP, with at least one hidden layer containing a finite number of neurons, can approximate any continuous function to a desirable degree of precision, given that the function is defined on compact subsets of real numbers.

What makes this theorem so compelling is its statement about arbitrary mapping. This means that an MLP can learn to map any input to any output, even when the relationship between them is exceptionally complex or non-linear, provided that there exists a function that describes this relationship. With the correct parameter settings, and a sufficient number of hidden neurons, the MLPs can approximate this function, regardless of how intricate or convoluted it may be.

This capability of MLPs is what sets the foundation for them to act as universal information detectors, able to parse through a vast array of data types and structures, and identify underlying patterns or correlations, however complex they may be. Notably, the application of this property finds significant

relevance in the analysis of neurobiological signals. The complexity and high dimensionality inherent in these signals necessitate a robust mechanism for information extraction. Neurobiological signals often present convoluted, non-linear patterns that evade simple statistical analyses, making MLPs particularly advantageous for this domain. By applying MLPs as universal information detectors, we enable the exploration of intricate and potentially meaningful relationships within these signals, promising to unlock new insights into neurobiological phenomena. This approach not only challenges the opacity of the 'black box', but it also presents an innovative way of deciphering the cryptic language of neurobiological data. The subsequent sections delve into the practical feasibility and potential of this approach.

In this light, the need for deploying supervised machine learning procedures in neuroscience research becomes clear. We can infer that the brain transforms complex multi-dimensional sensory input (along with its internal state) into motor outputs. By making the assumption that this mapping can be mathematically described, we can then use MLP to see if particular measured activity contributes to it. For example, if we assume that the brain creates an internal representation of the external world, we can attempt to decode that internal representation using machine learning (from a given set of measured neural data). If a classification attempt is successful, then we can confidently assume that there is information present in the signal about some aspect of the identity of the stimuli being classified. Further, we can interrogate more specific subsets of the data or of the environment to see how the classification performance varies to understand more precisely how it is encoded.

The practical utility of supervised machine learning procedures extends further, enabling researchers to unravel the subtleties of brain function and structure. For instance, analyzing patterns in functional Magnetic Resonance Imaging (fMRI) or Electroencephalogram (EEG) data using these advanced models could potentially expose novel insights about cognitive processes [2] and neural connectivity [3]. In addition, machine learning models could assist in the detection and diagnosis of neurological disorders by identifying distinctive patterns in neurobiological signals that may otherwise remain undetected by conventional analysis methods[[4], [5]].

However, the application of these models in neuroscience is not without its challenges. Among them, interpreting the result of a classification analysis is not a trivial task. Because MLP can approximate **any** arbitrary mapping, they may latch on to any aspect of the stimuli being classified. In other words, it is a mistaken assumption that the animal is experiencing only the stimuli that we are trying to classify, and any correlated stimuli will contribute to the decoding. Further, while MLP allow us to detect the presence of information, it is much more difficult to extract useful information about **how** the data is encoded. Thus, the development of approaches that combine the robustness of machine learning with enhanced interpretability is a burgeoning need in the field of neuroscience. Ultimately, supervised machine learning procedures could play a pivotal role in transforming our understanding of the brain's intricate workings.

In the world of neuroscience, numerous unanswered questions and unexplored territories persist where the application of machine learning could provide use. One such intriguing subject revolves around the role of the RSC in navigation, a task fundamental to survival, yet whose underlying neural mechanisms remain largely enigmatic. The RSC, an integral component of the brain's navigation system, has been a focal point of extensive research due to its suggested involvement in various aspects of spatial cognition [6]–[8]. However, the exact nature of its contribution is not fully understood, nor is the way that it encodes the position information.

Employing the universal information detection capabilities of MLPs on such complex data could provide fresh perspectives on the function of the RSC. Given the model's ability to approximate any continuous function, we are equipped with a tool that can potentially decode the complex patterns in the data, thereby elucidating the role of the RSC in navigation. Here, we will delve deeper into this intriguing question, demonstrating an implementation of supervised machine learning techniques in pursuit of answers within the multidimensional labyrinth of neurobiological signals.

Drawing on existing research, we propose to delve deeper into the multifaceted role of the RSC in spatial cognition and navigation. Prior research has established the presence of position-tuned cells in the RSC [9], as well as shown that it carries directional information [10]. Here, we wish to test the hypothesis that the RSC contributes significantly to path integration, a process through which external and internal information is integrated in order to encode the trajectory that an animal has taken.

This concept extends to the RSC's role in assimilating information about the present and target locations, as well as the relative directions and distances between them. We further suggest that the RSC could be integral in route planning, decision-making, and error correction during navigation. To test these hypotheses, we apply an MLP model to a series of analyses on GCaMP (green fluorescent calmodulin protein) data collected from the RSC during a navigation task. Our examination strives to untangle the complex, possibly non-linear, relationships present in this data.

Our experimental design incorporates a navigation task, where the animals were required to find their way to a goal location in the dark. Throughout these experiments, we collected optical fluorescence imaging (GCaMP) data from the RSC to capture the neuron activity as the animals navigated the environment. This dataset is high-dimensional and complex, embodying the challenges and opportunities of neurobiological data analysis. Applying the MLP model to this data allows us to decipher the underlying patterns and relationships that may explain the function of the RSC in navigation.

In summary, our research aims to harness the power of supervised machine learning, specifically MLPs, to parse through the complex GCaMP data gathered from the RSC during a navigation task. We propose to decipher the non-linear and

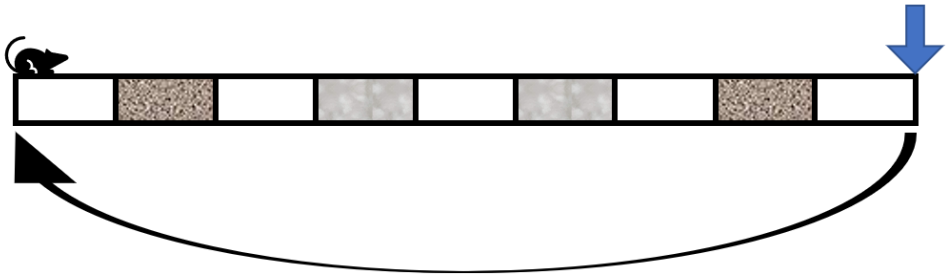

Fig. 1. **Navigation task** The brown textured tiles represent sandpaper texture, and the grey tiles represent cottonball texture, the blue arrow shows the location of the reward, and the black arrow shows the fact that this is set up on a wheel.

multidimensional patterns that could shed light on the role of the RSC in spatial cognition and navigation.

II. Materials and Methods

This article offers a summarized description of the experimental setup and data collection process used in the study conducted by Gianatti, M., Garvert, A. C., and Vervaeke, K. (2022), titled "Diverse long-range projections convey position information to the retrosplenial cortex" [11], from which we used the data for this paper. It should be noted that this summary is intended to provide a general understanding of the experimental design and does not encompass the full complexity and depth of the original methodology. For comprehensive insights into the exact experimental setup, recording system, and other procedural nuances, the reader is strongly encouraged to refer directly to the original publication.

A. Experimental setup

The experimental setup involved an investigation of activity within the RSC in response to tactile cues provided to mice. The mice navigated a track atop a polystyrene wheel, simulating a corridor, under complete darkness conditions. The wheel was a polystyrene cylinder, 157 cm in circumference and 10 cm wide.

Two tactile cues were used, namely sandpaper strips and white soft felt pads (furniture pads). These were positioned at specific intervals on the wheel: 42-47 cm and 122-128 cm for sandpaper strips and 63-68 cm and 100-106 cm for felt pads, as illustrated in Figure 1, and a water reward was given at the end of the track.

The test was designed to examine whether the RSC stored data regarding the position of the animals and how these positions might be impacted by the mice encountering different textures on the track. The assumption was that the accuracy of the internal position would reduce over the blank sections of the track due to cumulative errors, which would then be reset with the tactile cue of a texture change.

B. Data collection

Data was collected using a custom-built two-photon microscope (INSS), designed to accommodate the large running wheel. The microscope provided ample space under the objective. Images were acquired at a rate of 31 Hz with 512 x 512 pixel resolution, using SciScan (opensource, LabVIEW, National Instruments).

The recordings were taken in deep L1 and superficial L2/3 of agranular RSC (70-130 μm depth from the pial surface) and at 100-120μm below the pial surface for somatic recordings, corresponding to L2/3 of agranular RSC.

All experiments were conducted in total darkness within a sealed box made of black construction hardboard (Thorlabs, TB4) and black coated fabric (Thorlabs, BK5). The animals' running speed and absolute position on the wheel were calculated using a rotary encoder.

C. Data pre-processing

The pre-processing stage involved several steps to rectify brain movement and segment images. Brain movement was corrected by passing the images through a combination of custom-written Matlab scripts (NANSEN) and NoRMCorre [12].

Following the motion correction step, we proceeded with the selection of regions of interest (ROIs), each representing a distinct area of neuronal activity within the RSC (neurons). ROI selection was achieved using additional custom-written MATLAB scripts (NANSEN) and represents an image segmentation problem. Soma regions of interest (ROIs) were detected using a custom auto-segmentation method (NANSEN), followed by manual curation. This involved a visual inspection of each ROI to ensure accurate representation. In instances where ROIs overlapped, the overlapping sections were excluded to avoid data corruption.

Finally, we derived a time trace for each ROI, by averaging the luminance of pixels within, representing the activity over time within that particular ROI, providing a foundation for further analysis. Opting to maintain the integrity and full scope of the collected data, we used the raw, unfiltered traces for the subsequent steps. We elected not to apply further cleaning or signal processing steps, such as the OASIS algorithm [13] or other spike-detection coupled with an exponential decay kernel convolution, as we wanted to preserve as much of the original information as possible. We entrusted our chosen network to capably identify and extract relevant signals from potential background noise, thereby ensuring our analysis incorporated the fullest representation of the RSC's activity during the navigation task.

D. Machine learning analyses

The data traces were fed into a three-layer network (InputSize-100-100-100-OutputSize) with soft++ activation (c=2, k=1) [14]. The input size depends on the dataset and is the number of neurons, and the output size depends on the type of classification and is 1 for regression. The first two layers incorporated a 10% dropout. The batch size was set at 25, and the network was trained on 20 different data splits using the ADAM optimizer [15]. The learning rate schedule was linear, starting from a fixed value and gradually decreasing to zero over the epochs.

For the classification analyses, the initial learning rate was 0.5, and the network was trained for 100 epochs. The cross-entropy loss between the desired and given output distributions was minimized. For regression analysis, the starting learning rate was 0.001, the network was trained for 500 epochs, and the mean squared error between the desired and given output values was minimized. For each classification analysis, we also ran a control analysis, involving shuffling the labels before attempting to train a classifier on them.

III. Results

Our initial approach was to apply two binary classification tasks and a multiple (4-fold) classification task to understand what could be predicted directly from the data. We selected 100ms of data before the mouse interacted with any textured surface or reward location. Subsequently, we attempted to predict whether the texture was either sandpaper or cotton, if the mouse encountered the texture for the first or second time in a trial, and lastly, we endeavored to classify all location onsets from each other. This encompassed both tasks at once and the reward location. The promising results obtained from all three analyses across most datasets are illustrated in the following figures.

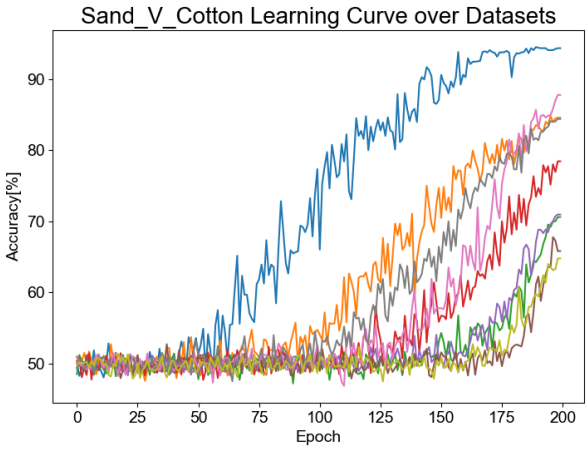

Fig. 2. **Stimulus identity learning curves** for 9 different datasets

It is common for representations of highly salient stimuli to be distributed across multiple different brain regions, and in this case, the textured patches would be particularly useful

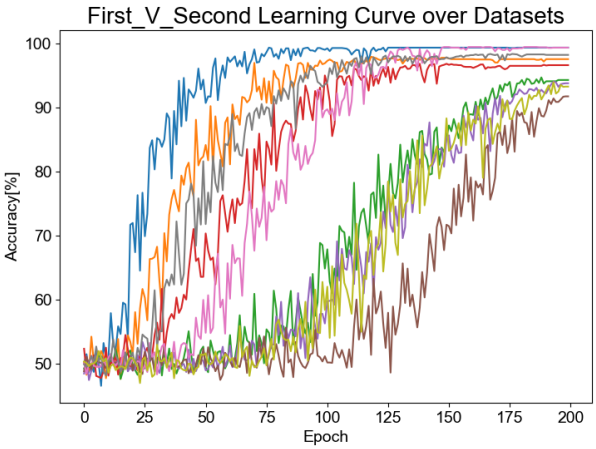

Fig. 3. **Encounter number learning curves** for 9 different datasets

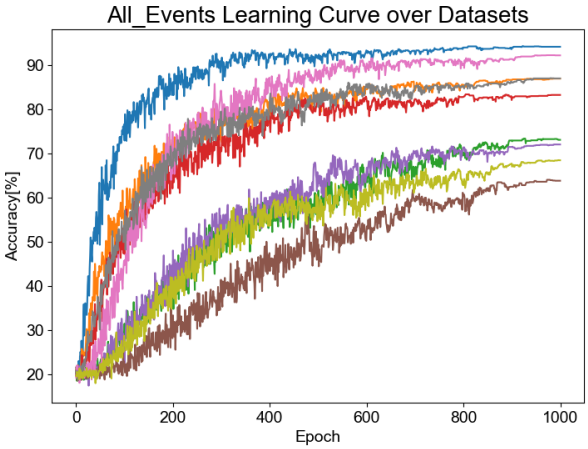

Fig. 4. **Encounter type learning curves** for 9 different datasets

in encoding spatial information. Thus, we show that it is possible to predict what kind of textured patch the animal is interacting with from the measured responses (Figure 2). There is a high variability across different datasets in how quickly the algorithm is able to extract the patterns, but in all cases, we obtained above chance level classification performance. Similarly, whether the animal is encountering a given texture for the first or second time in a trial is positionally relevant, so the classification presented in Figure 3 also seems to show positional encoding in the RSC. Finally, in Figure 4 we can see that there is a higher variability across datasets in the converged final performance when attempting to classify all relevant events in a trial (sandpaper 1 and 2, cotton 1 and 2, reward). This is somewhat expected due to the difficulty of the classification task (there are 5 classes leading to a theoretical chance level of 20%).

These results led to an important insight. If position information was indeed encoded in the data, it would render other variable predictions trivial because every other variable in our

experiment correlates 100% with the position on the track. In other words, the animal encounters the patches at the exact same positions during each trial, thus instead of using stimulus identity information to predict the stimulus, the MLP can rely solely on position information

To verify the predictability of position from the data, we used a comparable strategy. We selected non-overlapping segments of 200 samples from the data, and the expected output was the immediate subsequent position on the track. As this variable is continuous, we normalized the position data by dividing it by the maximum value and measured the Mean Squared Error (MSE). As shown in Figure 5, this rudimentary network was able to decode position information with less than 8 cm of error on a 157 cm track. Although the resolution is not particularly impressive, it is sufficient to distinguish all areas on the track from each other. Therefore, we can assert with high confidence that position information is encoded in the RSC. However, it remains uncertain whether any other type of information is encoded.

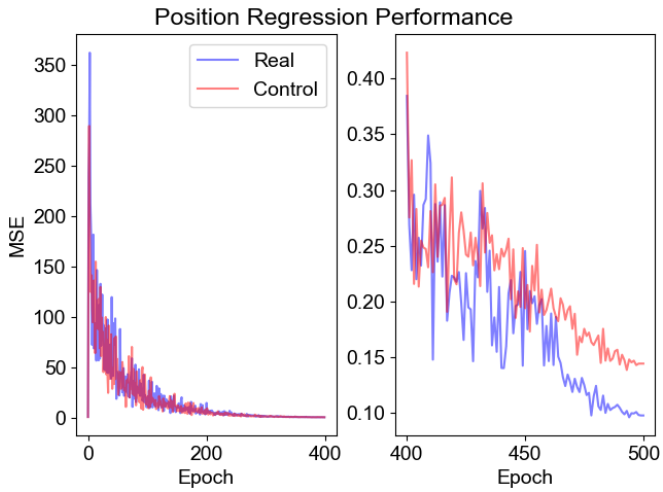

Fig. 5. **Mean Squared Error (MSE)** between the predicted and actual position of the mouse over the training epochs for one dataset. The red line represents a label shuffling control, while the blue line represents the real analysis. The panels display early (left) and late (right) stages of the training.

Finally, in our endeavor to determine whether the RSC utilizes textured patches to rectify its current prediction of position, we assessed the algorithm's error based on position, as demonstrated in Figure 6. If the patches indeed served as an error signal, we would anticipate a gradual deterioration of the correlation between the pattern and position in areas of the track without input, with a swift recovery once the animal encounters a textured patch. Such an effect would also likely manifest in the regression performance, given the algorithm's learning capacity relies on pattern-position consistency across trials. However, our findings did not demonstrate any such impact, as can be observed in Figure 6.

IV. Discussion and Conclusions

The primary novelties of this study are the feasibility of classification with a small network size, implications on the interpretability of machine learning results, and the utilization of raw neural data for these purposes. These outcomes of our research can be viewed from the lens of two prominent disciplines: neuroscience and computer science.

A. Neuroscience implications

A notable outcome of our research is that we were able to classify with such a small network. This finding might be indicative of efficient or relatively simple encoding mechanisms present within the real networks that we are studying. The nature of neural data is highly complex and dense, yet our results suggest that even simplified networks can effectively classify such data, opening new avenues for understanding the underlying principles of neural encoding. Further, the variability in classification results across different datasets speaks to the signal-to-noise ratio (SNR) in raw neural data. Where the classification performance was high, and achieved in a few epochs we can infer that the SNR was higher than in the cases where the classification problem was more difficult.

As for the implications concerning the interpretability of machine learning, our results can only suggest that the RSC contains spatial information. This assertion is due to the positional dependency of all events in our study. A further potential limitation is the correlation of space with time owing to the consistent running behavior across trials, which could lead to a similar interpretability issue.

B. Machine Learning implications

The successful extraction of meaningful information from raw neural data using a small network has significant implications for the field of computer science, particularly in the realm of machine learning and its engineering applications. Such a finding suggests potential for real-time or online analysis of complex data, making high-performing machine learning models more accessible and feasible in a broader range of applications.

However, our study also serves as a reminder about the potential pitfalls concerning the interpretability of machine learning models. If we had not performed the position regression tests, our conclusions about the nature of the data and the performance of the model could have been potentially erroneous. This underscores the critical importance of rigorous testing and careful interpretation in machine learning applications, especially when dealing with multifaceted and complex data such as neural signals.

C. Future research directions

A key direction for future research is the development and application of explainable AI techniques, which prioritize transparency and comprehensibility in machine learning models. This might be accomplished through feature perturbation analysis [16] or demixed principal component analysis (dPCA) [17]. Such methods could provide more granular insights into the mechanics of neural encoding and the behaviors of machine learning models, fostering a more nuanced understanding of these complex systems. They could also be

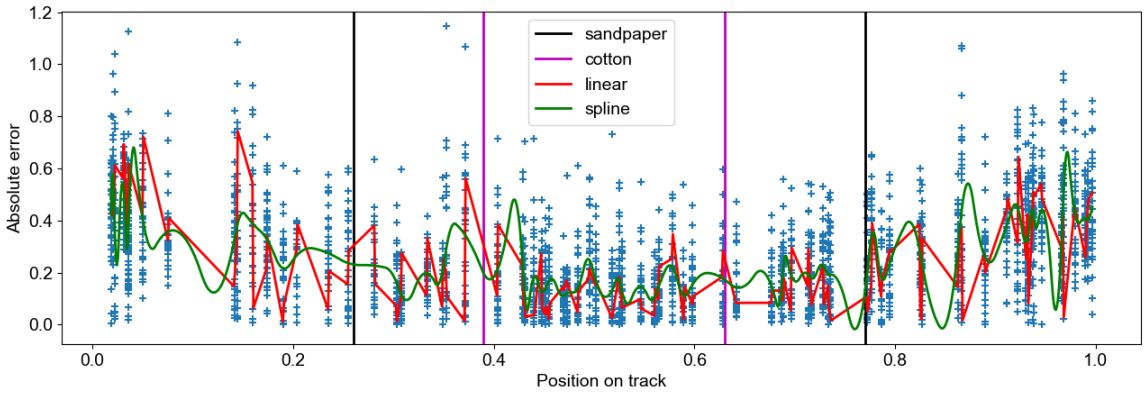

Fig. 6. **Mean Squared Error (MSE)** at each track position, presented as a percentage. The individual error on windows from different trials is represented by the blue crosses. The distribution of these points is due to the selection of non-overlapping and sequential windows, coupled with the animals' consistent speed along the track throughout trials. The vertical lines denote the textured patches' onset, whereas the red and green lines interpolate the potential error function from which these points were derived.

instrumental in advancing our ability to interpret machine learning results correctly, further bridging the gap between neuroscience and computer science.

D. Conclusion

We believe that our study sets a path for future investigations. The intersection between neuroscience and computer science, particularly with regards to machine learning applications, holds immense potential. It is our hope that future research in this realm will continue to bridge the gap between these two disciplines, ultimately leading to more robust, accurate, and interpretable machine learning models that can handle complex neural data and other intricate datasets.

E. Acknowledgments

The research leading to these results has received funding from: NO (Norway) Grants 2014-2021, under Project contract number 20/2020 (RO-NO-2019-0504), two grants from the Romanian National Authority for Scientific Research and Innovation, CNCS-UEFISCDI (codes ERA-NET-FLAG-ERA-ModelDXConsciousness, and ERANET-NEURON-Unscrambly), and a H2020 grant funded by the European Commission (grant agreement 952096, NEUROTWIN).

REFERENCES

[1] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[2] M. Saeidi, W. Karwowski, F. V. Farahani, K. Fiok, R. Taiar, P. Hancock, and A. Al-Juaid, "Neural decoding of eeg signals with machine learning: a systematic review," *Brain Sciences*, vol. 11, no. 11, p. 1525, 2021.

[3] L. B. Reid, R. Cunnington, R. N. Boyd, and S. E. Rose, "Surface-based fmri-driven diffusion tractography in the presence of significant brain pathology: a study linking structure and function in cerebral palsy," *PLoS One*, vol. 11, no. 8, p. e0159540, 2016.

[4] O. AlShorman, M. Masadeh, M. B. B. Heyat, F. Akhtar, H. Almahasneh, G. M. Ashraf, and A. Alexiou, "Frontal lobe real-time eeg analysis using machine learning techniques for mental stress detection," *Journal of Integrative Neuroscience*, vol. 21, no. 1, p. 20, 2022.

[5] T. Zhang, Q. Liao, D. Zhang, C. Zhang, J. Yan, R. Ngetich, J. Zhang, Z. Jin, and L. Li, "Predicting mci to ad conversation using integrated smri and rs-fmri: machine learning and graph theory approach," *Frontiers in Aging Neuroscience*, vol. 13, p. 688926, 2021.

[6] A. M. Miller, L. C. Vedder, L. M. Law, and D. M. Smith, "Cues, context, and long-term memory: the role of the retrosplenial cortex in spatial cognition," *Frontiers in human neuroscience*, vol. 8, p. 586, 2014.

[7] A. S. Mitchell, R. Czajkowski, N. Zhang, K. Jeffery, and A. J. Nelson, "Retrosplenial cortex and its role in spatial cognition," *Brain and neuroscience advances*, vol. 2, p. 2398212818757098, 2018.

[8] G. Cona and C. Scarpazza, "Where is the where in the brain? a meta-analysis of neuroimaging studies on spatial cognition," *Human brain mapping*, vol. 40, no. 6, pp. 1867–1886, 2019.

[9] D. Mao, S. Kandler, B. L. McNaughton, and V. Bonin, "Sparse orthogonal population representation of spatial context in the retrosplenial cortex," *Nature communications*, vol. 8, no. 1, p. 243, 2017.

[10] P.-Y. Jacob, G. Casali, L. Spieser, H. Page, D. Overington, and K. Jeffery, "An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex," *Nature neuroscience*, vol. 20, no. 2, pp. 173–175, 2017.

[11] M. Gianatti, A. C. Garvert, and K. Vervaeke, "Diverse long-range projections convey position information to the retrosplenial cortex." *bioRxiv*, pp. 2022–09, 2022.

[12] E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang *et al.*, "Simultaneous denoising, deconvolution, and demixing of calcium imaging data," *Neuron*, vol. 89, no. 2, pp. 285–299, 2016.

[13] J. Friedrich, P. Zhou, and L. Paninski, "Fast online deconvolution of calcium imaging data," *PLoS computational biology*, vol. 13, no. 3, p. e1005423, 2017.

[14] A. Ciuparu, A. Nagy-Dăbâcan, and R. C. Mureşan, "Soft++, a multi-parametric non-saturating non-linearity that improves convergence in deep neural architectures," *Neurocomputing*, vol. 384, pp. 376–388, 2020.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] H. Bârzan, A.-M. Ichim, V. V. Moca, and R. C. Mureşan, "Time-frequency representations of brain oscillations: which one is better?" *Frontiers in Neuroinformatics*, vol. 16, p. 871904, 2022.

[17] D. Kobak, W. Brendel, C. Constantinidis, C. E. Feierstein, A. Kepecs, Z. F. Mainen, X.-L. Qi, R. Romo, N. Uchida, and C. K. Machens, "Demixed principal component analysis of neural population data," *elife*, vol. 5, p. e10989, 2016.