Prediction of Autism in Mice through Behaviour Analysis using Computer Vision and Data Science

Reflection

Bachelor Applied Computer Science
Digital Innovation

Academic year 2022-2023

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

Nicky Embrechts

Table of Contents

1 INTRODUCTION

As I reach the end of my internship journey, I want to take a moment to reflect on the valuable experiences and lessons I have gained during this time. This reflection report serves as an opportunity for me to share my thoughts and insights from my internship at the AI Medical Sensing and Diagnostic Laboratory at the Holon Institute of Technology (HIT) in Holon, Israel.

During my internship, I had the chance to explore the exciting field of AI-driven healthcare research. From the beginning, I was fascinated by the potential of artificial intelligence to transform healthcare, and I was eager to contribute to this area.

In this report, I will discuss the projects I worked on, the challenges I faced, and the skills I developed. I hope that by sharing my experiences, I can provide valuable insights into how AI is used in healthcare and inspire others who are starting their own journeys.

Throughout the report, I will explore the specific tasks I undertook, the technical skills I learned, and the knowledge I gained in areas like data analysis, machine learning algorithms, and predictive modeling. I will also discuss the supportive and collaborative environment that encouraged my growth, the guidance I received from mentors, and the connections I made with professionals in the industry.

This report aims to convey the significant impact that my internship had on my personal and professional growth. I will reflect on the lessons I learned, the challenges I overcame, and the valuable insights I gained as I navigated the ever-changing landscape of AI in healthcare research.

2 SUBSTANTIVE REFLECTION

2.1 Background

During my internship, I had the opportunity to work on a cutting-edge research project focused on studying the biometric features of autism in mice. Autism is a complex neurodevelopmental disorder that affects millions of individuals worldwide. Understanding its underlying mechanisms and identifying potential biomarkers is crucial for developing effective diagnostic tools and treatment strategies.

Traditionally, the analysis of mouse behaviour in autism research has relied on manual observations, which can be subjective and prone to human error. Moreover, the process is labor-intensive and time-consuming, limiting the ability to analyze large datasets and extract meaningful insights.

To overcome these limitations, the research team recognized the need for an automated approach to classify and analyze biometric features in mice. This involved developing a sophisticated software tool capable of accurately and efficiently processing vast amounts of behavioural data. By leveraging advanced algorithms and machine learning techniques, we aimed to uncover patterns, correlations, and behavioural markers that may be indicative of autism in mice.

To overcome these limitations, the research team recognized the need for an automated approach to classify and analyze biometric features in mice. This involved developing a sophisticated software tool capable of accurately and efficiently processing vast amounts of behavioural data. By leveraging advanced algorithms and machine learning techniques, we aimed to uncover patterns, correlations, and behavioural markers that may be indicative of autism in mice.

2.2 Works

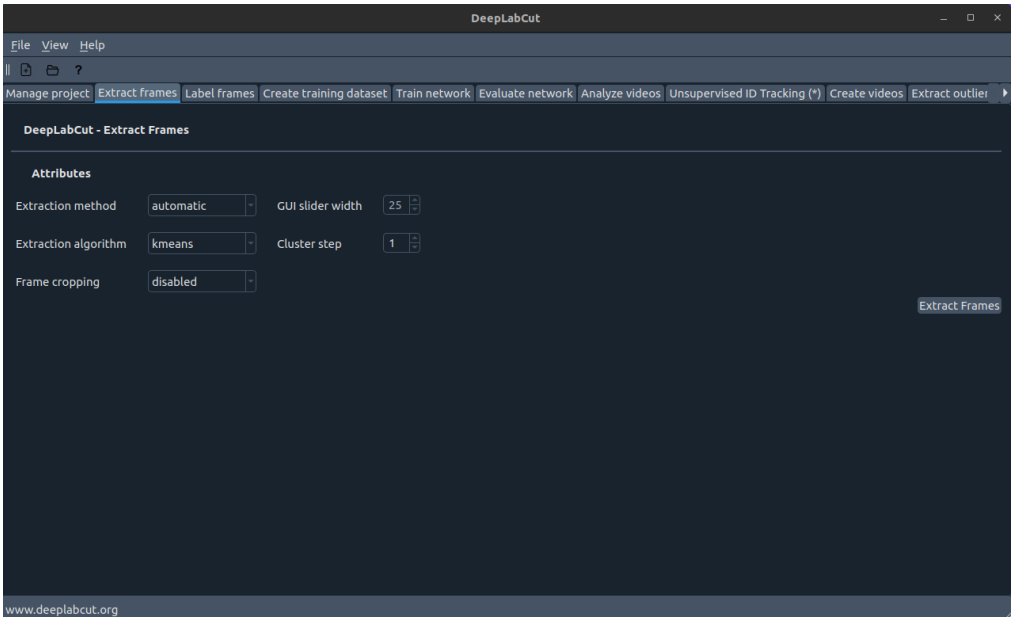
During my internship, my primary focus was on conducting research and exploring different software tools to address the challenge of analyzing biometric features in mice. I worked on developing and testing various software solutions to determine the most effective approach.

2.2.1 Skeleton extraction software

As a result of my research and experimentation, I developed two AI models utilizing computer vision techniques. The first model focused on extracting individual mouse joints, while the second model specialized in extracting joints from multiple mice within a single cage. These models, along with the DeepLabCut graphical user interface (GUI), formed a comprehensive solution for predicting and analyzing mouse poses.

DeepLabCut is a widely-used software tool in the field of behavioural analysis that provides a foundation for pose estimation tasks. Leveraging the capabilities of DeepLabCut, I customized and fine-tuned the models to accurately capture the joints of mice in various behavioural scenarios.

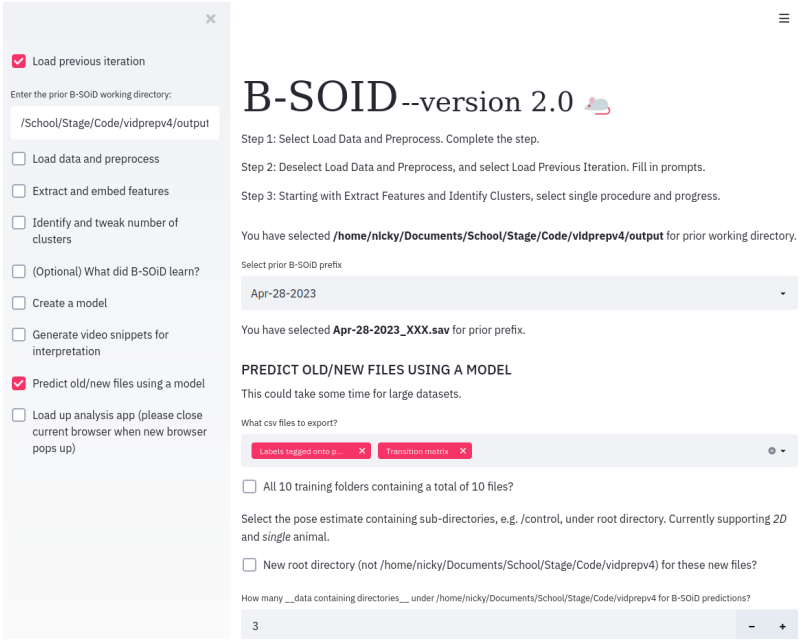
To ensure accessibility and ease of use, I also created a cloud-based notebook that allowed users to employ the AI models for making predictions or retraining them on their own datasets. This notebook provided a user-friendly interface to interact with the models, making it convenient for researchers to utilize the predictive power of the AI algorithms.



2.2.2 Behavioural feature and clustering software

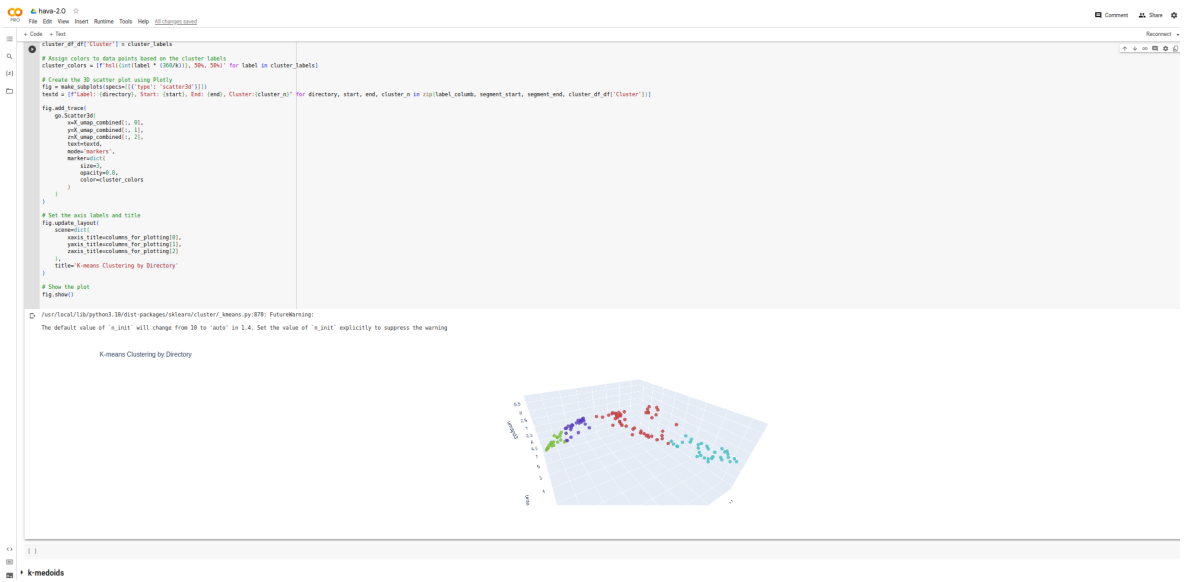
Another important contribution I made was an updated version of an outdated software called B-SOiD. This software was capable of unsupervised behaviour estimation using the joints extracted from the first set of software tools. However, the original software had outdated dependencies and did not provide sufficient data for the project's requirements. To address these limitations, I modified the source code to enable users to extract previously unavailable data and improved the software's performance by adding extra features to extract.

In addition to the core software tools, I developed various scripts to facilitate the data preparation process between DeepLabCut and B-SOiD. These scripts automated tasks such as data augmentation, feature extraction, dimension reduction, and data organization within a standardized data framework. By automating these processes, I aimed to streamline the data analysis pipeline and enhance overall efficiency.



2.2.3 Notebooks

To further support data analysis, I created notebooks specifically focused on clustering the data and deriving meaningful insights from these clusters. These analysis notebooks were instrumental in understanding the patterns and correlations within the collected data, enabling researchers to make informed conclusions and guide their further investigations.



2.2.4 Start of classifier

Finally, I embarked on developing a basic machine learning classifier capable of predicting the genotype of mice based on generalized, reduced, and clustered data. The purpose of this classifier was to utilize the processed data and leverage machine learning algorithms to make accurate genotype predictions.

However, it is important to note that the classifier is still in its early stages and requires further refinement and optimization. Due to time constraints, I was unable to fully complete the development and fine-tuning of the classifier. Nevertheless, the initial results showed promise, demonstrating the potential of using machine learning techniques in genotype prediction.

2.3 Future

Looking towards the future, there are several key suggestions and considerations for further improvement in the project. First and foremost, it is crucial to focus on obtaining a higher quality dataset for analysis. During my internship, I faced challenges with the dataset's quality, which required extensive time and effort to analyze and ensure its usefulness. By prioritizing the acquisition of a more reliable and comprehensive dataset, the subsequent analysis and research can be conducted more efficiently and effectively.

Another important aspect to address is the development of a streamlined and robust software tool. Currently, users are required to follow documentation on data processing and utilize multiple software tools. However, it would be ideal to create a containerized software that incorporates the optimal methodology discovered during the research phase. This containerized software should be designed to be user-friendly, particularly for non-technical individuals in the field of bio-sciences who may not have extensive IT

expertise. By simplifying the software interface and making it accessible to a broader audience, the adoption and utilization of the developed tools can be enhanced.

3 PERSONAL REFLECTION

3.1 Within the project

During my internship, under the guidance of Dr. Gaddi Blumrosen, I gained practical experience in machine learning and data science. I learned the importance of breaking down complex data problems into manageable components and acquired various techniques for problem decomposition. Additionally, I explored data augmentation techniques to enhance datasets and improve model performance. The internship provided opportunities to stay updated on state-of-the-art technologies in medical analysis and interact with renowned computer scientists. I developed skills in meticulous analysis and adjusting research trajectory based on intermediate results. Automating processes proved crucial in optimizing productivity and resource utilization. Overall, the internship enriched my technical abilities, instilled a passion for continuous learning, and fueled my drive to tackle complex problems.

3.2 Outside of the project

During my internship in Israel, I encountered various personal challenges that required me to adapt and step outside of my comfort zone. One significant challenge was the language barrier. Being in a country where I did not understand the language, everyday tasks such as using public transportation or grocery shopping became more difficult, and I had to rely on my communication skills to navigate these situations.

Additionally, I faced mental challenges due to missile strikes launched from Gaza into the regions where I lived and worked. I had to seek shelter multiple times, which was a daunting experience. The first time I heard the explosions from the Iron Dome intercepting the rockets in the sky, it was a startling moment that prompted me to quickly find the nearest bunker. It was disconcerting to witness other students acting as though it was just another day, which further emphasized the contrast in experiences. Although this situation lasted for a relatively short period of time, it left a lasting impression and tested my resilience.

On a technical level, I encountered a challenge with the secure public network setup at HIT. The network blocked access to certain profiles, particularly those associated with Linux operating systems. As a Linux user, I did not have direct internet access on my laptop. To overcome this obstacle, I set up a VPN connection to my home server in Belgium and created an exit node to access the internet. This not only provided me with internet access but also ensured a more secure and encrypted data stream. As a result, I could work from public spaces such as coffee shops with reduced concerns about data interception and potential hacking attempts. Although this technical issue fell outside the scope of my internship, I took the initiative to resolve it independently.

These personal and technical challenges during my internship in Israel not only tested my adaptability and problem-solving skills but also provided me with unique experiences that expanded my horizons and strengthened my resilience in unfamiliar situations.

4 Conclusion

In conclusion, my internship experience provided me with a comprehensive understanding of machine learning and data science, while also fostering critical skills in problem-solving and data analysis. Working under Dr. Blumrosen's guidance, I gained invaluable insights, had the opportunity to explore cutting-edge technologies, and engaged with prominent researchers in the field. These experiences have not only enhanced my technical abilities but have also instilled in me a passion for continuous learning and a drive to contribute innovative solutions to complex problems.