Master's Programme in Health Informatics
Spring Semester 2023
Degree thesis, 30 Credits

Direct Behaviour Prediction from Miniscope Calcium Imaging Using Convolutional Neural Networks

Author: Konstantinos Kalaitzidis

Author:	Konstantinos Kalaitzidis
Main Supervisor:	Professor Konstantinos Meletis, Department of Neuroscience, Karolinska Institutet
Main Supervisor:	PhD Student Maria Bampa, Department of Computer and Systems Sciences, Stockholm University
Co-supervisor	PhD Student Emil Wärnberg, Department of Neuroscience, Karolinska Institutet
Co-supervisor	Postdoctoral Researcher Ioannis Mantas, Department of Neuroscience, Karolinska Institutet
Examiner:	Professor Panagiotis Papapetrou, Department of Computer and Systems Sciences, Stockholm University

Affirmation

I hereby affirm that this Master thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified; nor has it been published.

Stockholm, May 2023

Konstantinos Kalaitzidis

Direct Behaviour Prediction from Miniscope Calcium Imaging Using Convolutional Neural Networks

Abstract

Background: Neurodegenerative diseases, including Parkinson's, continue to affect millions worldwide, driving neuroscience research to develop effective and personalised treatments. To study and understand the neural circuits in the human brain associated with the emergence and progression of these diseases, researchers are using neuroimaging techniques like calcium Imaging in disease-relevant animal models to link an organism's neural activity with its behaviour.

Aim: Calcium imaging techniques present various limitations for researchers, such as complex-to-use pipelines often limited to extracting cell body information without directly inferring any behavioural correlates. This thesis investigates the potential application of advanced deep learning techniques, such as convolutional neural networks, in improving calcium imaging analysis by reducing pre-processing requirements and directly arriving at behavioural correlations from animal neural activity.

Methods: In this study, previously collected calcium imaging datasets from behavioural assays of freely moving animals are repurposed and used to train a CNN-based tool called the BPNN (Behavioural Prediction Neural Network). Additionally, the performance of the BPNN is compared and evaluated with current state-of-the-art methods applied in neuroscience research.

Results: Several experiments were performed to evaluate the BPNN's capacity to predict behaviour compared to current methods. However, issues related to overfitting arose, which may have been caused by technical discrepancies or other biological artefacts produced during the calcium imaging recordings sessions. Despite this, the BPNN produced similar or better results in predicting animal behaviour, with an F1-score of 0.56 compared to the F1-score of 0.41 of an existing calcium imaging analysis tool concerning the same biological task.

Conclusion: The best-performing configuration of the BPNN model demonstrated a limited yet plausible ability to establish links between neural activity and specific animal behaviours, indicating the potential applicability of CNNs in behaviour prediction assignments. However, further research is required to address current technical and biological limitations to reaffirm the postulations of this study.

***Keywords**: Convolutional Neural Networks, Deep Learning, Calcium Imaging, Neuroinformatics*

Acknowledgements

First and foremost, thank you, Professor ***Konstantinos Meletis****, for the mindful support and rich opportunities to delve into extremely interesting research! It has been a cherished experience abundant in profound perspectives that I will carry close to me throughout my academic journey.*

My deepest gratitude to my co-supervisor ***Emil Wärnberg****, whose continuous support and oversight have made this work possible. Engaging with you has been joyful, thought-provoking, and intellectually illuminating.*

Next, I'm grateful to my main supervisor ***Maria Bampa****, for showing enthusiasm for my research topic and for providing me with valuable feedback when requested. Your feedback and assistance have been monumental for the development of this thesis!*

I also want to thank my co-supervisor ***Ioannis Mantas****, who has influenced me in areas beyond academia, such as fitness and living a healthy lifestyle. Every time I interact with you, I become richer.*

Yet, my gratitude extends further to ***Zacharias Wik, Vasiliki Skara, Moritz Weglage, Katharina Heining, Pierre Le Merre, Allessandro Contestabile****, and the entire* ***DMC Lab*** *for all the beautiful moments and meaningful conversations we've had over this time.*

Finally, I would like to devote this thesis to my ***family*** *as a token of gratitude for all the love and support they've given me to chase my dreams.*

Table of Contents

List of Abbreviations

1p	One-photon
2p	Two-photon
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
BG	Basal Ganglia
BPNN	Behaviour Prediction Neural Network
Ca^{2+}	Calcium Ions
CI	Calcium Imaging
CNMF	Constrained Non-negative Matrix Factorization
CNNs	Convolutional Neural Networks
DL	Deep Learning
DLC	Deep Lab Cut
FOV	Field of View
GCaMP	Genetically encoded calcium indicators
HI	Health Informatics
ICA	Independent Component Analysis
L-DOPA:	Levodopa
LSTM	Long-short Term Memory
MeSH	Medical Subject Headings
ML	Machine Learning
NNs	Neural Networks
PCA	Principal Component Analysis
PD	Parkinson’s disease
ResNet	Residual Network
ROI	Region of Interest
SNc	Substantia Nigra Compacta
SVMs	Support Vector Machines

List of Figures

List of Tables

1 Introduction

One of the biggest challenges in healthcare today is finding effective treatments for diseases that continue to affect millions of people worldwide. In neuroscience, researchers focus on understanding how the brain functions and affects behaviour in living organisms. They are doing this by mapping behaviour with neural activity in disease-relevant animal models during behavioural assays allowing them to study the brain in greater detail and develop potential therapies or diagnostic mechanisms.

In addition, recent advances in computer science are helping researchers analyse large-scale neuronal information using machine learning. The combination of brain research and machine learning push the boundaries of our understanding of the human brain with unprecedented potential and enables the development of new diagnosis tools and personalised therapies for neurodegenerative disorders, namely Parkinson's and Huntington's disease.

This thesis attempts to analyse neuronal activity captured during behavioural paradigms of mice by utilising the latest advancements in *deep learning (DL)*, like *convolutional neural networks (CNNs)*, to examine the extent to which direct animal behaviour prediction from neuro-imaging recordings is possible without following the standard and labour-intensive pre-processing steps of current state-of-the-art tools.

1.1 Motivation

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disease of unknown cause, characterised by a degeneration of dopamine-producing neurons in the *substantia nigra compact* (SNc)**,** a region in the *basal ganglia* (BG) that affects movement control, emotional regulation, and habit formation (1). PD was first clinically described by *James Parkinson* in 1817 (2) in the following terms:

"Involuntary tremulous motion, with lessened muscular power, in parts not in action and even when supported; with a propensity to bend the trunk forward, and to pass from a walking to a running pace: the senses and intellects being" (2). See Fig. 8.1 for PD illustration.

Tremendous progress was made in Parkinson's disease research during the second half of the 20th century. One of the most significant milestones occurred in the late 1950s through the mid-1960s when researchers discovered that decreased concentrations of the *dopamine* neurotransmitter in the SNc were prevalent in PD patients (3,4). This discovery led to the development of the first clinical trials of *levodopa* (L-DOPA), a drug that increases dopamine levels in the brain and helps alleviate PD symptoms, namely tremors, rigidity, and bradykinesia (5).

Today, PD is the second most common neurodegenerative disorder globally, affecting over 10 million people and one of the leading causes of motor impairment (6). In addition, recent studies argue that PD cases in the future will increase, potentially placing additional strain on the healthcare system (7). Despite extensive research efforts aimed at identifying the cause of PD and developing a cure, the underlying aetiology of the disease remains elusive (8). Moreover, the clinical diagnosis of PD is based on the presence of the disease's main motor symptoms; hence, finding biomarkers to identify PD in the *prodromal stage* (the disease stage before the emergence of motor symptoms) may enable personalised dopamine replacement therapies to initiate earlier (8–10) thus improving the patient's quality of life during the progression of the disease.

The prodromal or preclinical stage of PD is characterised by the onset of non-motor symptoms such as REM-sleep behaviour disorder, constipation, and hyposmia (11). To discover the underlying mechanisms of non-motor dysfunction in PD in the prodromal stage, animal models are developed to mimic the presence of the disease, monitor how it develops over time, and perform a variety of behavioural experiments to explore the role of specific neural circuits and their connection to PD (12). More specifically, neural circuits in the brain that regulate movement or cognitive behaviour, like the basal ganglia, are studied to investigate their role in shaping motor and cognitive behaviour within the context of PD (1). To visualise and interpret the activity of these neuron populations during behavioural experiments with mice, neuro-imaging techniques such as in-vivo calcium imaging are adopted. *Calcium imaging* (CI) works by mounting a miniature microscope, known as a *miniscope*, inside the mouse's brain to record fluorescence changes in cell activity while the mouse engages in behavioural tasks (13). After the experiment recording, a video depicting cell activity is analysed with calcium imaging analysis pipelines. The aim is to extract cell activity traces (or calcium traces) and associate them with the animal behaviour observed in the experiment (more on this neuroimaging technique in Sec. 2.1). However, as discussed by *Pnevmatikakis et al*. (14), CI analysis comes with several limitations that may hinder research efforts in interpreting relations between neural activity and behaviour. Such limitations derive from the miniscope's low spatial resolution, making it difficult to identify neurons, differentiate neurons that may be spatially overlapping, and decode neuronal activity from the captured calcium fluorescence (15).

Furthermore, as described in a study by *Dong Z et al*. (16), detecting individual neurons from video recordings using calcium trace extractions computationally expensive, particularly when large-scale recordings are considered. It requires a series of pre-processing steps performed by specialised imaging algorithms for denoising and motion-correcting the video recordings. Another bottleneck is the number of complicated parameters that need to be adjusted before processing, making it challenging for non-expert researchers in CI analysis to use these pipelines.

Over the recent decades, *Artificial Intelligence* (AI) has rapidly evolved, driven by computer science and advancements in mathematics. One of the key areas in AI is *Deep Learning* (DL), a subfield of *Machine Learning* (ML) which is concerned with a class of algorithms called *Artificial Neural Networks* (ANNs) modelled after the way nerve cells in the human brain process information (17). ANNs use multiple processing layers of interconnected nodes, called *artificial neurons* containing weighted values that allow for the transmission and processing of information through a network to perform overly complex non-linear computations like image and speech recognition, natural language processing, and more (17). As a result, ANNs have found widespread adoption in many domains, especially in the medical field, including clinical diagnosis, cancer prediction (18), image analysis and interpretation as well as drug development (19), to name a few.

Currently, studies have applied ML methods to improve CI analysis. However, the emphasis has been placed on improving individual cell segmentation (the process of identifying individual neurons) from raw CI recordings instead of mapping them directly to behaviour classifications (the process of correlating neural activity to animal behaviour) (14,16,20–29). On the other hand, *Convolutional Neural Networks* (CNNs), a type of deep learning algorithm, are particularly suitable for image analysis tasks (30,31) and could potentially be utilised to map large-scale CI recordings to behavioural classifications, thus assisting researchers in interpreting population-level (multiple neurons) dynamics. Moreover, after a CNN model is trained, optimised, and evaluated, neural activity patterns can be associated with specific behaviours of the animal, thereby enhancing the capacity for animal behaviour prediction without the parameter-intensive and multi-step process of current calcium imaging analysis pipelines (14,16,24,25,30). However, it is important to note that for a CNN model to produce reliable results, large amounts of data are required for training the model and avoiding overfitting (32).

One more advantage of using CNNs for this application is that the neuronal fluorescence from the recording's entire field of view (FOV) is utilised. Pixel values from each video frame encompass important neural information on cells, dendrites, and axons, whereas current methods are mostly limited to extracting cell body traces. More information extracted from each experiment leads to an ethical advantage compared to current methods, as calcium imaging recordings on animals are an invasive procedure that requires a surgical operation on the mouse’s head and consequently the animal’s sacrifice when the experiments are concluded.

This incremental improvement in extracting neuronal activity during behavioural tasks may enhance the potential discovery of new biomarkers that can lead to alternative therapies and the timely detection of various brain diseases.

1.2 Problem Description and Knowledge Gap

The central aim of systems neuroscience research is to link an organism's neural activity with its behaviour (33). Many tools and methods have been developed to enhance our understanding of how specific neuronal populations function and relate to observed behaviour in experiments. *Moore's Law* states that computer processing speeds double approximately every two years, highlighting the exponential increase in computational power (34). Regardless of this increase in power, the amount of data generated from neural recordings is continuously growing at a significant rate and may soon reach a scale of thousands or even millions of recording channels significantly (35).

The study of *Trautmann et al.* (36) addresses the challenge of the demanding data analysis required in a neuronal analysis task called *spike sorting* by extracting the firing rates of single neurons from extracellular recordings (37). Essentially, spike sorting is a pattern recognition problem (38) for analysing and classifying the action potentials (or electrical signals) generated from neurons in the brain. For interpretations of population-level neural activity, the authors propose an alternative approach that skips the labour-intensive pre-processing steps of spike sorting by applying a machine learning technique called *dimensionality reduction* on multiple neurons. This approach has been shown to produce comparable results without the computational overhead and complexity of the pre-processing steps applied in the original method (36). In this spirit, we postulated whether a similar application could extract population-level neural activity from the main brain circuit of interest for PD, the basal ganglia (39). To study dysfunction in the BG, behavioural paradigms of animal models are developed in controlled laboratory settings to simulate the presence of PD symptomatology and underlying functional mechanisms. However, how can we monitor the activity of neurons in living organisms while they participate in these behavioural tests?

Calcium imaging is a standard tool of choice for studying neural circuit dysfunctions in the BG as it allows for the recording of large neuron populations of interest *in vivo* (in a living organism) and over time (weeks or months) (40). Sec. 2.1 describes in detail how calcium imaging is employed for analysing neuronal data.

Several implementations of in vivo calcium imaging recordings require ML-based methods such as *principal component analysis* (PCA), *independent component analysis* (ICA), and a *constrained non-negative matrix factorisation* (CNMF) algorithm, to name a few, for reliably extracting neuronal information from behavioural experiments (29). The process of extracting neuronal information from calcium imaging recordings is called *calcium signal or calcium trace extraction* (41). It enables the acquisition and analysis of neurons based on their intracellular calcium concentration. High calcium concentrations indicate that neurons are activated.

Among the implementations in the field of calcium imaging analysis, there have been more approaches (26–28); however, the complexity of using these pipelines remains a bottleneck for most research teams. For example, each of the aforementioned algorithms requires heavy parameter tuning by trained experts in complicated software environments, limiting less-trained neuroscientists in calcium signal extraction from using the tools.

Soltanian-Zadeh S et al. (30) implemented a new pipeline that used CNNs to segment neurons from calcium imaging recordings by eliminating any manual step in cell segmentation as required by other pipelines, although no behaviour prediction is conducted. Despite this, the mentioned study is compelling evidence that CNNs have potential implementations in calcium imaging analysis tasks. Another interesting approach is that of *Etter et. al* (42), where they use a *Bayesian classifier* to infer behaviour from calcium imaging recordings. However, no neural networks, for instance, CNNs, were adopted in this approach. Moreover, it would be interesting to explore, with the recent advancements in DL, how neural networks can be harnessed in the task of inferring behaviour predictions from CI recordings.

In summary, to the best of our knowledge, limited research has been pursued in applying CNNs to infer behaviour from calcium imaging recordings. This creates a potential knowledge gap in developing a tool that can behaviour correlations among behaviour and acquired neural activity without the need for manual pre-processing intervention, which is currently common in alternative calcium imaging analysis methods.

1.3 Aims and Objectives

The *aim* of this thesis is to apply CNNs as an alternative method for population-level neural activity analysis of calcium imaging recordings, bypassing the current trace extraction steps of state-of-the-art methods. The overarching goal is to provide animal behaviour interpretations faster, directly, and while extracting more neuronal information than before, which may lead to an increased understanding of PD progression, diagnosis, and drug development. To reach this aim, the following *research objectives* were formulated:

- Pre-processing of raw calcium imaging recordings acquired from a cognitive task experiment conducted at the K. Meletis Group.
- Train a CNN model with calcium imaging recordings divided into training and validation parts.
- Optimise hyperparameters like the number of batch sizes, epochs, and dropout layers.
- Validate the model on new, unseen, before calcium imaging frames.
- Compare the model with current calcium imaging analysis tools.

1.4 Research Questions

The following *research questions* are formulated to address the *aim* and *objectives*:

1) *To what extent is direct behaviour prediction in mice possible from in-vivo calcium imaging videos by omitting the step of calcium trace extraction and using deep learning methods such as CNNs?*
2) *To what extent can we maximise the amount of neuronal information acquired from calcium imaging recordings to gain a better understanding of neural activity in brain circuits of interest?*

1.5 Limitations

The performance of a CNN model heavily depends on the amount and quality of the data used to train it. This study is limited to data obtained from calcium imaging recordings from one mice experiment. While recordings were acquired from multiple mice, it would be beneficial to validate the model using a collection of different behavioural assays and animal models.

Moreover, this study focuses on animal behaviour prediction based on the population-level activity of neurons obtained from calcium imaging recordings. Unlike current state-of-the-art calcium imaging analysis tools, this approach does not perform calcium trace extraction (14,16,25) and, as a result, is not ideal for research questions that explicitly require single-cell traces (identification of individual cells).

An important caveat to be addressed here is that the CNN model can only be trained and tested on calcium imaging recordings obtained from the same animal as the neuronal formations of each organism are unique, making it difficult to accurately predict behaviour across different animals. This raises the risk of overfitting the model, a phenomenon that occurs when an ML model is not trained with enough data (43). Despite this, it would be interesting to explore whether the model can be generalised and perform predictions across different animals.

Furthermore, this study is undertaken with limited time availability. It is also important to note that the author of this study is not a neuroscientist by training, so their conclusions may be subject to differing opinions. Nevertheless, this study has been supervised by neuroscientists, computational neuroscientists at the *K. Meletis Group* at the *Department of Neuroscience, Karolinska Institutet* and data scientists from the *Department of Computer and Systems Science, Stockholm University*, who have provided the author with guidance and support throughout the development of this thesis.

1.6 Relevance to Health Informatics

Health Informatics (HI) is an interprofessional field that examines how technology and big data can improve the quality of healthcare (44). In the context of HI, this study leverages recent advancements in deep learning, such as CNNs, the deep learning method with the most significant impact in HI (45), to develop an alternative calcium imaging analysis tool that may potentially be used for aiding medical research in identifying biomarkers related to the diagnosis and progression of neurodegenerative diseases such as PD. The scope of this study and its possible future extensions may become relevant in assisting the development of personalised treatments and interventions beyond Parkinson's disease to various other neurological and psychiatric conditions.

1.7 Structure of the Thesis

Chapter 2 presents an array of introductory terminology and scientific concepts necessary to comprehend this thesis. The reader is presented with a gentle introduction in calcium imaging, machine learning, deep learning, and an overview of related work in the research area. In Chapter 3, we outlay the methodology employed during the development of this thesis, including the methodology strategy and the technical trademarks embodying the proposed analysis tool. In Chapter 4, we render the results of the experimental procedures based on the predefined research questions while in Chapter 5, we discuss the outcomes of our research, including limitations, expectations, and future research trajectories. Finally, Chapter 6 provides a summary of the findings.

2 Extended Background

This Chapter provides background information on the relevant discipline areas and terminologies discussed in this thesis. Sec 2.1 focuses on explaining the basic concepts of calcium imaging. Sec 2.2 and Sec. 2.3 discuss the fields of machine learning, deep learning, and the applicability of CNNs in computer vision tasks. Sec. 2.4 provides the reader with an overview of the literature search strategy for finding potential knowledge gaps and discusses related work of CNNs applied in CI analysis. Lastly, in Sec 2.5, we discuss the significance of this thesis work compared to current solutions in calcium imaging analysis.

2.1 Calcium Imaging

Imaging properties of the brain are of great interest to neuroscientists as it allows for studying and understanding brain function. A range of techniques with different serving purposes is available for brain imaging, such as *electroencephalography* (EEG), *positron emission tomography* (PET), *magnetic resonance imaging* (MRI), and *functional magnetic resonance imaging* (fMRI) (46). However, none of these methods allows for monitoring single-cell or population-level neuronal dynamics (46).

Instead, one-photon and two-photon CI techniques have opened new frontiers for exploring deeper areas of the brain. This Section discusses the concept of calcium imaging and the common pre-processing steps found in calcium imaging analysis pipelines. We aim to give the reader an overview of the latest tools and their limitations, highlighting the need for a more effective approach to interpreting neural dynamics at the population level.

2.1.1 Calcium Indicators

How do we image brain activity in living animals? The answer lies in understanding the basic functionality of *genetically encoded calcium indicators* (GCaMP). Calcium indicators allow us to capture the neural activity of cell populations of interest in living animals that engage in various behaviour tasks over extended periods of time (47), however, delving deeply into the biological inner workings of calcium imaging is beyond the scope of this thesis.

Calcium ions (Ca^{2+}) are crucial for regulating neural activity. High intracellular calcium concentrations in neurons trigger electrical charges and the release of neurotransmitters. This biochemical process can be exploited as a proxy to indicate neural activity in animals (48). To image these electrical charges in neuron populations of interest under a microscope, protein indicators like GCaMP are injected into a living organism (see Fig. 2.1) and through a process called *DNA expression*, proteins are generated within the cells. When proteins bind to calcium ions, fluorescence is emitted, and with the use of a fluorescence microscope, researchers can visualise and track the fluorescence (48) or the *glowing* effect observed in activated neurons. As a result, researchers can visualise neuronal dynamics in living organisms and track their fluctuations over time during behavioural experiments (49).

2.1.2 Calcium Imaging Analysis Workflows

Two-photon (2p) calcium imaging is a high-resolution neuroimaging method for visualising neural activity in vivo. It captures three-dimensional images of subcellular dynamics that can reach $1mm$ in brain tissue (50). In comparison, one-photon (1p) calcium imaging has depth limitations and a lower resolution compared to 2p methods. However, 1p calcium imaging has the luxury of monitoring calcium dynamics in freely moving animals, while 2p approaches can only accommodate behavioural tasks where the animal's head is in a fixated position. In addition, 2p techniques, due to the nature of the head fixation requirement, have been reported to induce stress in animals and limit the variability of naturalistic experiments that can be conducted to study behavioural paradigms (51).

Although 1p CI may be more suitable for specific behaviour tasks, analysing CI data presents significant challenges that haven't been thoroughly addressed today. For example, the ever-growing amount of neuronal information generated from experiments makes it cumbersome, computationally expensive, and time-consuming for researchers to analyse. New analysis methods that can be operated reasonably fast and conveniently on regular computers instead of specialized high-performance equipment are needed (52). Alternative methods have been recently developed to compensate for these challenges with end-to-end open-source pipelines for CI analysis (14,16,25,53).

In Fig. 2.2, we illustrate a general overview of the main pre-processing steps undertaken in a CI analysis pipeline. Typically, *motion correction* is applied to minimise motion artefacts induced in recordings due to animal movement. Following this step, manual annotation from a trained expert or an algorithm is utilised for extracting spatiotemporal information of cell fluorescence traces and demixing possibly overlapping cells. The last step of CI analysis involves the deconvolution of the captured fluorescence traces of calcium signal activity (52). Steps 3 and 4, as portrayed in the Figure, involve the step of *calcium trace extraction* described in Sec. 1.2.

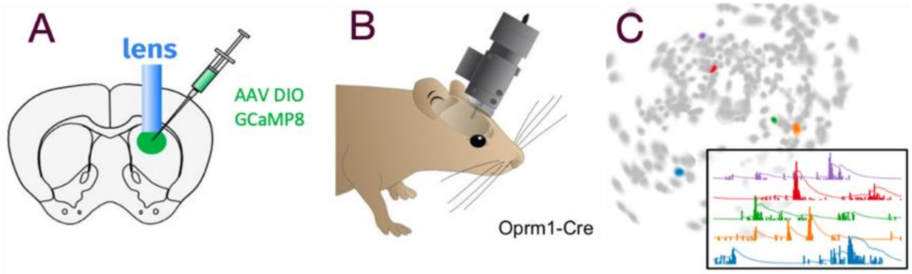


Figure 2.1 | Calcium imaging with 1p miniscopes in transgenic mice.

Note | The process of acquiring neural activity with a calcium imaging technique:

A. An AAV DIO GCaMP8 virus is injected into a transgenic Cre driver mouse for expressing the GCaMP8 fluorescent indicator.
B. A mouse with an integrated miniscope for recording neuronal activity from neurons expressing the GCaMP protein.
C. Field of view (FOV) from a miniscope recording showing activated neurons.

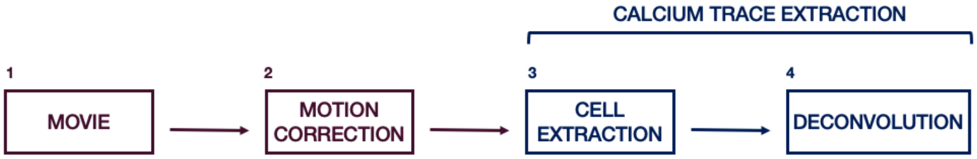


Figure 2.2 | A typical calcium imaging analysis pipeline.

Note | Adapted from Giovannucci et al., 2019 (52)

2.1.3 Recording Neural Activity in Behavioural Assays

Animal models are typically developed to simulate disease presence and compared to control animal groups to make plausible cause-and-effect correlations. They are appropriately conditioned for the corresponding behavioural experiment and recorded with neuroimaging techniques like calcium imaging to investigate factors that govern their behaviour under specific research requirements. This Section presents the *Arrow Maze* task from which calcium imaging recordings were retrieved to train the CNN tool, introduced in Chapter 3.

2.1.3.1 The “Arrow Maze” Task

In this thesis, we trained a CNN model using calcium imaging data collected from the *Arrow Maze* behavioural experiment. Throughout this experiment, water-restricted mice are recorded with one-photon (1p) CI. As described in Sec 2.1.1 and 2.1.2, a miniaturised microscope (i.e., “*miniscope*”) is mounted to the mouse’s head, allowing for the recording of fluorescence fluctuations for the duration of the task (see Fig. 2.1). Once the mice are prepared for the task, they're introduced into a maze of three corridors with waterspouts from which they receive water as a reward (illustrated in Fig. 2.3).

With this behavioural task, the goal is to monitor the activity of neuronal populations of interest in the basal ganglia and investigate their role in goal-directed behaviour and decision-making.

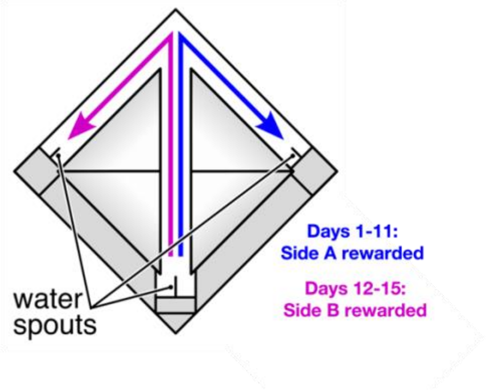


Figure 2.3 | Schematic of the Arrow Maze task.

Note | The experiment is parted into two stages as described in Weglage, 2022 (66):

- **Side A reward from days 1-11:** The mice must travel back and forth from the **initiation spout** located in the base of the main corridor to **Spout A** to obtain their reward. If **Spout B** was approached instead, no reward is attributed.
- **Side B reward from days 12-15 (Reversal):** The mice must travel from the **initiation spout** located in the base of the main corridor to **Spout B** to obtain their reward. If **Spout A** was approached instead, no reward has attributed.
- The duration of each experimental session (one per day for a total of 15 days) is about 20 minutes long.

The output of this experiment is a large amount of mouse behavioural data such as turning, running, stopping, rearing, and grooming. Also, tracking data of X and Y location coordinates are collected, as well as the calcium imaging recordings derived from the miniscope during the experiment.

2.1.3.2 Producing Behaviour Correlations from Calcium Imaging Recordings

In the previous section, we discussed how we could observe neural activity in mice during behavioural assays. The next challenge is to conduct behaviour correlations based on the recorded neural activity. In this thesis, we have leveraged behavioural labels generated from the *Arrow Maze* sessions, used to annotate the calcium imaging frames. The methods to generate these labels are as follows:

1. Tracking the mouse body (e.g., the base of the tail, and the head) using the *DeepLabCut* (DLC), a behavioural tracking tool for markerless pose estimation based on a transfer learning approach and deep neural networks (54).

2. Filtering the tracking data from the DLC using a particle filter algorithm ensures it is smooth and corrects outliers from tracking failures.
3. Use *a Hidden Markov Model* (55) with hand-coded transition and emission likelihoods to match the arrow-maze task's behavioural stages.

In Sec. 3.3.1, we explain how we use the behaviour labels obtained from this experiment, pre-process them and feed them as input to train the CNN model for making behavioural correlations of the mouse during task engagement. To understand this better, it is important to provide some background on deep learning and how convolutional neural networks can simplify pre-processing requirements in calcium imaging analysis, enabling us to perform behavioural correlations.

2.2 Machine Learning

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) and involves the development of algorithms or statistical models that are particularly successful in learning from raw data by extracting patterns or features to make predictions on new, unseen data (31). One of the numerous examples of ML algorithms is *logistic regression*, which can be used to predict whether a customer chooses to buy a particular product or not (31). Another example of a simple ML algorithm is *naives Bayes*, which can separate email into *spam* or *not spam* categories (56). Nowadays, ML algorithms are harnessed in various areas of society, such as web searches, content recommendations, in software for cameras as well as smartphones. (57). Many ML algorithms have different use cases depending on the problem they are meant to solve. However, ML algorithms are mainly categorised into these three distinct categories: *unsupervised learning*, *reinforcement learning*, and *supervised learning*.

Unsupervised learning **(UL)** algorithms are good at finding patterns and structures from raw input data without predefined labels or annotations by humans (56). For example, some unsupervised learning algorithms can group similar data points together, discovering underlying structures or patterns in the data, such as clusters (58).

Another form of deep learning is ***reinforcement learning*** **(RL),** which involves an algorithm that can learn from trial-and-error interactions with an environment without any intervention from an external operator (56). In reinforcement learning, the goal is to learn a policy that maximises the cumulative reward over time by selecting the best action at each step. The model must determine which actions to take and test over time to evaluate its performance (59).

Continuing, one of the most common forms of machine learning algorithms is ***supervised learning*** **(SL),** which involves training a model on a large set of similar data, such as images or text, where each training example is paired with a label or *target*. For example, a frame portraying a number of fluorescent neurons in a calcium imaging video consisting of an $\boldsymbol{X}$ number of frames can be associated with a vector of a numerically defined mouse behaviour label $\boldsymbol{y}$ such as *moving forward*, *turning left*, *turning right*, with $\boldsymbol{yi}$ being the label corresponding to a training example of an image $\boldsymbol{i}$ in the set of frames **X** (see Fig. 3.1 for an example of a CI frame). During training, the model learns to classify the images in the dataset according to their corresponding labels. A *loss function* measures the computation error between the predicted and actual labels, indicating how successful the model is in correctly classifying the image. Therefore, based on this likelihood, the model adjusts its parameters or *weights* accordingly to reduce the output error as much as possible. After the training is completed, the model is validated on new, unseen data to define how well it generalises (60).

2.2.1 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are a classification technique based on the statistical learning theory. In SVMs, a hyperplane divides the dataset into different classes by separating the data points with the maximum possible distance between them to minimise the risk of incorrect classifications of training and test examples (61). In Fig. 2.4, we can observe how the hyperplane separates the data points into two different classes. Even though SVMs minimise the risk of *overfitting* and require fewer training examples for class classifications, they are still considered a *black box model* as they do not reveal which training examples were used to learn for making predictions (61).

In this thesis, an SVM-based model is implemented to make behaviour predictions from extracted cell traces of calcium imaging videos to mimic the current processes in calcium imaging analysis, such as the one exemplified in Fig. 2.2. The goal of utilising the SVM model is to compare it with the proposed CNN tool in this thesis.

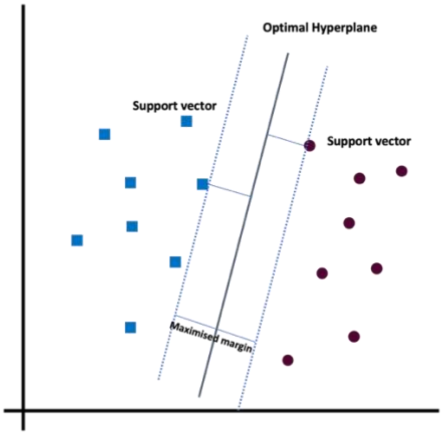


Figure 2.4 | A visualisation of an optimal hyperplane with an SVM algorithm.

2.3 Deep Learning

In the previous sections, we discussed ML and its distinct categories: *unsupervised*, *reinforcement*, and *supervised learning*. We also provided an overview of supervised learning algorithms, such as *support vector machines*, which are mentioned throughout this work. These ML algorithms have found applications in many areas of modern society, such as web searches, content filtering, personalised advertising, and more (60). More advanced ML systems can perform computer vision tasks like object recognition and image-to-text subscription. However, their ability to process raw data could be improved, as they require elaborate feature engineering to extract meaningful data from which ML models could train to make reliable predictions. For example, developing a model trained on the pixel values of an image requires creating an accustomed feature extractor that transforms the original data into useful representations from which the classifier can detect patterns. This bottleneck in conventional machine learning is called a *representation learning problem* (60).

Deep learning is a subfield of ML that utilises multiple-layer neural networks. It has gained popularity over the past decade for its ability to process and extract patterns from raw, high-dimensional data to solve previously challenging tasks. While traditional machine learning algorithms help identify patterns or objects in images, speech, or recommendation systems, they often require careful feature engineering and are challenged when extracting relevant features from raw data (62). For example, a deep learning model usually consists of multiple non-linear modules where each transforms the representation from raw input to a more complex representation, making it easier for the classifier to discriminate variations in the input data (60). In Fig. 2.7 we demonstrate a classic example of how an image is fed into a DL model as an array of pixel values and how the extracted features from the first layer led to further feature extractions from the next layer.

2.3.1 Artificial Neural Networks (ANNs)

In deep learning, *artificial neural networks* (ANNs or NNs) are a type of computational model inspired by the biological neural networks of the human brain (31). An ANN is a network-like structure composed of several layers (an *input layer*, *hidden layers*, and an *output layer*) of interconnected processing units called *neurons* or *nodes*. These neurons are connected to other neurons in the network with a specific *weight* value and perform computations to identify patterns in the input data, usually a set of images. During the training phase, the connections, or *synapses*, between the neurons are modulated by these weight values and adjusted based on the difference between the model's predicted and actual output (*error signal*). This process is called *backpropagation* (see Fig. 2.5) and increases the model's performance for a given task, such as identifying a particular object like a cat or a dog, by propagating backwards the error signal through the network to adjust the weights (31).

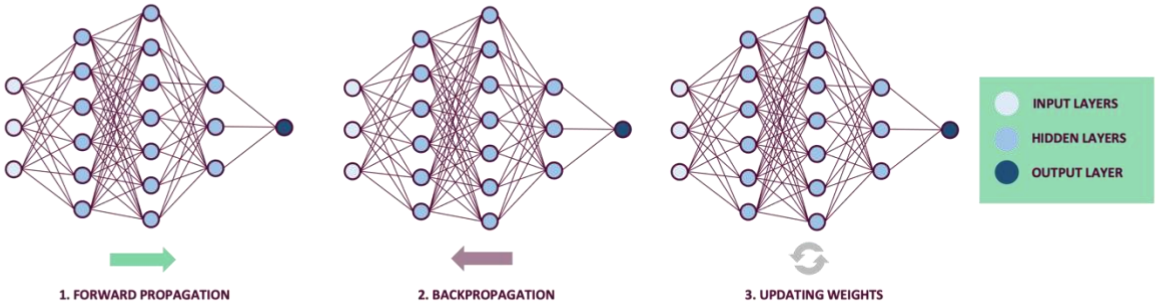


Figure 2.5 | Information propagating through the layers of a Neural Network.

Note | Adapted from (76). The process of information propagation in a Neural Network:

1. A batch of training data is forward propagated to compute the error signal.
2. The error signal is backpropagated through the network.
3. The weights of the nodes are updated with respect to the error signal.

2.3.1.1 Training a Neural Network

Non-linear transformations of non-linearities in the data help the model discover relationships between the inputs and outputs of the neurons in the network. The process of these non-linear transformations is a mathematical function called an *activation function* (31). A popular example used in CNNs is the *rectified linear unit* (ReLU) (see Fig. 2.6) that is mathematically represented as $\mathrm{f}(\mathrm{z}) = \max(0, \mathrm{z})$ (31) where $\boldsymbol{z}$ is the input data for the ReLU function. More precisely, when a value above $\mathbf{0}$ is passed into the ReLU function, it remains the same; otherwise, if it is negative, it is converted to $\mathbf{0}$, making it easier for the network to identify and extract patterns from the data. Afterwards, this output is fed as input to the neurons in the following network layer, and the process is repeated (31).

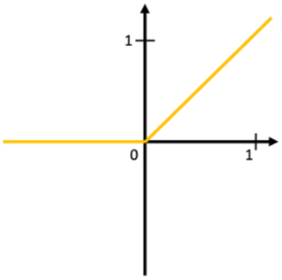


Figure 2.6 | Schematic of a ReLU function.

Other common activation functions include *sigmoid* and *softmax.* In brief, the sigmoid activation is typically used in *binary classification tasks* and maps the input data $\boldsymbol{z}$ to an output value between $\mathbf{0}$ and $\mathbf{1}$ (31). The softmax activation function is mainly used at the output layer of the neural network for *multi-class classification tasks* to represent the prediction probability of an $\boldsymbol{n}$ number of classes (31).

Before the training phase is initiated, the input data is partitioned into training and validation sets. The training set is usually larger than the validation set (for example, split to 80% training - 20% validation sets). During the training phase, the network's weights are adjusted to minimise the difference between the predicted and actual output using an optimisation algorithm called *gradient descent* (63).

Once the model is trained on the training set, its performance is evaluated on the validation set by checking how well it makes correct predictions from unseen data. However, this 80-20 splitting arrangement, the model may have limited validation data points to learn from, which may result in poor model performance due to a phenomenon called *overfitting.* Overfitting occurs when the difference between the training error in the training dataset and the validation error in the validation dataset is significant. For example, the model becomes too good at learning from the training data but struggles to learn from the validation data (31). On the contrary, *underfitting,* is the phenomenon where the model lacks the capacity to learn from the training data (64).

Early stopping is a technique used to decrease the difference between the training and validation errors by stopping the optimisation algorithm when the validation error has not improved after a certain amount of time (31). In addition, a dataset-splitting technique called *cross-validation* (or *k-fold cross-validation*) can also be applied to improve performance. Here, $\boldsymbol{k}$ is the number of non-overlapping subsets (or *folds*) of the input data $\boldsymbol{z}$ and $\boldsymbol{k}-\mathbf{1}$ folds are used for training the model, with the remaining fold used for validation. This process is repeated several times depending on the number of $\boldsymbol{k}$ folds, and afterwards, the performance is averaged across all trials (31).

When training a neural network, several other components are crucial in ensuring that the network learns well on the training data and can generalise similarly well on the validation data:

1. Epochs: Refers to a single pass the model completes throughout the entire dataset to update its weights (31).

2. Batch size: Represents the number of training examples used in one iteration before updating the model's weights. It is essential to decide on a correct batch size number, as a small number limits the variations in the data from which the model can identify patterns. In contrast, a larger batch size number may result in significant computational processing costs (31).

3. Loss function: A loss function $\mathbf{L}$ is used to evaluate how close the predicted output $\boldsymbol{y}$ is to the actual output $\boldsymbol{z}$. In this thesis work, we use *categorical cross-entropy loss* (65), which is used for making multi-class classifications of two or more labels, for example, *turning left*, *turning right*, *and moving forward*.
4. Learning rate: It is the pace at which the weights in the network are updated. The most popular method, and the one employed in this study, is the *Adam optimiser* (66).
5. Dropout: A regularisation technique prevents overfitting by randomly removing neurons and their connections from the network during training (67).

Overall, ANNs are a powerful tool for solving complex problems in many areas. However, they have several limitations. First, their black-box nature makes it challenging to understand how they make their predictions, which is an issue in cases where transparency and explainability are essential. In addition, they are prone to overfitting by learning too much from the training data and generalising poorly on the new, unseen data (31). Lastly, they also require substantial amounts of labelled data to train effectively, which can be laborious and time-consuming (31).

2.3.2 Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs, or ConvNets) are a type of deep learning algorithm widely used in computer vision tasks, such as *object recognition* and *image classification* (60). Unlike other machine learning algorithms, such as *logistic regression*, and *support vector machines*, CNNs can process array-type data, such as pixel values from a 2D image, and extract high-level features. In other words, they can automatically learn and identify patterns and features from an image without explicit feature engineering as in other machine learning algorithms. A CNN is comprised of a series of layers, including the *input layer*, *convolutional layers*, *pooling layers*, and *fully connected layers*. The architecture of a typical CNN is shown in Fig. 2.7.

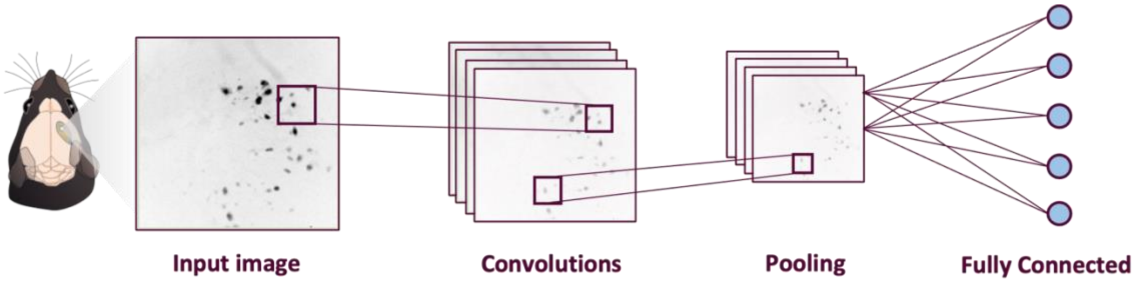


Figure 2.7 | A simple overview of the different types of layers in a Convolutional Network.

Note | This Figure showcases one frame from a 1-photon calcium imaging recording as the input image to the CNN model (adapted from (76)).

The input layer receives the input data, which can be an image ***I*** with height, width, and the number of channels (one channel for each colour) as shown in Fig. 2.8. The number of channels depends on whether the image is grayscale (one colour) or RGB (red, green, and blue).

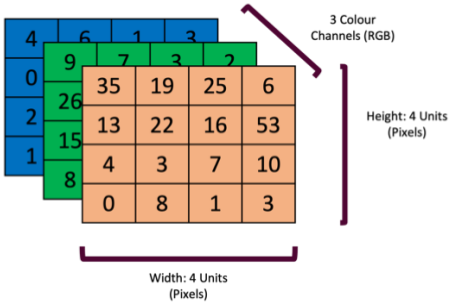


Figure 2.8 | Illustrating an RGB image as an example of an input image to a CNN network.

Note | An image and its pixel values in a *height x width* arrangement consisting of three channel dimensions (RGB).

In the *convolutional layer*, a filter (otherwise referred to as a *kernel*) performs an element-wise multiplication operation (*Hadamard Product*, see Fig. 2.9) called a *convolution* to extract high-level features such as edges or corners from the input image. The *filter size* depends on the dimensions of the input image (or the *input vector*). More precisely, a filter size of **F x F** is applied to an image of **C** colour channels to give us a filter of **F x F x C** volume that performs convolutions on the input image to give an output called a *feature map* **O x O x K** where **K** is the number of applied operations the filter made on the image and **O** the dimension of the feature map output. Stride **S** is the number of pixels values the filter of size **F x F** moves across the image to perform the convolution operation. Afterwards, the feature map becomes the input for the next convolutional layer, and the process repeats. Usually, multiple convolutional layers are required to extract complex features from the image (31).

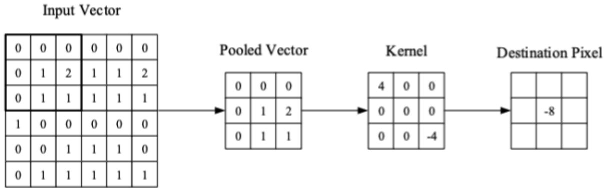


Figure 2.9 | A visual representation of a convolutional layer.

Note | From O’Shea et al. (68)

The *pooling layers* (*Max pooling* or *Average pooling*) reduce the spatial size of the convolved feature maps by applying *dimensionality reduction* with the goal of minimising the computational power required to continue processing the image. The dimensionality reduction method works by merging similar features from the image and helps the network reduce the possibility of overfitting by improving its generalisation ability. Max and average pooling are two common pooling operations (31,60).

The *fully connected layers* perform the final classification task based on the extracted features. The extracted features from these layers are fed into a *flatten layer*, which converts the values into a one-dimensional vector. This vector is then passed through a feed-forward neural network to perform the final prediction. Finally, the *softmax* activation function provides the probabilistic distribution of the predicted labels. As in regular neural networks, backpropagation is applied after each training session to adjust the network weights based on the difference between the predicted and actual output (see Fig. 2.7 for a visual representation of a fully connected layer within an overall CNN architecture).

2.4 Related Work

Sec. 1.2 mentions that limited research has been directed towards predicting animal behaviour from calcium imaging recordings using deep learning tools such as CNNs. To arrive at this conclusion, a comprehensive literature review was organised to assess the current state of research in the field and identify potential knowledge gaps. In Sec 2.4.1, we illustrate the methodology adopted for this literature analysis. At the same time, Sec. 2.4.2 focuses on current approaches in CI analysis, while in Sec. 2.4.3, we comment on how CNNs have been employed for improving calcium imaging analysis workflows.

2.4.1 Literature review

The literature review approach has been influenced by Karolinska Institutet Library guidelines (69). The first step of the literature review was to determine appropriate search terms likely to yield the most relevant publications. The *PICO* (Problem, Intervention, Comparison, Outcome) framework (70), presented in Table 8.1 in the Appendix, illustrates how from the research questions outlined in Sec. 1.4, search blocks to find relevant literature on online databases were generated.

For this search, a multitude of databases were considered, including *PubMed* (71), *Web of Science* (72), *IEEE Xplore* (73), *ACM Digital Library* (74), and *Google Scholar* (75). Upon reviewing the findings of *Falagas et al.* (76), we selected the *PubMed* and *Web of Science* online databases as they're more relevant for medical research or clinical topics and excluded *Google Scholar* due to the possible presence of incomplete or outdated information. As for computational material, the choice of *IEEE Xplore* has been preferred over *ACM Digital Library*. The former contains close to twice the number of publications of the latter. Consequently, the primary databases selected for searching relevant literature are *PubMed*, *Web of Science*, and *IEEE Xplore*.

An example of a search block used to search for available literature on the selected online databases is presented here:

("calcium imaging" AND ("deep learning" OR "convolutional neural networks" OR CNN))

Moreover, this literature review was conducted in February 2023. The search blocks were adapted to meet the corresponding search requirements for each database (one example is the use of Medical Subject Heading (MeSH) terms (77) in PubMed). In addition to the search queries, irrelevant publications were filtered out from the results if they weren't open access, in English, and published between 2013 and 2023, most of which (75%) were published between 2019 and 2022.

Table 8.2 in the Appendix demonstrates the number of records identified per database using the corresponding search block queries. Additionally, the list of selected publications is cross-referenced with publications found via backward citation chaining (78). As a result, this thesis work is influenced by literature findings identified by structured (academic database search) and unstructured (manual search) approaches.

Selection of findings

After duplicate findings were removed with the Mendeley Desktop software, 65 publications remained for further analysis. Two screening rounds were conducted to filter critical literature relevant to this work. The selection process of these findings has been summarised in the PRISMA flow diagram depicted in Fig. 8.2 in the Appendix.

Synthesising literature findings

The final 14 publications were selected to provide an overview of the field in calcium imaging techniques and any applications encompassing deep learning-based implementations such as CNNs. Most of the publications selected (75%) were published in 2019 or later (25,79–86), indicating a strong interest among researchers in exploring possible ways to enhance calcium imaging. The main criteria for choosing these publications were based on CNN or other deep learning implementations applied on calcium imaging data and secondly on other relevant techniques.

Moreover, the selected publications address the field of calcium imaging from different angles, such as the development of cell segmentation techniques with deep learning from one-photon or two-photon data (24,83,84,86,87), behaviour analysis or prediction from two-photon data (81,82), and analysis pipelines from one or two-photon data (85,88–90). Even though the behavioural experiments related to this thesis work contain one-photon calcium imaging recordings, for this literature review, we have encompassed publications from both one-photon and two-photon implementations to gather more information on possible existing CI approaches with deep learning. From the selected publications, about 50% concern one-photon CI (85,86,89–92).

Even more limited is the number of publications that involve one-photon CI with any deep learning implementation. Only two publications (84,88) refer to this approach but concern only cell identification and not behaviour prediction from one-photon CI recordings which is the focus of this study. Nonetheless, this multitude selection of publications encompassing behaviour analysis, cell segmentation, and CI pipelines for either one-photon or two-photon calcium imaging has provided us with a strong understanding of the current state-of-the-art solutions in the field, while at the same time, augmented our belief that a potential knowledge gap exists in the area of behaviour prediction from one-photon calcium imaging recordings.

In addition, the use of CNNs from many of these publications (30,82,83,88,91,93,94) fuels us with even more motivation to pursue the investigation of this knowledge gap, as it stands as strong evidence that CNNs can be trained on CI recordings to infer possible behavioural interpretations. Table 8.3 in the Appendix summarises the content of the final 14 selected publications from the PRISMA flow diagram.

2.4.2 Current Approaches in Calcium Imaging Analysis

Various unsupervised and supervised methods of machine learning techniques have been utilised for source extraction (the task of extracting cells from calcium imaging recordings). A worth-noting supervised technique was recently applied by *Apthorpe et al.* 2016 (95), where the authors demonstrated using a deep neural network for detecting neurons from CI recordings (95). Moreover, a *constrained nonnegative matrix factorisation* (CNMF) approach was proposed by *Pnevmatikakis et al., 2016* (96), allowing for the extraction of spatiotemporal information in active cells from CI recordings. With this algorithm, it is possible to demix spatial overlapping cells with minimum parameter tuning (96). To further highlight developments in CI analysis, in this paper by *Giovannucci et al.,* 2017, the authors developed a pipeline that processes fluorescence activity in real-time during experimental sessions (97). In a following paper by the same group (52), the authors developed *CaImAn*, an open-source tool for one-photon and two-photon CI analysis that can execute both offline (after an experiment) and online analysis (by streaming data during the experiment) while using moderate infrastructure such as a personal laptop.

More publications have surfaced on the aforementioned research work (98–100). However, most of these tools are complex for use by non-trained experts and require deeper knowledge of the parameters that must be tuned prior to the execution of the analysis pipeline. If the user is not an expert in the corresponding tool, any incorrect parameter setting may affect the credibility of the analysis outcome. With this need in mind, *Minian*, an open-source CI analysis pipeline, was recently developed to make CI analysis easier for non-experts to set up and execute (101). However, end users still find CI analysis pipelines difficult to implement, time-consuming, and tricky to infer behaviour interpretations from.

2.4.3 CNN Applications in Calcium Imaging Analysis

Recently, convolutional neural networks have emerged as a powerful tool for analysing calcium imaging data. In this Section, we briefly outline some methods identified from the literature search and have been deemed relevant to this thesis work. By elaborating on the use cases of CNNs in calcium imaging analysis, we hope to give the reader a first impression of how CNNs are relevant and appropriate for conducting and improving behavioural analysis on large-scale calcium imaging data.

Commonly, human experts manually annotate the cell bodies displayed in calcium imaging movies with the aim of constructing a training dataset from which a CNN can learn to recognise neuronal activity patterns and perform classifications on new, unlabelled calcium imaging recordings (52,102,103). The manual annotation process may be time-consuming, but in addition to providing the model with data to train on, it also provides a good benchmark for comparing the performance of the CNN model to that of human experts (103). These developments demonstrate the potential for further automaticity in calcium imaging analysis processes using deep learning methods like CNNs (57).

An open-source pipeline for calcium imaging analysis developed by *Giovannucci et al.* 2019 (52) introduces a CNN model trained by human experts to recognise neurons from streamed calcium imaging data. The CNN model discussed in this paper is trained to recognise the soma of each cell based on its spatial footprint in the recording and distinguish neurons from non-neurons, with the authors claiming it can accurately classify neurons and generalise across different calcium imaging datasets.

Another application of CNNs for calcium imaging analysis is demonstrated in this research work by *Yung et al.* 2019 (104). The authors have used a *Residual Network* (or ResNet) CNN architecture (a pre-trained CNN model with promising performance in object detection tasks (105)). Their implementation was trained on a small size of calcium imaging datasets across different days where the data from each day was randomly partitioned with a five-fold data separation process that is otherwise known as *cross-validation* (briefly described in Sec. 2.3.1.1) to achieve similar distributions between training and validation datasets. The authors applied a *transfer learning strategy* by adopting a pre-trained ResNet-18 model to address any limitations related to the small CI dataset. Transfer learning is a technique where a pre-trained CNN is used for a newly defined task. This approach may reduce overfitting possibilities, time, and resources from training a CNN model from scratch or when the training set size is limited or inadequate (106).

Furthermore, in the research work by *Denis et al.* 2020, *DeepCINAC*, a deep-learning pipeline developed for calcium imaging analysis, is another noteworthy tool that analyses calcium movies by using a CNN model. This implementation aims to infer neuronal activity from CI recordings without human intervention for cell annotation or parameter setting. The model was initially trained on manual annotations conducted by 4 human experts and processed by a CNN model with an attention mechanism, a method for selective focus on relevant parts of the input data to increase model performance, and a bidirectional *long-short term memory* (LSTM) structure that allows for capturing sequence dependencies in the training data making the model more capable of visual context comprehension. DeepCINAC can provide a classification output of cell activity ranging from 0 to 1, which refers to the cell activation level. It also provides cell type classifications of *interneurons, pyramidal,* or *noisy* cells.

2.5 Beyond the State of the Art

Neural recording methods, such as calcium imaging and their corresponding analysis pipelines, allow us to link animal behaviour to neural activity. However, as highlighted in Sec 2.4.2, some of the most popular and openly-available CI pipelines (98–100) have a range of limitations (i.e. complex use of the tool for non-experts, time-consuming processing, the requirement of deep knowledge of related parameters and how to set them in the pipeline) (101).

As a first step to addressing these limitations, we conducted a literature search to study current processes in calcium imaging analysis. Most approaches (30,82,83,88,91,93,94) use different ML or DL algorithms, namely CNNs, to extract cells from calcium imaging videos. In contrast, we aim to skip this cell extraction step and feed the CI video recordings directly to a CNN to infer behavioural aspects from the experiment.

The CI data utilised in this work for training and testing the CNN model implementation discussed in the following Chapters originates from a previous research study by *Weglage et al.* (107) at the *K. Meletis Group, Department of Neuroscience,* *Karolinska Institutet*. A driving force for pursuing this thesis work is to demonstrate a technical solution that overcomes the current CI analysis limitations and enables researchers from non-technical backgrounds to use a tool that can enhance them in inferring behaviour correlations from behavioural experiments.

3 Methods

In this chapter, we delve into the research approach of choice, the techniques for collecting and processing data, and the CNN model's training, optimisation, and assessment for attaining the best possible research outcomes within the given development timeframe of this study. Additionally, we have highlighted the ethical considerations that were considered throughout the development of this work.

3.1 Research Approach

When undertaking a research objective, deciding on a suitable research strategy, and evaluating this decision against other potentially applicable approaches is imperative. Chiefly, choosing the correct research approach enhances our ability to answer the formulated research questions. In scientific research, a research strategy is a process of designing a framework that outlines methods and procedures to address research questions and evaluate the output (or *research artefact*) produced (108).

Two research strategies were considered to establish that our approach is well-defined and appropriate for investigating the research questions addressed in Sec. 1.4: the *experiment and design science research strategies*. The following Section shortly demonstrates each research strategy and a corresponding rationale for selecting or excluding each approach.

3.1.1 Experiment Research Strategy

An *experiment research strategy*, as defined by *Johannesson* and *Perjons* (109), involves enacting studies to test or prove a hypothesis and evaluate the effectiveness of the tested method in solving a particular problem. In even more detail, this strategy identifies relationships and measures the effects of one or more independent variables (also called experimental variables) on one or more dependent variables (110). Examples are provided in Table 3.1 below.

Table 3.1 | Examples of relationships between the independent and dependent variable

Independent Variables	**Dependent Variables (Outcome)**
Patient	Diagnosed with a disease / Not diagnosed with a disease
Customer	Buys product / Does not buy product
Animal	Performs behaviour A / Does not perform behaviour A

In this machine learning project, it is important to discuss how ML experiments differ from other experiments described in the literature. ML experiments are complex and require running the experiment multiple times under different configurations. This study aims to test the hypothesis that a CNN model can be trained to predict animal behaviour directly from calcium imaging recordings without the need for calcium trace extractions. To achieve this, we designed technical processes that use previously collected calcium imaging data as *independent* variables to predict behavioural outcomes using CNNs. We conducted multiple experiments under various conditions and parameter configurations, and we introduced benchmark models to compare and evaluate the effectiveness of our deep learning approach, as required in any machine learning experiment.

Given the time limitations for developing this thesis and considering the definition of the experiment research strategy, we have opted for this research methodology compared to the design science research strategy as it permits more flexibility in designing experimental research processes.

3.1.2 Design Science Research Strategy

A *design science research strategy* is a problem-solving paradigm for creating new methods, models, and constructs (called *artefacts*) to solve a real-world problem and enhance human knowledge (108).

In the context of this work, the creation of the CNN-based deep learning pipeline, constituted of a multitude of machine learning libraries, is the *artefact*. The real-world problem we aim to solve is improving the current calcium imaging analysis pipelines using CNNs to extract all possible neural information from animal-based behavioural assays. Moreover, the design science research strategy consists of various development phases comprising the *DSR Process* (111) (see Fig. 8.3 in the Appendix). Even though this research strategy may be a more comprehensive and suitable approach for this study, it has not been selected due to time limitations.

3.2 Experimental Setup

3.2.1 Environment

To develop the model described in this thesis, we utilised hardware resources made available from the K. Meletis Group. A regular computer was provided to develop the code for the CNN model, and a high-performance accelerated computer was utilised for training and assessing the performance of the CNN model, allowing for much faster processing times (see Appendix Table 8.4 for the hardware specifications of both computers used in this study). For pre-processing, data analysis, and building the CNN model, we used the *Python* programming language (v. 3.10.9) (112) and the *JupyterLab* code editor (113) in an isolated *conda environment* (114) where all the project's components, such as dependencies and packages are encapsulated. For version control, we used *Git* (115) and uploaded the code to *GitHub* (115), an open-source website for code development (source code available here: link to an external website). Additionally, we used the packages and libraries mentioned in Table 3.2 below.

Table 3.2 | Packages/Libraries utilised in the study.

Package/Library	**Type**	**Version**	**Purpose**
Pandas	Library	1.5.3	Data manipulation and analysis
NumPy	Library	1.24.2	For scientific computing, including working with arrays and matrices
Matplotlib	Library	3.7.1	Data visualization
Seaborn	Package	0.12.2	Data visualization
TensorFlow	Library	2.11.0	For building and training machine learning models
Keras	Package	2.11.0	For building and training deep learning models
CUDA Toolkit	Package	11.5	For speeding up the training process of the DL model by using a GPU

3.2.2 Datasets

The data used to train the CNN model originated from a behavioural experiment conducted before the commencement of this thesis work and concerned three Oprm1+ Cre mice (see Sec. 2.1.3.1 for more experiment details). In particular, the different types of datasets generated from these experimental sessions and their purpose are displayed in Table 3.3.

Table 3.3 | Choice of datasets and purpose of use.

Dataset	**Type**	**Description**	**Purpose**
Behaviour Segmentation	.h5	Segmentation of each frame to one behaviour class	Retrieve behaviour labels
Alignment data	.csv	Contains reward delivery information and assists with synchronizing calcium and behavioural recordings.	Align behaviour labels with calcium imaging frames
Calcium Imaging video	.nwb	Contains the video recording derived from the miniscope device.	Input data for the CNN model

Calcium Imaging Recordings

Fig. 3.1 illustrates 5 random calcium imaging frames from a single experimental session with corresponding behaviour label annotations. These image-label pairs are provided as input to the CNN model for training. More specifically, these frames are produced from a miniscope capturing about 25000-frames-long image series from a $1mm$ field of view, a resolution of $350\ x\ 400$ pixels and acquired at a frame rate of $20\ Hz$. To train the CNN model, we tried two different methods. The first involved training and validating one video from a single behavioural session. The second involved using multiple videos from various behavioural sessions involving the same animal.

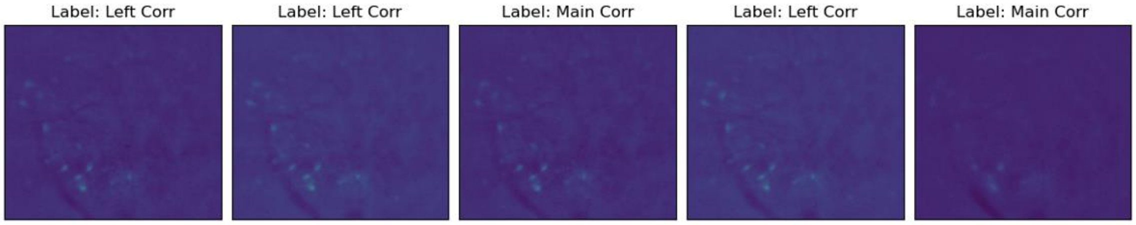


Figure 3.1 | Five randomly chosen frames from a calcium imaging video annotated with their corresponding behaviour label.

3.3 The Behaviour Prediction Neural Network (BPNN)

Our objective was to create a CNN model capable of interpreting behaviour from calcium imaging recordings without requiring the step of calcium trace extraction that is present in current pipelines. For this purpose, we developed the *Behaviour Prediction Neural Network* (BPNN), a CNN model for simple image classification on frames from calcium imaging recordings. The BPNN tool uses a *Sequential architecture* (116) to extract neuronal features from calcium imaging frames of one animal, associate them with assigned behaviour labels (such as *rearing*, *grooming*, *turning*, *running*, *stopping*, and others), and perform behaviour correlations. Fig. 3.2 provides an overview of where the BPNN pipeline fits in the overall workflow of experimental behavioural analysis.

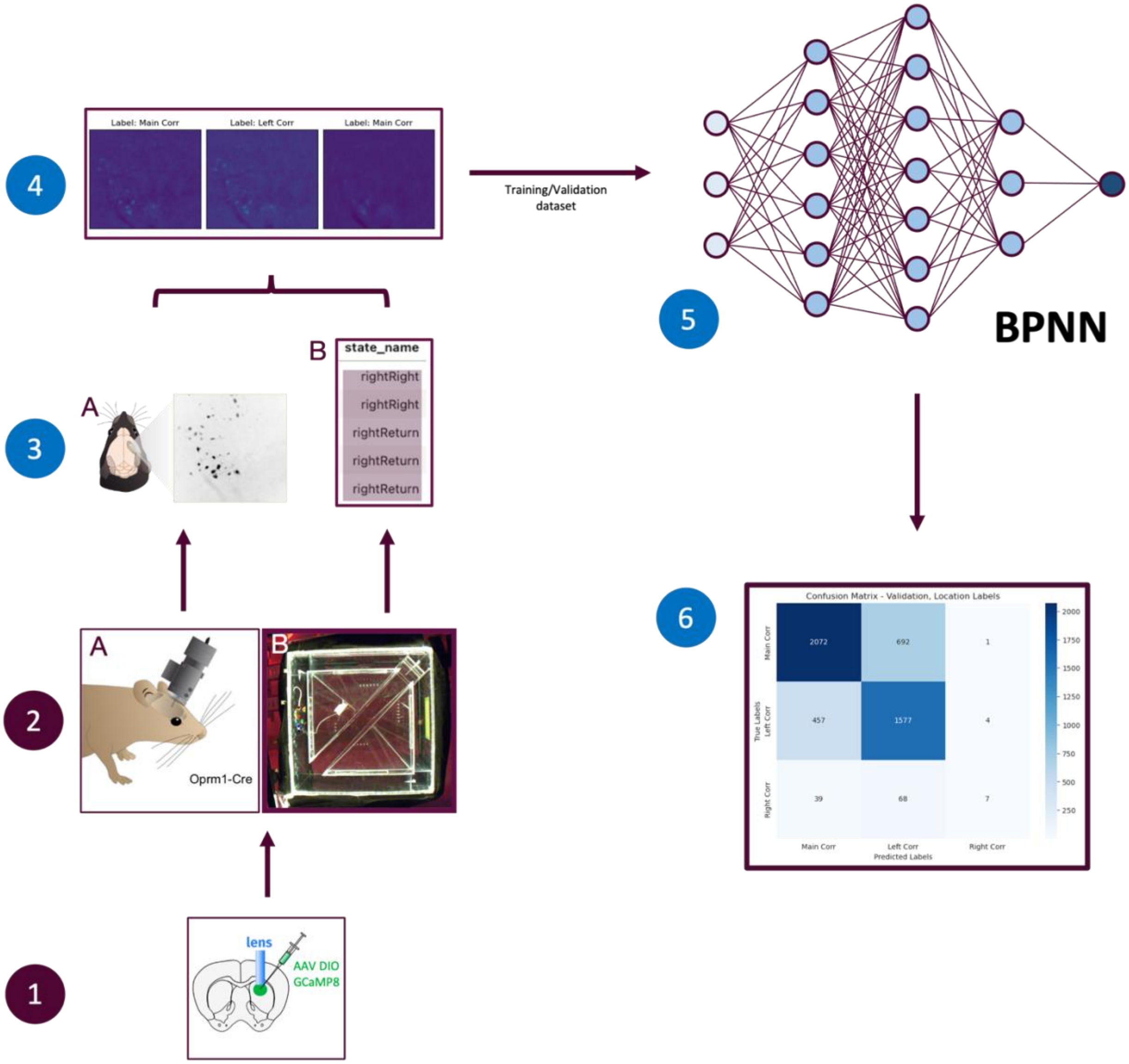


Figure 3.2 | Illustration of the BPNN within a proposed behavioural analysis workflow.

Note | Steps 1-2 relate to current processes of acquiring neural activity from calcium imaging. Steps 3-6 relate to the process involving the BPNN tool and how it processes neural data.
1. Injection of a virus responsible for expressing the GCaMP8 fluorescence indicator.
2. A. A mouse with an integrated miniscope for recording neural activity during behavioural assays.
B. An upward view of the Arrow Maze experiment (see Sec. 2.1.3.1) filmed by a camera while the mouse is performing the experiment.
3. Temporal alignment for calcium imaging frames (A) showing activated neurons with the behaviour annotation dataset (B).
4. Creating pairs of individual calcium imaging frames and the corresponding behaviour label at the specific time point during the experiment. This is used as the input dataset for the BPNN.
5. Split of the dataset into training and validation parts. The model is trained on the train set and evaluated on the validation set.
6. The BPNN produces an output in the form of a confusion matrix to demonstrate its ability to predict true labels from the dataset.

3.3.1 Data Pre-processing

The BPNN tool has a primary advantage over other calcium imaging analysis tools (14,16,25,67,72,73) in that it requires minimal pre-processing of raw calcium imaging videos. However, before the calcium imaging video can be fed into the CNN model for training, some pre-processing is still necessary. This tool primarily targets computational neuroscientists, bioinformaticians, and researchers interested in behavioural analysis of neural recordings. To cater to the users' needs, the pre-processing stage of the tool is automated, requiring the users to provide only the originally required files as input.

Step 1: Pre-processing of Calcium Imaging Recordings

The pre-processing stage begins by loading the Behaviour Segmentation, Alignment, and Calcium *Video* datasets into the pipeline. The miniscope (for recording calcium imaging videos) is detached and reattached between experimental sessions of the same animal across different days. This process may produce motion artefacts in the field of view, potentially interfering with the model's ability to learn from the images and detect patterns. To train the BPNN model, multiple calcium videos are selected, concatenated, and cropped based on a pre-defined set of ***X*** and ***Y*** coordinates to ensure the same field of view is extracted from all video recordings.

Step 2: Label Selection

The next step is to temporally align the *Behaviour Segmentation* and *Alignment* datasets such that each frame from the concatenated calcium video has a corresponding behaviour label. Initially, individual behaviour labels generated from the Hidden Markov model (as described in Sec. 2.1.3.2) are assigned to the corresponding calcium frame based on the temporal alignment. A filtering process is applied to identify all individual behaviour states from the corresponding recording session. Depending on the type of experiment, labels can be merged with other labels or even removed from the final input dataset, thus ensuring that the model is trained only on calcium imaging frames of interest. Lastly, each label is assigned a numerical ID to facilitate easier recognition by the model.

The class imbalance of the different behaviour classes is also inspected to contextual associations of the mouse’s behaviour during each experimental session. Figures 3.3, 3.4 and 3.5 show examples of behaviour label distributions from varying behavioural experimental assays originating from the Arrow Maze set-up.

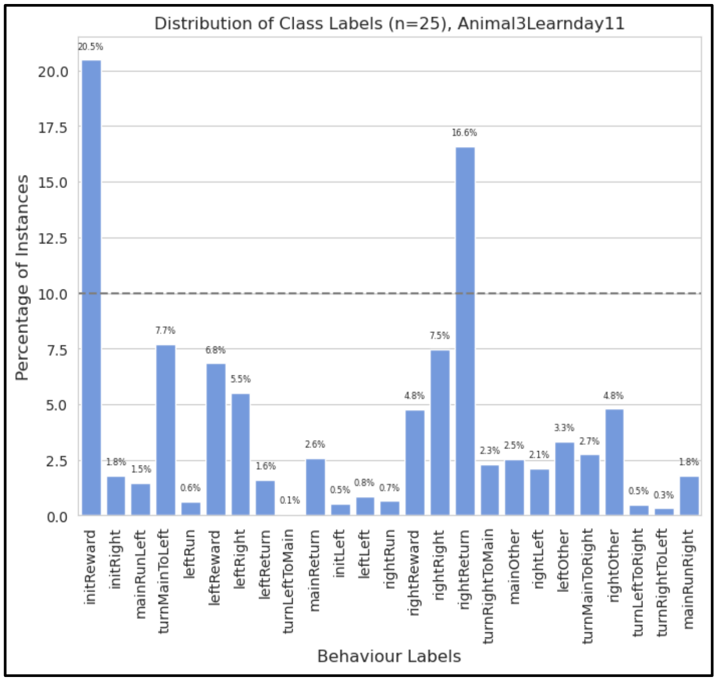


Figure 3.3 | Class balance of 25 behaviour labels from the experimental session of Animal 3 (Day 11).

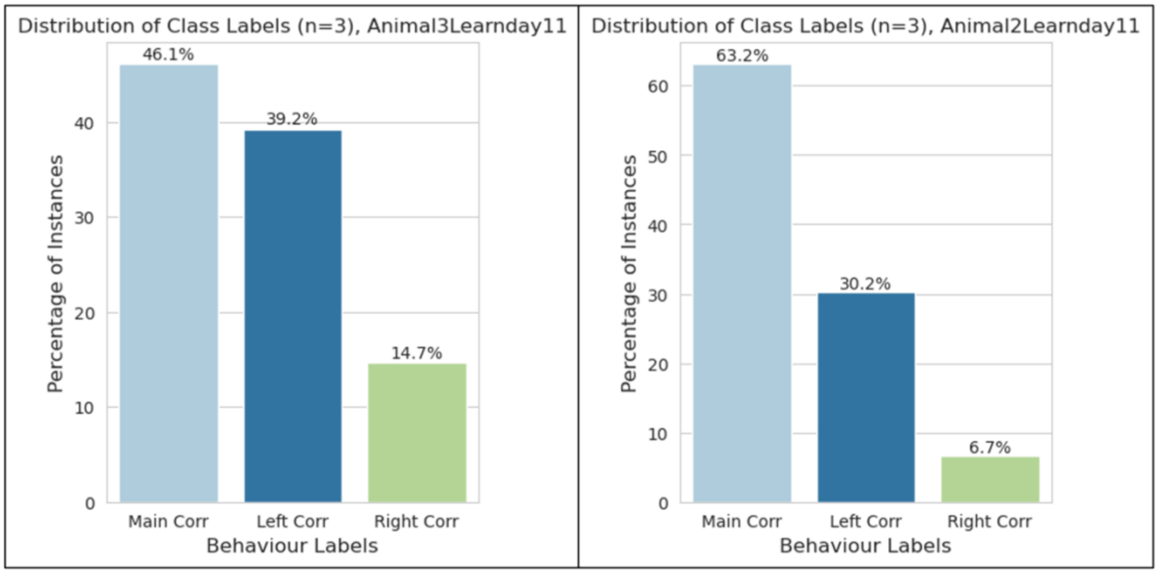


Figure 3.4 | Class balance of 3 location-based labels from two different animals from the same corresponding experimental day (Day 11).

Note | The labels shown in Fig. 3.3 have been merged into the 3 *spatial* (location-type) behaviour labels describing where the mouse is in the Arrow Maze experiment e.g., the *Main Corridor*, *Left Corridor*, or *Right Corridor*. This label configuration is sometimes employed to observe where the mouse spends most of its time during the experiment to understand the extent to which the mouse has learned the task. As observed from the Fig. 3.4, most of the behaviour labels are the Main and Left Corridor labels, indicating that the mouse has learned the task to a certain extent (for example, travelling from the *Initiation Spout* to the *Reward Spout* – see Sec. 2.1.3.1 for more information on the Arrow Maze experiment).

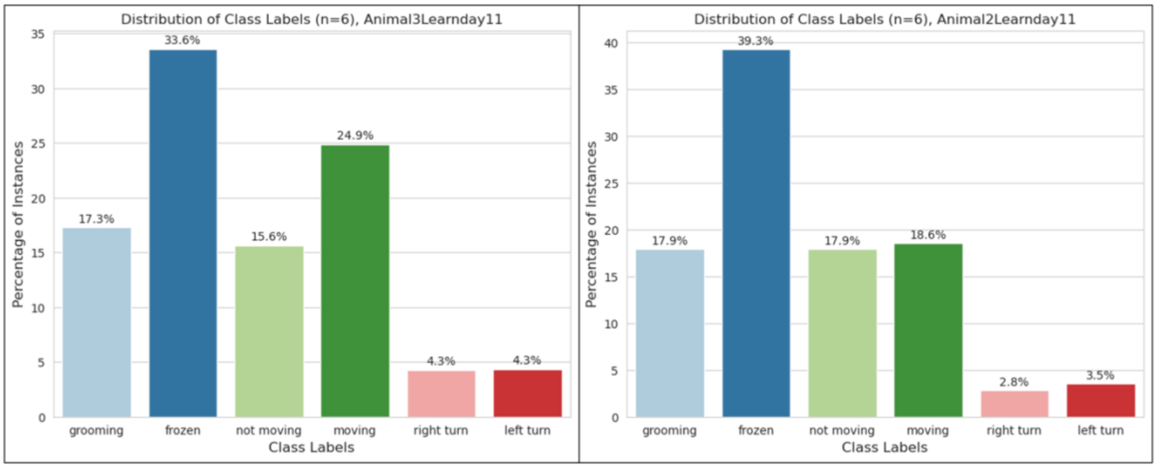
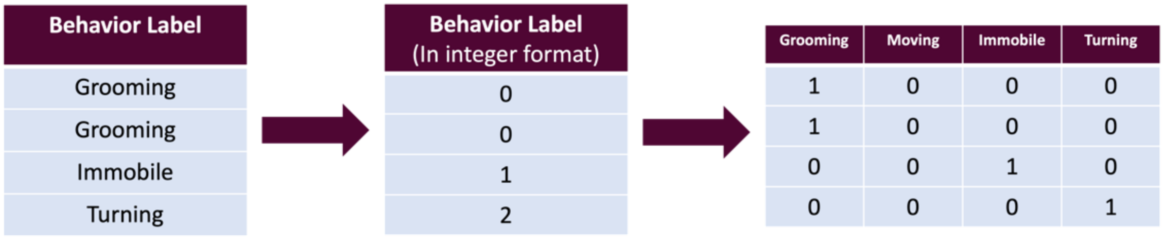


Figure 3.5 | Class balance of 6 different behaviour labels from the Arrow Maze experiment (Day 11).

Note | Day 11, Animal 3 (Left), Animal 2 (Right). The labels depicted above reflect behavioural mice information during the Arrow Maze task in contrast to the location-type labels shown in Fig. 3.4. They don't depict where the mouse is (spatial labels) but rather what it is doing (behaviour labels) such as grooming, frozen (standing still), not moving (staying in the same region in the maze), moving, turning right and turning left.

Step 3: One-hot Encoding

Finally, the distinct number of behaviour labels is converted to *categorical values* with a process called *one-hot encoding*. One-hot encoding converts the integer-encoded class labels into a binary matrix representation, where each row corresponds to a sample and each column corresponds to a class label (117). Fig. 3.6 exemplifies how one-hot encoding is applied in the context of behavioural labels.



Behavior Label
Grooming
Grooming
Immobile
Turning

Behavior Label (In integer format)
0
0
1
2

Grooming	Moving	Immobile	Turning
1	0	0	0
1	0	0	0
0	0	1	0
0	0	0	1

Figure 3.6 | One-hot encoding applied on behaviour labels.

Note | One-hot encoding allows for categorical data to be represented as numerical data making it easier for an ML model to distinguish different behaviour classes.

Step 4: Data Verification

The next pre-processing step is data verification of the calcium imaging frames and the assigned behaviour labels. Five random calcium imaging frames and their corresponding behaviour labels are visualised to ensure the alignment process has been performed correctly (see Fig. 3.1).

Step 5: Dataset Splitting

The final pre-processing step involves splitting the data into training and validation sets. Fig. 3.7 demonstrates how the datasets are split with the K-fold cross-validation method validation (see Sec. 2.3.1.1 for more information on the cross-validation technique).

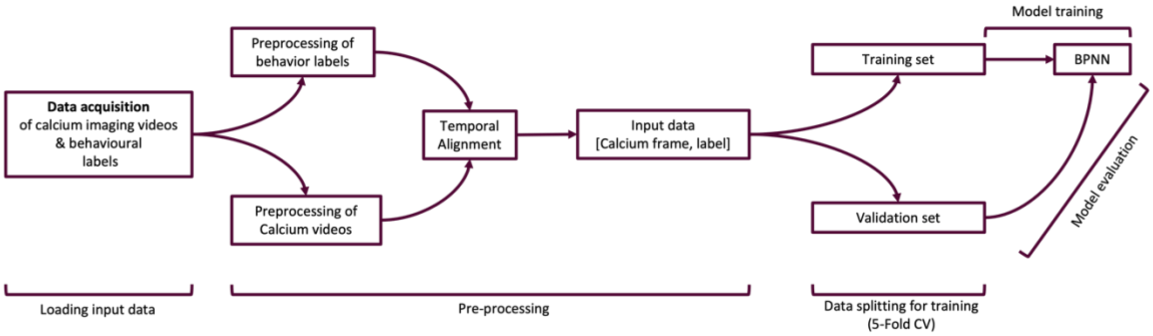

Model training
Preprocessing of behavior labels
Training set
BPNN
Data acquisition of calcium imaging videos & behavioural labels
Temporal Alignment
Input data [Calcium frame, label]
Model evaluation
Preprocessing of Calcium videos
Validation set
Loading input data
Pre-processing
Data splitting for training (5-Fold CV)

Figure 3.7 | BPNN data flow schematic.

Note | Schematic summarising steps 1 to 5 as discussed above of the data flow from loading the calcium imaging videos and behavioural labels to model training and evaluation.

3.3.2 BPNN Architecture

The BPNN pipeline contains a convolutional neural network with a *Sequential architecture* commonly used for image classification tasks (118). The input shape is $393\ x\ 444\ x\ 1$, indicating that the images we input into the model are 393 pixels in height and 444 pixels in width, with "1" representing a single channel of grayscale images.

In addition, the model incorporates two convolutional layers with 32 and 64 filters, each with a filter size $3\ x\ 3$. Each convolutional layer utilises a ReLU activation function to introduce non-linearity into the model. The output of each convolutional layer is then passed through a max pooling layer with a pool size of $2\ x\ 2$. This pooling layer helps reduce the spatial dimensions of the output and captures the most important information from each layer.

The Flatten layer receives the output from the second max pooling layer and transforms it into a 1D vector, allowing the fully connected layers to process the extracted features. Next, the output from the flatten layer is fed into a dropout layer, which helps prevent overfitting by randomly disabling a certain proportion of nodes.

Following the dropout layer is a fully connected (Dense) layer with 128 units and a ReLU activation function.

Lastly, the output layer comprises several units corresponding to the specific number of classes present in the problem. Additionally, it includes a softmax activation function, which enables multi-class classification by producing a probability distribution over the classes. Fig. 3.8 visualises all the layers of the network architecture mentioned above. Fig. 8.4 in Appendix depicts Fig. 3.8 in a more top-down format.

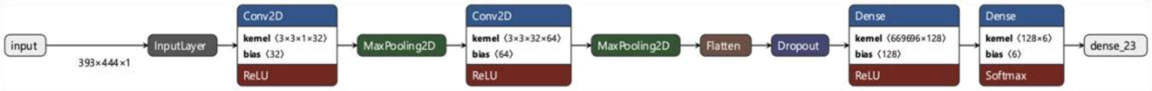


Figure 3.8 | The basic architecture of the BPNN model.

3.3.2.1 Adding the Element of Time in the BPNN model

When neurons are activated, the fluorescence that is captured from the miniscope device fluctuates over a period t, where t is the entire time duration of the corresponding calcium imaging recording session. To improve the model's ability to identify and extract distinct patterns of neural activity and perform neurobehavioural correlations, we modified the number of channel dimensions that can be used during the training process from 1 channel dimension (the typical setting for grayscale images) to 3 or 5 channel dimensions (see Sec. 2.3.2 for more on channel dimensions).

The input data in the BPNNt (t for time), receives a sum of different grayscale images that are sequentially related (one frame after the other). For example, the CNN model receives an input image Ii where i, is the ith frame of the input dataset I. When adding the element of time t, the model will also receive the images $Ii - 1$ and $Ii + 1$ (for 3 channel dimensions) and the sequence of frames, in this case, would be $Ii - 1,\ Ii,$ $Ii + 1$. This sequence of images allows the CNN model to capture fluorescence changes that may be relevant in specific animal behaviours making it easier for the network to recognise and extract features and patterns in the dataset and correlate them to behavioural.

3.3.3 The Behaviour Prediction Support Vector Machines (BPSVM) Model

The BPSVM model can categorise extracted calcium traces into different behaviour groups and was developed to serve as a benchmark for comparison with the BPNN model. We hypothesised that the BPSVM model will outperform the BPNN model as it utilises support vector machines (SVMs), a current state-of-the-art method for associating neural activity with behaviour from calcium imaging movies.

If the BPNN model produces a similar or better performance than the BPSVM, it indicates that CNNs are a suitable deep-learning method for conducting behaviour correlations from calcium imaging recordings without the time-consuming pre-processing step of extracting calcium traces. In addition, similarly to the BPNNt, a modified version of the BPSVM was created, the BPSVMt, to account for differences in the calcium trace signals and to make fair comparisons with the BPNNt.

3.3.4 Model Training

Throughout the BPNN training phase, we aimed to test various configurations and determine the optimal combination of parameters for a CNN model that could perform as well as or even better than BPSVM. Therefore, it is crucial to carefully select the loss function, optimiser, evaluation metrics, and regularisation techniques when developing a model training routine, as they all significantly impact the model's performance. We have outlined the selected parameters and our reasoning behind them in Table 3.4.

Table 3.4 | Choice of training parameters used in the BPNN.

Parameter	Type	Purpose
Loss function	Categorical cross-entropy	For multi-class classifications (two or more labels)
Optimizer	Adam	Suitable for a wide range of neural network architectures
Evaluation Metrics	Accuracy, Loss, F1-Score	For monitoring training progress and overfit/underfit potentiality
Regularization Techniques	Dropout	To prevent overfitting

Training Protocol

While developing the BPNN model, we carefully considered different training protocols. To ensure a fair comparison between our implementation and the benchmark solution, we applied the same parameters for both models. These included training, validating, and testing both models using the same input data and data-splitting methods, such as K-Fold cross-validation. We also ensured that both models were trained with the same number of behaviour labels. Additionally, we estimated the chance-level accuracy for both models to better evaluate their performance compared to the chance outcome. See Table 3.5 for more information on the training protocol parameters.

The choice of input datasets also influenced the training protocol. Ultimately, we concatenated 3 calcium imaging videos acquired from the same animal but from different days and implemented the 5-fold cross-validation method for splitting the dataset into training and validation parts. Furthermore, during training, a background removal process was carried out on each calcium imaging frame to enhance the model's ability to detect patterns from the calcium imaging frames. This operation enabled the CNN model to better differentiate between the foreground and background elements in the frames, thereby allowing it to extract relevant features for learning associations between neuronal activity and behavioural labels. In addition, after the background removal process, the image pixel values were normalised to a range between the minimum and maximum pixel values. Normalisation ensures that the pixel values of the images are within a standard range, which is essential for reducing the impact of variations in brightness and contrast among the frames. This step was necessary to standardise the input data and achieve more stable training results.

Table 3.5 | Training protocol parameters

Training protocol parameters	**BPNN**	**BPSVM**
Data Splitting Method	5-Fold CV	5-Fold CV
Early Stopping	Yes	*
Epochs	50	*
Number of labels	25 / 3 / 6	6
Calcium Imaging videos	3 concatenated / 1	3 concatenated / 1
Chance level prediction result	Yes	Yes
Batch-size	32	*
Animals	Animal 2/**3**	Animal 2/**3**
Training days	Days 8, 9, 10, **11**	Days 8, 9, 10, **11**

* Not applicable

Note | For training the BPNN or BPSVM models with one calcium imaging recording, the experimental session from Day 11 of Animal 3 (highlighted in bold) was used. Animal 3 was observed by the experimenters to have learned the task more effectively than Animals 1 and 2 on the corresponding day (Day 11), hence the reason for choosing this dataset for training the models to perform neurobehavioural correlations instead of others. It is possible to train the model on any experimental session from any animal. However, researchers need to account for incorrect behaviour to calcium activity correlations if the corresponding experimental session concerns earlier days (Days 1, 2, 3, for example) when the mouse is still learning the task. To compare the best-performing model across animals, we executed the most optimal configuration with calcium imaging data on Day 11 from Animal 2, as shown in Sec. 4.4.

3.3.5 Model Evaluation

To verify that the model has been adequately trained to make plausible associations between calcium imaging data and interoperations of either spatial or functional activity during the behavioural task, it must be compared against current well-established solutions practised at the K. Meletis Group. Our approach has been to test the efficacy of the BPSVM model in predicting behaviour in one animal under certain conditions and compare it with the BPNN model's efficacy for the same animal and under the same conditions.

3.3.5.1 Evaluation Metrics

To measure the performance of the BPNN model, the *accuracy, loss,* and *F1-score* metrics have been chosen. Accuracy indicates the proportion of correctly classified examples to the total number of examples.

$$Acc = \frac{tp+tn}{tp+fp+tn+fn} \quad (119)$$

Fig. 3.4 shows that most label instances are the Main and Left corridor labels. Given the context of the Arrow Maze experiment described in Sec. 2.1.3.1, this distribution indicates that the mouse is moving most of the time between these two corridors indicating that it has learned the task to a certain extent. However, the accuracy metric may not be the most appropriate for imbalanced datasets. For this reason, we have also employed the F1-score, which is the harmonic mean of *precision* and *recall*.

$$F1 - score = \frac{2*p*r}{p+r} \quad (119)$$

Precision refers to the number of true positive predictions divided by the total number of positive predictions. At the same time, *recall* is the number of true positive predictions divided by the total number of positive examples in the dataset. *Loss*, on the other hand, measures the difference between the predicted behaviour labels and the actual behaviour labels. It is adopted for optimisation purposes as a metric during the training stage of the CNNs, by penalising the model for incorrect predictions made with higher confidence (120).

Other evaluation methods are *confusion matrices* (shown in Fig. 3.9), a helpful tool in visualising a model's performance by looking at how many correct label classifications have been made compared to the actual labels to depict whether the model has been trained well enough to predict behaviour from unseen data points.

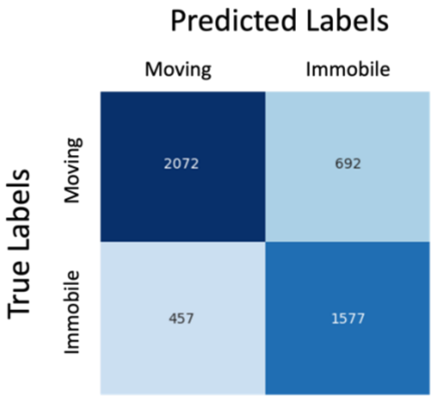


Figure 3.9 | An example of a confusion matrix applied to mice behavioural data.

Note | This example of a confusion matrix illustrates that a DL model is able to correctly predict the label *Moving* for 2072 data samples (True Positive) and was able to incorrectly predict the label *Immobile* for 692 samples (False Positive). For the label *Moving* the model has been able to accurately predict the behaviour with 75% accuracy.

3.4 Ethical Considerations

Calcium imaging is a neuroimaging technique that enables the investigation of neuronal activity of single or population-level neuronal structures of interest in living animals (121). However, applying a recording mechanism, like a miniscope for one-photon or two-photon calcium imaging, necessitates an invasive surgical operation on the animal's head. Furthermore, these animals are sacrificed after their study has been completed.

Additional animal experiments are typically required to produce enough training data for a CNN model to learn from and extract useful biological interpretations. Nevertheless, repurposing data from previous experiments to produce the required data may offer a possible solution to this ethical problem. In this study, for example, we did not conduct invasive surgical operations on animals for the purposes of the thesis, nor did we initiate any animal testing on behalf of it. Instead, we repurposed existing data from previously conducted experiments for which researchers had already acquired appropriate ethical permits. In the future, having sufficiently large and well-managed data from calcium imaging recordings, either from past or ongoing experiments, could help train CNNs more effectively and potentially limit, to a certain extent, the need for animal testing.

It is worth noting, however, that potential bias can arise with all machine learning models. For example, erroneous label classifications may result in a model that predicts animal behaviour inaccurately. Hence, trained experts must ensure that ground truth labelling is cross-checked before inputting into a CNN model.

4 Results

In this Section, we present the results from the three different challenges developed to assess the model's capability of providing behavioural correlations from calcium imaging data. The first challenge concerned the BPNN's ability to perform behaviour correlations from one calcium imaging video with three different behaviour label configurations (see Sec. 3.3.1 for more information on the label class balance). This experiment allowed us to select the most optimal behaviour label configuration for the next model executions.

The second challenge required the BPNN to identify behaviour correlations from a larger calcium video consisting of three different calcium imaging videos originating from the same animal and recorded across different days (for example, experimental sessions performed on days 8, 9, and 10). This would allow us to observe how the BPNN's ability to perform behaviour correlations improves with more training data from across different days.

Furthermore, the third challenge involved modifying the BPNN to account for time relevance during training to ensure the model can identify patterns from the fluctuations of the fluorescently activated neurons. In this case, we ran the BPNN twice, once for multiple calcium recording datasets (Days 8, 9, and 10) and once for an individual recording (Day 11). Finally, the BPNN is compared to chance-level configurations throughout these experiments by shuffling the behaviour labels and the BPSVM, our benchmark model developed to assess the model's performance compared to current alternatives. In addition, we demonstrate the performance of our best performing model when executed with data from another animal (Animal 2) to assess its performance across animals.

4.1 Evaluating Performance Across Different Label Configurations

The first challenge for testing the BPNN's ability to perform behaviour correlations is to use the pair of frame-label information from one calcium imaging movie. We trained the model on one calcium imaging recording with three different behavioural datasets. On all three model executions, we adopted the *5-Fold cross-validation* method for splitting the dataset into training and validation sets. Finally, we added an *early stopping* mechanism to prevent the model from overfitting on the training data and standardised the number of epochs to 50 (meaning 50 single passes the model completes throughout the entire dataset to update its weights).

Table 4.1 provides an overview of this experiment's specifications. The number of 25 class labels indicated in Table 4.1 involve all the defined behaviours in the *Animal3learday11* experimental session; the number of 3 (merged-spatial) class labels refers to the merging of the previous 25 behaviours to 3 new labels based on the corridor of the *Arrow Maze* task on which they occur, and 6 (behaviour) labels refer to another set of behaviour annotations that include mouse activity such as *grooming*, *frozen*, *not moving*, *moving*, *turning right, and turning left* (Sec. 3.3.1 provides a more extensive elaboration on the origin and meaning of these labels). Figures 4.1, 4.2, and 4.3 illustrate the averaged accuracy and loss results across 5-Folds of the training and validation data for *Exp_1.1*, *Exp_1.2* and *Exp_1.3*.

Table 4.1 | Experiment specifications.

Model Execution ID	No. of Videos	Experimental Session(s)	No. of Class Labels	Train/Validation Split Method
Exp_1.1	1	Animal3learnday11	25 (All)	5-Fold CV
Exp_1.2	1	Animal3learnday11	3 (Merged - Spatial)	5-Fold CV
Exp_1.3	1	Animal3learnday11	6 (Behaviour)	5-Fold CV

Note | The notation "Animal#learnday#" refers to the behavioural experimental session from which calcium imaging recording have been used as input for training and validating the BPNN model.

Exp_1.1: Training the BPNN with one CI recording of 25 labels.

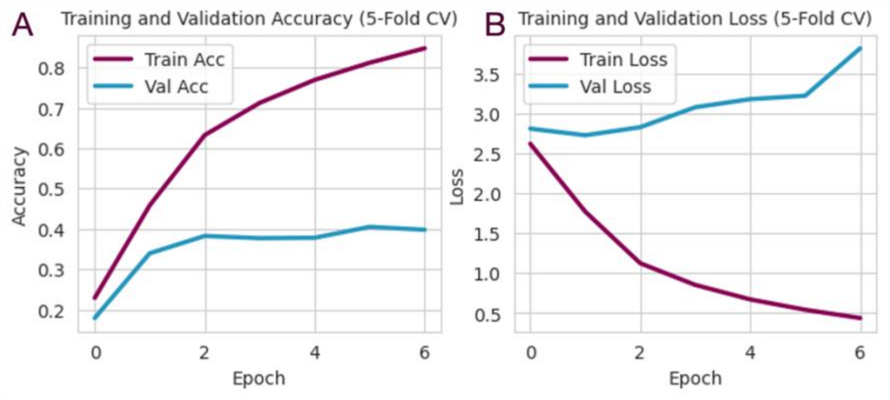


Figure 4.1 | Accuracy (A) and Loss (B) for the training and validation data averaged across 5 folds with 25 labels (BPNN, 1 CI video).

Exp_1.2: Training the BPNN with one CI recording of 3 labels (merged from 25).

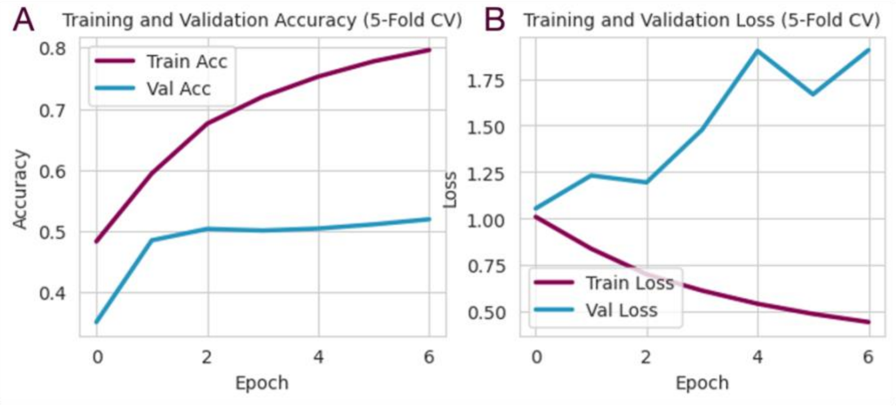


Figure 4.2 | Accuracy (A) and Loss (B) for training and validation data averaged across 5 folds with 3 labels (BPNN, 1 CI video).

Exp_1.3: Training the BPNN with one CI recording of 6 labels.

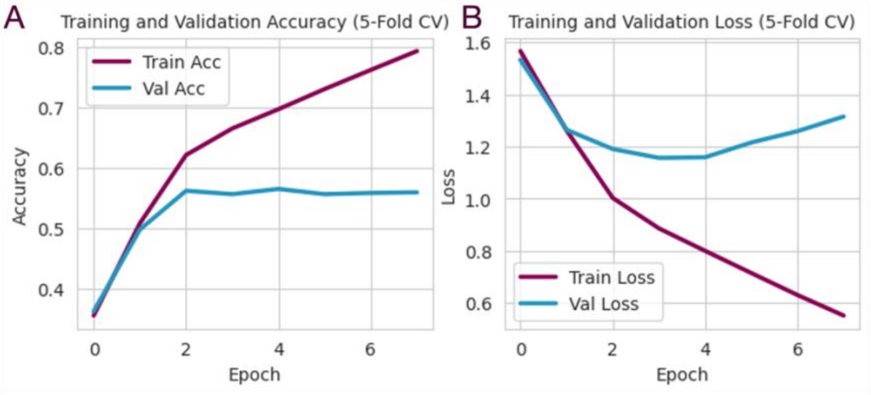


Figure 4.3 | Accuracy (A) and Loss (B) for the training and validation data averaged across 5 folds with 6 labels (BPNN, 1 CI video).

Table 4.2 summarises the BPNN tool's performance following the execution of the specified parameters above. The label configuration that performs the best is highlighted in **bold font**. It was found that although the 3 spatial labels configuration produced a the highest F1-score **(0.56)**, using 6 labels with a score of **0.55** is a better option overall due to having more behavioural classifications from which prediction can be made from the BPNN model. As a result, the label configuration from *Exp_1.3* was selected for following model executions. The poorest performing configuration was observed when using all behaviour labels in *Exp_1.1* to train the BPNN model, which produced an F1 score of **0.39.** Despite the best-performing configuration being *Exp_1.3*, the overall results suggest that the BPNN model is prone to overfitting on the training data, making it challenging to produce accurate predictions on unseen data from in the validation set.

Table 4.2 | Comparing the evaluation metrics across different label configurations.

Experiment ID	Label Configuration	Training Accuracy (%)	Validation Accuracy (%)	F1-score
Exp_1.1	All Labels (25)	84.73	39.77	0.39
Exp_1.2	Spatial Labels (3)	79.60	51.89	0.56
Exp_1.3	**Behaviour Labels (6)**	**79.25**	**55.89**	**0.55**

4.2 Training with Multiple Calcium Imaging Recordings

In Sec. 4.1, we observed that the BPNN model was overfitting, resulting in lower validation accuracy than training accuracy. To address this problem, we concatenated multiple calcium imaging videos from the same animal across different experimental sessions (Days 8, 9, and 10) to provide the model with more data to train and validate on. The same label configuration that performed the best in the previous experiment (*Exp_1.3*) was applied. In Figures 4.4, 4.5, and 4.6 we demonstrate the BPNN's performance using multiple calcium imaging videos, the chance-level accuracy when shuffling the labels, and compare it to the BPSVM model, the current benchmark comparison.

Table 4.3 | Experiment specifications for executing the BPNN and BPSVM models on larger datasets.

Model	Model Execution ID	No. of Videos	Experimental Session(s)	No. of Class Labels	Train/Validation Split Method
BPNN	Exp_2.1	3	Animal 3, days 8, 9, 10	6 (Behaviour)	5-Fold CV
BPNN Chance	Exp_2.2	3	Animal 3, days 8, 9, 10	6 (Behaviour)	5-Fold CV
BPSVM	Exp_2.3	3	Animal 3, days 8, 9, 10	6 (Behaviour)	5-Fold CV
BPSVM Chance	Exp_2.4	3	Animal 3, days 8, 9, 10	6 (Behaviour)	5-Fold CV

Exp_2.1: BPNN results

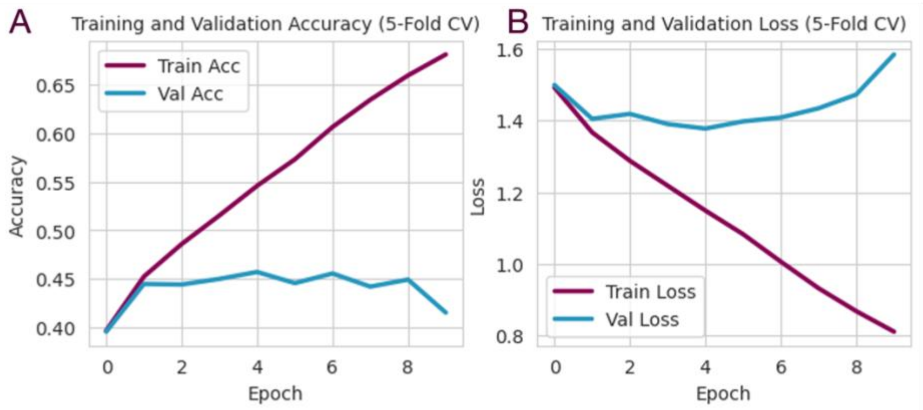


Figure 4.4 | Accuracy (A) and Loss (B) for the training and validation data averaged across 5 folds (BPNN, 3 CI videos).

Exp_2.2: BPNN Chance-level results

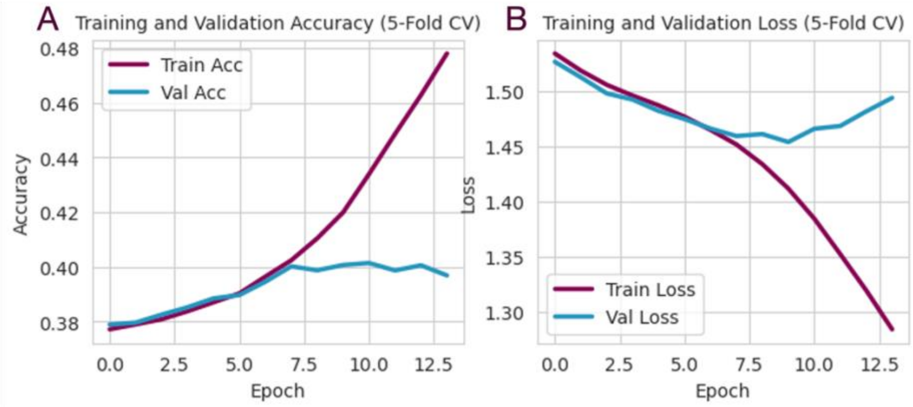


Figure 4.5 | Chance-level Accuracy (A) and Loss (B) for training and validation data averaged across 5 folds (BPNN chance, 3 CI videos).

Exp_2.3 / Exp_2.4: BPSVM and Chance-level results

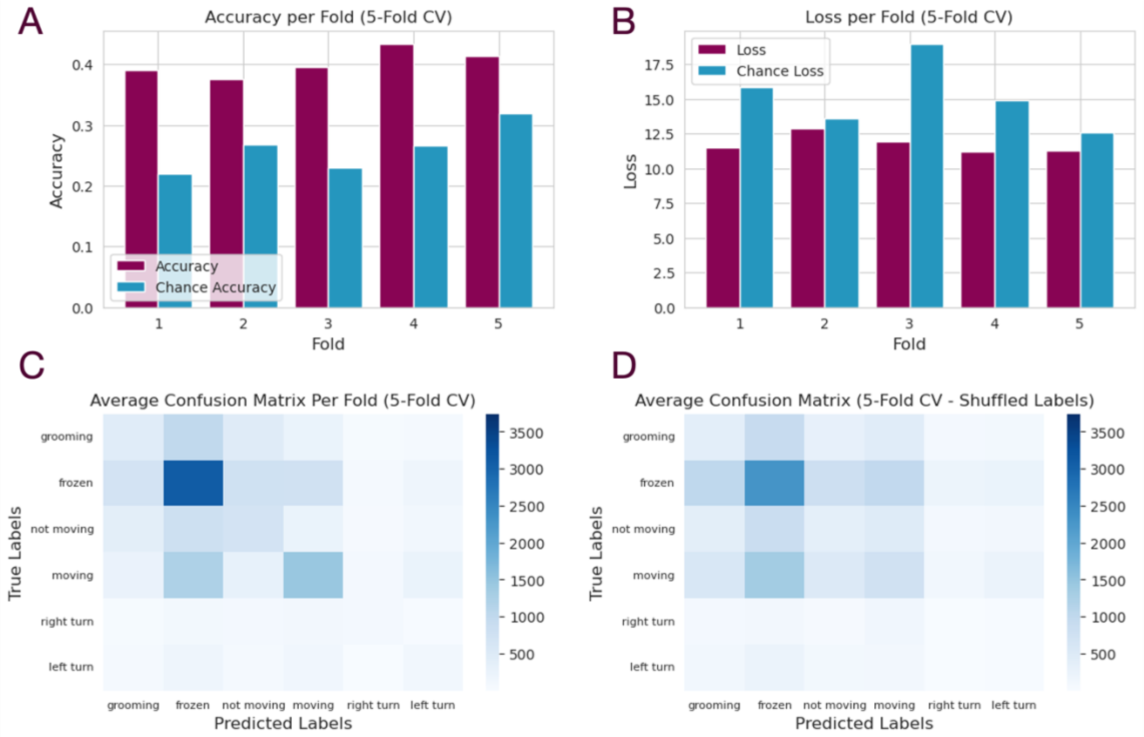


Figure 4.6 | Accuracy, loss, and confusion matrices for BPSVM and BPSVM Chance (3 CI videos).

A. Accuracy and chance-level accuracy averaged over 5 folds.
B. Loss and chance loss averaged over 5 folds.
C. Confusion matrix for 6 behaviour labels
D. Confusion matrix for 6 shuffled behaviour labels

Tables 4.4 and 4.5 summarise the performance of the BPNN and BPSVM tools, following the execution of the specified parameters found in Table 4.3 above. The model that performs the best is highlighted in **bold font**. Contrary to initial expectations, it was found that with more data the BPNN model achieved a validation accuracy of around **41%,** considerably less than the **55%** accuracy observed in *Exp_1.3* (Table 4.2) where the BPNN model was trained on one calcium imaging recording. Conversely, the BPSVM tool performed slightly worse than the BPNN tool, leaving room for the interpretation that the CNN model can perform similarly well with a support vector machine model. Both models achieve better performance than their chance-level counterparts.

Table 4.4 | Comparing the BPNN and Chance-level evaluation metrics.

Model	Training Accuracy (%)	Validation Accuracy (%)	F1-score
BPNN	**68.15**	**41.50**	**0.44**
BPNN Chance	47.80	39.69	0.40

Table 4.5 | Comparing the BPSVM and chance level evaluation metrics.

Model	Accuracy (%)	F1-score
BPSVM	39.66	0.40
BPSVM Chance	28.94	0.26

4.3 Adding Time in the BPNN and BPSVM Models

The next challenge is to investigate the model's ability to arrive at behavioural correlations when the time element is introduced in the training process. For this reason, we have developed a refined version of the BPNN model called *BPNNt*, which can take a multitude of frames before and after the current frame that is being processed to account for fluctuations in the fluorescent neurons recorded in the calcium video (See Sec. 3.3.2.1 for more information on the BPNNt).

4.3.1 Comparing the BPNNt and BPSVMt Across Multiple Recordings

To ensure that we can compare the results of this model with its previous versions, we developed the *BPSVMt*, a similar configuration to BPSVM. We continued training and validating our model using the same CI and behavioural label datasets as shown in Table 4.3 (3 calcium videos with 6-label behaviour datasets each).

Table 4.6 | Experiment specifications for executing the BPNNt and BPSVMt

Model	Model Execution ID	No. of videos	Experimental Session(s)	No. of Class Labels	Train/Validation Split Method
BPNNt	Exp_3.1	3	Animal3, days 8, 9, 10	6 (Behaviour)	5-Fold CV
BPNNt Chance	Exp_3.2	3	Animal3, days 8, 9, 10	6 (Behaviour)	5-Fold CV
BPSVMt	Exp_3.3	3	Animal3, days 8, 9, 10	6 (Behaviour)	5-Fold CV
BPSVMt Chance	Exp_3.4	3	Animal3, days 8, 9, 10	6 (Behaviour)	5-Fold CV

Exp_3.1: BPNNt results

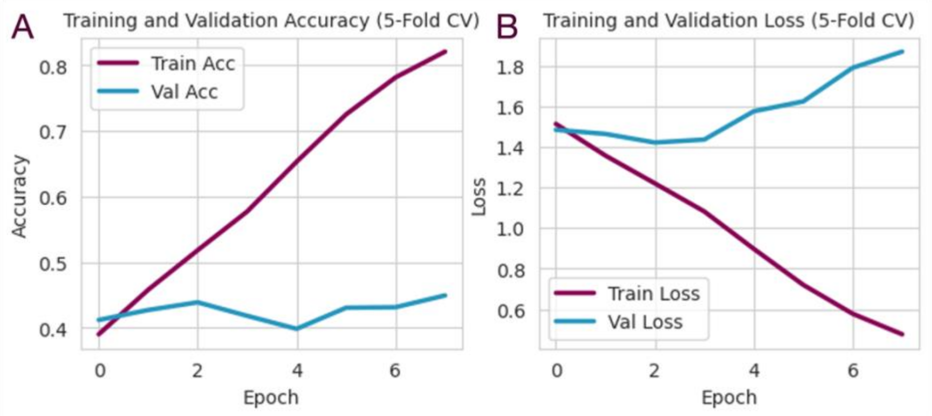


Figure 4.7 | Accuracy (A) and Loss (B) for training and validation data averaged across 5 folds (BPNNt, 3 CI videos).

Exp_3.2: BPNNt Chance-level results

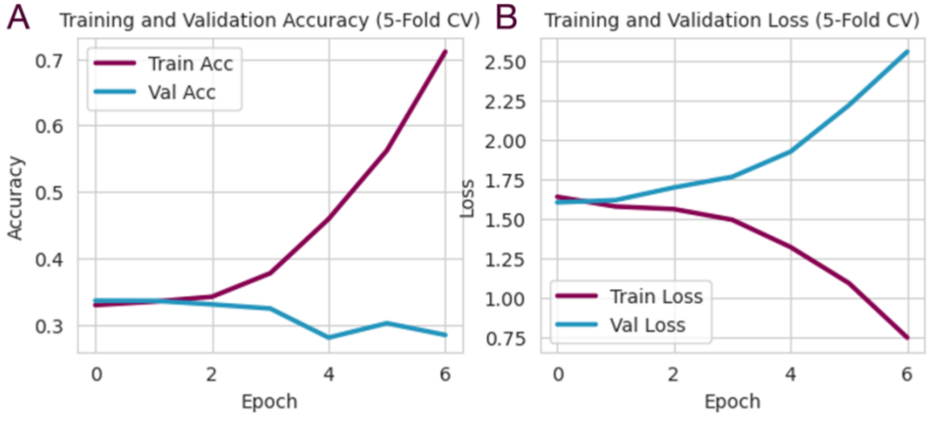


Figure 4.8 | Accuracy (A) and Loss (B) for training and validation data averaged across 5 folds (BPNNt, 3 CI videos).

Exp_3.3 / Exp_3.4: BPSVMt and Chance-level results

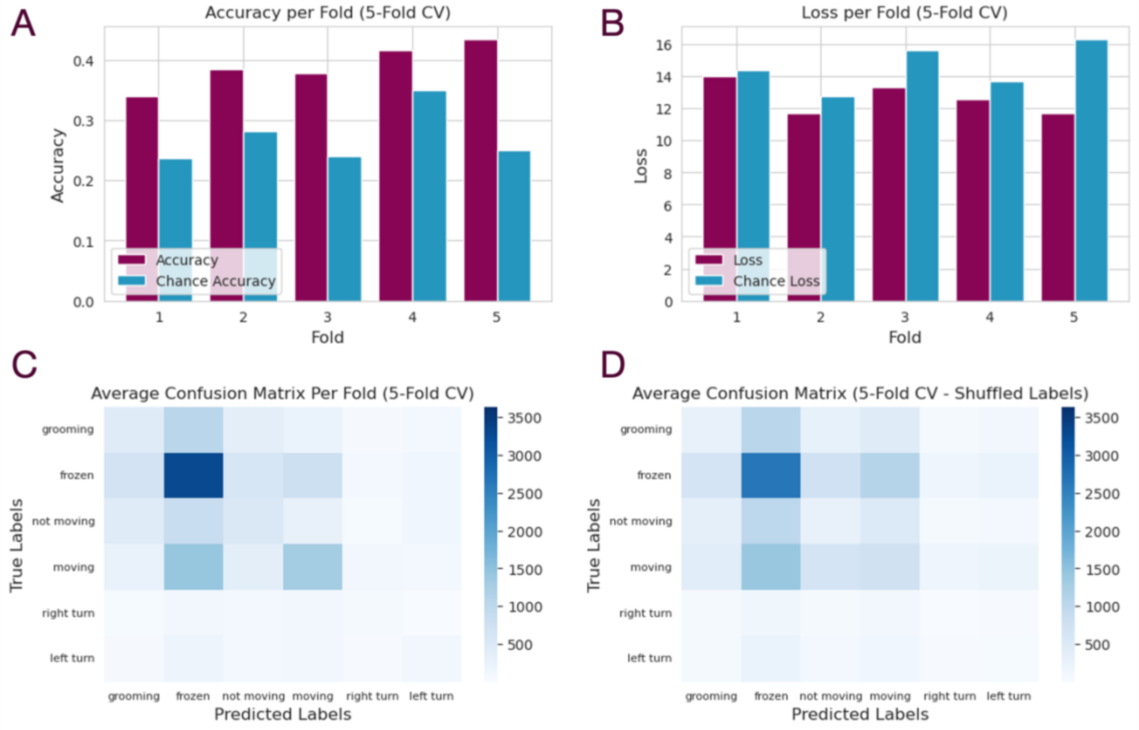


Figure 4.9 | Accuracy, loss, and confusion matrix for BPSVMt and BPSVMt Chance with shuffled labels (BPSVMt, 3 CI videos).

A. Accuracy and chance-level accuracy averaged over 5 folds.
B. Loss and chance loss averaged over 5 folds.
C. Confusion matrix for 6 behaviour labels
D. Confusion matrix for 6 shuffled behaviour labels

The performance of BPNNt and BPSVMt tools have been summarised in Tables 4.7 and 4.8. By incorporating the time element in training with 3 CI videos, the BPNNt achieved a slightly higher F1-score of **0.46** compared to BPNN’s **0.44** shown in Table 4.4. The BPSVMt managed to get an F1-score of **0.39**, which is lower than BPNNt's performance; however, the difference is negligible since the validation accuracy is relatively low on both occasions. This may suggest that training the BPNN model with data from different experimental days decreases overall performance.

Table 4.7 | Comparing the BPNNt and Chance-level evaluation metrics.

Model	Training Accuracy (%)	Validation Accuracy (%)	F1-score
BPNNt	**82.07**	**44.94**	**0.46**
BPNNt-Chance	41.05	36.87	0.34

Table 4.8 | Comparing the BPSVMt and Chance-level evaluation metrics.

Model	Accuracy (%)	F1-score
BPSVMt	41.05	0.39
BPSVMt-Chance	28.26	0.27

4.3.2 Comparing the BPNNt and BPSVMt on one Recording

A final set of experiments was conducted for testing the BPNNt and BPSVMt tools on a single calcium imaging recording. As in previous experiments, the experiment's specifications are presented in Table 4.9, and the results are depicted in Figures 4.10, 4.11, 4.12 and 4.13. Lastly, we have included the evaluation metrics results in Tables 4.10 and 4.11.

Table 4.9 | Experiment specifications for executing the BPNNt and BPSVMt models.

Model	Model Execution ID	No. of Videos	Experimental Session(s)	No. of Class Labels	Train/Validation Split Method
BPNNt	Exp_4.1	1	Animal3, day11	6 (Behaviour)	5-Fold CV
BPNNt Chance	Exp_4.2	1	Animal3, day11	6 (Behaviour)	5-Fold CV
BPSVMt	Exp_4.3	1	Animal3, day11	6 (Behaviour)	5-Fold CV
BPSVMt Chance	Exp_4.4	1	Animal3, day11	6 (Behaviour)	5-Fold CV

Exp_4.1: BPNNt results

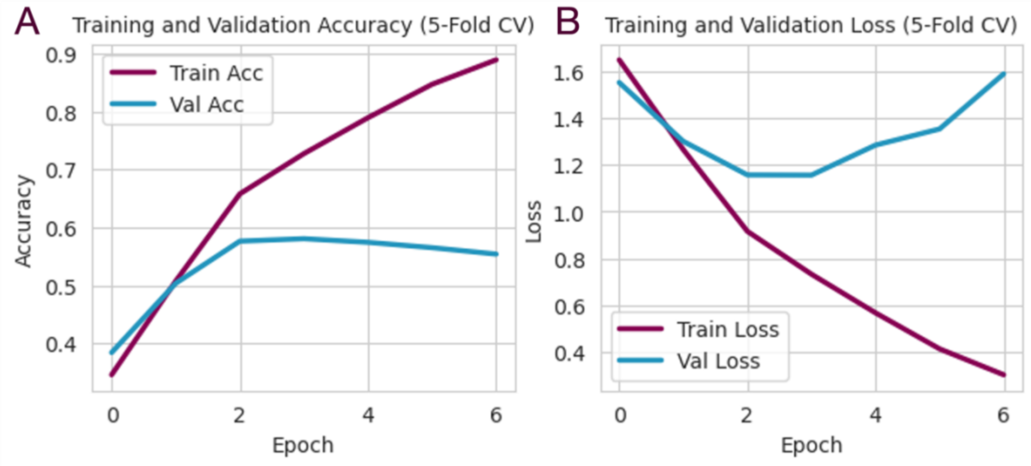


Figure 4.10 | Accuracy (A) and Loss (B) for training and validation data averaged across 5 folds (BPNNt, 1 CI video).

Exp_4.2: BPNNt Chance-level

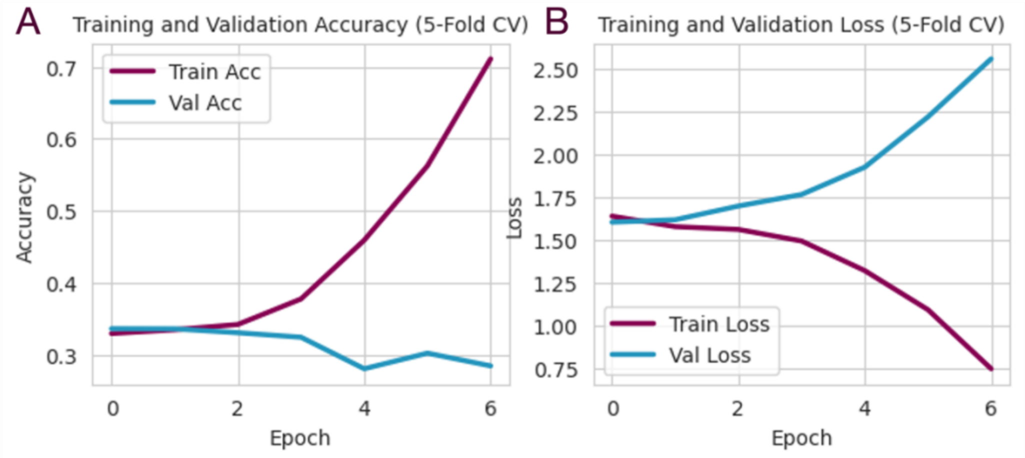


Figure 4.11 | Chance accuracy (A) and chance loss (B) for training and validation data averaged across 5 folds (BPNNt chance, 1 CI video).

Exp_4.3 / Exp_4.4: BPSVMt and Chance-level results

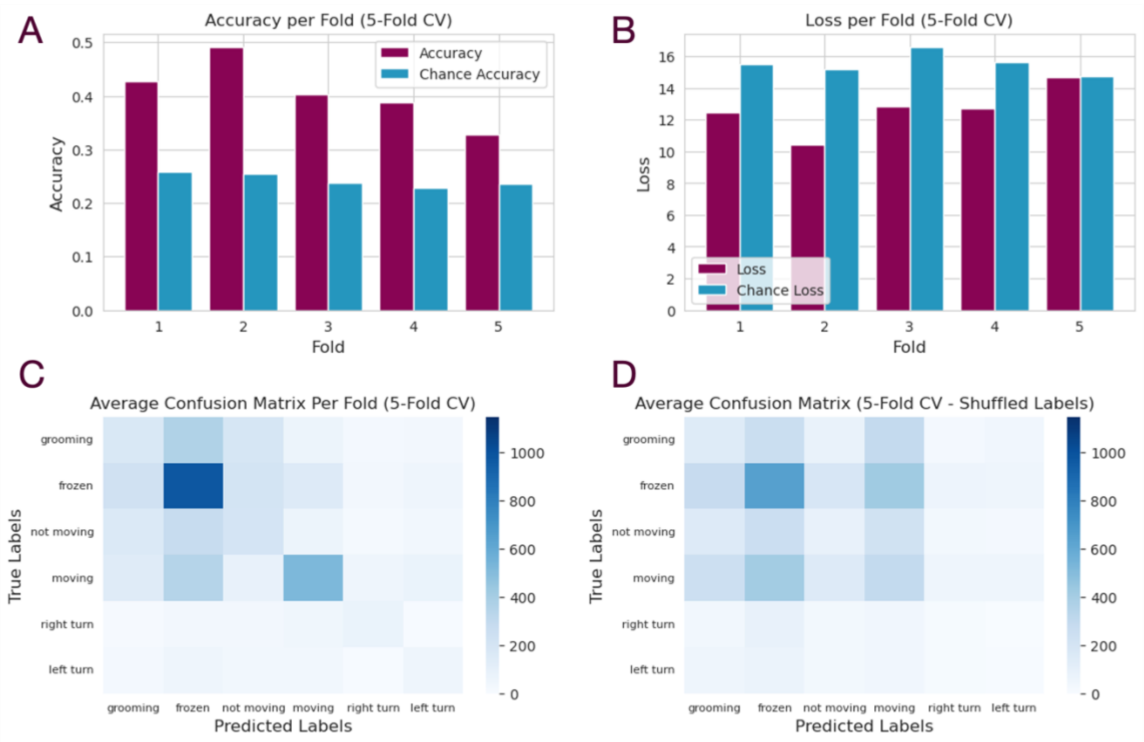


Figure 4.12 | Accuracy, loss, and confusion matrix for BPSVMt and BPSVMt-chance with shuffled labels.

A. Accuracy and chance-level accuracy averaged over 5 folds.
B. Loss and chance loss averaged over 5 folds.
C. Confusion matrix for 6 behaviour labels
D. Confusion matrix for 6 shuffled behaviour labels

The version of model that performs the best is highlighted in **bold font** in Table 4.10. Compared to executing the BPNNt on multiple videos, training only on one video resulted in the highest F1-score across all model configurations (**0.56)**. The BPSVMt, when executed for predicting behaviour labels from one calcium imaging video, managed to achieve an F1-score of **0.41**, which is comparatively lower than the BPNNt.

Table 4.10 | Comparing the BPNNt and chance level evaluation metrics.

Model	Training Accuracy (%)	Validation Accuracy (%)	F1-score
BPNNt	**88.90**	**55.39**	**0.56**
BPNNt-Chance	71.08	28.42	0.27

E.

Table 4.11 | Comparing the BPSVMt and chance level evaluation metrics.

Model	Accuracy (%)	F1-score
BPSVMt	42.94	0.41
BPSVMt Chance	25.14	0.24

4.4 BPNN Best Performing Configuration

Table 4.10 depicts the best-performing model configuration of the BPNN, the BPNNt when trained with one calcium imaging video (Day 11). Fig. 4.13A, shows the confusion matrix of this model configuration and the F1-score per label. It can be observed from the figure that the BPNNt, under these training conditions, is able to predict with higher accuracy the movements *Frozen* (**0.61**) and *Moving* (**0.71**) from the rest of the labels (*Grooming, Not Moving, Right Turn, and Left Turn*) which have a significantly lower representation in the dataset as seen in Fig. 4.13B. Interestingly, the less-represented labels are still moderately predicted (for example, the *Right Turn* with a low representation of **4.3%** out of all labels, achieves an F1-score of **0.48**). In addition, the model seems to confuse the labels *grooming* and *frozen.* A biological explanation for this could be that these two behaviours are physically similar, potentially challenging the model to distinguish them (See Fig. 8.5 in the Appendix).

Continuing, Figure 4.14, depicts the precision, recall, and F1-score of each class label in comparison to chance-level performance. Especially for the *left turn* and *right turn* labels, the BPNN model performs many-fold times better than chance-level reflecting the interpretation that it is easier for the model to differentiate left from right turns.

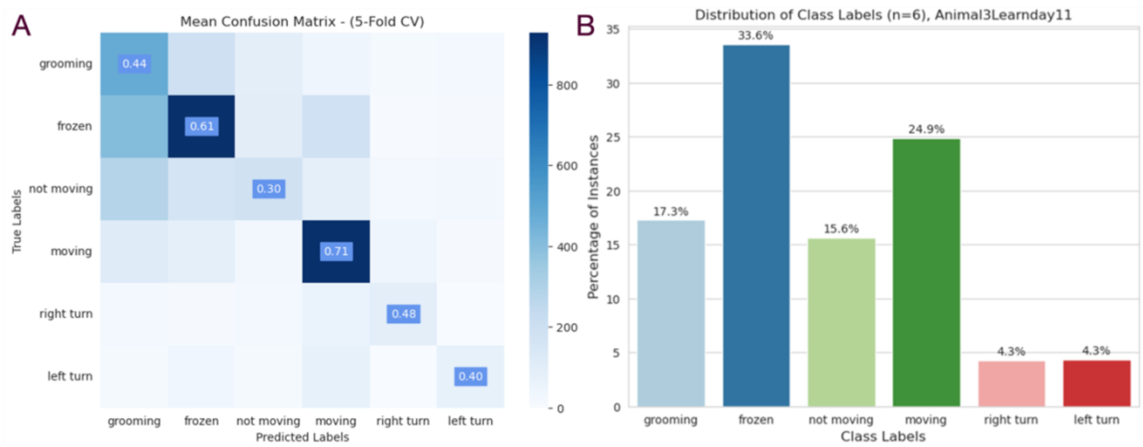


Figure 4.13 | Confusion matrix on the BPNNt when provided with one CI video as input with 6 behaviour labels.

Note | **A:** This is the configuration of the BPNN that provides the best possible output out of all illustrated configurations. The confusion matrix allows us to visualise to what extent the model has been able to correctly predict mouse behaviour based on the provided data for validation. The F1-score per label has been added to demonstrate which behaviour classes the BPNNt is performing the best and for which ones the weakest.
B: The label class distribution from the calcium imaging dataset of Animal 3 and the experimental session from Day 11 is provided to give more context on how the percentage of each behaviour label (high or low representation) affects the performance (high or low F1-score) of the model.

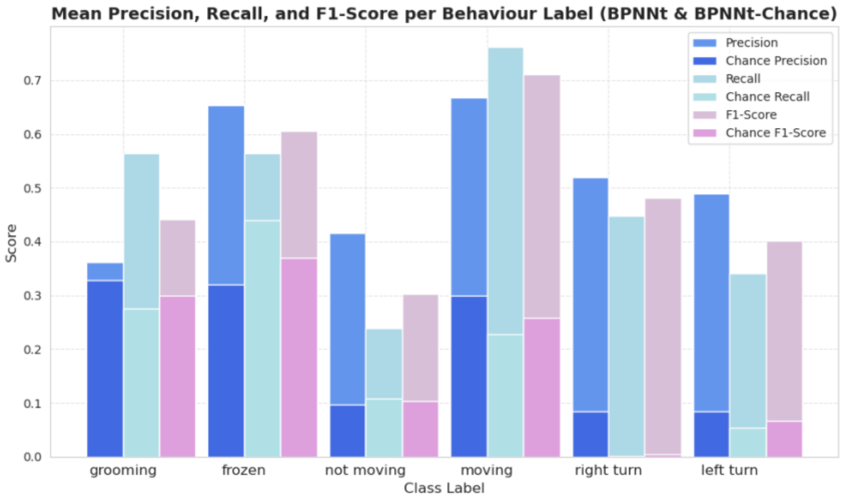


Figure 4.14 | Mean precision, recall, and F1-score per class label of the best performing model (BPNNt & BPNNt-Chance).

4.4.1 Assessing the Best Model Configuration on a Different Mouse

To test the model's performance across animals, we used a calcium imaging dataset from a different experimental session involving Animal 2, which performed the same behavioural task. For the same learning day (Day 11) of the experiment, Animal 2 has the following behavioural distribution (Fig, 3.5, Right). Also, in the same Figure, the reader can appreciate the differences in the behaviour distribution of both animals reflecting on the extent to which they perform the task correctly.

When executing the BPNNt, under the same specifications as in *Exp_4.1* (Table 4.9), we found that the BPNN for Animal 2 underperforms compared to Animal. The training accuracy is **76.94**, compared to **88.90,** and the validation accuracy is **42.85,** compared to **55.39** for Animal 3.

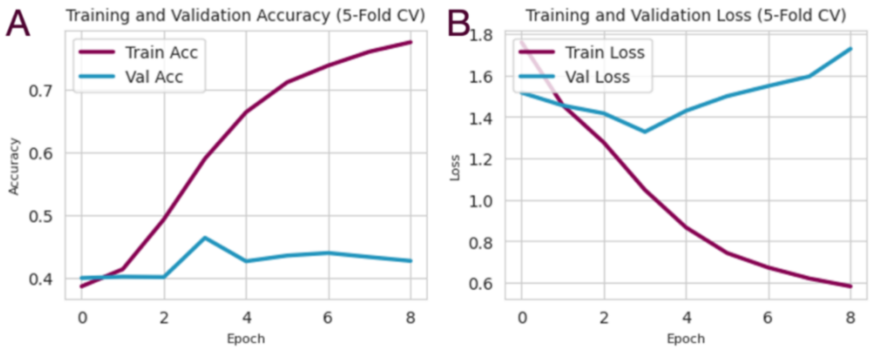


Figure 4.15 | Accuracy (A) and Loss (B) for training and validation data averaged across 5 folds (BPNNt, 1 CI video, Animal 2).

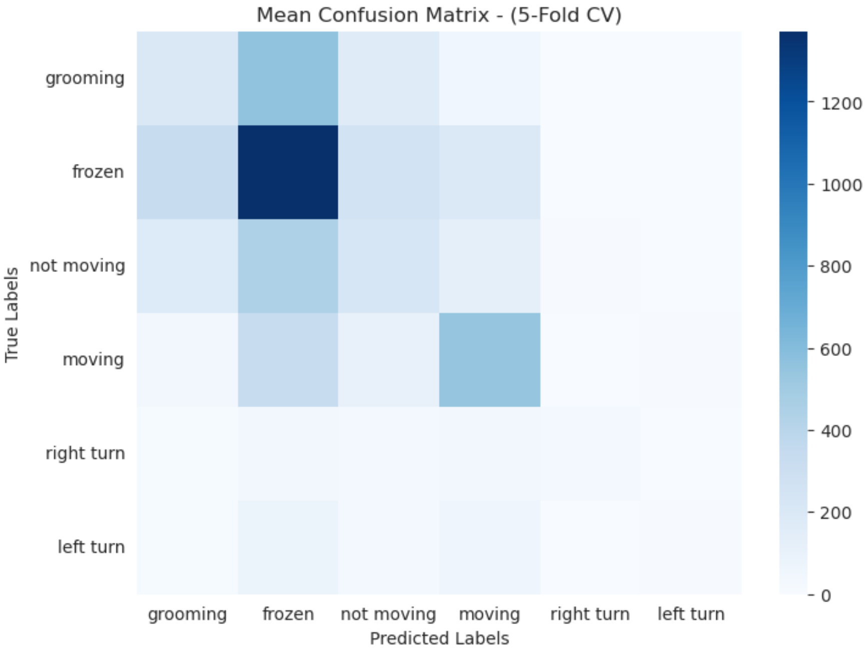


Figure 4.16 | Confusion matrix on the BPNNt when provided with one CI video as input from Animal 2 with 6 behaviour label classifications.

5 Discussion

The purpose of this thesis is to develop a new tool that utilizes convolutional neural networks to analyse calcium imaging recordings. We have called this tool, the BPNN (Behaviour Prediction Neural Network) which aims to establish correlations between animal behaviour and neural activity during experiments that investigate PD-related brain circuits, specifically the Basal Ganglia. For this study, Parkinson's disease has been mentioned as a use case of a potential disease area where the BPNN tool could deliver potential impact by improving calcium imaging analysis and assisting the discovery process of potential disease biomarkers relevant for PD diagnosis or PD treatment.

Most calcium imaging analysis techniques use machine learning algorithms to aid researchers interpret neural activity from animal behavioural experiments, either in freely moving assays or fixated experiments (50). However, these pipelines often require users to define several parameters to ensure high-quality and valid results. To the best of our knowledge, limited research has been pursued in inferring behaviour correlations directly from CI movies leading to a potential knowledge gap. We postulate that there is a lack of a simple CI analysis tool that can utilise the latest advancements in deep learning, namely CNNs, to predict animal behaviour without requiring researchers to understand complex pipelines and how to set their required parameters. Some of these parameters include motion artefact correction, baseline, and peak detection parameters for identifying fluorescence signals, and defining ROI for detecting neuronal activity (14,26,92).

To address this knowledge gap, the BPNN requires the user to provide the raw CI recording and the corresponding behaviour labels of the experimental session for which they wish conduct analysis. To revisit the research objectives mentioned in Sec. 1.3, in this thesis, we pre-process raw CI recordings acquired from a behavioural experiment (experiment details in Sec. 2.1.3.1) by developing a simple Sequential architecture commonly used in image classification tasks (118) and trained the model using high-performance computational equipment available at the K. Meletis Group (see Table 8.4 for hardware specifications). Continuing, we trained the model using different combinations of CI data configurations, such as multiple or individual CI recordings, and separated them into training and validation parts using the 5-Fold cross-validation train-split method. We further optimised the CNN model by experimenting with different architectures such as VGG16 (122), including transfer learning approaches such as a ResNet 50 (123) and InceptionV3 (124) without noteworthy success. We also compared the performance of the BPNN with that of the BPSVM, an SVM-based model currently used in the K. Meletis Group for associating calcium traces with behaviour. To better articulate the results in relation to the defined aims and objectives, we outlay the research questions, as formulated in Sec. 1.4, and the extent to which they have been addressed in this study:

1) *To what extent is direct behaviour prediction in mice possible from in-vivo calcium imaging videos by omitting the step of calcium trace extraction and using deep learning methods such as CNNs?*
The BPNN model achieved moderate results in predicting mice behaviour from CI recordings; however, as indicated in Fig. 3.5, left, the BPNN is more successful in predicting the most represented behaviours of the input dataset such as Frozen (33.6% of all labels) and Moving (24.9% of all labels). In addition, Table 4.10 shows that the BPNN tool achieved the highest overall F1 score **(0.56)** when trained only on one CI recording. For the first series of model runs (*Exp_1.1*, *Exp_1.2*, and *Exp_1.3* as shown in Table 4.1), only data from the 11th day of learning were utilised for training the BPNN model, as the mouse performed the task better than the previous days. Fig. 3.4 validates this observation by illustrating the behaviour label distribution during the experimental session. Most labels concern the *main* and *left corridors* as the mouse navigates the Arrow Maze from the *Initiation Spout* to *Spout A*, suggesting that it has learned the task. However, the BPNN tool showed signs of overfitting on the training data and poorly performing on the smaller validation dataset (See Fig. 4.1, 4.2, and 4.3).

We tried to overcome the limitation by training the model with more CI training data. Specifically, we concatenated CI recordings from different experimental sessions of the same mouse across multiple days (see Table 4.3). However, we were surprised to find that this resulted in a decrease in the performance of the BPNN from an F1-score of **0.55** to **0.44** when using more data (Table 4.4). As a result, training the model with more data does not improve its performance. In Sec. 4.4, we demonstrate that the best configuration of the BPNN produces an F1-score of **0.56**, however, this score needs to be considered in accordance with two parameters: the F1-score per class label, and the percentage of instances per class label.

When observing the results in Fig. 4.13, it can be concluded that the BPNN model can confidently predict, to a large extent, behaviours such as *Frozen* and *Moving*, which have the highest representation in the data and moderately predict labels such as *left turn* and *right turn* that even though are less represented in the dataset, still achieve a moderate F1-score.

2) *To what extent can we maximise the amount of neuronal information acquired from calcium imaging recordings to gain a better understanding of neural activity in brain circuits of interest?*

Current CI analysis tools that employ CNNs (52,95) are focused on cell body identification in calcium imaging recordings. However, our approach takes advantage of all the pixels of the input frame with the aim of training the CNN model to identify neuronal activity from axonal and dendrite fluorescence. Exploiting the entire field of view from raw CI input may allow us to better understand the functional mechanisms in a circuit of interest such as the Basal Ganglia. As a result, the BPNN truly maximises the amount of neuronal information that can be utilised to train a CNN model and potentially arrive at behavioural correlations; however, further research is required for interpreting and validating this statement.

5.1 Main Findings and Evaluation of Results

A series of experiments are demonstrated in Chapter 4 to evaluate the performance of the BPNN. These experiments include comparisons among the BPNN, the BPSVM, and the chance-level output of each model. Both models were trained on the same input data configurations to ensure that the results were consistent and comparable. The input dataset was split into training and validation sets using the 5-Fold cross-validation method as described in Sec. 3.3.1 allowing us to acquire a more reliable estimate of the model's results.

The architecture utilised for the BPNN across all experiments has been the Sequential architecture as described in Sec. 3.3.2. For each experiment showcased in Chapter 4, we have provided tables of the execution specifications such as the kind of input data, the number of class labels, the data splitting method employed, as well as the number of videos included for training the CNN model. The evaluation metrics used across all experiments are *Accuracy*, *Loss*, *F1-score*, and *confusion matrices*.

Initially, in Sec. 4.1, we tested the BPNN on one CI recording with three different label configurations to identify which labels produce the most reliable results. We proceeded with the 6-behaviour label configuration that achieved an F1-score of **0.54** compared to **0.55** of the 3-behaviour label configurations due to the former being able to predict more behaviours from the same calcium imaging dataset. However, as is clearly shown in Fig. 4.3, the BPNN overfits on the training data and performs poorly on the validation data. To address the overfitting problem, we trained the model with 3 concatenated CI videos from different experimental sessions of the same animal. We tested its performance to account for any increase in the validation accuracy and the F1 score. However, and to our surprise, the F1-score averaged lower (**0.44**) in this experiment (Table 4.4) compared to the first experiment (**0.55**) (Table 4.2), leading us to conclude that either the behaviour labels that correspond to specific neuronal activation patterns differ across sessions or that that raw CI video of the different experimental sessions contains conflicting motion artefacts or other displacements as a result from the physical detachment and reattachment of the miniscope in the animal's head. This can also be interpreted by the training accuracy curve depicted in Fig. 4.4, showing the model struggling to learn from the training data compared to the previous experiment 1.3 (Fig. 4.3), where the training curve increases more rapidly.

Another approach was taken to minimise the potential of overfitting by adding the time dimension in the input shape of the dataset. This modification was introduced in the BPNN and adopted on the BPSVM for comparison reasons and is showcased in Sec. 4.3.2. The reason for adopting this new approach was to capture the sequence of fluctuations in cell fluorescence present in the frames of the CI video. However, as illustrated in Tables 4.4 and 4.7, a moderate increase in the F1-score was only identified (from **0.44** to **0.46**), which enhanced the suspicions that the overfitting issue cannot be addressed by increasing the size of the dataset by incorporating videos across different experimental sessions nor by including the time dimension in the input shape. Ultimately, the best results were yielded from using the time dimension only on one CI recording (see Sec. 4.3.2).

It is particularly worth noting that the BPNN averaged higher results than the BPSVM in the same experiments (e.g., executions with one or multiple calcium videos and with the modification of adding the time dimension in the input shape). The most significant difference between the two models is demonstrated in Sec. 4.3.2, where the BPSVM averaged an F1-score of **0.41** and the BPNN an F1-score of **0.56,** leaving room for interpretation that the BPNN can perform similarly well or slightly better from a support vector machine model under the same testing conditions. In a further comparison, we validated that both BPNN and BSVMS models perform better than their chance-level counterparts throughout the experiments presented in Sec. 4.2 and 4.3.

Furthermore, we conducted Exp_4.1 with a different CI dataset from another animal to inspect performance across animals. In Fig. 3.5, we can see that Animal 2 has a lower representation of the *moving* behaviour state compared to Animal 3, resulting in a lower validation accuracy score of **42.85,** compared to **55.39** in Animal 3. This indicates that the BPNN's success may depend on the mouse's behaviour; for example, if the mouse performs all types of behaviour on a similar frequency, then the dataset has a better representation distribution of the behaviour labels, potentially yielding better results.

Overall, the CNN implementation struggles to achieve a high validation accuracy on unseen data samples resulting in low-performance metrics. Increasing the input data size from one calcium video to 3 negatively impacts the model's performance. Adding the time dimension into the BPNN when training on multiple videos does not substantially improve the performance (from **0.44** to **0.46**) nor when training on a single video (from **0.55** to **0.56**).

5.2 The Novelty of the Thesis

Several research groups have attempted to use the latest advancements in ML and DL to improve CI analysis workflows (52,101,126–128). While their implementations have focused on enhancing the calcium trace extraction process or using CNNs to identify cell bodies, this thesis takes an alternative approach to CI image analysis. We have achieved moderate performance by using CNNs to interpret the behaviour without conducting the calcium trace extraction step; however, more work is required to exhaust the potential of this method. Nonetheless, to the best of our knowledge, this is the first indication of a method that employs CNNs to analyse CI data to predict animal behaviour. Compared to the benchmark model, the BPSVM, the BPNN achieves similar or better results under the same testing conditions in all experimental runs, indicating that it can infer behaviour from CI recordings without performing calcium trace extraction. A second novel feature of this study is that the CNN model is trained to extract patterns from neuronal activity across the entire frame, including additional neuronal information that may be included in dendrites and axons.

5.3 Limitations

One limitation of this study is that the model's performance depends on the biological research question that defines the structure of the behavioural experiment. If the experiment results in the mouse performing behaviours of high variance (see Fig. 3.3), this imbalance in the class representation can cause the model to overfit the training data, making it challenging to accurately predict behaviour from the validation data. While data augmentation can help address overfitting in many ML problems, it's essential to avoid it in this case, as modifying the CI data in any way will result in a loss of biological relevance and information.

On a more technical level, when preparing the animal for a behavioural assay, researchers attach and detach the miniscope to the mouse's head before and after each session, respectively (as described in Sec. 2.1.3). However, the miniscope’s field of view may vary, which can, in turn, affect the BPNN's ability to extract patterns and features from the CI data. A process for aligning the field of view across different days has been applied to address this limitation. Despite this, differences in focus or brightness in the lens may still challenge the BPNN's ability to recognise the same field of view across sessions, potentially explaining the underperformance of the BPNN when concatenating multiple videos together, as illustrated in Sec. 4.2. To address this; attempts were made to improve the model's performance, such as introducing more dropout layers to the Sequential architecture or using a Keras tuner to define the best architecture hyperparameters automatically (129). However, these approaches were unsuccessful, as the Keras tuner proposed the same architecture that was already in use.

5.4 Future Research

This study has moderately exhibited the ability of CNN models to be employed in image processing tasks such as recognising neuronal activation patterns in CI data and producing correlates with behavioural labels. However, there are possible directions that could deepen the investigation of CNN applicability in this task. For example, due to time limitations, this study hasn't elaborately explored applying a transfer learning approach in depth. Yet, a transfer learning approach could improve the model's ability to generalise on unseen data and increase performance. Another possible action point would be to use a different kind of CI training data, such as two-photon (2p) calcium imaging, where recording in greater depths is possible.

Continuing Parkinson’s is a complex neurodegenerative disease of unknown cause, and using neural activity from populations of interest in the Basal Ganglia may be a daunting application for testing the BPNN’s efficacy. A future recommendation would be to utilise calcium imaging data from brain areas where neural activity related to behavioural output is more investigated. This may make it easier for the model to extract patterns and features and arrive at behavioural correlations.

6 Conclusions

Neurodegenerative diseases, such as Parkinson's disease, affect millions of people globally and are the centre of focus in neuroscience research as no available drug can halt the degeneration of dopamine cells. To delve into the neural circuits relevant to Parkinson's disease, researchers employ behavioural assays to study the manifestation of disease symptoms in animal models. To image neural activity during these behavioural assays, calcium imaging, one of many neuroimaging techniques, allows us to study the activity of specific neuronal populations and their relation to behaviour during task engagement. However, calcium imaging techniques present several limitations that challenge researchers from effectively establishing behavioural correlations such as difficulty in demixing and acquiring spatially overlapping cells, low spatial resolution from the miniscope, and complex parameter set-up. To address these challenges, we explored minimising the burden of pre-processing CI videos by directly providing the raw CI recording to a CNN. In accordance with this aim, a series of calcium imaging datasets acquired from previously completed behavioural assays were repurposed, minimally pre-processed, and used as input to train the proposed CNN tool, the BPNN (Behaviour Prediction Neural Network).

Furthermore, we developed an additional tool, the BPSVM (Behaviour Prediction Support Vector Machines), to compare and evaluate the BPNN's performance with state-of-the-art methods. Multiple experiments were conducted to test the BPNN's ability to predict behaviour; however, we met challenges relating to overfitting. Adding more data to the model worsens the model's performance, which we hypothesise is the result of either technical inconsistencies as the result of the displacement of the miniscopes in the mouses head, or that the neural activity tunings to behaviour change between experimental sessions. However, as illustrated in Fig. 4.13, the best-performing model achieved an F1-Score of **0.56** and was able to produce behaviour correlates of the mouse's activity, such as "*frozen*" with an F1-score of **0.61** and "*moving*" with an F1-score of **0.71**, indicating that there is potential for CNNs to predict behaviour in animals during assays.

All in all, utilising CNNs as an alternative method for analysing behaviour in animal assays is novel idea that to the best of our knowledge has not surfaced in the literature. However, the several technical challenges we encountered during the development of this tool mandates further research to establish the extent to which CNNs are applicable for predicting behaviour from calcium imaging recordings.

7 References

1. Johns P. Parkinson’s disease. Clinical Neuroscience [Internet]. 2014 Jan 1 [cited 2023 Feb 3];163–79. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780443103216000138
2. Parkinson J. An Essay on the Shaking Palsy. J Neuropsychiatry Clin Neurosci [Internet]. 2002;14(2):223–36. Available from: https://neuro.psychiatryonline.org/doi/abs/10.1176/jnp.14.2.223
3. MONTAGU KA. Catechol Compounds in Rat Tissues and in Brains of Different Animals. Nature. 1957 Aug;180(4579):244–5.
4. CARLSSON A, LINDQVIST M, MAGNUSSON T, WALDECK B. On the Presence of 3-Hydroxytyramine in Brain. Science (1979). 1958 Feb 28;127(3296):471–471.
5. L-Dopa for Parkinsonism. New England Journal of Medicine. 1968 Mar 14;278(11):630–630.
6. Understanding Parkinsons - Statistics. [cited 2023 Feb 3]; Available from: https://www.parkinson.org/understanding-parkinsons/statistics
7. Kowal SL, Dall TM, Chakrabarti R, Storm M V., Jain A. The current and projected economic burden of Parkinson’s disease in the United States. Movement Disorders. 2013 Mar;28(3):311–8.
8. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, et al. Parkinson disease. Nat Rev Dis Primers. 2017 Mar 23;3(1):17013.
9. Johns P. Parkinson’s disease. In: Clinical Neuroscience. Elsevier; 2014. p. 163–79.
10. Pellicano C, Benincasa D, Pisani V, Buttarelli FR, Giovannelli M, Pontieri FE. Prodromal non-motor symptoms of Parkinson’s disease. Neuropsychiatr Dis Treat. 2007 Feb;3(1):145–51.
11. Mahlknecht P, Seppi K, Poewe W. The Concept of Prodromal Parkinson’s Disease. J Parkinsons Dis [Internet]. 2015 Oct 17;5(4):681–97. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JPD-150685
12. Balleine BW. Animal models of action control and cognitive dysfunction in Parkinson’s disease. In 2022. p. 227–55.
13. Iwasaki S, Ikegaya Y. In vivo one-photon confocal calcium imaging of neuronal activity from the mouse neocortex. J Integr Neurosci. 2018 Sep 12;17(3–4):671–8.
14. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, et al. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron. 2016 Jan;89(2):285–99.
15. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, et al. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron [Internet]. 2016 Jan;89(2):285–99. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0896627315010843
16. Dong Z, Mau W, Feng Y, Pennington ZT, Chen L, Zaki Y, et al. Minian, an open-source miniscope analysis pipeline. Elife. 2022 Jun 1;11.

17. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. PLoS One. 2019 Feb 19;14(2):e0212356.
18. Lee CW, Park JA. Assessment of HIV/AIDS-related health performance using an artificial neural network. Information & Management. 2001 Feb;38(4):231–8.
19. Sordo M. Introduction to neural networks in healthcare. Open Clinical Document. 2002;
20. Reynolds S, Abrahamsson T, Schuck R, Jesper Sjöström P, Schultz SR, Dragotti PL. Able: An activity-based level set segmentation algorithm for two-photon calcium imaging data. eNeuro. 2017 Sep 1;4(5).
21. Hattori R, Komiyama T. PatchWarp: Corrections of non-uniform image distortions in two-photon calcium imaging data by patchwork affine transformations. Cell Reports Methods. 2022 May 23;2(5).
22. Bao Y, Soltanian-Zadeh S, Farsiu S, Gong Y. Segmentation of neurons from fluorescence calcium recordings beyond real time. Nat Mach Intell [Internet]. 2021 May 20;3(7):590–600. Available from: https://www.nature.com/articles/s42256-021-00342-x
23. Gauthier JL, Koay SA, Nieh EH, Tank DW, Pillow JW, Charles AS. Detecting and correcting false transients in calcium imaging. Nat Methods. 2022;19(4):470–8.
24. Sità L, Brondi M, Lagomarsino de Leon Roig P, Curreli S, Panniello M, Vecchia D, et al. A deep-learning approach for online cell identification and trace extraction in functional two-photon calcium imaging. Nat Commun [Internet]. 2022 Mar 22;13(1):1529. Available from: https://www.nature.com/articles/s41467-022-29180-0
25. Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, et al. CaImAn an open source tool for scalable calcium imaging data analysis. Kleinfeld D, King AJ, editors. Elife [Internet]. 2019;8:e38173. Available from: https://doi.org/10.7554/eLife.38173
26. Giovannucci A, Friedrich J, Kaufman M, Churchland A, Chklovskii D, Paninski L, et al. OnACID: Online Analysis of Calcium Imaging Data in Real Time. bioRxiv [Internet]. 2017 Jan 1;193383. Available from: http://biorxiv.org/content/early/2017/10/02/193383.abstract
27. Inan H, Schmuckermair C, Tasci T, Ahanonu BO, Hernandez O, Lecoq J, et al. Fast and statistically robust cell extraction from large-scale neural calcium imaging datasets. bioRxiv [Internet]. 2021 Jan 1;2021.03.24.436279. Available from: http://biorxiv.org/content/early/2021/03/27/2021.03.24.436279.abstract
28. Jewell S, Witten D. Exact spike train inference via ℓ_{0} optimization. Ann Appl Stat. 2018 Dec 1;12(4).
29. Mukamel EA, Nimmerjahn A, Schnitzer MJ. Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data. Neuron. 2009 Sep;63(6):747–60.
30. Soltanian-Zadeh S, Sahingur K, Blau S, Gong Y, Farsiu S. Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. Proceedings of the National Academy of Sciences [Internet]. 2019 Apr 23;116(17):8554–63. Available from: https://pnas.org/doi/full/10.1073/pnas.1812995116
31. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press;

32. Santos CFG dos, Papa JP. Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. 2022 Jan 10;

33. Rizzolatti G, Fabbri-Destro M, Caruana F, Avanzini P. System neuroscience: Past, present, and future. CNS Neurosci Ther [Internet]. 2018 Aug 1;24(8):685–93. Available from: https://doi.org/10.1111/cns.12997

34. Moore GE. Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. IEEE Solid-State Circuits Society Newsletter. 2006;11(3):33–5.

35. Stevenson IH, Kording KP. How advances in neural recording affect data analysis. Nat Neurosci [Internet]. 2011 Feb 26;14(2):139–42. Available from: http://www.nature.com/articles/nn.2731

36. Trautmann EM, Stavisky SD, Lahiri S, Ames KC, Kaufman MT, O'Shea DJ, et al. Accurate Estimation of Neural Population Dynamics without Spike Sorting. Neuron [Internet]. 2019;103(2):292-308.e4. Available from: https://www.sciencedirect.com/science/article/pii/S0896627319304283

37. Rey HG, Pedreira C, Quian Quiroga R. Past, present and future of spike sorting techniques. Brain Res Bull [Internet]. 2015;119:106–17. Available from: https://www.sciencedirect.com/science/article/pii/S0361923015000684

38. Pachitariu M, Stringer C, Harris KD. Robustness of Spike Deconvolution for Neuronal Calcium Imaging. J Neurosci. 2018 Sep;38(37):7976–85.

39. DeLong MR. Primate models of movement disorders of basal ganglia origin. Trends Neurosci. 1990 Jul;13(7):281–5.

40. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature. 2013 Jul 18;499(7458):295–300.

41. Tran LM, Mocle AJ, Ramsaran AI, Jacob AD, Frankland PW, Josselyn SA. Automated Curation of CNMF-E-Extracted ROI Spatial Footprints and Calcium Traces Using Open-Source AutoML Tools. Front Neural Circuits. 2020 Jul 15;14.

42. Etter G, Manseau F, Williams S. A Probabilistic Framework for Decoding Behavior From in vivo Calcium Imaging Data. Front Neural Circuits. 2020;14:19.

43. Santos CFG dos, Papa JP. Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. 2022 Jan 10;

44. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. IEEE J Biomed Health Inform. 2017 Jan;21(1):4–21.

45. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. IEEE J Biomed Health Inform. 2017 Jan;21(1):4–21.

46. Abbas W, Masip D. Computational Methods for Neuron Segmentation in Two-Photon Calcium Imaging Data: A Survey. Applied Sciences [Internet]. 2022 Jul;12(14):6876. Available from: https://www.mdpi.com/2076-3417/12/14/6876

47. Flusberg BA, Nimmerjahn A, Cocker ED, Mukamel EA, Barretto RPJ, Ko TH, et al. High-speed, miniaturized fluorescence microscopy in freely moving mice. Nat Methods. 2008 Nov 5;5(11):935–8.

48. Svoboda K, Denk W, Kleinfeld D, Tank DW. In vivo dendritic calcium dynamics in neocortical pyramidal neurons. Nature. 1997 Jan;385(6612):161–5.

49. Russell JT. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. Br J Pharmacol. 2011 Aug;163(8):1605–25.

50. Stamatakis AM, Resendez SL, Chen KS, Favero M, Liang-Guallpa J, Nassi JJ, et al. Miniature microscopes for manipulating and recording in vivo brain activity. Microscopy. 2021 Oct 5;70(5):399–414.

51. Juczewski K, Koussa JA, Kesner AJ, Lee JO, Lovinger DM. Stress and behavioral correlates in the head-fixed method: stress measurements, habituation dynamics, locomotion, and motor-skill learning in mice. Sci Rep. 2020 Jul 22;10(1):12245.

52. Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, et al. CaImAn an open source tool for scalable calcium imaging data analysis. Kleinfeld D, King AJ, editors. Elife [Internet]. 2019;8:e38173–e38173. Available from: https://doi.org/10.7554/eLife.38173

53. Lu J, Li C, Singh-Alvarado J, Zhou ZC, Fröhlich F, Mooney R, et al. MIN1PIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline. Cell Rep. 2018 Jun;23(12):3673–84.

54. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci. 2018 Sep 20;21(9):1281–9.

55. Franzese M, Iuliano A. Hidden markov models. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. 2018 Jan 1;1–3:753–62.

56. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press;

57. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436–44.

58. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. 2020 Jun 17;

59. Sutton RS& BAG. Reinforcement Learning: An Introduction . Second Edi. MIT Press, Cambridge, Massachusetts, 2018;

60. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436–44.

61. Farquad MAH, Ravi V, Raju SB. Churn prediction using comprehensible support vector machine: An analytical CRM application. Appl Soft Comput. 2014 Jun;19:31–40.

62. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–44.

63. Ruder S. An overview of gradient descent optimization algorithms. 2016 Sep 15;

64. Bashir D, Montanez GD, Sehra S, Segura PS, Lauw J. An Information-Theoretic Perspective on Overfitting and Underfitting. 2020 Oct 12;

65. tf.keras.losses.CategoricalCrossentropy [Internet]. [cited 2023 May 17]. Available from: https://www.tensorflow.org/api_docs/python/tf/keras/losses/CategoricalCrossentropy

66. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014 Dec 22;

67. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res. 2014 Jan;15(1):1929–58.

68. O’Shea K, Nash R. An Introduction to Convolutional Neural Networks. Int J Res Appl Sci Eng Technol [Internet]. 2015 Nov 26 [cited 2023 May 22];10(12):943–7. Available from: https://arxiv.org/abs/1511.08458v2

69. Mendeley Reference Manager for Desktop [Internet]. [cited 2023 Mar 2]. Available from: https://www.mendeley.com/download-reference-manager/windows

70. Eldawlatly A, Alshehri H, Alqahtani A, Ahmad A, Al-Dammas F, Marzouk A. Appearance of Population, Intervention, Comparison, and Outcome as research question in the title of articles of three different anesthesia journals: A pilot study. Saudi J Anaesth. 2018;12(2):283.

71. PubMed [Internet]. [cited 2023 Feb 20]. Available from: https://pubmed.ncbi.nlm.nih.gov/

72. Web of Science [Internet]. [cited 2023 Feb 20]. Available from: https://www.webofscience.com/wos/woscc/basic-search

73. IEEE Xplore [Internet]. [cited 2023 Feb 20]. Available from: https://ieeexplore.ieee.org/Xplore/home.jsp

74. ACM Digital Library [Internet]. [cited 2023 Feb 20]. Available from: https://dl.acm.org/

75. Google Scholar [Internet]. [cited 2023 Feb 20]. Available from: https://scholar.google.com/

76. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. The FASEB Journal. 2008 Feb 20;22(2):338–42.

77. Medical Subject Headings [Internet]. Available from: https://www.nlm.nih.gov/mesh/meshhome.html#:~:text=The%20Medical%20Subject%20Headings%20(MeSH,biomedical%20and%20health%2Drelated%20information.

78. Citation Chaining [Internet]. [cited 2023 Feb 21]. Available from: https://guides.erau.edu/citation-chaining

79. Soltanian-Zadeh S, Sahingur K, Blau S, Gong Y, Farsiu S. Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. Proc Natl Acad Sci U S A. 2019;116(17):8554–63.

80. Sità L, Brondi M, Lagomarsino de Leon Roig P, Curreli S, Panniello M, Vecchia D, et al. A deep-learning approach for online cell identification and trace extraction in functional two-photon calcium imaging. Nat Commun. 2022 Dec 1;13(1).

81. Li C, Chan DCW, Yang X, Ke Y, Yung WH. Prediction of Forelimb Reach Results From Motor Cortex Activities Based on Calcium Imaging and Deep Learning. Front Cell Neurosci [Internet]. 2019 Mar;13. Available from: https://www.frontiersin.org/article/10.3389/fncel.2019.00088/full

82. Arac A, Zhao PP, Dobkin BH, Carmichael ST, Golshani P. DeepBehavior: A Deep Learning Toolbox for Automated Analysis of Animal and Human Behavior Imaging Data. Front Syst Neurosci. 2019;13.

83. Denis J, Dard RF, Quiroli E, Cossart R, Picardo MA. DeepCINAC: A Deep-Learning-Based Python Toolbox for Inferring Calcium Imaging Neuronal Activity Based on Movie Visualization. eNeuro. 2020 Jul;7(4):ENEURO.0038–20.2020.

84. Taniguchi M, Tezuka T, Vergara P, Srinivasan S, Hosokawa T, Cherasse Y, et al. Open-Source Software for Real-time Calcium Imaging and Synchronized Neuron Firing Detection. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) [Internet]. IEEE; 2021. p. 2997–3003. Available from: https://ieeexplore.ieee.org/document/9629611/

85. Dong Z, Mau W, Feng Y, Pennington ZT, Chen L, Zaki Y, et al. Minian, an open-source miniscope analysis pipeline. Elife. 2022 Jun;11.

86. Spaen Q, Asín-Achá R, Chettih SN, Minderer M, Harvey C, Hochbaum DS, et al. HNCcorr: A Novel Combinatorial Approach for Cell Identification in Calcium-Imaging Movies. eNeuro [Internet]. 2017 Mar;6(2):ENEURO.0304–18.2019. Available from: https://www.eneuro.org/lookup/doi/10.1523/ENEURO.0304-18.2019

87. Soltanian-Zadeh S, Sahingur K, Blau S, Gong Y, Farsiu S. Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning. Proceedings of the National Academy of Sciences [Internet]. 2019 Apr;116(17):8554–63. Available from: https://pnas.org/doi/full/10.1073/pnas.1812995116

88. Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, et al. CaImAn an open source tool for scalable calcium imaging data analysis. Kleinfeld D, King AJ, editors. Elife [Internet]. 2019;8:e38173–e38173. Available from: https://doi.org/10.7554/eLife.38173

89. Zhou P, Resendez SL, Rodriguez-Romaguera J, Jimenez JC, Neufeld SQ, Giovannucci A, et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. Elife. 2018 Feb;7.

90. Lu J, Li C, Singh-Alvarado J, Zhou ZC, Fröhlich F, Mooney R, et al. MIN1PIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline. Cell Rep. 2018 Jun;23(12):3673–84.

91. Taniguchi M, Tezuka T, Vergara P, Srinivasan S, Hosokawa T, Cherasse Y, et al. Open-Source Software for Real-time Calcium Imaging and Synchronized Neuron Firing Detection. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) [Internet]. IEEE; 2021. p. 2997–3003. Available from: https://ieeexplore.ieee.org/document/9629611/

92. Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, et al. CaImAn an open source tool for scalable calcium imaging data analysis. Kleinfeld D, King AJ, editors. Elife [Internet]. 2019;8:e38173–e38173. Available from: https://doi.org/10.7554/eLife.38173

93. Sità L, Brondi M, Lagomarsino de Leon Roig P, Curreli S, Panniello M, Vecchia D, et al. A deep-learning approach for online cell identification and trace extraction in functional two-photon calcium imaging. Nat Commun [Internet]. 2022 Mar 22;13(1):1529. Available from: https://www.nature.com/articles/s41467-022-29180-0

94. Li C, Chan DCW, Yang X, Ke Y, Yung WH. Prediction of Forelimb Reach Results From Motor Cortex Activities Based on Calcium Imaging and Deep Learning. Front Cell Neurosci. 2019 Mar 12;13.

95. Apthorpe N, Riordan A, Aguilar R, Homann J, Gu Y, Tank D, et al. Automatic Neuron Detection in Calcium Imaging Data Using Convolutional Networks. 2016;

96. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, et al. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron. 2016;89(2):285–99.

97. Zhou P, Resendez SL, Rodriguez-Romaguera J, Jimenez JC, Neufeld SQ, Giovannucci A, et al. OnACID: Online Analysis of Calcium Imaging Data in Real Time. Kleinfeld D, King AJ, editors. Elife [Internet]. 2017 Apr;13(5):285–93. Available from: https://doi.org/10.7554/eLife.38173

98. Giovannucci A, Friedrich J, Kaufman M, Churchland A, Chklovskii D, Paninski L, et al. OnACID: Online Analysis of Calcium Imaging Data in Real Time&lt;a id=&quot;xref-fn-1-1&quot; class=&quot;xref-author-notes&quot; href=&quot;#fn-1&quot;&gt;*&lt;/a&gt; bioRxiv [Internet]. 2017 Jan 1;193383. Available from: http://biorxiv.org/content/early/2017/10/02/193383.abstract

99. Zhou P, Resendez SL, Rodriguez-Romaguera J, Jimenez JC, Neufeld SQ, Giovannucci A, et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. Elife. 2018 Feb 22;7.

100. Taniguchi M, Tezuka T, Vergara P, Srinivasan S, Hosokawa T, Cherasse Y, et al. Open-Source Software for Real-time Calcium Imaging and Synchronized Neuron Firing Detection. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021. p. 2997–3003.

101. Dong Z, Mau W, Feng Y, Pennington ZT, Chen L, Zaki Y, et al. Minian, an open-source miniscope analysis pipeline. Elife. 2022 Jun;11.

102. Arac A, Zhao PP, Dobkin BH, Carmichael ST, Golshani P. DeepBehavior: A Deep Learning Toolbox for Automated Analysis of Animal and Human Behavior Imaging Data. Front Syst Neurosci. 2019;13.

103. Denis J, Dard RF, Quiroli E, Cossart R, Picardo MA. DeepCINAC: A Deep-Learning-Based Python Toolbox for Inferring Calcium Imaging Neuronal Activity Based on Movie Visualization. eNeuro. 2020 Jul;7(4):ENEURO.0038–20.2020.

104. Li C, Chan DCW, Yang X, Ke Y, Yung WH. Prediction of Forelimb Reach Results From Motor Cortex Activities Based on Calcium Imaging and Deep Learning. Front Cell Neurosci [Internet]. 2019 Mar;13. Available from: https://www.frontiersin.org/article/10.3389/fncel.2019.00088/full

105. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015 Dec 10;

106. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. 2019 Nov 6;

107. Weglage M, Wärnberg E, Lazaridis I, Calvigioni D, Tzortzi O, Meletis K. Complete representation of action space and value in all dorsal striatal pathways. Cell Rep. 2021 Jul;36(4):109437.

108. Johannesson P, Perjons E. An Introduction to Design Science [Internet]. Cham: Springer International Publishing; 2014. Available from: https://link.springer.com/10.1007/978-3-319-10632-8

109. Johannesson P, Perjons E. An Introduction to Design Science. Cham: Springer International Publishing; 2014.

110. Ross SM, Morrison G. Experimental Research Methods. 2003.

111. vom Brocke J, Hevner A, Maedche A. Introduction to Design Science Research. In 2020. p. 1–13.

112. Welcome to Python.org [Internet]. [cited 2023 May 6]. Available from: https://www.python.org/
113. Project Jupyter | Home [Internet]. [cited 2023 May 6]. Available from: https://jupyter.org/
114. Conda — conda documentation [Internet]. [cited 2023 May 6]. Available from: https://docs.conda.io/en/latest/
115. GitHub: Let’s build from here · GitHub [Internet]. [cited 2023 May 6]. Available from: https://github.com/
116. The Sequential model [Internet]. [cited 2023 May 10]. Available from: https://keras.io/guides/sequential_model/
117. One Hot Encoding in Machine Learning - GeeksforGeeks [Internet]. [cited 2023 May 10]. Available from: https://www.geeksforgeeks.org/ml-one-hot-encoding-of-datasets-in-python/
118. The Sequential model [Internet]. [cited 2023 May 19]. Available from: https://keras.io/guides/sequential_model/
119. M H, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process. 2015 Mar 31;5(2):01–11.
120. Janocha K, Czarnecki WM. On Loss Functions for Deep Neural Networks in Classification. 2017 Feb 18;
121. Robbins M, Christensen CN, Kaminski CF, Zlatic M. Calcium imaging analysis – how far have we come? F1000Res. 2021 Aug 26;10:258.
122. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014 Sep 4;
123. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015 Dec 10;
124. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2015 Dec 1;
125. Russell JT. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. Br J Pharmacol. 2011 Aug;163(8):1605–25.
126. Spaen Q, Hochbaum DS, Asín-Achá R. HNCcorr: A Novel Combinatorial Approach for Cell Identification in Calcium-Imaging Movies. 2017 Mar 6;
127. Friedrich J, Zhou P, Paninski L. Fast online deconvolution of calcium imaging data. PLoS Comput Biol. 2017 Mar;13(3):e1005423.
128. Klibisz A, Rose D, Eicholtz M, Blundon J, Zakharenko S. Fast, Simple Calcium Imaging Segmentation with Fully Convolutional Networks. In 2017. p. 285–93.
129. KerasTuner [Internet]. [cited 2023 May 22]. Available from: https://keras.io/keras_tuner/
130. Zafar S, Yaddanapudi SS. Parkinson Disease. 2023.

8 Appendices

8.1 Examiner Response

Examiner: (Recommendations for the author):

Is it movement execution or movement ideation/initiation? Good to make this consistent.
On some occasions, neurons fire directly when a movement is executed (e.g., a left turn). However, it has been observed that certain neuronal patterns fire before movement execution as well. As mentioned in Sec. 3.3.1, Step 2, the behaviour labels are temporally aligned with the calcium imaging frames. Consequently, the neuronal firing that is associated with each behaviour label is directly time dependent. The BPNN model considers the frame-label pairs when training and as a result, its predictions only take into consideration neuronal firing patterns that relate to direct movement manifestations observed in the Arrow Maze task. It is however interesting to explore in the future the extent to which this alternative implementation of calcium imaging analysis enriches our understanding of how neural activity relates to behaviour on an initiation-basis which would require the consideration of dynamic time-bound neurobehavioural associations.

Explain what labels are you using for building your classifiers and make this very explicit when the results are presented.
In Sec. 3.3.1, Step 2, we describe the 3 different label configurations reported in this study. In the captions of Figures 3.4 and 3.5, we have added descriptions for the Merged-Spatial and Behaviour label configurations. For training the classifiers, we mention in page 41 that the Behaviour-type labels are used. In addition, Tables 4.3, 4.6, and 4.9 specify the use of Behaviour-type labels for the corresponding training session.

Your first research question only has one sub-question. You should either have at least two sub-questions or just merge them to a single question.
In Sec. 1.4, we have merged the sub-question with the parent research question and have updated the text in the Discussion where the research questions are addressed.

Providing the results on accuracy/f1-score on all labels together (as an average) when the dataset is imbalanced may be misleading. Even if the dataset is balanced this can also be misleading. Good to have the precision/recall per class label so that one can see for which classes the classifier is well-trained and for which ones it is weak.
We have added Figures 4.13 and 4.14 to address this recommendation. Indeed, in Fig. 4.13A, we can observe the F1-score per label. Fig. 4.13B, provides the context of the

label representation of the specific dataset (Animal3Learnday11). It can be concluded that the BPNN can perform moderate to confident predictions on the validation data for labels *Frozen* and *Moving.* In addition, given that for labels such as *Left Turn* and *Right Turn*, the number of instances is significantly lower (**4.3%** each compared to **33.6**% and **24.9%** for Frozen and Moving labels respectively), the F1-scores of **0.48** and **0.40** can be remarked as promising.

8.2 Extended Background

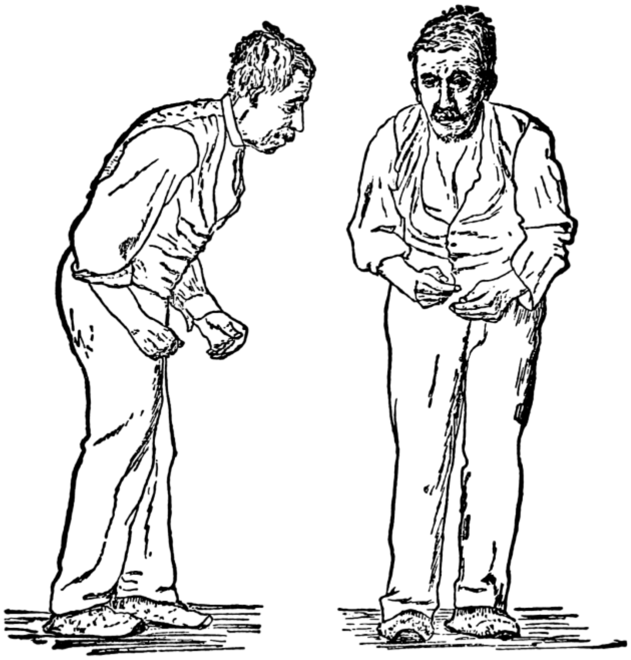


Figure 8.1 | PD symptoms observed in patients.

Note | Illustration by Sir William Richard Gowers, neurologist, researcher, and artist drawn in 1886 (130).

8.2.1 Literature Review

8.2.1.1 PICO Framework

This thesis addresses the complexity of current calcium imaging analysis workflows. To investigate existing solutions in the field, the search terms *"calcium imaging", "calcium activity", "calcium trace*",* and *"calcium transient*"* were selected. A potential path in alleviating the problem of complex pipelines in processing large-scale calcium imaging recordings is the application of deep learning techniques. The keywords *"deep learning", "convolutional neural network*",* and *CNN** were selected. More keywords were considered in the *Comparison* and *Outcome* PICO components; however, their inclusion was found to limit the range of found publications. Therefore, keywords generated from the *Problem* and *Intervention* components were chosen to ensure comprehensive coverage of relevant publications.

Table 8.1 | The use of the PICO framework to generate keywords from research questions and form search blocks for searching literature on online databases.

PICO Component	Abstract component	Keyword(s) generated
Problem	Complex calcium imaging analysis workflows	*"calcium imaging", "calcium activity", "calcium trace*", "calcium transient*"*
Intervention	Use of Deep Learning advances such as CNNs to improve analysis	*"deep learning", "convolutional neural network*", CNN**
Comparison	Performance comparison to current calcium imaging analysis methods	*compar*, optimi*, accuracy*
Outcome	Less complex, faster, less pre-processing steps, potentially more neural information related to PD	*neuroscience, neur*, "parkinson's disease", PD, parkinson**

Note | Truncations and quotations have been added where necessary. The Asterix (*) symbol at the end of a keyword allows for multiple letter endings for that specific word, e.g., trace* can include the word "traces" in the search. Quotations ("") are employed to search for multiple-word phrases such as "calcium imaging", "deep learning", etc. More information on truncations and quotations can be found on the KI Library website (link to an external website) (118).

Table 8.2 | Total literature findings from database search

Platform/Method	Search Block	Number of Records	Link
1. PubMed	("calcium imaging" OR "calcium trace*") AND (deep learning [MeSH Terms] OR " convolutional neural networks" OR CNN)	10	(link to an external website)
2. Web of Science	TS=("calcium imaging" OR "calcium trace*" OR "calcium transient*") AND (TS=(deep learning) OR TS=("convolutional neural networks") OR TS=(CNN))	34	(link to an external website)
3. IEEE Xplore	("calcium imaging" AND ("deep learning" OR "convolutional neural networks" OR CNN))	7	(link to an external website)

4. Manual search	*	32	-
Total number of entries found after removal of duplicates	-	**73**	-

***Method** 4 does not require a search block as publications were identified via manual search methods such as backward citation chaining.

Note | The Number of retrieved records per database is the result of the search block applied and the use of filters (open access text or with access provided by the institution (Karolinska Institutet or Stockholm University), year of publication, and published in the English language). Links are provided to reproduce these findings. As this literature review was conducted in February 2023, more publications may be found. Findings from unstructured searches have also been included and displayed in the bottom row of the table.

8.2.1.2 Selection of Findings

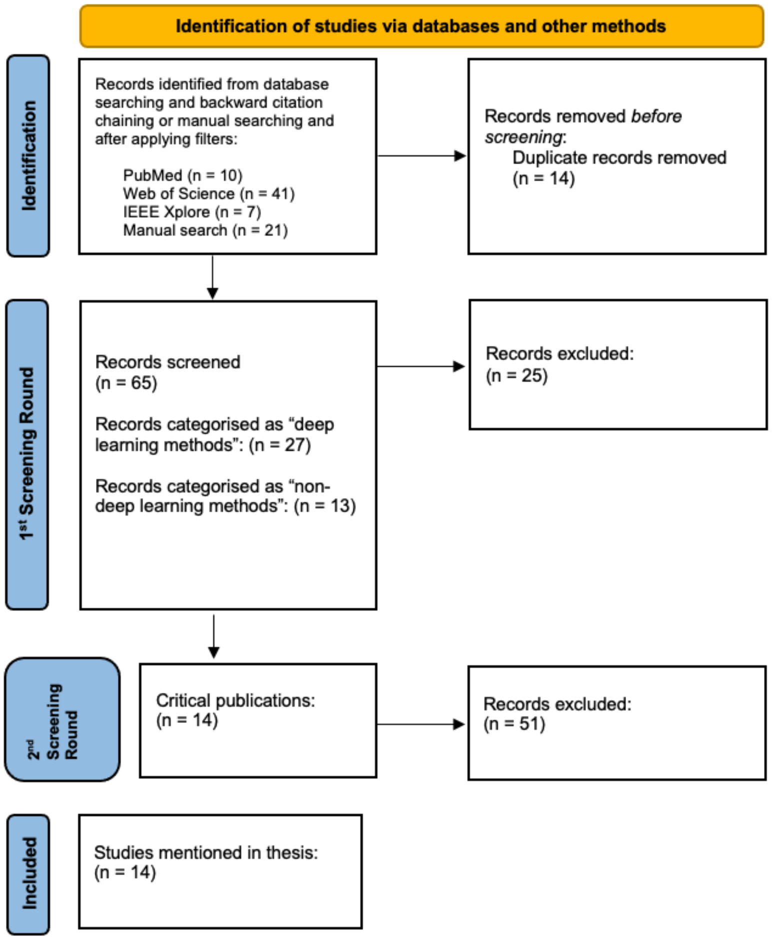

Identification of studies via databases and other methods
Identification
Records identified from database searching and backward citation chaining or manual searching and after applying filters:
PubMed (n = 10)
Web of Science (n = 41)
IEEE Xplore (n = 7)
Manual search (n = 21)
Records removed *before screening*:
Duplicate records removed (n = 14)
1st Screening Round
Records screened (n = 65)
Records categorised as "deep learning methods": (n = 27)
Records categorised as "non-deep learning methods": (n = 13)
Records excluded: (n = 25)
2nd Screening Round
Critical publications: (n = 14)
Records excluded: (n = 51)
Included
Studies mentioned in thesis: (n = 14)

Figure 8.2 | A PRISMA flow diagram representing the selection process of critical literature.

Note | More information on PRISMA flow diagrams can be found here: (link to an external website)

1st Screening round:

A two-step filtering process was conducted to categorize these findings based on their titles and abstract; the first step focused on identifying publications related to current calcium imaging analysis processes (classified as “non-deep learning methods”) while the second step focused on filtering findings relevant to deep learning implementations (classified as “deep learning methods”) in the context of behaviour prediction or cell segmentation from 1p or 2p calcium imaging recordings. A total number of 27 publications were identified in the first step and a total number of 13 were found in the second step of the screening process. The rest (25 findings) were excluded from the literature review.

2nd Screening round:

The second screening round was conducted on the remaining 40 publications where the focus was set on categorizing them according to article type (original article, methods article, conference paper, review article) and differentiating the articles into categories based on their content such as cell segmentation with DL (one-photon or two-photon), behaviour prediction (one photon or two photon), behaviour analysis with DL, and calcium imaging analysis pipelines. Finally, after the conclusion of the second screening process, a total of **14** publications were defined as critical literature based.

Table 8.3 | Summary of selected publications for this study.

Ref	Year	Area of interest	1p or 2p?	DL model	Behaviour Prediction from one photon calcium imaging?
(71)	2022	Cell segmentation	2p	CNN	No
(63)	2022	CI analysis	1p	No	No
(62)	2021	Cell segmentation	1p	CNN, RNN	No
(61)	2020	Cell segmentation	2p	CNN	No
(30)	2019	Cell segmentation	2p	CNN	No
(59)	2019	Behaviour analysis	2p	CNN	No
(60)	2019	Behaviour analysis	2p	CNN	No
(66)	2019	CI analysis	Both	CNN	No
(64)	2019	Cell segmentation	Both	No	No
(68)	2018	CI analysis	1p	No	No
(120)	2018	CI analysis	1p	No	No
(121)	2016	CI analysis	Both	No	No

Note | This list is sorted from newest to oldest publications.
Abbreviations: 1p = one-photon, 1p = two-photon, CI = calcium imaging, CNN = Convolutional Neural Networks

8.3 Methods

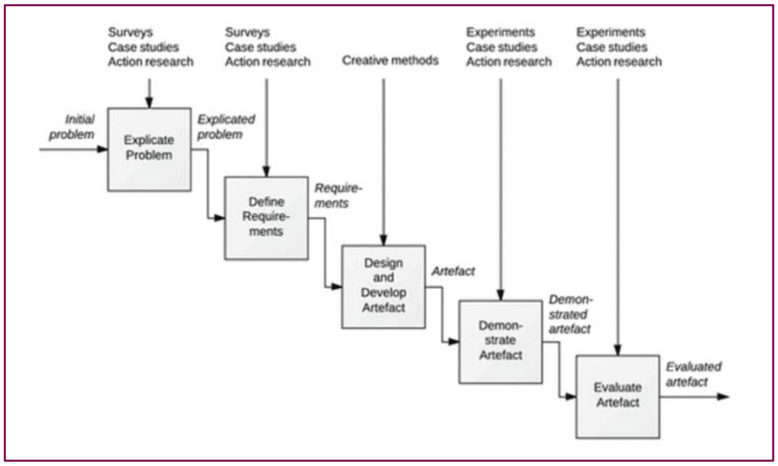


Figure 8.3 | The DSR Process (108)

8.3.1.1 Experimental Setup

Table 8.4 | Hardware specifications of computations resources employed for this study.

Computer	GPU	CPU	Memory	OS
2017 iMac	Radeon Pro 8 GB	4,2 GHz Quad-Core Intel Core i7	32 GB	macOS Ventura
DMC Lab High-performance Computer	NVIDIA GeForce RTX 3090	11th Gen Intel Core i9, 3.50GHz x 16	128 GB	Ubuntu 22.04.2 LTS

8.3.1.2 BPNN Model Architecture

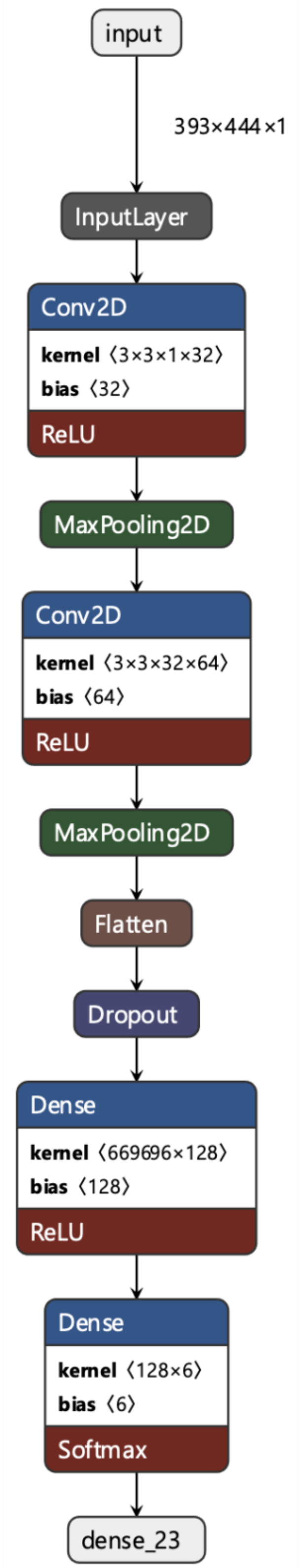


Figure 8.4 | An architectural visualisation of the basic version of the BPNN.

8.4 Results

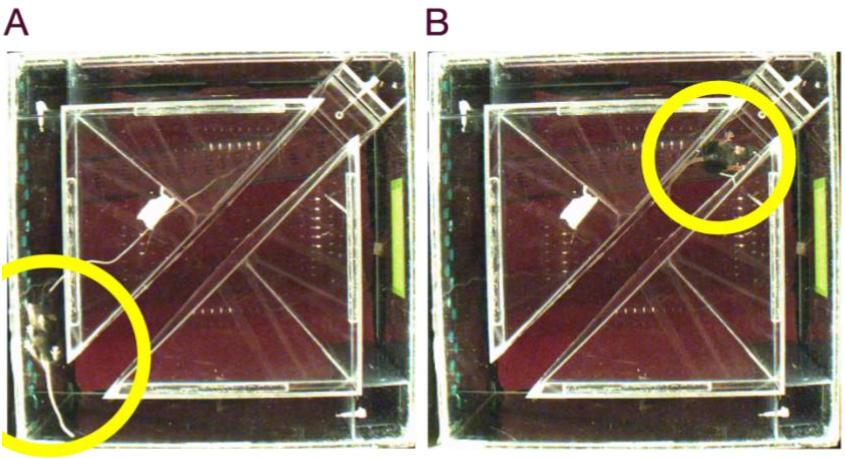


Figure 8.5 | Comparing behaviour labels Frozen (A) and Grooming (B).

Note | In cases where the mouse is standing still, it may be challenging for the BPNN model to distinguish the two behaviours from each other in comparison to movements such as *moving*, *left turn*, *right turn* which are easier distinguished.