Machine Learning Driven Detection and Pathology Validation of Parkinson's Disease in Rodents: A Preliminary Model

Ying Han
Lake Forest College
Lake Forest, Illinois 60045

ABSTRACT

Animal models are crucial for neurological research as they allow researchers to elucidate complex neurological processes and develop treatments through controlled experimental trials. Most neurological disorders research using animal models involves two critical steps: evaluation of behavioral symptoms of the animal models and histological validation of pathologies. These two steps both require multiple human judges' exquisite expertise in neurological diseases, for instance, behavioral shifts and cellular changes in anatomical structure. Such human-dependent methodologies are time-consuming, labor-intensive, and susceptible to human bias, undermining the robustness and reproducibility of animal research. In this project, I aimed to improve the efficiency, objectivity, and accuracy of animal model research by applying emerging artificial intelligence (AI) technologies to the analysis of behavioral and pathological data, focusing on the mouse models of Parkinson's disease (PD). Using an AI tool DeepLabCut, I implemented an automatic body-pose extraction pipeline for the analysis of free-roaming behavior of PD mice. I also established a semi-automatic analysis pipeline for quantifying the dopaminergic neuron degeneration in the brain tissues of PD mice, using a machine-learning workflow QUINT. These AI-based pipelines significantly reduced human intervention in data analyses. Furthermore, using this semi-automatic analysis pipeline, I found a significant correlation between the extent of dopaminergic neuronal degeneration, a prime pathological feature in PD mice, and the severity of motor dysfunction computed from their behavioral data. The finding suggests that the progression of pathological processes may be reliably inferred from behavioral data. These results set a strong foundation for future endeavors to develop AI-based algorithms that are capable of computing behavioral symptoms and pathology scores directly from video recordings and brain images, a powerful tool for animal model research.

ACKNOWLEDGEMENT

I want to express my sincere gratitude to Dr. Eun Jung Hwang, my advisor at Rosalind Franklin University, who led me into research with patience and kindness. The past two years working in the Hwang Lab has been one of the most challenging but also the best experiences I have had during my undergraduate years. I am extremely fortunate to be able to work with you. Your curiosity, professionalism, and intelligence are the role model guidance that will motivate me for the rest of my career.

I would like to thank Professor Sara Jamshidi, my computer science "day one". Like a Professor and like friends, your enthusiasm makes coding so much fun that I can step into the door of computer science. Thank you for advising me during this challenging year!

I would also like to thank Dr. Shubhik DebBurman, my LFC advisor, who made everything possible starting from the COVID-19 time. I also like to thank Dr. Rebecca Delvanthal, Professor Chad McCraken, and Dr. Knuckles Craig for your guidance inside and outside the classroom.

Additionally, I would like to thank my lab manager Sayli Korde. Sayli is one of the most responsible and organized people I've met, and the best lab manager I can ask for. Thank you for your endless help with all the lab skills training, and for being so approachable and funny. I also like to thank Ama Owusu-Ofori, Gloria Bae, Terry Cierrncione, Sophia Cinqurmani, Francis Ongkodjojo, Nadine Pickering, and Bowen Lian for helping me with my project and making our lab welcoming. I want to thank Jack Farrell for picking me up from the lab to campus many times, so I don't melt after 8 hours work.

Lastly, I want to acknowledge my family. My family shaped who I am today, and I live with eternal gratitude towards my family.

Thank you for everyone who offered me a smile, a great laugh, company, a moment of tolerance, or even an argument. You make LFC a great time.

*This author wrote this paper as a senior thesis under the direction of Dr. Eun Jung Hwang.

List of Abbreviations

PD: Parkinson's Disease

DA: Dopamine or dopaminergic

GFP: Green Fluorescent Protein

DAPI: 6-diamidino-2-phenylindole

AI: Artificial intelligence

VTA: Ventral Tegmental Area

SNpc: Substantial Nigra Pars Compacta

MPTP: 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine

6-OHDA: 6-hydroxydopamine

IHC: Immunohistochemistry

TH: Tyrosine Hydroxylase

ML: Machine Learning

DLC: DeepLabCut

FPS: Frame Per Second

PBS: Phosphate-buffered Saline

PFA: Paraformaldehyde

INTRODUTION

According to the 2024 World Health Organization report, one in three people will be affected by a neurological disorder at some point in their lifetime (WHO Media Team, 2024). This stark statistic highlights the pressing need to deepen our understanding of neurological disorders in order to improve disease prevention, treatment, and rehabilitation. Animal models have played pivotal roles in this endeavor as they allow in-depth investigation of brain pathogenesis associated with neurological disorders at multiple stages of disease progression, as well as underlying mechanisms at molecular and cellular levels. For example, MPTP-induced parkinsonian primate models have provided valuable insights into disease progression including the formation of Lewy bodies in dopaminergic (DA) neurons, synaptic alterations, mitochondrial dysfunction, and neuroinflammation/immune responses (Mat Taib & Mustapha, 2020). Animal models are also critical in diagnostic biomarkers development, therapeutic targets identification, and novel treatment testing. For instance, common PD treatment L-DOPA and recently developed deep brain stimulation were both developed and verified on primate models before clinical trials (Pereira & Aziz, 2006).

Most neurological disorders lack known biomarkers (e.g., antigens in blood) that readily and reliably provide information about the patho-

genic processes or response to treatments. As a consequence, such information is mainly inferred from clinical behavior observations, a practice that prompts human judge bias and error. PD symptoms in animal models are scored based on observable symptoms like movement patterns and body postures. For example, open-field tests are used to assess general motor activities, based on simple features such as travel distance, and movement speed. The pole test is used to assess bradykinesia, and the cylinder test is used to measure sensory-motor coordination (Glajch et al., 2012). A reliable assessment requires a combination of tests, as well as a group of human judges with expertise who can interpret behavioral nuances, which requires both time and money (Asakawa et al., 2016). Besides, although human eyes can capture behavioral deviation, our visual sensitivity is limited. This limitation makes it difficult to study the early stages of diseases, a time usually associated with subtle symptoms, but the time to expect the highest intervention efficiency (Gaenslen & Berg, 2010). Furthermore, a prolonged duration of assessments could lead to fatigue in human judges and cause observer bias and low reproducibility.

These limitations also apply to the validation of disease models or treatment efficacy, which relies on behavioral assays and post-mortem brain histology. Such histological assessment of pathogenesis relies on human judges' expert knowledge for differentiating anatomical structures and identifying disease-indicating features from brain tissue sections. For example, DA neuron quantification in assessing neurodegenerations in PD has been done manually under microscopes (Wakamatsu et al., 2008). Thus, the histological assay also suffers from the same issues as behavior assessment, posing general challenges in achieving the desirable level of efficiency and objectivity in animal model research.

To increase the sensitivity, effectiveness, and reproducibility of neurological disease research using animal models, we propose to use artificial intelligence (AI) technologies to outperform human judges in behavioral and histological analyses. In particular, we aim to apply AI algorithms to PD research using mice the dominant animal model used in the field.

1. Dopamine Neuron and Parkinson's Disease

DA neurons are typically medium-sized neurons with elongated axons located in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNpc) of the midbrain (Dugan et al., 2011). These neurons synthesize DA, a monoamine neuromodulator that is involved in many brain functions. Although DA neurons only account for less than 1% of total neuron populations in animals (Chinta & Andersen, 2005), they play crucial roles in the reward system, motor function, working memory, motivation (Wise, 2004), emotion, learning (Redgrave & Gurney, 2006), and so on. They are also implicated in a diverse range of neurological and psychiatric disorders including PD, addiction, obsessive-compulsive disorder, and schizophrenia (Girault & Greengard, 2004). Beyond the central nervous system, DA receptors are also present in the kidneys, lungs, and blood vessels, contributing to pancreatic endocrine and insulin regulation (Volkow et al., 2008). These multifaceted involvements of DA highlight their significance in maintaining animal health.

The progressive loss of nigrostriatal innervation – DA neurons projecting their axons from the SNpc to the striatum - causes PD (Dickson, 2012). The majority of the existing research reports that the first PD motor symptom occurs only after ~50% of the DA neurons are lost, while higher thresholds have been also reported. Because DA neurons distinctively carry pigments called neuromelanin, which only forms when free DAs are present in the cytosol (Fedorow et al., 2005), DA-depleted brain display a loss of pigments in the SNpc and VTA region (Figure 1).

DA neurons are divided into three categories based on their anatomical locations: VTA - A10, SNpc - A9, and retrorubral field - A8 (German & Manaye, 1993). These anatomical divisions of neurons are also related to their projection targets and functional specificities, distinguishing DA neurons from other monoaminergic projections such as norepinephrine and serotonin (Avery & Krichmar, 2017). A9 DA neurons in the SNpc project to the dorsal striatum through the nigrostriatal pathway, which releases DA that modulates D1 and D2 receptors, each selectively expressed in striatal neurons in the "go" and "no-go" pathway, respectively (McGregor & Nelson, 2019). So A9 neurons are especially important for regulating voluntary motor control (Aubert et al., 1997) (Figure 2). In contrast, DA released from VTA and the retrorubral field play roles in regulating emotion, reward, addiction, and other cognitive functions. Besides anatomical structures, a better understanding of DA neurons' genetic profiles can more comprehensively elucidate DA neurons' diverse functionality. A recent study identified three subtypes of DA neurons based on their genetic profiles and their distinctive functional roles: *Aldh1a1*+, *Calb1*+, and *Vglul*+. A subset of the *Aldh1a1*+ subtype (*Anxa1*+) that projects to the dorsal striatum displays increased activity patterns that are temporally locked to the movement acceleration phase. The other subtypes projecting to the ventral striatum display deceleration-locked activity patterns in addition to reward responses. Therefore, diverse subtypes of DA neurons, each with a distinct molecular profile, show different anatomical projection targets, and functional roles (Azcorra et al., 2022). In particular, a loss of *Anxa1*$^{+}$ neurons might account for significant motor deficits in PD.

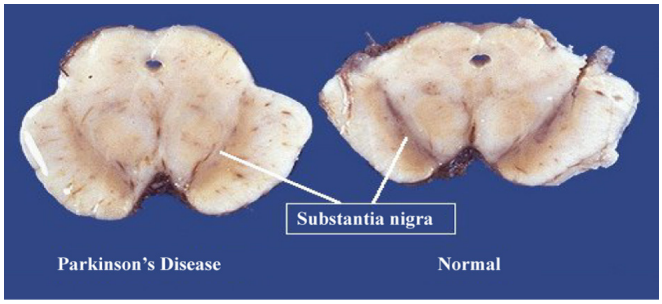


Figure 1: Substantia nigra pigment comparison between a PD patient and a healthy individual (Neveen A, 2019).

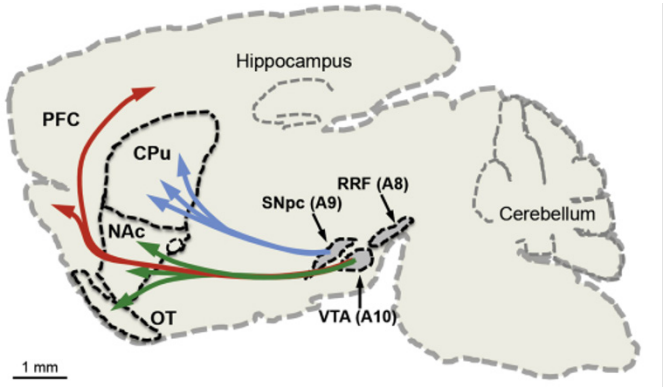


Figure 2: A8, A9, and A10 DA neuron clusters and their projections to caudate putamen (i.e., striatum), nucleus accumbens, olfactory tubercle, and prefrontal cortex (Luo & Huang, 2016)

2. Rodent Models of Parkinson's Disease

2.1 Genetic PD model

Normal aggregation of alpha-synuclein is beneficial and essential for an organism, but excessive alpha-synuclein aggregation will lead to PD (Fields et al., 2019). Alpha-synuclein's function is not clearly identified yet, but existing research suggests its role in maintaining the supply of synaptic vesicles in the presynaptic terminals and regulating the release of DA. Thus, the misfolding of alpha-synuclein will cause abnormal activities in these regards (Fields et al., 2019). There are multiple genetic mouse models that show alpha-synuclein aggregation. These models play a critical role in understanding the molecular processes underlying the aggregation of alpha-synuclein and their effects on behavioral functions.

The first gene discovered to underlie PD was the SNCA gene that encode alpha-synuclein protein, which is linked to familial PD (Polymeropoulos et al., 1997). SNCA is associated with an early-onset of PD. Mutations such as duplication, triplication, and point mutation of SNCA lead to the synthesis and misfolding of alpha-synuclein (Srinivasan et al., 2021). SNCA-mutated PD mice have provided evidence in understanding the altered DA release and uptake (Villar□Piqué et al., 2015), olfactory dysfunction (Uemura et al., 2021), and rapid eye movement sleep disorders (Zhao et al., 2020) in PD. Besides SNCA mutation, alkaline phosphatase-related gene mutations, and ubiquitin-proteasome system-related gene mutations are also used in PD research (Pan et al., 2008). These genetic PD models have made significant contributions to the understanding of pathogenesis in PD (Siddiqui et al., 2016).

2.2 Neurotoxin PD Model

1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) and 6-hydroxydopamine (6-OHDA) are commonly used chemical neurotoxins for inducing DA depletion and parkinsonian symptoms in animal models.

MPTP is a specific and reproducible neurotoxin that damages the nigrostriatal system. Its effectiveness has been shown in monkeys, mice, and rats (Tieu, 2011). MPTP can easily cross the blood-brain barrier and accumulate in DA neurons through DA transporters, so it can be injected subcutaneously in rodents. The accumulation of MPTP leads to ATP reduction and oxidative stress increment in neurons, thus causing cell death and neuroinflammation (Jackson-Lewis & Przedborski, 2007). While the MPTP method is robust and easy, the amount of MPTP administered needs to be delicately controlled because rodents are resistant to lower doses, but are susceptible to death with higher doses. Additionally, studies suggest that certain mouse strains can show recovery from acute motor deficits after MPTP injection within a few days (Sedelis et al., 2001). Therefore, MPTP is less ideal to use in long-term characterizations of the disease.

In contrast, 6-hydroxydopamine (6-OHDA) is an analog to DA and norepinephrine that depletes DA neurons by producing oxidative stress and reactive oxygen species within DA neurons (Tieu, 2011). It does not cross the brain-blood barrier, so it must be injected intracranially to affect neurons. Although the administration of 6-OHDA is more complicated, it allows specific targeting of DA depletion in the brain. For instance, small brain regions like the medial forebrain bundle can be targeted with 6-OHDA and leads to robust PD induction in mice. This cannot be done with genetic PD mice model or MPTP mice model. Additionally, injecting 6-OHDA into only one hemisphere of the brain allows animal models to mimic the asymmetrical nature of nigrostriatal dopaminergic innervation of PD progression in humans (Bové & Perier, 2012).

In the current study, we use 6-OHDA to induce DA neuronal degeneration and PD symptoms in mice. 6-OHDA enables us to target axons in the striatum that are projected from SNpc in only one hemisphere, so our mice model can emulate the asymmetric pathogenesis of human PD.

2.3 Behavior Analysis in PD Mice

Studying PD phenotypes in mice requires a combination of behavior observations and histological assessment. It is important to obtain valid animal behavior data because histological assessment can only be done after mice are sacrificed. I will discuss a few behavioral assays that are commonly used to assess the degree of motor deficits in PD mice compared to normal control mice below.

Open-field Test

The open-field test is the most widely used assay that examines the animal's free locomotive activity. Across the literature and various mouse models,PD mice exhibit a decrease in locomotor activities in open-field tests (Taylor et al., 2010), except for A53T Synuclein Transgenics mice (Unger et al., 2006). During open-field tests, animals are placed in an open-field arena, where they freely roam for typically 10 minutes while being video recorded with an overhead camera. Human judges can analyze the animals' general motor features such as total travel distance, average movement speed, and freezing behaviors from the recorded videos. Simple measures like the time the animal spends around the edge of the box can also be used to measure the level of anxiety associated with PD (Seibenhener & Wooten, 2015). Besides using video recording, infrared beam arrays are another way to track the animal's locomotion. This method requires outfitting the open-field chamber with infrared beam sensors. As the animal moves and interrupts the photobeam, the infrared beam system can estimate the animal's location and movement pattern (Klein et al., 2022). Photo-beam open-field recording requires less computation and less storage space with a more automated measurement of basic locomotor activities, but it does not allow the tracking of multiple animals if one intents to study animal interactions. While both methods can serve the same movement-tracking purpose, the video-recorded open-field test is more advantageous in providing comprehensive and detailed data for more complex body postures and movement patterns. Video-recorded open-field tests are also more suitable for recording multiple animals when assessing social interactions.

Grip Strength Test

The grip strength test is typically used to diagnose central nervous system disorder, It assesses the animal's neuromuscular functions. PD mice tend to show a decrease in grip strength (Tillerson et al., 2002). In the test, the animal's paws are placed on a wire grip, which they naturally hold on to. The human observers gently pull the rodent's tail backwards and observe the gripping strength through the dynamometer connected with the wire grip. The maximal gripping strength before the rodent releases its paw from the grid will be recorded and used for neuromuscular function assessment.

Pole Test

The pole test is widely used for assessing bradykinesia (slowness of the movement) and unilateral brain lesions in PD mice. During the test, mice will be placed facing upward on a pole. The time that mice spent descending from the pole, and the number of rotations during descending motion are recorded. PD mice are shown to take longer to descend, exhibiting bradykinesia (Mitsumoto et al., 1998). Unilateral brain-lesioned mice additionally exhibit ipsilateral – the same direction as the lesioned brain hemisphere- turning when descending.

Drug-Induced Ipsilateral Rotation

Amphetamine is a stimulant drug that increases DA release. In humans, it is used to treat ADHD (Heal et al., 2013). In unilaterally lesioned PD mice (i.e., DA degeneration manifests in only one hemisphere), their body movements on the contralateral side of the lesioned hemisphere are impaired. The administration of amphetamine induces an excessive release of DA and leads to hyperactivity of the non-lesion sides of the brain. This hyperactivity of the non-lesion side of the brain will induce a dramatic ipsilateral turning behavior (Iancu et al., 2005). Conventionally, turning behaviors are counted by human judges' visual inspections, posing a risk of subjective biases and errors.

Human judges play essential roles in PD diagnosis. Trained judges can use their expert knowledge to make valuable judgments about animal behavior patterns, thus detecting behavior changes and altering disease symptoms. However, even expert judges have limited visual sensitivity to subtle changes. Besides, manual measurements are also prone to biases and errors. To improve the efficacy of animal behavior analysis, scientists have developed artificial intelligence technology tools to detect and analyze animal behavior with higher sensitivity and objectivity, ideally to reduce the number of behavior assays and time spent for a single diagnosis. For instance, DeepLabCut is an automated video-tracking tool that comprehensively tracks animals' body parts thus allowing researchers to analyze animal behavior patterns in depth.

3. Artificial Intelligence

3.1 Introduction of Artificial Intelligence

Artificial Intelligence (AI) has grown dramatically in the past

twenty years and has transformed the way we live. The goal of AI is to simulate human intelligence to solve real-life problems. This rapid development portrays the current technological advancements and humanity's growing desire to develop automated assistants that make life easier. According to John McCarthy, the development of an AI system can be either software, hardware, or both. AI's key application elements are natural language processing, expert systems, robotics, intelligent agents, and computational intelligence (Collins et al., 2021) (Figure 3).

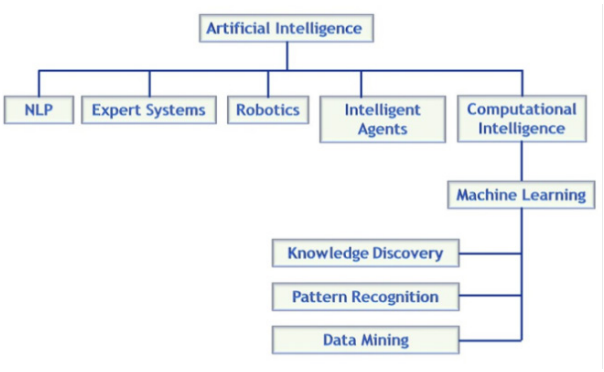


Figure 3: Key elements of the AI family tree. Source: Saudi Aramco

Natural language processing (NLP) is a machine-learning technology that enables computers to interpret, process, and comprehend human languages. Natural language processing (Nadkarni et al., 2011) has been used to develop chatbots like virtual nursing assistants. This technology takes tedious tasks off nurse's shoulders so they can spend time in patient care. Nurses still interact with online clients but only clients request to speak with a real nurse. An expert system is a computer program that is refined and sufficient to provide diagnosis or treatment insights for disease. DXplain developed by Massachusetts General Hospital/Harvard Medical School Laboratory of Computer Science is based on over 2600 diseases and 5700 clinical findings. It has become a staple use in hospitals and medical schools for clinical education (Barnett, 1987). However, despite AI's rapid development, human clinical AI technology is not yet independent, as it must be used in combination with human doctors. In contrast, robotic-assisted stereotaxic surgery has been used for microinjection and even craniotomy in animal research (Ball et al., 2021). This robotics substantially automates the surgery process compared to the traditional manual approach. Yet, human supervision is still required and necessary. Lastly, an intelligent agent is a program that can make decisions based on the environment, user input, and experience (Engelbrecht, 2023). IBM for Oncology has been used in real life. This program can combine a patient's file, clinical expertise, external research, and data to identify and rank the treatment plans and options.

Various AI elements have become tools in healthcare. From backstage paperwork to face-to-face patient care, the collaboration between AI and human efforts improves the efficiency of different processes. Most AI applications are built on large datasets, thus requiring excellent efficient data processing and analyzing techniques, like machine learning.

3.2 Machine Learning Techniques

Machine learning (ML) is the most popular technique in AI, where statistical models make inferences based on datasets going so far as to mimic human decision-making in very specific instances. The goal of machine learning is to interpret and learn from data. Without explicit coding, ML will play the knowledge-discovery and data-mining role to derive values from data. ML can automatically improve itself over time with more training and data (Nasteski, 2017). With the rising demands for data processing, machine learning has been applied widely in our daily lives, such as image recognition and speech recognition from personalized recommendations on platforms such as Google, Spotify, and Netflix.

3.2.1 Supervised Machine Learning

Supervised ML learns patterns from data based on input-output pairs. "Supervised" is derived from the fact that output for training data is given. To train a supervised ML algorithm, input data are split into training and testing datasets. The training dataset will include labels that indicate classifications, such as male vs. female, big vs. small, or 1 vs. 0. This labeled dataset is used to train the algorithms. The algorithm will pick up data pattern differences relative to their labeling. After sufficient training and evaluation, the algorithm is used to predict the test dataset. Adjustments might be made if prediction performance is not ideal (Figure 4) (Nasteski, 2017).

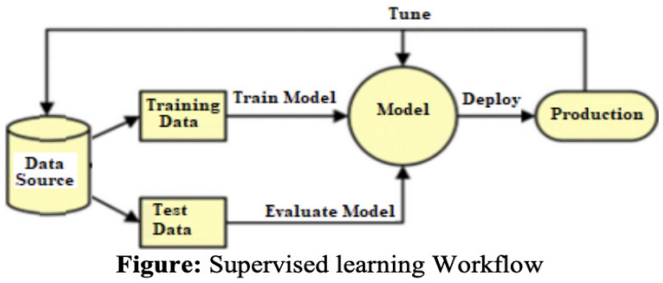


Figure 4: Supervised Learning Workflow. Source: https://pub.aimind.so/supervised-machine-learning-algorithms-and-techniques-explained-in-depth-2a904fe77ee3

3.2.2 Supervised Machine Learning: Linear Regression

Ordinary least squares regression (OLS) is the simplest regression method. Its goal is to capture the best-fit linear relationship (coefficients and intercept) between dependent and independent variables. The linear model calculates the coefficients (*a, b, c, d…*) of each independent variable (*x1, x2, x3, x4…*). The weighted function ($f(x) = ax1 + bx2 + cx3 + dx4…$) represents the optimal linear relationship. To evaluate its performance, the sum of squared error is used. The difference between the predicted data point and the actual data point is known as the residual or error. Taking the sum of the squared residual/error of different weighted functions select the best linear unbiased estimator (BLUE) of a linear relationship. Because OLS regression relies on minimizing the *residual sum of squares*, the performance of these models are typically evaluated by the *mean squared error* – an equivalent matric. The lower the *mean squared error*, the better the regression model (Figure 5). Even though the vast majority of relationships in the natural world are far more complex than linear relationships, nonlinear relationships have linear components. Because ordinary least squares regression has an optimal, closed-form solution, it provides better guarantees than any other ML model of discovering a linear component (Hastie et al., 2017).

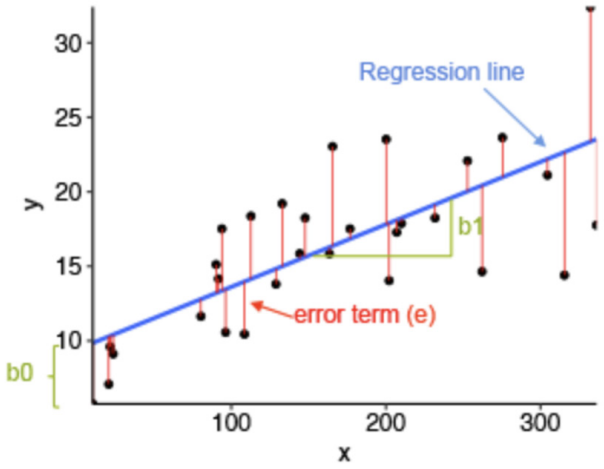


Figure 5: Linear Regression and its residual error. Source: https://rpubs.com/cuborican/regression

3.2.3 Supervised Machine Learning: Feedforward Neural Network

A feedforward neural network is a more complicated form of supervised machine learning. It is used when many independent variables are used to predict dependent variables. Different from linear regression, a neural network is used for nonlinear relationship prediction. In the brain, each neuron has its role in information processing, and they propagate signals back and forth. In a neural network, we articulate neurons of individual roles. The leftmost layer is the input layer that takes input features, and its size is typically the size of input data. After this layer, the hidden layer(s) will transform the input layer's data using the weights of the hidden layer's neurons Often, the more complicated a dataset is, the more hidden layers and neurons it will require. This computation propagates till data passes from the input layer to the last hidden layer, where a non-linear activation function will map current data into a known range, so the output layer can receive the information and transform it into output values. With competitive training, the weights of each hidden layer's neuron can be tuned to obtain a "better" model (Figure 6) (Hastie et al., 2017).

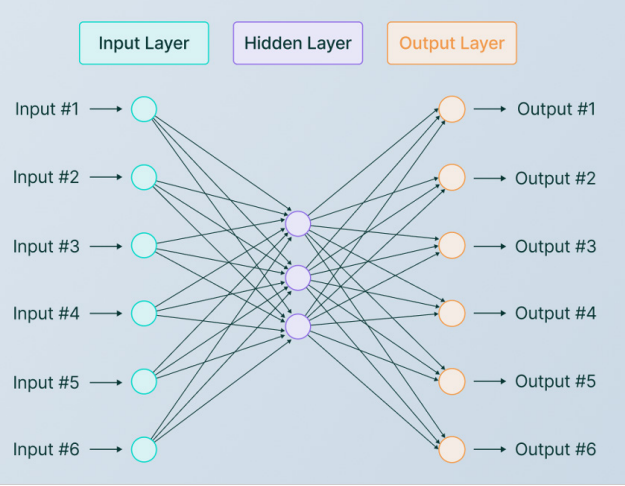


Figure 6: Neural Network Model Architecture. Source: https://www.v7labs.com/blog/neural-network-architectures-guide

Continuous training of neural networks has the potential to recognize and model complex relationships from the dataset. However, neural networks are highly prone to overfitting, thus it requires a large data set, if not thousands, or even millions of data points to minimize the risk of overfitting. Besides, there are some other disadvantages when applying supervised neural networks: 1) neural networks take time and money, computing power, and the tuning of the algorithm; 2) neural networks' algorithms are mysterious, it is yet to know the mechanism of neural network's decision-making process. Although neural networks might show incredible results, their unknown nature should be taken cautiously especially if it is used in real life.

4. Artificial Intelligence in Biomedical Research

4.1 AI Application in Human PD Diagnosis

Recent research has leveraged machine learning (ML) techniques to detect PD symptoms in human individuals, aiming to enable earlier diagnosis and monitoring of disease progression. These ML algorithms have demonstrated promising results in analyzing various motor PD symptoms for diagnosis, particularly using data from wearable sensors that capture relevant movement patterns from different body parts. Keijsers et al. used six triaxial accelerometers (devices that provide simultaneous acceleration measurements in three orthogonal directions) that can be attached to different human PD patients' body parts to collect their motor movement data (both upper arms, both upper legs, the wrist of the most affected body side, and the trunk) in a simulated house for 3 hours while PD patients conducting daily activities. With the comprehensive data collected from each sensor, Keijsers et al. characterized the mean segment velocity, mean segment velocity when moving, percentage of moving, and the percentage of dominant frequency per body part. They developed a neural network with two hidden layers, inputting all body parts' characterized data. As a result, this neural network achieved a 97% of sensitivity and specificity in distinguishing the on and off state of the disease (Keijsers et al., 2006). This study reveals how integrated analysis of motor movement patterns can facilitate the detection of PD with the help of ML. Interestingly, another group of researchers investigated whether using less wearable devices to characterize one specific PD symptom can be used in PD detection. Hssayeni et al. estimated tremors using two sensor trackers (one on the wrist and one on the ankle) and correlated it with PD diagnosis under a more natural environment. Data was collected when participants conducted daily activities like walking, chopping vegetables, and getting groceries. By using a deep learning model, they showed the highest up-to-date correlation between PD diagnosis and tremor (Hssayeni et al., 2019). These studies show the potential of ML in improving the efficiency and sensitivity in diagnosing PD based on the patient's motor behavior.

Building upon these findings, Aich et al. (2020) developed a method for identifying PD patients without wearable sensors. They applied 3D motion capture techniques to analyze the gait of participants without wearing any sensors. The 3D motion analysis system can capture participants' motor activities based on participants' video recordings in the natural environment. After obtaining data using 3D motion analysis, Aich et al. (2020) applied feature selection techniques to determine what motor features are significantly related to the disease and developed a ML classification algorithm based on these identified motor features. Their model was able to achieve an accuracy of over 98.56% in identifying PD patients (Aich et al., 2020). This study demonstrates the potential of ML in automating PD diagnosis through non-invasive techniques in the natural environment.

4.2 AI Application in Animal Research

4.2.1 Behavioral Data Analysis: DeepLabCut

AI technologies have been introduced and actively utilized in animal research. For instance, DeepLabCut (DLC) is an open-source software package used for tracking animal body parts and estimating poses without markers using deep learning. This program utilizes a combination of feature detectors (ResNets, readout layers, MobileNetV2s, etc) to obtain a state-of-the-art algorithm performing pose estimation (Nath et al., 2019) at an inexpensive cost. DLC processes video data from camera recording and performs body parts labeling and tracking based on a small set of user input. As a result, it exports files with x and y coordinates for each body part and timestamps. DLC can be applied to a wide range of organisms. DLC was used in Hayakawa et al's research in studying crickets' circadian rhythms. DLC successfully recognized and labeled crickets' six body parts with high confidence during daily activities, thus providing evidence showing crickets' diurnal rhythms (Hayakawa et al, 2024). Studying animal behaviors in the wild can interrupt animal's natural behavior and cause danger to the researcher. DLC can help with this issue. Wiltshire et al utilized DLC to label chimpanzees and bonobos' free-living behaviors, offering valuable kinematic data in studying wild primates (Wiltshire et al., 2023). DLC utilizes an elegant graphic user interface (GUI) that is easy to use, all the training can be done by mouse-clicking, and no coding knowledge is required to use DLC (Figure 7).

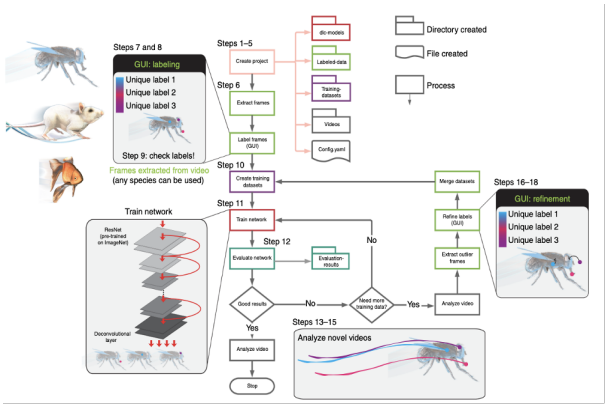


Figure 7: DeepLabCut

4.2.2 Histology Data Analysis: QUINT

Besides behavioral observation, the analysis of the animal brain tissue is also an important component in neurological disease research because they can provide critical pathological evidence of the disease. The common approach to studying the pathological processes of neurodegenerative disorders is to visualize specific molecule aggregates or loss by applying immunohistochemistry or other staining techniques on sectioned brain slices. The processed brain tissues are observed under a microscope to detect molecular/structural changes. Such brain tissue analyses involve two essential tasks: 1) anatomical identification of different structures and 2) protein/target molecule quantification. These two steps rely on domain-expert decisions and a large amount of manual operation, which makes the task difficult and time-consuming. Addressing these challenges in histology, a recently developed QUINT workflow provides semi-automatic feature quantification and analysis tools. Its efficacy has been demonstrated in many research. For instance, QUINT quantified the co-labeling SNCA and alpha-synuclein and revealed the revealed alpha-synuclein's accumulation specificity in selected neuron types. QUINT is also capable of labeling multiple targets of interest (Geertsma et al., 2024). In Gurdon et al.'s research on Alzheimer's disease, QUINT was used to label different neurons, microglial, reactive astrocytes, all nuclei, and beta-amyloid 1-42 pathology. These targets of interest are quantified in Alzheimer's disease mouse models, revealing that the variation of age at onset and symptoms are related to cell compositions (Gurdon et al. 2023). QUINT workflow is a powerful tool with high efficiency in tissue registering and quantifying the pathogenic features in animal research.

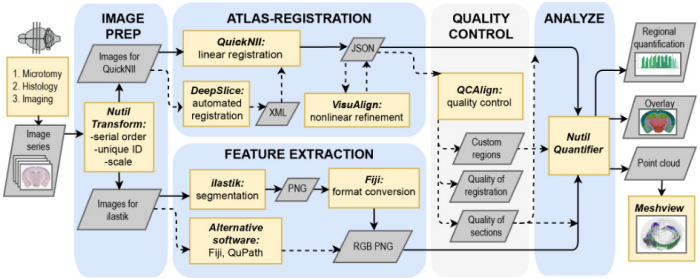


Figure 8: QUINT workflow

GOAL

I aim to build an AI-based analysis pipeline that semi-automatically processes behavior and histology data from PD mouse models, to improve efficiency, objectivity, and precision in the research of neurological disorders using animal model.

SPECIFIC AIMS

1) Establish a surgical protocol to generate PD mice.

2) Develop an automatic pipeline to analyze PD mouse movement data using DLC.

3) Establish an immunohistochemistry (IHC) protocol for histological assay of PD pathology.

4) Develop a semi-automatic pipeline to analyze IHC data using QUINT.

5) Develop an ML algorithm to diagnose PD from movement data.

Chapter 1: PD Mice Generation and Behavior Data Collection Methods

Animals

Both male (N = 16) and female (N = 5) C57/B16J WT mice (13 – 55 weeks old at the time of the experiment) were used. Mice were kept on a 12 h light/dark cycle with food and water available ad libitum. All procedures were approved by the Rosalind Franklin University Institutional Animal Care and Use Committee (IACUC) and followed the NIH guidelines.

Open-Field Test

We used an open-field test to collect the free-roaming behavior data to examine the movement patterns of PD mice compared to control mice. A mouse was placed in the center of a box (50 cm * 40 cm_* 33cm) and was allowed to freely roam for 10 minutes. An overhead camera (AV Alvium 1800 U-291c, Allied Vision) records the mouse in the open-field arena. The video is recorded at 60 frames per second and was saved in an mp4 format for subsequent analysis. After the test, the mouse was placed back in their home cage.

Behavioral Data Collection Timeline

The open-field test was conducted five times for each mouse: Day -3 (3 days before the surgery day), Day -2, Day 7, Day 14, and Day 21. Day -3 was used to help the mouse to acclimate to the open-field environment. Day -2 was to collect control data. Day 7, Day 14, and Day 21 data were to track the progression of the disease following a PD-inducing surgery. The five test sessions were referred to as acclimation, control, pd1, pd2, and pd3, respectively (Figure 9).

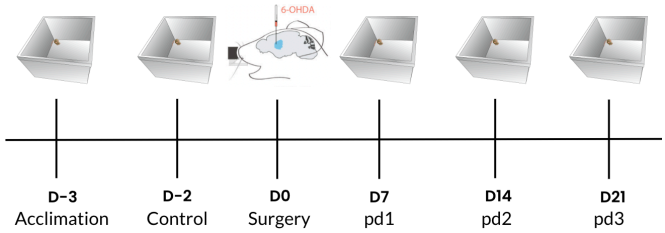


Figure 9: Kinematic Data Collection Timeline\

Surgery

On the surgery day, mice were injected with the assigned substances: 6-OHDA or saline vehicle. Before injecting 6-OHDA, mice received premedication made of desipramine and pargyline 15 minutes before the surgery started. Desipramine is a 5HT/norepinephrine (NE) uptake inhibitor, so it protects 5HT and NE neurons and thus increases 6-OHDA's selectivity DA neurons. Pargyline is a monoamine oxidase inhibitor, thus it increases the sensitivity of DA neurons' axon terminals to 6-OHDA and increases their vulnerability (Thiele et al., 2012). Detailed recipes and surgery procedures can be found below. After surgery, mice went through a 7-day post-operative recovery period with water and food ad libitum. 6-OHDA-injected mice were also supplied with diet gel to prevent dehydration and malnutrition.

Drug Preparation

- Premedication

Desipramine HCI was administered at 25 mg/kg and pargyline at 5 mg/kg. Both desipramine and pargyline were dissolved in normal saline at a concentration of 2.5 mg/ml and 5 mg/ml, respectively. Therefore, both drugs were administered at 10 ml/kg. 10 ml of premedication was made each time and was stored in a -80 Celsius-degree refrigerator until use. The amount needed is calculated as below:

Desipramine correction factor = molecular weight of desipramine HCl/ molecular weight of desipramine

*Desipramine HCl needed = Desired concentration * desired volume * correction factor*

*2.5 mg/ml * 10.0 ml * (302.84/266.38) = 28.43 mg*

Pargyline correction factor = molecular weight of pargyline HCl/ molecular weight of pargyline

*pargyline needed = Desired concentration * desired volume * correction facto*

*0.5 mg/ml * 10.0 ml * (195.69/159.23) = 6.15 mg*

28.45 mg of desipramine.HCI and 6.15 mg of pargyline.HCl was mixed with 8 ml of sterile normal saline (0.9%) in a 10 ml glass beaker. The mixture was vortexed and heated at 45 degrees Celsius till everything was dissolved. Then, drops of 1M NaOH were added till the pH became 7.4. Finally, sterile saline (0.9%) was added to the mixture to make it 10 ml in total. The solution was aliquoted, labeled, and stored in the -80 Celsius degree freezer until use.

- Vehicle (0.9% saline and 0.02% ascorbic acid mixture)

0.2 g of ascorbic acid was mixed with 1L of sterile saline (0.9%), and stored in the

-80 Celsius degree freezer until use.

- 6-OHDA

6-OHDA.HBr is light-sensitive and heat-sensitive. It has an expiration duration of 6 hours. Thus, 6-OHDA.HBr should be made fresh before the surgery and put on ice at all times. If the 6-OHDA solution turns brown during the 6-hour interval, it is an indication of oxidation. If it happens, a new batch of 6-OHDA should be made. The formulation for making 6-OHDA is below:

6-OHDA correction factor = molecular weight of 6-OHDA.HBr/ molecular weight of 6-OHDA

*6-OHDA. HBr needed = Desired concentration * desired volume * correction factor15.0 mg/ml * 0.5 ml * (250.09/170.19) = 11.03mg*

*33.7 mg/ml * 0.5 ml * (250.09/170.19) = 24.76 mg*

The desired amount of 6-OHDA was quickly mixed with 0.5ml of the aforementioned vehicle solution in a light-blocking tube. After vortexing, the solution was kept on ice immediately. The time was recorded to use the solution within 6 hours.

Surgery Procedure

1) 15 minutes after premedication administration, place the mouse in the anesthesia-induction chamber and anesthetize it with 2% isoflurane vaporized in 100% oxygen. The lack of toe reflex is an indicator of sufficient anesthesia. Take the mouse out of the induction chamber, and place it on the warm surgery plate. Make sure mouse's airway is sealed with the isoflurane nose mask.

2) Check the mouse's toe reflex before proceeding and inject an analgesic meloxicam (10 mg/kg) subcutaneously. Apply a generous amount of eye lubricant ointment to cover the mouse's eyes to avoid corneal drying. Adjust the isoflurane concentration to 1.5%.

3) Apply lidocaine on the tip of the earbars and secure them onto two sides of the skull in front of the mouse's ears. Stabilize the mouse head on the stereotaxic frame by adjusting the heights of the earbars and the bite bar (Figure 10b).

4) Shave the top of the mouse's head and disinfect the scalp three times by directly applying beta iodine and 70% ethanol.

5) Check the mouse's response to the toe pinch and incise the scalp with sterile surgical scissors. Gently clean the exposed skull by picking up hair. Apply 0.5 ml of bupivacaine on the skull and leave it on for 5 minutes for local anesthesia to be induced.

6) After 5 minutes of waiting time, gently scrap the periosteum off with a scalpel blade and wash the surface with sterile saline. Clean any blood, hair, or periosteum tissues on the exposed skull to ensure a clean injection area.

7) Gently detach the tendons attached to the posterior part of the skull and create a clean "pocket" between the skull and the skin near the neck. If the skull bleeds, use surgifoam to cover the area and wait for 5 minutes. After removing surgifoam, clean the exposed skull with saline and ethanol, and wait for it to dry.

8) Locate bregma using the manipulator attached to the stereotaxic frame and set the position reader of the manipulator to zero. Locate the lambda (Figure 10a). Adjust the anterior (A) and posterior (P) alignment of the brain until the depth difference between bregma and lambda is less than 0.10 mm.

9) Use the points 2.00 mm left and 2.00 mm right to the bregma for the lateral alignment of the brain. Align the brain until the depth difference between the left and right is less than 0.10 mm. The lateral adjustments of the brain might change anterior (A) and posterior (P) alignments, thus AP alignment should be re-checked.

10) After the brain is aligned in both directions, move the manipulator to target the injection site and mark a "dot" with a sterilized marker.

11) Before drilling the skull above the injection site, check the mouse's toe pinch response. If there is no response, carefully drill the skull with a dental drill. If bleeding starts, apply surgifoam for 5 minutes and clean the area with saline.

12) Fill a glass pipette with mineral oil and attach it to the nanoinjector. Flush 1000 nl of mineral oil out of the glass pipette using the nanoinjector pump.

13) Turn down the microscope light to low, and dispense 6-OHDA on the parafilm-wrapped surface to load into the pipette.

14) Load the desired amount of 6-OHDA into the glass pipette attached to the nanoinjector. Move the pipette to the injection site and lower it down to the desired depth to dispense 6-OHDA at 25 nl/min. While injecting 6-OHDA, the microscope light should be turned off com-

pletely to avoid light-induced degradation.

15) When the injection is complete, wait for 5 minutes before slowly retracting the pipette from the brain. Close the injection hole with a small amount of dental wax and cover it with a drop of medical-grade adhesive.

16) When the adhesive is dry, disinfect the exposed skull again and cover it with a thin layer of adhesive. Wait for the adhesive to dry completely.

17) Apply cement (dental acrylic and jet fluid mixture) to cover the exposed skull and the mouse's ID will be written on top of the dried cement.

18) Turn off the isoflurane, take the mouse off the stereotaxic frame, and place it into a warm cage with easy access to water and diet gel. If the surgery took longer than 2 hours, inject saline subcutaneously to prevent dehydration.

19) The mouse should be monitored until it regains its consciousness from anesthesia. It will be transported back to the biological resource facility housing afterward.

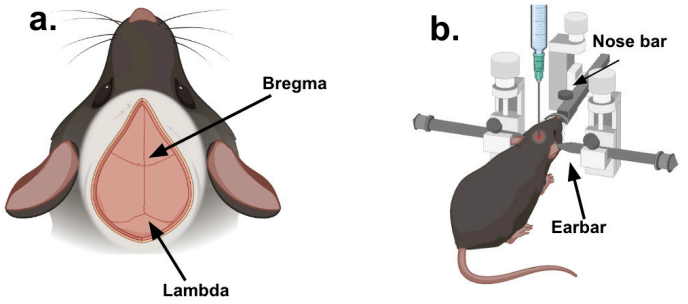


Figure 10: Demonstration of intracranial injection. (a) the location of bregma and lambda; (b) simplified stereotaxic frame and intra-cranial injection setup.

Results

A total of 25 mice received either 6-OHDA (n = 20) or vehicle injection (n = 5) at various doses (Table 1). 15 mice (10 of 6-OHDA mice, 5 of vehicle mice) received an injection in the ventral striatum (STR) in the right hemisphere, and 10 received an injection in the dorsal STR in the right hemisphere. Two mice (0.008 mg 6-OHDA in the ventral STR) died after pd3 data collection. Four mice (0.009 mg 6-OHDA in the dorsal STR) died within 3 days after the surgery. We suspect the mortality is increased with both 6-OHDA dose and mouse age. No vehicle mice were found dead. In total, we completed data collection from 21 mice.

Mouse	Sex	Age	Dose (mg)	Injection Site	Found Dead	Incomplete Data
774	M	30	0.00	R (0.6A, 2.0L, 4.0V)		
775	M	30	0.00	R (0.6A, 2.0L, 4.0V)		
776	M	30	0.00	R (0.6A, 2.0L, 4.0V)		
777	M	33	0.00	R (0.6A, 2.0L, 4.0V)		
778	M	33	0.00	R (0.6A, 2.0L, 4.0V)		
635	F	20	0.004	R (0.6A, 2.0L, 4.0V)		
636	F	20	0.004	R (0.6A, 2.0L, 4.0V)		
642	M	20	0.004	R (0.6A, 2.0L, 4.0V)		
648	M	25	0.004	R (0.6A, 2.0L, 4.0V)		
651	F	22	0.004	R (0.6A, 2.0L, 4.0V)		
724	F	19	0.008	R (0.6A, 2.0L, 4.0V)		
725	F	19	0.008	R (0.6A, 2.0L, 4.0V)		
726	M	24	0.008	R (0.6A, 2.0L, 4.0V)		
727	M	25	0.008	R (0.6A, 2.0L, 4.0V)	x	
728	M	25	0.008	R (0.6A, 2.0L, 4.0V)	x	
789	M	43	0.009	R (0.6A, 2.0L, 3.0V)		
790	M	43	0.009	R (0.6A, 2.0L, 3.0V)		
791	M	43	0.009	R (0.6A, 2.0L, 3.0V)		
663	M	43	0.009	R (0.6A, 2.0L, 3.0V)		
669	M	43	0.009	R (0.6A, 2.0L, 3.0V)		
891	M	55	0.009	R (0.6A, 2.0L, 3.0V)		
709	M	43	0.009	R (0.6A, 2.0L, 3.0V)	x	x
792	M	47	0.009	R (0.6A, 2.0L, 3.0V)	x	x
888	M	51	0.009	R (0.6A, 2.0L, 3.0V)	x	x
889	M	50	0.009	R (0.6A, 2.0L, 3.0V)	x	x

Table 1: Mouse ID, sex, age (in weeks), 6-OHDA dose received, injection coordinates, whether they died during the experiment, and whether only partial data were collected.

Chapter 2: Behavioral Data Analysis using DeepLabCut

Methods

DeepLabCut

In order to analyze the open-field data more parametrically, efficiently, and precisely, I built a DLC pipeline to automatically track the animal's body parts of interest and produce the body position data. The DLC pipeline was customized to capture the x-y coordinates of the mouse's snout, left ear, right ear, nape, tail base, and tail tip, so that the overall body postures of the mouse and their specific body parts' movement data can be collected.

To train a DLC model, a few open-field video samples (recorded with non-experimental mice) were used as training data. DLC automatically generates relatively diverse qualities of frames for manual labeling. In those automatically selected frames, the snout, left ear, right ear, nape, tail base, and tail tip were manually labeled by a human experimenter. After labeling, DLC merged labeled frames with the rest of the frames and split them into training and testing datasets. These two datasets are processed using ResNet, a convolutional neural network that supports hundreds or thousands of convolutional layers. As the DLC ResNet has been pre-trained to segment and detect objects with numerous real-world data from wide range of animals, it only needs to be refined to detect the specific body parts labeled in the training and testing data. After this training, if the model did not yield the desired level of accuracy, frames with poor accuracy were extracted to be re-labeled so that the ResNet could be further fine-tuned. This training step was performed only in one mouse and the same trained network was used for all other mice without further training. Thus, each open-field video file was simply loaded to DLC and processed into a CSV file which contains the x-y coordinates (units) of each of the six body parts in every video frame.

Data Analysis

To analyze the body position data in the DLC CSV file (i.e., kinematic data), an open-source Python library PyRat was utilized. Pyrat computes gross locomotion features such as total travel distance (m) and mean travel speed (cm/s) from DLC-processed CSV files. To compute median speed (cm/s), hyperactivity frequency and total freezing duration (s), I wrote custom code utilizing the open-source code in Pyrat. A spatial filter of 64.03 cm (the longest distance between 2 points within the open-field area) was applied to remove motion artifacts. Each video is analyzed at a 60 frames per second base.

Total travel distance(m) is calculated as the sum of the Euclidean distance of the tail-base position between two consecutive time frames. Median speed (cm/s) is the median of all speed values computed between consecutive frames. The frequency of hyperactivity is defined as the number of incidences where speeds exceed 75% percentile of the overall speed distribution across all mice, continuously for more than 30 frames (equivalent to 0.5 seconds). The total duration of freezing is

quantified as the cumulative time of each freezing duration in which two consecutive frames' dislocation is less than 0.1cm for longer than 4.5 seconds.

To compute the hyperactivity frequency, the overall distribution of all movement speeds was first calculated across all data, and the 75% percentile speed was determined to be 8.47 m/sec (Figure 11).

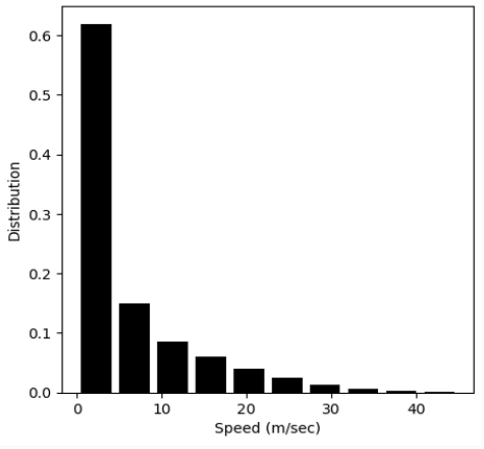


Figure 11: Distribution of speed.

Freezing is computed when the movement distance from a previous frame is shorter than 0.001cm for a continuous interval. To determine the most informative freezing duration for identifying PD mice, the total duration of freezing was computed using different durations: 2.5s, 3s, 3.5s, 4s, 4.5s, 5s, and 5.5s. Then, I conducted a LASSO (Least Absolute Shrinkage and Selection Operator) feature selection test to determine the significance of these different freezing thresholds in identifying PD mice. The result indicates that the highest significance is achieved by using a threshold of 4.5s (Figure 11).

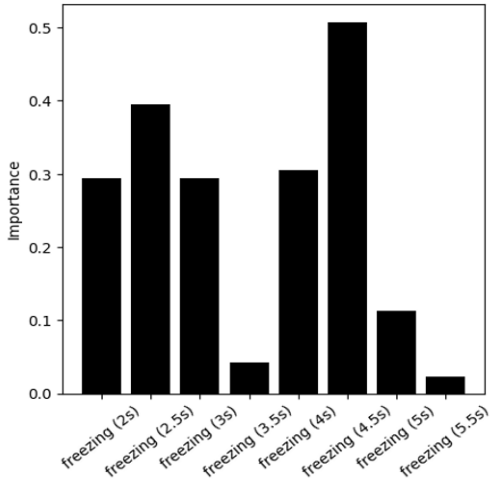
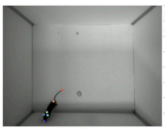


Figure 12: Importance of freezing duration by LASSO.

Results

All the open-field test data were analyzed through the DLC pipeline to generate CSV files containing the x-y coordinates of six body parts for each time frame. Figure 13 shows an example frame with the six body parts labeled by DLC and a sample CSV file.



bodyparts	Snout	Snout	Snout	R_Ear	R_Ear	R_Ear
coords	x	y	likelihood	x	y	likelihood
0	1012.224854	1324.229004	0.9978247881	952.8920898	1282.341797	0.9997335076
1	1007.873169	1328.048828	0.9978619218	951.0184937	1285.223877	0.9994893074
2	1002.736816	1332.069824	0.9974673986	946.7733765	1285.785889	0.9984240532
3	999.3208618	1334.765625	0.9952434897	944.1177979	1285.726563	0.9929785132
4	996.4650269	1336.649414	0.9968857169	940.0735474	1287.731445	0.9963204861
5	991.6109009	1338.416382	0.9899673462	935.8883667	1287.210205	0.9941077232

Figure 13: DLC-processed frame and CSV file sample.

Using the kinematic data (i.e., times series of body parts' positions) from the DLC output CSV file, four movement features were computed: total travel distance, median speed, hyperactivity frequency, and freezing duration.

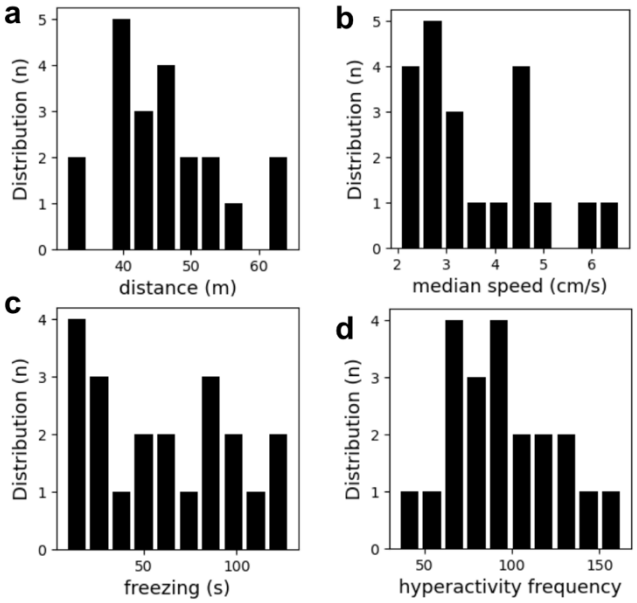


Figure 14: Features baseline distribution per feature.

Figure 14 shows the distributions of the four movement features from the control session (1 day before 6-OHDA or vehicle injections) across 21 mice. The large variability among animals in the control session indicates that the baseline motor performance highly varies among different mice. Thus, the analysis of drug or vehicle treatment effects should take this inter-subject variability of the baseline into account. To address this, for each mouse, the mean of each feature was computed across all open-field test sessions (control, pd1, pd2, and pd3). This mean was subtracted from the raw movement feature value to adjust their baseline variabilities. For example, to calculate the mean-adjusted distance in pd1 for mouse 651, the mean distance across control, pd1, pd2, and pd3 sessions of mouse 651 was subtracted from this mouse's original distance in pd1 session, i.e., 651 *pd1 distance - 651 mean distance*.

The baseline-adjusted motor features changed over time in both 6-OHDA-treated PD and vehicle-treated mice in general. The total travel distance significantly increased from the control (*mean* = -4.079, *SD* = 9.815) to pd3 (*mean* = 5.567, *SD* = 8.491) sessions in the vehicle mice ($p < 0.01$*), while no differences were observed among PD mice (control: *mean* = 2.795, *SD* = 8.596; pd3: *mean* = -1.799, *SD* = 5.048). Both PD ($p < 0.05$*) and vehicle ($p < 0.01$) mice showed a significant decrease in the median speed when comparing pd1 (vehicle: *mean* = -0.188, *SD* = 0.332; PD: *mean* = -0.0794, *SD* = 0.843), pd2 (vehicle: *mean* = -0.182, *SD* = 0.231; PD: *mean* = -0.137, *SD* = 0.666), and pd3 (vehicle: *mean* = -0.432, *SD* = 0.154; PD: *mean* = -0.317, *SD* = 0.474) against control sessions (vehicle: *mean* = 1.045, *SD* = 0.378; PD: *mean* = 0.771, *SD* = 0.859). Consistent with the decline in the median speed, freezing durations increased in both PD ($p < 0.01$*) and control ($p < 0.05$*) mice from control (vehicle: *mean* = -34.928, *SD* = 23.810; PD: *mean* = -21.062, *SD* = 32.482) to pd3 (vehicle: *mean* = 22.155, *SD* = 16.465; PD: *mean* = 10.251, *SD* = 23.139) session. Similarly, the hyperactivity frequency significantly decreased from control (vehicle: *mean* = 28.3,

SD = 6.763; PD: *mean* = 25.156, *SD* = 21.072) to pd3 (vehicle: *mean* = -11.1, *SD* = 6.26; PD: *mean* = -10.156, *SD* = 9.935) session in both groups (Figure 15). These results suggest that most movement features exhibited similar trends over time between the PD and control groups, potentially reflecting longitudinal changes from behavioral habituation to the open-field arena with repetitive exposures, rather than PD-specific motor deficits.

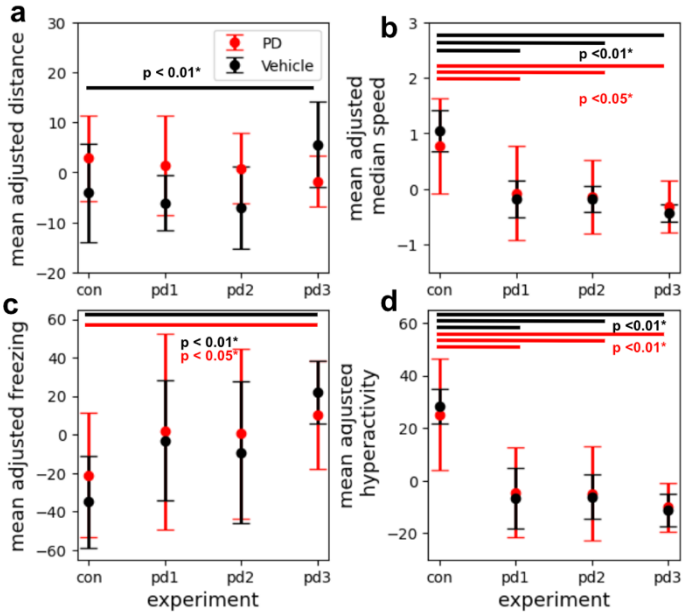


Figure 15: Feature Progression in PD and healthy mice.

Despite the common habituation-like behavior, mice may show motor functions that change with the dose of 6-OHDA. To evaluate this possibility, the relationship between the motor features at PD3 and the dose of 6-OHDA was assessed (Figure 16). Within PD mice, mice treated with 0.009 mg of 6-OHDA showed significantly lower median speed (0.004: *mean* = 0.0883, *SD* = 0.176; 0.009: *mean* = -0.569, *SD* = 0.408; $p < 0.05$*) and hyperactivity frequency (0.004: *mean* = 0.0883, *SD* = 0.176; 0.009: *mean* = -0.569, *SD* = 0.408; $p < 0.01$*) compared to those treated with 0.004 mg 6-OHDA.

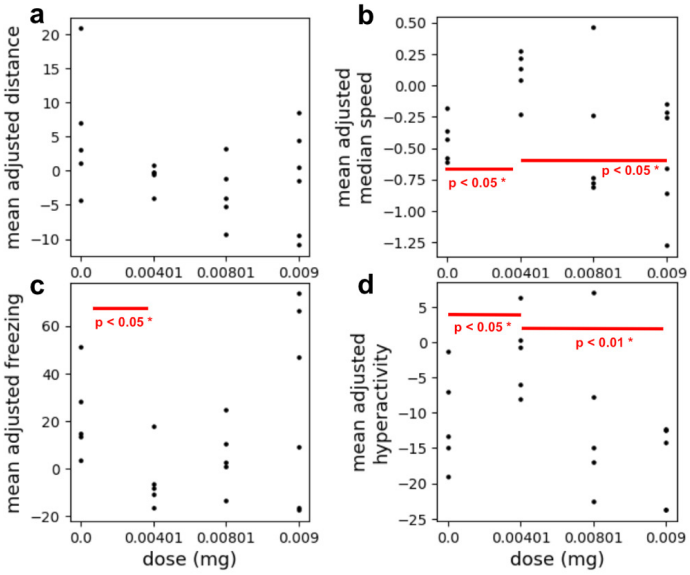


Figure 16: Feature vs. 6-OHDA dose from pd3 data.

Discussions

The DLC pipeline automated the locomotive activities data extraction and significantly facilitated the movement features analysis. From the automatically extracted locomotive data, gross motor features were computed for each open-field test session. Likely reflecting habituation effects to the open-field arena from repeated exposures, median speed, freezing duration, and the hyperactivity frequency decreased over time, but this change did not differ between PD mice and vehicle mice. However, within PD mice, certain motor functions such as median speed and hyperactivity frequency were deteriorated in a 6-OHDA dose-dependent manner three weeks after 6-OHDA injection. Surprisingly, hyperactivity frequency is not a common symptom examined in previous PD research. Our finding suggests hyperactivity frequency might be a new potential behavior predictor in PD diagnosis in mouse models. Nevertheless, gross movement features were not robust predictors that clearly distinguish the PD mice from the vehicle mice. This limitation suggests that there might be a loss of information when the full kinematic data are summarized into a few specific movement features. This further motivates us to study the complete kinematic data for discriminating between the PD and vehicle mice in the future chapter.

Chapter 3: Quantification of PD Pathology Using QUINT Workflow

Methods

Perfusion and Histology

Once behavioral data collection was complete, mice were anesthetized in 2% isoflurane chambers, followed by intraperitoneal injection of ketamine/xylazine mixture injection (10 mg/kg). After ensuring the absence of its response to toe pinches, transcardial perfusion was performed with 25 ml of 1% PBS followed by 25 ml of 4% PFA. The mouse brain was then extracted and immersed in 4% PFA for fixation for at least 24 hours before being transferred to 30% sucrose solution. After at least 24 hours' immersion in sucrose, the brain was sectioned in the coronal plane using a freezing sliding microtome at a thickness of 30 µm (Figure 17). Coronal sections of the brain were either mounted on adhesive slides immediately sectioning, or stored in 1% PBS in a 4 Celsius degree refrigerator for future mounting.

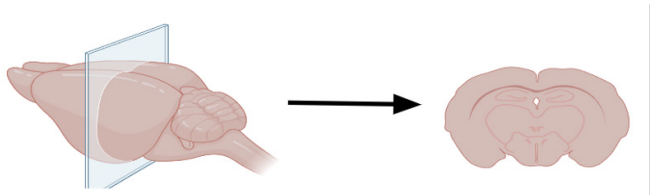


Figure 17: Visual demonstration of brain sectioning

Immunohistochemistry and Result

Immunohistochemistry (IHC) is a protein visualization technique used to target proteins of interest by using selective antibodies. To visualize dopamine neurons, tyrosine hydroxylase (TH), an enzyme that catalyzes the synthesis of dopamine neuromodulators, which are exclusively present in DA neurons, was targeted. Sectioned brain tissues can be used for IHC immediately after mounting or stored at -80 Celsius degree refrigerators until IHC procedure. On IHC day 1, if IHC tissues were retried from -80 Celsius degree refrigerators, they had to be thawed and washed with 1% phosphate-buffered saline (PBS). Then, they were incubated in the blocking buffer (1% normal goat serum, 1% bovine serum albumin in 0.3% PBS-Triton X-100) for 1 hour. After blocking, tissues were incubated in primary antibody solution (1/500 ThermoFisher Tyrosine Hydroxylase Polyclonal Primary Antibody in 1% BSA, 0.3% PBS-Triton X-100) in the -4 Celsius degree fridges overnight. On IHC day 2, overnight-incubated tissues were washed with BSA washing buffer (0.25% BSA in 0.3% PBS-Triton X-100) three times each lasting 5 minutes. Subsequently, tissues were incubated in the secondary antibody solution (1/500 Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488) Secondary Anti-

body in 1% BSA, 0.3% PBS-Triton X-100) in the dark for one hour. The secondary antibody tagged the TH molecules bonded with the primary antibody with green fluorescence. After incubation, tissues were washed with BSA washing buffer (0.25% BSA in 0.3% PBS-Triton X-100) three times, 5 minutes each again. Finally, brain tissues were covered with coverslips and 6-diamidino-2-phenylindole (DAPI) mounting medium. Before fluorescence imaging, slides were ensured to be completely dry. Prepared slides were scanned under a Leica DMB6 Thunder Imager at 5X magnification. For the new QUINT model, images were scanned under 20X magnification.

QUINT Workflow

To systematically analyze scanned IHC-processes images, I built a histology analysis pipeline using the two main steps in the QUNIT workflow: 1) registration of brain sections to a reference atlas, and 2) classification/detection of features of interest, such as labeled cells or structures. The registration of brain sections was done in QuickNll, where histological images were aligned against the standardized Allen Institute of Brain Science Mouse Brain Common Coordinates Atlas. This step also allows manual registration of vertical and horizontal brain sectioning angles, to align the anatomical structures to the reference atlas. Finer and non-linear refinement of brain slice registration was done in VisualAlign. This program enabled users to adjust the alignments of each brain tissues individually by dragging standard atlas brain outline.

Upon the completion of histology registration, ilastik was employed to detect TH⁺ neurons. Similar to DLC, I first labeled features of interest (green fluorescence cells) for training. The trained model was then used to detect the features in new brain tissues automatically. By combining the registration and feature detection in the brain section images, the spatial distribution of the features of interest can be automatically quantified in Nutil. This semi-automatic approach to pathological tissue processing is systematic, inexpensive, and efficient.

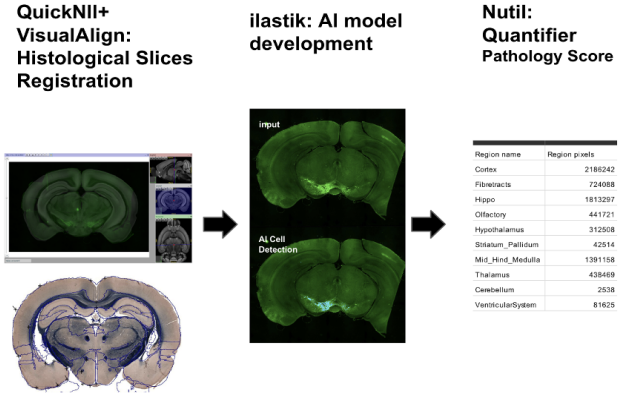


Figure 18: QUINT pipeline.

Results

Brain tissues between 2.5 mm to 3.7 mm posterior to bregma from each mouse were processed using IHC, and imaged using a 5X objective. In the sample image below (Figure 19), the green fluorescence signals represent TH⁺ neurons, while the blue DAPI signals mark all nuclei in the brain slices. A clear asymmetry in the TH signal is observed between the hemisphere injected with 6-OHDA and the non-injected hemisphere. This asymmetry indicates the loss of DA neurons in the injection side under the effect of 6-OHDA.

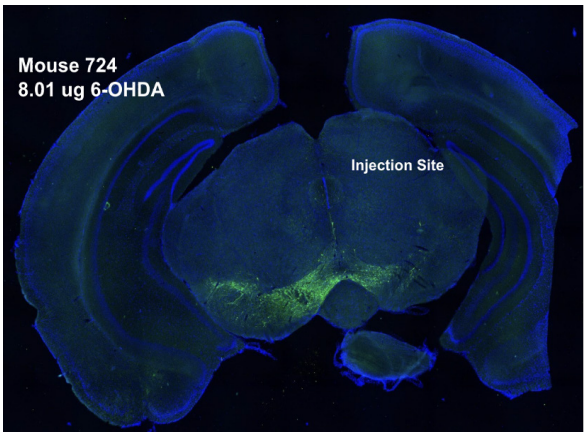


Figure 19: IHC sample image.

A casual inspection of the TH asymmetry suggests 6-OHDA dependence of DA neuron depletion (Figure 20). The higher the 6-OHDA dose, the more visual asymmetry appears to be. This observation motivated a systematical quantification of the extent of asymmetry to link the pathology and motor symptoms that vary with the 6-OHDA doses.

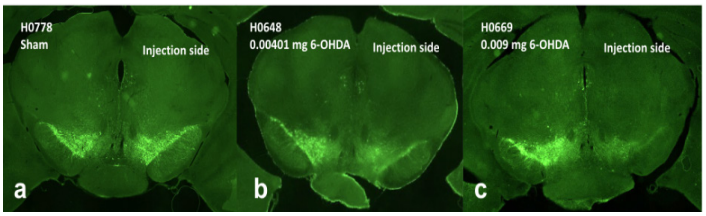


Figure 20: TH asymmetry vs. 6-OHDA dose.

Using the QUNIT pipeline, I systematically registered and automatically quantified the number of TH⁺ neurons in the brain section. I computed the ratio in the number of TH⁺ neurons between the right and left hemispheres. As all mice were injected with 6-OHDA or vehicle solution in the right hemisphere, the ratio is expected to be less than 1 for 6-OHDA mice. A larger DA degeneration will result in a ratio closer to zero. Consistent with this prediction, a significant negative correlation between 6-OHDA dose and the symmetry ratio (i.e., larger loss of DA neurons in the injected hemisphere) was observed (0 mg : *mean* = 0.983, *SD* = 0.0784; 0.00401 mg : *mean* = 0.76, *SD* = 0.254; 0.00801 mg : *mean* = 0.838, *SD* = 0.116; 0.009 mg : *mean* = 0.337, *SD* = 0.229; r = -0.577, p = 0.015*) (Figure 21). Thus, the earlier finding of the dose-dependent increase of impairment in some movement features is likely related to this dose-dependent increase of DA neurons. Indeed, a correlation between DA neuron loss and hyperactivities was found significant (r = 0.511, p < 0.015*) (Figure 22).

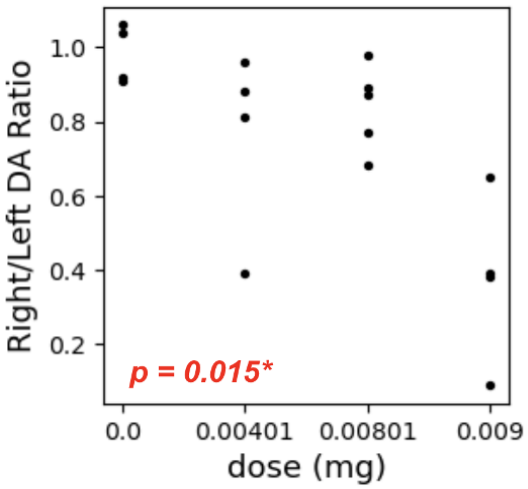


Figure 21: 6-OHDA dose vs. DA Neuron Depletion.

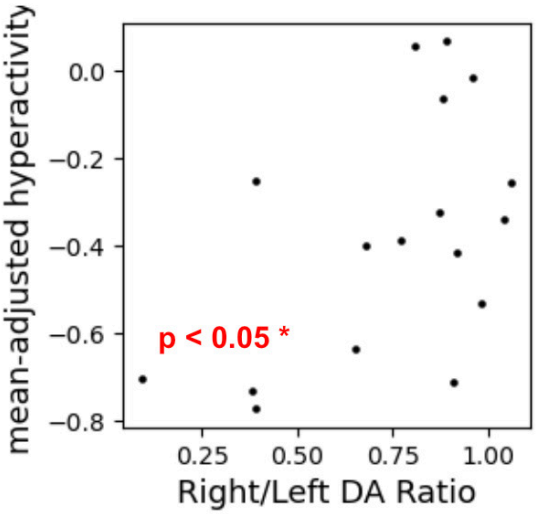


Figure 22: DA Neuron Depletion vs. mean-adjusted Hyperactivity.

This implementation of QUINT significantly increased my efficiency in TH⁺, DA neuron quantifications and provided data in line with the behavior analysis to some extent. However, it is crucial to evaluate the performance of QUINT model. Thus, I evaluated its performance by comparing QUINT quantification vs. human manual quantification in two aspects: 1) efficiency: how much DA neurons labeled by the human experimenter are also detected by QUINT; 2) accuracy: among all the QUINT detected neurons, how many of them are DA neurons labeled by the human experimenter.

QUINT outcome and human counting outcome were compared among three types of images: 1) high signal-to-noise ratio (SNR) that allows the human experiment to identify distinct DA neurons easily; 2) low SNR (Figure 23), and 3) saturated images that makes it difficult for the human experimenter to tease apart individual DA neurons from DA neuronal clusters (Figure 24). Across these three types of images, QUINT yielded high accuracy. However, the efficiency deteriorates for low SNR or saturated images (Table 2).

Tissue	False Positive	False Negative		Human Detected	Efficiency	Accuracy
Clear, distinct	14	60	636	682	91%	97.8%
Weak DA neuron	0	117	100	217	46%	100%
Clustered Neuron	28	752	143	867	16.5%	80.4%

Table 2: Manual evaluation of QUINT performance.

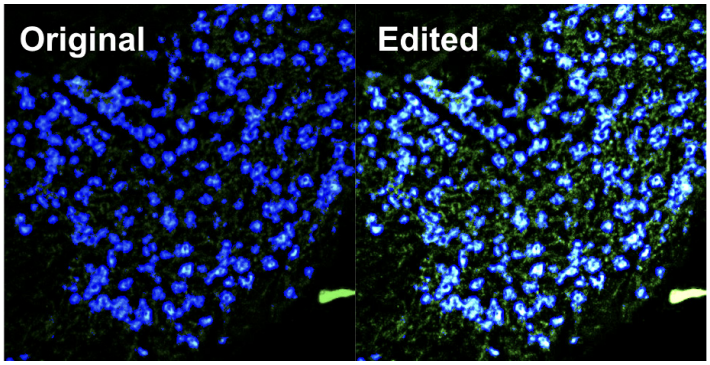


Figure 23: QUINT detection on weak neurons. QUINT is not able to detect weak signals from input images (left: raw input; right: edited to show weak signals).

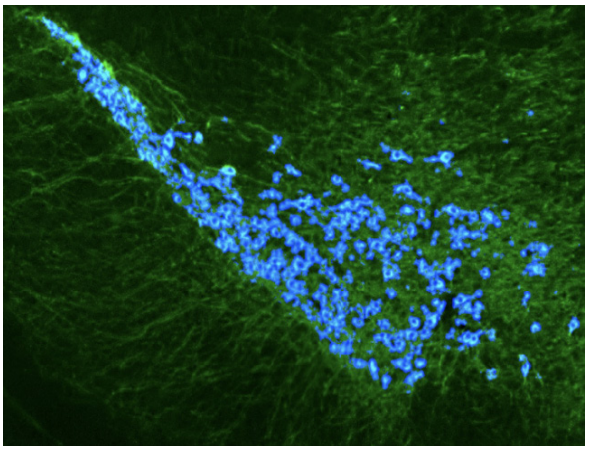


Figure 24: QUINT detection on clustered neurons.

The SNR of images can be improved by using a high magnification objective and longer exposure time during microscope imaging. To examine how much the QUINT-based automatic quantification can be improved by optimizing imaging procedures, I re-imaged low SRN tissues using a 20X objective. When exporting images, I also adjusted the contrast level and intensity of signals to enhance the visibility of weaker signals. Additionally, a diverse set of images was intentionally selected

to increase the reliability of the new model in quantifying a wide range of tissues. The resulting images exhibited improved sharpness and clarity, as shown on the QUINT training GUI (Figure 25). During manual labeling in this GUI, I labeled a larger number of weak and clustered neurons to ensure a sufficient input size.

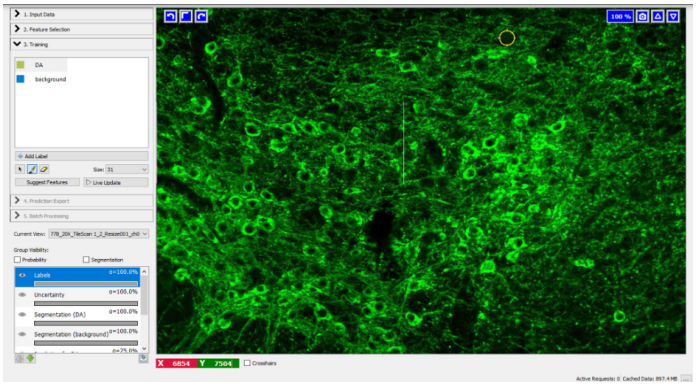


Figure 25: GUI of QUINT training page with 20X images.

While this new model is still undergoing trained and optimization, preliminary results have showed a significant improvement in detecting weak and clustered DA neurons, as shown in Table 3.

Tissue	False Positive	False Negative	QUINT Detected	Human Detected	Efficiency	Accuracy
Good	24	13	477	465	97.41%	94.96%
Weak	9	27	316	334	91.90%	97.20%
Clustered	16	40	263	287	93.90%	86.10%

Table 3: Manual evaluation of new QUINT model performance.

Discussions

The QUINT workflow is a powerful tool that automates our neuron quantification process. With a trained model, QUINT was capable of quantifying roughly 6-8 brain slices within half an hour, which is highly efficient compared to human operation, typically requiring ~5 hours. In our experiment, QUINT provided pathological evidence that supports our hypothesis about the 6-OHDA dose-dependent DA neuron loss and motor function impairments. While the current QUINT model performs less efficiently when detecting weak and saturated signals, we foresee great potential in further optimizing the performance of this AI-based analysis pipeline after improved pre-processing of images and model-training strategies.

Applying this new model after it is trained is expected to help us obtain more accurate quantifications. The combination of DA neuron loss quantification results and movement feature analysis suggests a more nuanced PD symptom detection that can be predictive in terms of pathological scores (e.g. the extent of DA neuron loss), captured by kinematic data. This potential model will go beyond the current binary classification algorithms, providing specific pathology evaluation for disease progression tracking and intervention strategy designing and adjustments.

Chapter 4: Linear Regression Model Development

Methods

To evaluate whether gross motor features can be reliable for PD disease diagnosis, a linear regression model was tested to classify whether mice were injected with 6-OHDA (i.e., PD mice) or not (non-PD mice), using the DLC and PyRat processed motor features extracted from the open-field test data as independent variables. The linear regression will provide us with a quantitative measure of the strength of each independent variable in detecting PD mice. To determine which motor features are significantly related to PD diagnosis, an Ordinary Least Square (OLS) test was conducted with each feature as the single independent variable. Using all identified significant movement features, a multiple linear regression model was built and evaluated with 1000 iterations of bootstrapping. The statistics of accuracy, precision, recall, R-squared, and mean squared error (MSE) were assessed from these bootstrapped samples. Accuracy evaluates the overall prediction (i.e., classification) accuracy. Precision provides insights into the model's ability to correctly identify correct and positive predictions and help minimize the risk of false positives, while recall evaluates the risk of false negatives. The ideal recall and precision values are both 1. R-squared is the coefficient of determination, which indicates the relevance between the dependent variable and independent variables. MSE assesses the differences between predictions and real data. Considering the imbalanced samples in our dataset (0:1 = 26:16; 0 represents the non-PD and 1 represents PD), a better-than-random model should obtain an accuracy above 0.619 (26/42 = 0.619). A perfect prediction model will have an MSE of 0 and an R-squared of 1.

Results

The total dataset used for the linear regression analysis includes 16 PD videos and 26 Non-PD videos. OLS revealed the total travel distance (r = -0.0183, p = 0.042) and hyperactivity frequency (r = -0.0288, p = 0.001) are significant predictors for PD diagnosis. These two features were used as input and the treatment conditions (PD versus non-PD) as output in the linear regression model, where the chance level accuracy was set to 0.61 (26/42 = 0.61).

Figure 25 shows a bootstrap example in which the linear regression model successfully classified 31 out of 42 videos, achieving an accuracy of 0.74. With 1000 bootstrapping, the model obtained a mean accuracy of 0.777, a mean precision of 0.805, a mean recall of 0.494, a mean R-square of 0.395, and a mean MSE of 0.139 (Figure 27). The low recall alarms the tendency for false negatives. This result suggests that input features used in the linear regression model are not sufficiently separable between the PD and non-PD mice to achieve the perfect classification.

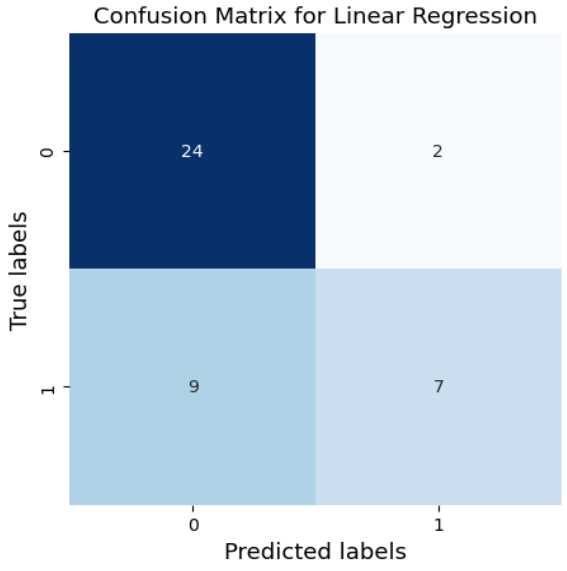


Figure 26: Confusion Matrix for linear regression.

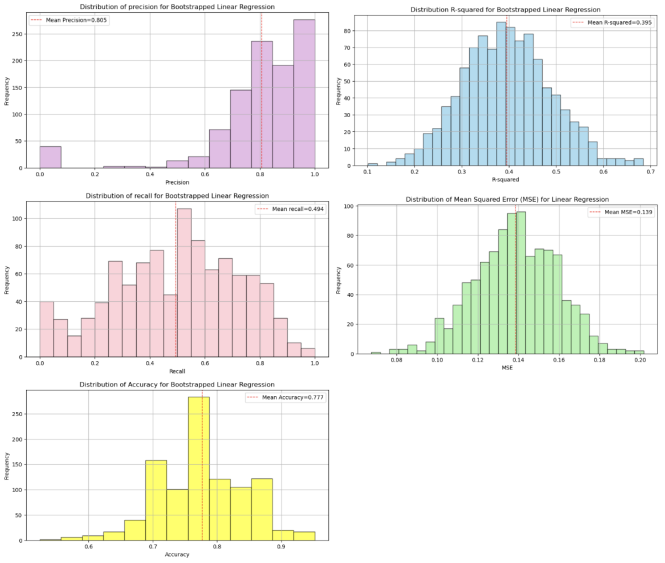


Figure 27: Linear regression bootstrapping evaluation.

Figure 28 shows which mice are misclassified by the linear regression model. This analysis indicates that the imperfect classification accuracy might be partly due to the inclusion of lowest-dose 6-OHDA samples, which exhibited motor features that are less discernable from the control condition, as shown in my earlier behavior analysis.

mouse	dose	prediction	experiment	label	video time (min)	dist (m)	median speed (cm/s)	freezing (s)	# of speedy movement
635	4.01	0	pd3	1	10.001667	0.764168	0.215797	-16.158333	6.25
774	0.00	1	pd3	0	10.002222	3.064842	-0.610996	3.445833	-19.00
791	9.00	0	pd3	1	10.002222	4.503407	-0.211187	-17.300000	-12.50
651	4.01	0	pd3	1	10.000556	-0.226731	0.273859	-6.600000	0.25
775	0.00	1	pd3	0	10.002222	-4.291538	-0.426063	51.104167	-13.25
642	4.01	0	pd3	1	10.001389	-0.259085	0.043865	-8.266667	-0.75
728	8.01	0	pd3	1	10.001389	-1.076996	-0.241641	2.533333	-7.75
669	9.00	0	pd3	1	10.002222	8.475097	-0.149554	-16.145833	-12.25
636	4.01	0	pd3	1	10.001667	-0.521155	-0.227760	18.050000	-8.00
724	8.01	0	pd3	1	10.000278	3.249837	0.462373	-13.295833	7.00
648	4.01	0	pd3	1	10.001111	-4.046249	0.135851	-10.745833	-6.00

Figure 28: Linear regression misclassified mice.

Discussions

While the linear regression model yields better-than-chance level accuracy and precision in identifying PD from non-PD mice, the low mean recall rate suggests a high number of false negatives. Notably, all five PD mice injected with the lowest dose of 6-OHDA were misclassified as non-PD mice, indicating that the selected motor features as inputs to the linear regressor are insufficient in detecting subtle symptoms. In other words, our highly reduced, low-dimensional movement features may not capture the full characteristics of PD movements that distinguish them from non-PD behaviors. To address this, I applied a neural network approach to detecting PD mice in the next chapter, using the full time series of kinematic data instead of the processed features.

Chapter 5: Ensemble Neural Network Development

Methods

I developed a preliminary ensemble neural network, combining the kinematic data from six individual body parts tracked by the DLC pipeline. First, individual base neural networks were constructed to classify PD mice from their kinematic data of each body part separately. Subsequently, the ensemble neural network combined the six predictions from these base neural networks as input to a logistic neural network.

Given the necessity for a large sample size for developing a neural network model, each 10-minute open-field video was segmented into 20 short clips (each 30 second-long clip) and used those short clips as individual samples. To further increase the sample size, videos recorded at pd1 and pd2 sessions were incorporated into the data pool, except for 2 videos with unexpected frame loss. The total sample size increased to 920 non-PD and 958 PD samples. The total data set was split into 90% training data and 10% testing data. Training data was used to tune the model while testing data was used for evaluating the performance of the trained model.

All base neural networks have the same input layer structure (n = 3601, including x and y coordinates of the body part at 1800 timeframes). A neural network using each body part data was separately trained and evaluated with one iteration of bootstrapping. The output layers consists of two output nodes: each representing the probability of being PD (i.e., 1) and non-PD (i.e., 0). The chance-level accuracy was set to 0.51 (i.e., chance level = 958/ (958+920) = 0.51).

Results

All base neural networks achieved a higher-than-random accuracy in predicting PD, where the right ear (accuracy = 0.706) and the snout (accuracy = 0.704) have the highest accuracies (Table 4). Accuracies, recalls and precision are promising across all six models.

Body Part	Input Layer	Hidden Layer	Output Layer	Accuracy	Recall	Precision	MSE
Tail Base	3601	(180,20)	2	0.627	0.66	0.659	0.373
Tail Tip	3601	(180,20)	2	0.569	0.506	0.633	0.431
Left Ear	3601	(600,30)	2	0.585	0.62	0.62	0.415
Right Ear	3601	(900,50)	2	0.706	0.669	0.776	0.294
Snout	3601	(600,30)	2	0.705	0.666	0.768	0.295
Nape	3601		2	0.695	0.648	0.758	0.305

Table 4: The structure and performance of base neural networks

Discussions

An ensemble neural network that combines the six base models is expected to perform better than any of individual base neural networks. Due to the time limit, an ensemble network model has not been fully tested. Nevertheless, my preliminary model tuned after only one iteration produced a promising result. An increased number of iterations of the current model is expected to yield a better performance in detecting PD mice from their rich movement data.

DISCUSSIONS

Neurological diseases can impair one's motor and cognitive functions, reducing quality of life significantly. There is an urgent call for a deeper understanding of these diseases thus treatments and interventions can be discovered. A fundamental challenge of elucidating the disease mechanisms is the fact that studies in humans are correlative in nature except for clinical trials. Therefore, animal model systems that enable controlled perturbations to exacerbate or ameliorate pathogenesis are needed to establish causality. For instance, PD animal models induced genetically or chemically have demonstrated their importance in our understanding of the molecular/cellular processes underlying Lewy body formation, synaptic modifications, neuroinflammation and other PD-related pathophysiology. A substantial, remaining challenge in those models is to identify how those pathophysiology result in behavioral symptoms. Currently, diagnosing PD symptoms in animal models is time-consuming, labor-intensive, and subjective, heavily reliant on human judges. The

objective of this thesis was to develop an AI system that automatically scores PD symptoms in mouse model from their motor data. My research findings show that 1) the extent of motor impairment is positively correlated with the extent of DA neuron loss, 2) the machine-learning-based tools, DeepLabCut and QUINT workflow can be utilized to automatically and robustly extract the animal model's motor behavior and pathology data, 3) an ensemble neural network that combines base networks trained with the time series of each body part position is a promising model that can accurately detect animals with PD symptoms from their motor behavior data.

Motion and Pathology Analysis

Machine learning algorithms are actively utilized in PD diagnosis in human subjects. These algorithms detect PD-related features in the subjects' voice data, gait data, and handwriting data (Mei et al., 2021). Research that examines motor behavior typically uses wearable devices to sense the subject's fine movements (Sotirakis et al., 2023). In contrast, the use of machine-learning technologies in animal model research has been far more limited despite the same need for accurate and sensitive detections of behavioral symptoms. In my thesis, I applied a machine learning technique to analyzing free-roaming kinematic data collected in the natural environment to detect PD motor symptoms and decern PD mice from non-PD mice.

I observed an overall increment in the total travel distance during the open-field test performed 3 weeks after injecting 6-OHDA or vehicle solution. Additionally, reduction in the median speed, and the frequency of hyperactivity, and an increment in freezing durations were observed in both 6-OHDA and vehicle mice Therefore, the change of these motor features over time do not differ between PD and vehicle mice. Thus, these apparent time-dependent changes that are common between the two treatment groups are likely to reflect the animals' habituation to the open-field environment from repetitive exposures which have been shown previously across different mice breeds (Bolivar et al., 2000). Nevertheless, some features show larger amounts of changes in high-dose 6-OHDA mice than vehicle or low-dose 6-OHDA mice, indicating that DA degeneration-related hypoactivity accentuated the habituation effect.

The QUINT workflow automated the DA neuron detection and quantification in the histology assay. The quantification of DA degeneration revealed a dose-dependent reduction in the frequency of hyperactivity. This finding that DA neuronal degeneration impairs mice's ability to execute fast movements is consistent with the known PD symptoms in humans. Herz and Brown's proposed an increase in energetic cost in movements among PD patients (Herz & Brown, 2023) (Figure 29). Even when the expected peak force is lowered for PD patients, their energetic cost is higher than healthy individuals. In addition, Montgomery and Nuessen suggest a speed-accuracy trade-off system that was reduced at an initial gain in movements and offset movement time among PD patients (Montgomery & Nuessen, 1990). They found that human PD subjects have longer reaction time and movement time, compared to healthy human subjects. With medication, PD subjects were able to shorten reaction time, movement time. The ability for one to execute maximal movement speed was improved with medication.

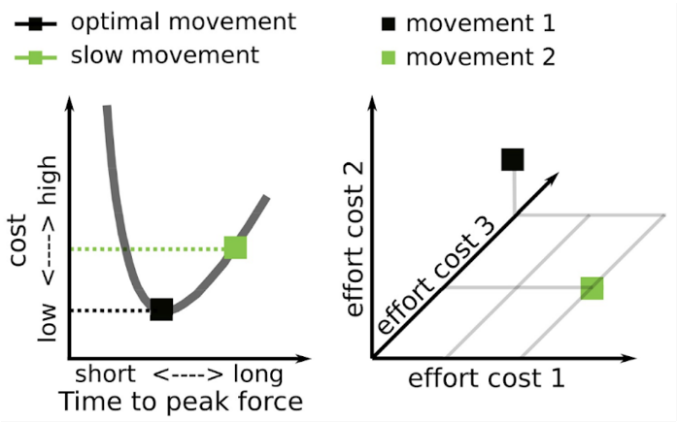


Figure 29: Effort cost of movements (Herz & Brown, 2023).

Overall, our results revealed 1) an uncommonly studied motor feature – hyperactivity, might be a new predictor in PD diagnosis in animal research, 2) greater DA neuron loss induces greater motor impairment.

Ensemble Machine Learning

I built a classifier that discriminate PD mice from non-PD mice by using a linear regression model and an ensemble neural network. Although our linear model using gross movement features show their potential in distinguishing between PD and non-PD mice, its performance was highly limited, likely due to a loss of information in a few motor features extracted from the rich movement data. Thus, I devised an ensemble neural network, aiming to improve the classification accuracy by using the full time series data of six body part positions. I first developed 6 base neural networks, each using the locomotor activities of a single body part as input signal. With only one iteration of each model, all of base networks exhibited better-than-chance performance in classification. These results suggest that the ensemble network utilizing the kinematic data from all six body parts might perform far more accurately after a proper optimization through multiple iterations.

LIMITATIONS

In the current study, movement features are computed only based on the base of the tail, assumed to be the most reliably labeled body part. Other body parts exhibit more dynamic movements during open-field tests, potentially reflecting PD symptoms that cannot be detected from the tail base movement. We computed only four movement features. Notably, hyperactivity frequency is not a feature commonly used in previous studies, but it unexpectedly turns out to be the most significant feature informing whether a mouse exhibiting PD. It is very possible that other less-explored motor features can be relevant to PD progression. For example, previous studies (Kim et al., 2018) showed the importance of gait changes in PD.

Second, the inconsistency of IHC quality across brain tissues deteriorates the automatic detection and quantification of DA neurons. Despite the same IHC procedure and recipes, variations in the IHC quality demonstrate human experimental errors' impact. Future investigations should optimize human IHC operations to avoid such performance-affecting variability.

Addressing the data size and bias issue is a significant challenge in the machine-learning model development. Machine learning is all about data. Small data size will lead to overfitting while biased data will yield a biased algorithm. Underrepresented population data tend to experience failed prediction or marginalization during training and prediction (Richardson, 2022).

FUTURE STUDIES

I plan to collect anterior brain sections from PD mice to investigate the relationship between the exact location and amount of the 6-OHDA injection and motor/pathological symptoms. As discussed in the introduction, ventral DA neuron projections are related to both cognitive and motor dysfunction. Dorsal projections are more specific to motor dysfunction. Examining the injection site will reveal the specific anatomical locus that corresponds to specific symptoms. I am currently developing the new QUINT model. Preliminary data from the new model show improved efficiency and sensitivity. I anticipate that further training of the new model will establish a more accurate DA neuron quantification system that can ideally remove human intervention.

I also aim to finish the development of the preliminary ensemble neural network by increasing the number of iterations and adjusting the model architectures. Building on the existing 90% training and 10% testing data splitting structure, splitting ~5% optimization data set will help significantly in fine-tuning each base neural network after using training and testing dataset, to yield better model performance. Developed base models should generate predictions per video clip. By integrating all predictions per video (n = ~20), a threshold (e.g. the number of PD predictions vs. the number of non-PD predictions) will be used to determine whether a video exhibits PD features. Following the logistics, each video will obtain 6 predictions, one per body part. From there, a final logistic or linear regression model can be used to integrate these 6 predictions, thus generating a final prediction of the video (Figure 30). With this ensemble model, the final model is expected to eliminate the prediction bias and error of each neural network and create an optimal model for PD diagnosis.

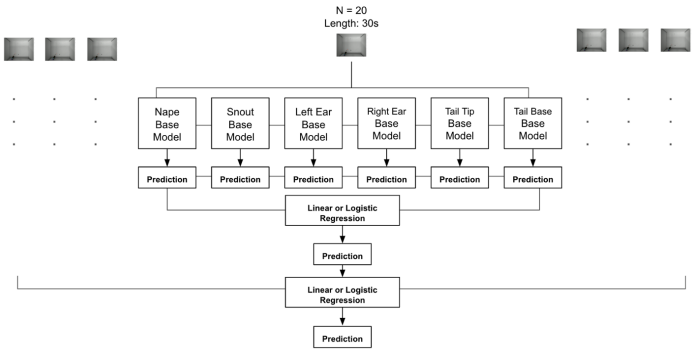


Figure 30: Future ensemble neural network structure per video

Finally, the ultimate goal of this project is to predict the pathological score of each mouse based on locomotive data. I will employ established machine-learning classifier such as time-series classifiers, deep-learning neural networks, and VVG to accomplish this goal. These advanced techniques could potentially capture the complexity of locomotive data better than the ensemble network to predict the corresponding pathological score in PD mice.

CONCLUSION

By employing AI-based models to acquire and analyze kinematic data from rodent models and to validate pathogenesis, my research established an analysis pipeline that reduces human intervention and potentially increases the objectivity and precision in detecting PD phenotypes. My finding that with more DA neuron loss, mice exhibit more sever motor dysfunction affirms the premise that pathological scores may be predicted from locomotive data collected in a natural environment. The preliminary neural network models that uses the full time series of kinematic data, instead of a few pre-defined motor feature, showed promising performance in detecting PD mice. By refining the base neural networks and combining them in an ensemble network is expected to enhance the classification performance. These results and outlook lay a strong foundation for the future endeavors to develop a system that automatically computes fine-grained symptom scores from movement videos.

References

Aich, S., Youn, J., Chakraborty, S., Pradhan, P. M., Park, S., & Park, J. (2020). A Supervised Machine Learning Approach to Detect the On/Off State in Parkinson's Disease Using Wearable Based Gait Signals. *Diagnostics*, *10*(6). https://doi.org/10.3390/diagnostics10060421

Aubert, I., Brana, C., Pellevoisin, C., Giros, B., Caille, I., Carles, D., Vital, C., & Bloch, B. (1997). Molecular anatomy of the development of the human substantia nigra. *The Journal of Comparative Neurology*, *379*(1), 72–87. https://doi.org/10.1002/(sici)1096-9861(19970303)379:1<72::aid-cne5>3.0.co;2-f

Asakawa, T., Fang, H., Sugiyama, K., Nozaki, T., Kobayashi, S., Hong, Z., Suzuki, K., Mori, N., Yang, Y., Hua, F., Ding, G., Wen, G., Namba, H., & Xia, Y. (2016). Human behavioral assessments in current research of parkinson's disease. *Neuroscience & Biobehavioral Reviews*, *68*, 741–772. https://doi.org/10.1016/j.neubiorev.2016.06.036

Avery, M. C., & Krichmar, J. L. (2017). Neuromodulatory systems and their interactions: A review of models, theories, and experiments. *Frontiers in Neural Circuits*, *11*. https://doi.org/10.3389/fncir.2017.00108

Azcorra, M., Gaertner, Z., Davidson, C., Hayes, C. K., Ramakrishnan, C., Fenno, L., Kim, Y. S., Deisseroth, K., Longnecker, R., Awatramani, R., & Dombeck, D. A. (2022). *Unique Functional Responses Differentially Map onto Genetic Subtypes of Dopamine Neurons*. https://doi.org/10.1101/2022.12.19.521076

Ball, T., González-Martínez, J., Zemmar, A., Sweid, A., Chandra, S., Van-Sickle, D., Neimat, J. S., Jabbour, P., & Wu, C. (2021). Robotic applications in cranial neurosurgery: Current and future. *Operative Neurosurgery*, *21*(6), 371–379. https://doi.org/10.1093/ons/opab217

Barnett, G. O. (1987). DXplain. *JAMA*, *258*(1), 67. https://doi.org/10.1001/jama.1987.03400010071030

Billingsley, K. J., Bandres-Ciga, S., Saez-Atienzar, S., & Singleton, A. B. (2018). Genetic risk factors in parkinson's disease. *Cell and Tissue Research*, *373*(1), 9–20. https://doi.org/10.1007/s00441-018-2817-y

Bolivar, V. J., Caldarone, B. J., Reilly, A. A., & Flaherty, L. (2000). Habituation of Activity in an Open Field: A Survey of Inbred Strains and F1 Hybrids. *Behavior Genetics*, *30*(4), 285–293. https://doi.org/10.1023/a:1026545316455

Bové, J., & Perier, C. (2012). Neurotoxin-based models of parkinson's disease. *Neuroscience*, *211*, 51–76. https://doi.org/10.1016/j.neuroscience.2011.10.057

Cheng, H., Ulane, C. M., & Burke, R. E. (2010). Clinical progression in parkinson disease and the neurobiology of axons. *Annals of Neurology*, *67*(6), 715–725. https://doi.org/10.1002/ana.21995

Chinta, S. J., & Andersen, J. K. (2005). Dopaminergic neurons. *The International Journal of Biochemistry & Cell Biology*, *37*(5), 942–946. https://doi.org/10.1016/j.biocel.2004.09.009

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021a). Artificial Intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021b). Artificial Intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Dickson, D. W. (2012). Parkinson's disease and parkinsonism: Neuropathology. *Cold Spring Harbor Perspectives in Medicine*, *2*(8). https://doi.org/10.1101/cshperspect.a009258

Dugan, J. P., Stratton, A., Riley, H. P., Farmer, W. T., & Mastick, G. S. (2011). Midbrain dopaminergic axons are guided longitudinally through the diencephalon by slit/robo signals. *Molecular and Cellular Neuroscience*, *46*(1), 347–356. https://doi.org/10.1016/j.mcn.2010.11.003

Engelbrecht, D. (2023). Unity ML-Agents. *Introduction to Unity ML-Agents*, 87–135. https://doi.org/10.1007/978-1-4842-8998-3_6

Fedorow, H., Tribl, F., Halliday, G., Gerlach, M., Riederer, P., & Double, K. (2005). Neuromelanin in human dopamine neurons: Comparison with peripheral melanins and relevance to parkinson's disease. *Progress in Neurobiology*, *75*(2), 109–124. https://doi.org/10.1016/j.pneurobio.2005.02.001

Fields, C. R., Bengoa-Vergniory, N., & Wade-Martins, R. (2019). Targeting alpha-synuclein as a therapy for parkinson's disease. *Frontiers in Molecular Neuroscience*, *12*. https://doi.org/10.3389/fnmol.2019.00299

Gaenslen, A., & Berg, D. (2010). Early diagnosis of parkinson's disease. *International Review of Neurobiology*, 81–92. https://doi.org/10.1016/s0074-7742(10)90006-8

Geertsma, H. M., Fisk, Z. A., Sauline, L., Prigent, A., Kurgat, K., Callaghan, S. M., Henderson, M. X., & Rousseaux, M. W. (2024). A topographical atlas of α-synuclein dosage and cell type-specific expression in adult mouse brain and peripheral organs. *Npj Parkinson's Disease*, *10*(1), 1-14. https://doi.org/10.1038/s41531-024-00672-8

German, D. C., & Manaye, K. F. (1993). Midbrain dopaminergic neurons (nuclei A8, A9, and A10): Three-dimensional reconstruction in the rat. *Journal of Comparative Neurology*, *331*(3), 297–309. https://doi.org/10.1002/cne.903310302

Girault, J.-A., & Greengard, P. (2004). The neurobiology of dopamine signaling. *Archives of Neurology*, *61*(5), 641. https://doi.org/10.1001/archneur.61.5.641

Glajch, K. E., Fleming, S. M., Surmeier, D. J., & Osten, P. (2012). Sensorimotor assessment of the unilateral 6-hydroxydopamine mouse model of parkinson's disease. *Behavioural Brain Research*, *230*(2), 309–316. https://doi.org/10.1016/j.bbr.2011.12.007

Goetz, C. G. (2012). Unified Parkinson's Disease Rating Scale (UPDRS) and Movement Disorder Society Revision of the UPDRS (MDS-UPDRS). *Rating Scales in Parkinson's Disease*, 62–83. https://doi.org/10.1093/med/9780199783106.003.0112

Gurdon B, Yates SC, Csucs G, Groeneboom N, Hadad N, Telpoukhovskaia M, Oullette A, Oullette T, O'Connell K, Singh S, Murdy T, Merchant E, Bjerke I, Kleven H, Schlegel U, Leergaard T, Puchades M, Bjaalie J, Kaczorowski C. Detecting the effect of genetic diversity on brain composition in an Alzheimer's disease mouse model. BioRxiv. Feb 2023. https://doi.org/10.1101/2023.02.27.530226

Hastie, T., Tibshirani, R., & Friedman, J. H. (2017). *The elements of Statistical Learning: Data Mining, Inference, and prediction*. Springer.

Hayakawa, S., Kataoka, K., Yamamoto, M., Asahi, T., & Suzuki, T. (2024). DeepLabCut-based daily behavioural and posture analysis in a Cricket. *Biology open*, bio.060237. Advance online publication. https://doi.org/10.1242/bio.060237

Herz, D. M., & Brown, P. (2023). Moving, fast and slow: Behavioural insights into Bradykinesia in parkinson's disease. *Brain*, *146*(9), 3576–3586. https://doi.org/10.1093/brain/awad069

Hssayeni, M. D., Jimenez-Shahed, J., Burack, M. A., & Ghoraani, B. (2019). Wearable sensors for estimation of Parkinsonian Tremor Severity During Free Body Movements. *Sensors*, *19*(19), 4215. https://doi.org/10.3390/s19194215

Iancu, R., Mohapel, P., Brundin, P., & Paul, G. (2005). Behavioral characterization of a unilateral 6-OHDA-lesion model of parkinson's disease in mice. *Behavioural Brain Research*, *162*(1), 1–10. https://doi.org/10.1016/j.bbr.2005.02.023

Jackson-Lewis, V., & Przedborski, S. (2007). Protocol for the MPTP mouse model of parkinson's disease. *Nature Protocols*, *2*(1), 141–151. https://doi.org/10.1038/nprot.2006.342

Keijsers, L. W., Horstink, I. M., & Gielen, A. M. (2006). Ambulatory motor assessment in Parkinson's disease. *Movement Disorders*, *21*(1), 34-44. https://doi.org/10.1002/mds.20633

Kim, S. M., Kim, D. H., Yang, Y., Ha, S. W., & Han, J. H. (2018). Gait patterns in parkinson's disease with or without cognitive impairment. *Dementia and Neurocognitive Disorders*, *17*(2), 57. https://doi.org/10.12779/dnd.2018.17.2.57

Klein, C. J., Budiman, T., Homberg, J. R., Verma, D., Keijer, J., & van Schothorst, E. M. (2022). Measuring locomotor activity and behavioral aspects of rodents living in the home-cage. *Frontiers in Behavioral Neuroscience*, *16*. https://doi.org/10.3389/fnbeh.2022.877323

Kordower, J. H., Olanow, C. W., Dodiya, H. B., Chu, Y., Beach, T. G., Adler, C. H., Halliday, G. M., & Bartus, R. T. (2013). Disease duration and the integrity of the nigrostriatal system in parkinson's disease. *Brain*, *136*(8), 2419–2431. https://doi.org/10.1093/brain/awt192

Kumar, N. (2009). The Sydney Multicenter Study of Parkinson's Disease: The inevitability of dementia at 20 years. *Yearbook of Neurology and Neurosurgery*, *2009*, 94–95. https://doi.org/10.1016/s0513-5117(09)79027-1

Luo, S. X., & Huang, E. J. (2016). Dopaminergic neurons and brain reward pathways. *The American Journal of Pathology*, *186*(3), 478–488. https://doi.org/10.1016/j.ajpath.2015.09.023

Mat Taib, C. N., & Mustapha, M. (2020). MPTP-induced mouse model of parkinson's disease: A promising direction of therapeutic strategies. *Bosnian Journal of Basic Medical Sciences*. https://doi.org/10.17305/bjbms.2020.5181

McGregor, M. M., & Nelson, A. B. (2019). Circuit mechanisms of parkinson's disease. *Neuron*, *101*(6), 1042–1056. https://doi.org/10.1016/j.neuron.2019.03.004

Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of parkinson's disease: A review of literature. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.633752

Mitsumoto, Y., Watanabe, A., Mori, A., & Koga, N. (1998). Spontaneous regeneration of nigrostriatal dopaminergic neurons in MPTP-treated C57BL/6 mice. *Biochemical and Biophysical Research Communications*, *248*(3), 660–663. https://doi.org/10.1006/bbrc.1998.8986

Montgomery, E. B., & Nuessen, J. (1990). The movement speed/accuracy operator in parkinson's disease. *Neurology*, *40*(2), 269–269. https://doi.org/10.1212/wnl.40.2.269

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Neveen A, S. (2019). Mesenchymal stem cell based therapy for parkinson's disease. *International Journal of Stem Cell Research & Therapy*, *6*(1). https://doi.org/10.23937/2469-570x/1410062

Obeso, J. A., Rodriguez-Oroz, M. C., Rodriguez, M., Lanciego, J. L., Artieda, J., Gonzalo, N., & Olanow, C. W. (2000). Pathophysiology of the basal ganglia in parkinson's disease. *Trends in Neurosciences*, *23*. https://doi.org/10.1016/s1471-1931(00)00028-8

Pan, T., Kondo, S., Le, W., & Jankovic, J. (2008). The role of autophagy-lysosome pathway in neurodegeneration associated with parkinson's disease. *Brain*, *131*(8), 1969–1978. https://doi.org/10.1093/brain/awm318

Pavese, N., & Brooks, D. J. (2009a). Imaging neurodegeneration in parkinson's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1792*(7), 722–729. https://doi.org/10.1016/j.bbadis.2008.10.003

Pereira, E. A., & Aziz, T. Z. (2006). Parkinson's disease and primate research: Past, present, and future. *Postgraduate Medical Journal*, *82*(967), 293–299. https://doi.org/10.1136/pgmj.2005.041194

Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E. S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W. G., Lazzarini, A. M., Duvoisin, R. C., Di Iorio, G., Golbe, L. I., & Nussbaum, R. L. (1997). Mutation in the α-synuclein gene identified in families with parkinson's disease. *Science*, *276*(5321), 2045–2047. https://doi.org/10.1126/science.276.5321.2045

Redgrave, P., & Gurney, K. (2006). The short-latency dopamine signal: A role in discovering novel actions? *Nature Reviews Neuroscience*, *7*(12), 967–975. https://doi.org/10.1038/nrn2022

Richardson, A. (2022). Biased data lead to biased algorithms. *Canadian Medical Association Journal*, *194*(9). https://doi.org/10.1503/cmaj.80860

Sedelis, M., Schwarting, R. K. W., & Huston, J. P. (2001a). Behavioral phenotyping of the MPTP mouse model of parkinson's disease. *Behavioural Brain Research*, *125*(1–2), 109–125. https://doi.org/10.1016/s0166-4328(01)00309-6

Seibenhener, M. L., & Wooten, M. C. (2015). Use of the open field maze to measure locomotor and anxiety-like behavior in mice. *Journal of Visualized Experiments*, (96). https://doi.org/10.3791/52434

Siddiqui, I. J., Pervaiz, N., & Abbasi, A. A. (2016). The parkinson disease gene SNCA: Evolutionary and structural insights with pathological implication. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep24475

Sotirakis, C., Su, Z., Brzezicki, M. A., Conway, N., Tarassenko, L., FitzGerald, J. J., & Antoniades, C. A. (2023). Identification of motor progression in parkinson's disease using wearable sensors and machine learning. *Npj Parkinson's Disease*, *9*(1). https://doi.org/10.1038/s41531-023-00581-2

Srinivasan, E., Chandrasekhar, G., Chandrasekar, P., Anbarasu, K., Vickram, A. S., Karunakaran, R., Rajasekaran, R., & Srikumar, P. S. (2021). Alpha-synuclein aggregation in parkinson's disease. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.736978

Taylor, T. N., Greene, J. G., & Miller, G. W. (2010). Behavioral phenotyping of mouse models of parkinson's disease. *Behavioural Brain Research*, *211*(1), 1–10. https://doi.org/10.1016/j.bbr.2010.03.004

Thiele, S. L., Warre, R., & Nash, J. E. (2012). Development of a unilaterally-lesioned 6-OHDA mouse model of parkinson's disease. *Journal of Visualized Experiments*, (60). https://doi.org/10.3791/3234-v

Tieu, K. (2011). A guide to neurotoxic animal models of parkinson's disease. *Cold Spring Harbor Perspectives in Medicine*, *1*(1). https://doi.org/10.1101/cshperspect.a009316

Tillerson, J. L., Caudle, W. M., Reverón, M. E., & Miller, G. W. (2002). Detection of behavioral impairments correlated to neurochemical deficits in mice treated with moderate doses of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. *Experimental Neurology*, *178*(1), 80–90. https://doi.org/10.1006/exnr.2002.8021

Uemura, N., Ueda, J., Yoshihara, T., Ikuno, M., Uemura, M. T., Yamakado, H., Asano, M., Trojanowski, J. Q., & Takahashi, R. (2021). α□synuclein spread from olfactory bulb causes hyposmia, anxiety, and memory loss in bac□*snca* mice. *Movement Disorders*, *36*(9), 2036–2047. https://doi.org/10.1002/mds.28512

Unger, E. L., Eve, D. J., Perez, X. A., Reichenbach, D. K., Xu, Y., Lee, M. K., & Andrews, A. M. (2006). Locomotor hyperactivity and alterations in dopamine neurotransmission are associated with overexpression of A53T mutant human α-synuclein in mice. *Neurobiology of Disease*, *21*(2), 431–443. https://doi.org/10.1016/j.nbd.2005.08.005

Villar□Piqué, A., Lopes da Fonseca, T., & Outeiro, T. F. (2015). Structure, function and toxicity of alpha□synuclein: The bermuda triangle in Synucleinopathies. *Journal of Neurochemistry*, *139*(S1), 240–255. https://doi.org/10.1111/jnc.13249

Volkow, N. D., Wang, G.-J., Fowler, J. S., & Telang, F. (2008). Overlapping neuronal circuits in addiction and obesity: Evidence of systems pathology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1507), 3191–3200. https://doi.org/10.1098/rstb.2008.0107

Wakamatsu, M., Ishii, A., Iwata, S., Sakagami, J., Ukai, Y., Ono, M., Kanbe, D., Muramatsu, S., Kobayashi, K., Iwatsubo, T., & Yoshimoto, M. (2008). Selective loss of nigral dopamine neurons induced by overexpression of truncated human α-synuclein in mice. *Neurobiology of Aging*, *29*(4), 574–585. https://doi.org/10.1016/j.neurobiolaging.2006.11.017

WHO Media Team. (2024, March 14). *Over 1 in 3 people affected by neurological conditions, the leading cause of illness and disability worldwide*. World Health Organization. https://www.who.int/news/item/14-03-2024-over-1-in-3-people-affected-by-neurological-conditions--the-leading-cause-of-illness-and-disability-worldwide#:~:text=%C2%A9-,Over%201%20in%203%20people%20affected%20by%20neurological%20conditions%2C%20the,of%20illness%20and%20disability%20worldwide&text=A%20major%20new%20study%20released,living%20with%20a%20neurological%20condition.

Wiltshire, C., Lewis□Cheetham, J., Komedová, V., Matsuzawa, T., Graham, K. E., & Hobaiter, C. (2023). deepwild: Application of the pose estimation tool deeplabcut for behaviour tracking in wild chimpanzees and bonobos. *Journal of Animal Ecology*, *92*(8), 1560–1574. https://doi.org/10.1111/1365-2656.13932

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, *5*(6), 483–494. https://doi.org/10.1038/nrn1406

Yates, S. C., Groeneboom, N. E., Coello, C., Lichtenthaler, S. F., Kuhn, P.-H., Demuth, H.-U., Hartlage-Rübsamen, M., Roßner, S., Leergaard, T., Kreshuk, A., Puchades, M. A., & Bjaalie, J. G. (2019). QUINT: Workflow for quantification and spatial analysis of features in histological

images from Rodent Brain. *Frontiers in Neuroinformatics*, *13*. https://doi.org/10.3389/fninf.2019.00075

Zhao, A., Li, Y., Niu, M., Li, G., Luo, N., Zhou, L., Kang, W., & Liu, J. (2020). SNCA hypomethylation in rapid eye movement sleep behavior disorder is a potential biomarker for parkinson's disease. *Journal of Parkinson's Disease*, *10*(3), 1023–1031. https://doi.org/10.3233/jpd-201912