Automated Methodologies to Assess Welfare of Rhesus Macaques (Macaca mulatta) in a Breeding Colony

Giulia Ciminelli

A thesis submitted for the degree of Doctor of Philosophy

Biosciences Institute

Newcastle University

November 2023

Abstract

This thesis endeavours to explore the application of machine learning-based methodologies to gain valuable insights into the behaviour of captive Rhesus macaques. The primary objectives are to enhance their welfare and streamline management practices. Rhesus macaques hold significant importance in biomedical research, making their welfare a paramount concern in such environments. The subjects of this study reside within a breeding colony in the UK, serving as a source of individuals for neuroscientific laboratories. These macaques are grouped into breeding and juvenile cohorts, with continuous surveillance via a comprehensive CCTV system at the Centre of Macaques, comprising cameras within each enclosure.

Efficient macaque management is crucial not only for ensuring their well-being but also for facilitating successful breeding programs and the subsequent supply of animals to research laboratories. However, the process of guaranteeing both their welfare and the collection of informative behaviour data necessitates specialized personnel, incurring significant time and financial costs.

This study presents three distinct implementations of computer vision-based pipelines that leverage video footage from the existing CCTV system and researcher-recorded videos to autonomously collect and analyse behaviour of interest. This innovative approach not only saves time in data collection and analysis but also extracts previously unattainable information for the facility's staff.

The first project focuses on temperament assessment in macaques, employing three models based on object detection and pose estimation. These models analyse pre-existing researcher-recorded videos designed for the same purpose.

The second project investigates diverse foraging patterns based on varying food mixes provided. This automated methodology utilizes object detection to calculate the percentage of macaques engaged in foraging activities from CCTV footage.

The third project aims to evaluate two enrichment items, one food-based and one non-food-based. The automated methodology relies on object detection algorithms to extract data concerning the number of individuals interacting with these objects over specified time intervals.

These projects harness existing camera infrastructure and furnish invaluable insights into neophobia, temperament, preferred food choices, and enrichment planning, all of which inform crucial management decisions. Consequently, this thesis underscores the advantages of employing such methodologies and illustrates their potential for broader application, with the capacity to enhance the welfare and management of non-human primates in similar facilities.

Acknowledgments

I would like to express my profound gratitude to the individuals who have been instrumental in guiding and supporting me throughout my PhD journey.

First and foremost, I extend my appreciation to my supervisors: Dr Claire Witham, Professor Melissa Bateson, and Dr Stephen McGough. Claire introduced me to the fascinating realm of automated behaviour detection and shared her passion for macaques. Melissa has been a constant source of guidance and inspiration, resembling a pillar of support in my academic journey.

I am also deeply thankful to my panel members, Dr Tom Smulders, and Professor Andrew Jackson, who reviewed my progress over the years, offering invaluable guidance and assistance.

My gratitude extends to my fellow PhD colleagues, with a special acknowledgment to David Massey, who generously shared his knowledge and enthusiasm about macaques and collaborated with me during a significant portion of my PhD. A heartfelt thank you goes to Tim-Joshua Andres, with whom I shared the highs and lows of my PhD journey.

I am immensely appreciative of the Leverhulme Trust, the University of Newcastle, and the Centre for Macaques for their support, which made this project possible.

My friends and family deserve a special mention for their presence throughout these four years. Their curiosity and interest in my project have been invaluable sources of support and motivation.

Lastly, a special thanks to the macaques at CFM whose unpredictable and charming behaviours made my work both challenging and incredibly rewarding.

Table Of Contents

Automated Macaque Behaviour

The Automated Macaque Behaviour repository is available at https://github.com/GiuliaCiminelli/AutomatedMacaqueBehaviour and contains four folders, each tailored to a specific project:

- Temperament Test: This folder includes Python scripts for extracting information from three models: two based on DeepLabCut and one based on YOLACT.
- Foraging: This directory is subdivided into three sub-folders, each aligning with a distinct project. Every sub-folder contains Python and R scripts for extracting data on the number of macaques foraging.
- Enrichment_Yolcat: This folder includes Python and R scripts designed for extracting movement information from a YOLACT-based model trained to recognize a white tank, used as enrichment.
- Enrichment_Yolo: This folder contains Python and R scripts for counting the percentage of macaques interacting with an enrichment puzzle. The YOLO-based model can be accessed within the 'Model' sub-folder.

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Overview

Macaques play an essential role in medical and scientific research because of their similarity to humans in physiology, reproduction, cognition and social complexity (Phillips *et al.*, 2014). However, this social complexity, makes their management a challenge in captivity. Many research facilities housing captive animals must balance the need for standardized housing conditions, essential for treatments, husbandry, and sample collection, with the desire for a more varied environment. This typically includes multi-sex and multi-age groups, presence of enrichments, and a diverse food selection. All of these factors contribute to the macaques' welfare.

Studies have shown that providing a social environment that mimics the behavioural biology of macaques in their natural habitat not only enhances the welfare of these animals but also positively influences the reproducibility and consistency of research outcomes (Hannibal *et al.*, 2017). However, the implementation, control, and management of a complex environment for macaques can be time-consuming. In this thesis, I explore how recent advances in machine learning can contribute to the collection and analysis of data, which, holds the potential to enhance the management of macaques residing in a breeding colony under varying conditions.

1.2 Motivation and Problems

The welfare of laboratory non-human primates (NHPs) is one of the most challenging and contested issues concerning animals used in biomedical research. In laboratories, as well as in any other facilities hosting captive animals, the personnel have the legal and moral obligation to guarantee a high standard of animal welfare. Observing behaviour is one of the most used methods to assess animal welfare. However, conducting behavioural observations is time consuming to implement and it requires trained personnel (Porto et al., 2013; Mathis et al., 2018; Weinstein, 2018). To collect data on specific behaviours of interest, a human observer may need to watch hours of video and observe individuals for extended periods, which can be tiring and affect data reliability (Young et al., 1979; Mathis et al., 2018; Bateson and Martin, 2021; Kennedy, 2022). On the other hand, different observers can score the same behaviour differently adding inconsistency and bias to the data collected (Young *et al.*, 1979; Tuyttens *et al.*, 2014; Kennedy, 2022). Moreover, watching for behaviours that are infrequent but important can be very tedious for the human observer and some behaviours may not be visible at all.

Automated video analyses have previously found application in the evaluation and enhancement of animal welfare. Notable examples include the utilization of optical flow in the assessment of movement patterns in chickens and other farm animals, as well as the deployment of pixel difference methods in macaque studies (Dawkins, Cain and Roberts, 2012; Gronskyte *et al.*, 2016; Green, 2018; Wurtz *et al.*, 2019). Although these methodologies have proven effective in detecting animal movement, they have been unable to capture more detailed behaviours. Consequently, they are limited in their capacity to investigate more intricate welfare-related inquiries, such as the examination of interactions with objects and food. However, the advancements in machine learning and computer vision, have the potential to solve some of these challenges (Blumrosen, Hawellek and Pesaran, 2017; Kennedy, 2022). In the last 20 years, artificial intelligence algorithms have replaced many human tasks, and several toolboxes based on computer vision have been implemented to identify body parts and objects in videos of humans and other animals (i.e. Blanco Negrete et al., 2021; Bolya et al., n.d.; Labuguen et al., 2019). Video-based systems are consistent, effective, and allow for data recording for future study (Rushen, Chapinal and de Passillé, 2012; Okinda et al., 2020). Moreover, these approaches are able to gather more data recording the environment and the individuals within it, save time in data analyses and collection, and are cost effective (Pimm et al., 2015; Barnard et al., 2016; Mathis et al., 2018; Weinstein, 2018). Computer vision-based technologies also remove the need for sensors or marks, making them non-invasive and non-intrusive, without affecting individuals' behaviours (Line et al., 1990; Rushen, Chapinal and de Passillé, 2012; Porto et al., 2013; Aroeira et al., 2016; Chen, Zhu and Norton, 2021). On the other hand, it is crucial to acknowledge the potential risks associated with employing automated technologies for welfare assessment and enhancement. These technologies are susceptible to failure, often requiring external validation, and may not always capture all meaningful indicators of animal welfare (Dawkins, 2021a; Tuyttens, Molento and Benaissa, 2022). This limitation could result in a biased focus solely on the data that these technologies are capable of collecting.

Despite the rising availability of machine learning algorithms and increasing interest from the animal behaviour research community, there are still many challenges in using computer vision methodologies to extract valuable data. One major limitation is the restricted amount of animal data available for algorithm training compared to the vast amount of human data (Labuguen et al., 2021; Li et al., 2022). Additionally, NHPs, as many other species, may assume unusual postures, have similar features, or be occluded by objects or structures in their environment, making individual identification and behaviour detection challenging (Vidal et al., 2021). These are only some of the main limitations concerning the use of computer vision methodologies to study animal behaviour.

1.3 Animal Welfare: Current Measures and Strategies

The definition of animal welfare has been the subject of extensive research and assessing it has proven to be a very challenging task. Although various definitions and approaches have been proposed over the past decades, there are still only general guidelines and a broad definition of animal welfare. Animal welfare emerged as a young science in 1965 with the Brambell report and has since evolved, particularly with the concept of animals as sentient individuals, which emphasized the need to identify their needs and experiences (Brambell, 1965; Millman *et al.*, 2004). This led to the development of the Five Freedoms by the FAWC in 1993, which include freedom from thirst, hunger, and malnutrition; freedom from discomfort; freedom from pain, injury, and disease; freedom to express normal behaviour; and freedom from fear and distress (Millman *et al.*, 2004; Duncan, 2006; Carenzi and Verga, 2009).

Contrary to these definitions, Dawkins has advocated for defining welfare without relying on consciousness or related concepts. This perspective favours a viewpoint where an animal with good welfare is deemed healthy and satisfied with its needs fulfilled (Dawkins, 2021b). Another recent approach involves the 5 domains model proposed by Mellor (Mellor, 2017; Mellor *et al.*, 2020). Each of these domains (nutrition, environment, health, behaviour and mental state) serves to highlight specific areas pertinent to both the assessment and management of animal welfare with a central focus on prioritizing animal feelings (Mellor, 2017). Continuing the evolution of animal welfare definitions, the concept of quality of life has emerged as a significant milestone, aiming to foster a positive welfare state. The UK Farm Animal Welfare Council introduced the idea of "a life not worth living," which pertains to comprehensive lists of potential positive impacts and their generation (Yeates, 2011; Mellor, 2016).

In laboratories and breeding colonies that house NHPs, an animal's welfare is deemed satisfactory if it meets the following criteria: (1) it is in good physical health, (2) it can participate in a diverse range of advantageous species-typical activities and demonstrates minimal abnormal behaviour, (3) it is not consistently under distress, and (4) it can competently adjust to daily modifications in its social and physical environment by fulfilling its own requirements (Novak and Suomi, 1988). However, it is crucial to recognize that for non-human primates (NHPs) to effectively serve their designated roles in medical research, a proactive, concerted effort is necessary. This effort should focus on implementing refinement strategies to minimize harm and enhance welfare throughout their lives, ensuring their well-being is consistently upheld (Buchanan-Smith *et al.*, 2023).

The Directive 2010/63/EU mandates that macaques used in experimental procedures must be born from animals within licensed, self-sustaining colonies to minimize the impact on wild populations. This thesis will focus on the welfare of NHPs within breeding colonies, which are facilities designed to house multiple groups of the same species (in this case rhesus macaques) with the aim of producing enough individuals so that some can be removed for use in scientific research (Ha and Sussman, 2023). This condition differs from the laboratory setting since the animals are housed in larger groups and are less likely to be involved in biomedical experimental procedures. However, it is important to note that breeding colonies also differ from zoo settings, where primates are typically able to live in more natural environments with access to outdoor space and stable social groups (Coleman *et al.*, 2023). As a result, these facilities have specific characteristics and welfare needs and must find a balance between them and achieving the aim of breeding and maintaining the animal population.

Ensuring appropriate housing for NHPs in captivity is a crucial aspect of their welfare (Ha and Sussman, 2023). Enclosures should be sufficiently spacious to accommodate the animals comfortably, yet still be easy to clean and provide a clear view of each individual. Additionally, the enclosures should have structures that allow the animals to engage in natural behaviours such as resting, climbing, exploring, and foraging. Therefore, the presence of structural enrichments, visual barriers to hide and avoid conspecifics, substrates that promote foraging and prosocial behaviour, and a lighting system that mimics natural lighting conditions as closely as possible are all essential requirements (Bayne *et al.*, 1992; Blois-Heulin and Jubin, 2004; Crast, Bloomsmith and Jonesteller, 2015; Ha and Sussman, 2023).

Once the enclosures are appropriate, the next challenge is the formation of social groups. Choosing individuals to form a breeding group needs to take into account many factors, such as genetic proximity, available space, and social compatibility (Albert, Jeong and Barabási, 2000; Kanthaswamy *et al.*, 2006; Coleman and Novak, 2017; McCowan, Beisner and Hannibal, 2018; Beisner *et al.*, 2023). Forming groups is a continuous challenge in a breeding colony, as it is necessary to avoid inbreeding and maintain genetic diversity. To achieve this, it is crucial to dismantle and create new social groups, and to remove and introduce individuals from already stable ones. However, these practices can destabilise the group's social structure and lead to aggression, poor welfare, and injuries. Once a stable group is formed, it needs to be maintained, assuring the welfare of the individuals in it. This can be another challenge in a breeding colony where the animals have no access to outdoor spaces and live in a controlled environment. Non-human primates are known for their intelligence, and this means that living in an environment lacking complexity can lead to boredom and apathy (Buchanan-Smith, 2011a). Therefore, it is essential to provide environmental enrichment to keep the primates mentally and physically stimulated and enhance their overall welfare. These are only some of the main issues that can affect the welfare of macaques in a breeding colony, and as such, great care and expertise are required to manage the group dynamics of a breeding colony and ensure the welfare of the NHPs housed within it.

Despite the numerous approaches and definitions proposed, evaluating the welfare of primates remains a complex task. However, ensuring good welfare is a fundamental objective for all institutions that house NHPs, and evaluating whether this objective has been achieved is a crucial aspect of animal management. One of the most commonly used techniques to assess primate welfare is measuring their behaviour. This involves identifying and limiting abnormal behaviours while promoting positive emotional states. Abnormal behaviours are classified as those behaviours that deviate from the norm for that species, both qualitatively and quantitatively (Lutz and Baker, 2023). These behaviours can vary in type and frequency, some can have little impact on the animal welfare, while others can negatively affect the animal welfare (Lutz and Baker, 2023). Certain abnormal behaviours exhibited by animals may indicate that their captive environment is lacking the necessary characteristics to fulfil their behavioural needs (Rushen and Mason, 2006). For this reason, it is essential to provide a stimulating environment that promotes the manifestation of a large number of species-specific behaviours. This can be achieved through providing adequate and stimulating housing conditions, as well as ensuring the opportunity for socialization with conspecifics, delivering predictable husbandry procedures, and limiting location changes (Waitt and Buchanan-Smith, 2001b; Bellanca and Crockett, 2002; Lutz *et al.*, 2007a; Lutz and Baker, 2023). The most commonly employed strategies to not only reduce abnormal behaviours but also promote species-typical ones are structural enrichments. Nonetheless, the effectiveness of this approach is not always certain and is influenced by factors such as the animal's history, welfare status, and the type of enrichment provided (Lutz and Baker, 2023). In addition, what worked for a while will not necessary work forever. When an animal becomes habituated to an enrichment, it may lose its purpose, and its presence may no longer be beneficial. To prevent this, it is crucial to rotate and change the enrichments regularly and to be able to identify when they are no longer engaging for the animals.

Nonhuman primates in their natural habitat exhibit an innate curiosity towards their surroundings, leading them to investigate and handle a diverse array of objects. Often, such exploratory behaviours are motivated by their search for sustenance. In the wild, free-ranging primates devote a substantial portion of their time to foraging and hunting, whereas those in captivity are usually provided with only one or two meals a day, leading to a decline in their foraging behaviour (Lutz and Novak, 2005). Consequently, providing foraging opportunities throughout food-based enrichment, different foraging mixes, and spreading food within the enclosure can be beneficial. Increasing the time spent by animals on foraging has been demonstrated to reduce abnormal behaviour, prevent the development of stereotypic behaviours, and alleviate boredom (Lutz and Novak, 2005; Kemp, 2023).

While the environment plays a crucial role in primate behaviour, individual temperament and personality also have an impact. Robinson and Weiss emphasized the need to investigate the connection between personality and welfare in NHPs since primate personality traits can influence the psychological impact of enrichment, such as fear and boredom. (Robinson and Weiss, 2023). It is also essential to consider neophobia when introducing new stimuli, whether it be food, objects, or situations, to prevent incidents and tensions within the group (Hoy, Murray and Tribe, 2009; Lutz and Baker, 2023). Moreover, individual personality has been shown to be linked with their welfare, and it is used to assess their emotional and physiological state, as well as their ability to cope with new situations and to form affiliative relationships (Weinstein and Capitanio, 2008; Weiss *et al.*, 2011; Gottlieb, Capitanio and McCowan, 2013a).

In conclusion, the welfare of non-human primates in captivity is a multifactorial issue that requires thoughtful consideration of multiple factors. To optimize NHPs welfare, it is imperative to provide opportunities for socialization, environmental enrichment, physical activity, and cognitive stimulation, while also systematically measuring their effects on the animals.

1.4 Rhesus Macaque Welfare

This next section reviews the welfare of rhesus macaques in breeding colonies. Macaques welfare comprises numerous factors, such as physical health, access to species-typical activities, minimal abnormal behaviour, and the capacity to adapt to daily changes in social and physical environment (Hannibal *et al.*, 2017; Lutz and Baker, 2023).

1.4.1 Macaque Behaviour in Captivity: Concerns and Needs

Rhesus macaques (*Macaca mulatta*) are social animals and live in groups of varying size, typically ranging from 10 to more than 200 individuals (Hasan *et al.*, 2013; K N Balasubramaniam *et al.*, 2014). Because of their extensive adaptability, this macaque species can thrive in diverse geographical regions characterized by varying climates and ecological conditions (Jaman and Huffman, 2013). Their hierarchical structure is typically regulated by dominant males and females, who use aggression and displacement to gain priority access to resources such as food, water, and mates (Lutz, Well and Novak, 2003; Thierry, Singh and Kaumanns, 2004; C. K. Lutz and Novak, 2005; Lutz *et al.*, 2007a; Beisner and Isbell, 2011a; Gottlieb, Coleman and McCowan, 2013). Macaques are diurnal and spend much of their time foraging for food, which includes a variety of fruits, seeds, and leaves, as well as insects, and small animals (Yeager, 1996; Hill, 1997; Hanya, 2004a). As in many other primate species, communication is an important aspect of rhesus macaque behaviour; they use a variety of vocalizations, facial expressions, and body postures to communicate with one another and regulate interactions within their group. These signals are used in several circumstances, in both affiliative and submissive communications, between same sex individuals and in male-female interactions, as well as in mother-infant dyads (Maestripieri and Wallen, 1997). Although rhesus macaques can live in a range of environments, including tropical forests and urban areas, adapting to life in captivity remains a challenge (Lutz and Novak, 2005).

Due to their highly intelligent and social nature, complex behaviour and cognitive needs, it is necessary to provide captive rhesus macaques with opportunities for environmental enrichment, socialization, and physical activity. These measures are essential for mitigating boredom and enhancing the overall welfare of this specie. Despite the numerous efforts, different housing conditions and group settings can lead to abnormal behaviours that can vary from mild (pacing, flipping, self-sucking, etc) to extreme forms (self-injury behaviours). These situations can cause an increase in aggression in a species which already controls its social system with more frequent and more severe aggression than most other macaque species (Lutz, Well and Novak, 2003; Thierry, Singh and Kaumanns, 2004; C. K. Lutz and Novak, 2005; Lutz *et al.*, 2007a; Beisner and Isbell, 2011a; Gottlieb, Coleman and McCowan, 2013). While some modifications in management have been made to decrease aggressive behaviours in captivity (Pruetz and Isbell, 2000; Beisner and Isbell, 2011a), there are still some factors that increase aggression in laboratory monkeys, such as, the impossibility to escape threatening situations, the introduction or removal of individuals from social groups, veterinary interventions and food competition (Beisner and Isbell, 2011a; Beisner *et al.*, 2015; McCowan, Beisner and Hannibal, 2018).

1.4.2 Macaques in Research

Non-human primates (NHPs), such as macaques, have been utilized in a variety of scientific fields, including neuroscience, infectious diseases, and behavioural research (Gardner and Luciw, 2008; Isoda, Atsushi and Ninomiya, 2018; Chiou *et al.*, 2020). Due to their similarities to humans, their availability and suitability for use in controlled laboratory environments, studies on NHPs have the potential to significantly contribute to scientific knowledge (Passingham, 2009; Caselli and Chelazzi, 2011; Uno, Uehara and Yamazaki, 2016; Hannibal *et al.*, 2017). Macaques, in particular, have played a vital role in biomedical research, leading to Nobel Prizes, new treatments for Alzheimer's and Parkinson's diseases, and ground-breaking insights into various aspects of neuroscience (Wiesel, 1982; Norrby and Prusiner, 2007; Capitanio and Emborg, 2008; Roelfsema and Treue, 2014; Wang, 2019). Additionally, these species are utilized in the development and testing of new treatments for human diseases, as well as vaccine development and drug testing. Notably, macaques have been instrumental in understanding and developing vaccines for SARS-CoV-2, as they exhibit COVID-19 symptoms, providing a valuable model for investigating the virus's pathophysiology and developing therapeutic and prophylactic interventions (Salguero *et al.*, 2021; Urano *et al.*, 2021).

Despite the fact that the use of non-human primates (NHPs) in biomedical research is comparatively small, comprising 0.08% of all animals used for scientific purposes in the UK, their use is still essential for scientific and medical progress (Mitchell *et al.*, 2021). However, the use of NHPs in research is controversial due to ethical concerns. For example, the utilization of chimpanzees and other apes in medical research has been widely discussed in the past century. In the United States, there has been a substantial decline in the use of these species over the past few decades of the 20th century, with many companies discontinuing their use since the early 2000s (Turner, 2023). Similarly, in Europe, a significant trend away from the employment of chimpanzees in research emerged during the 1990s and early 2000s, culminating in a complete ban on ape usage in research across the continent (Turner, 2023).

Efforts have been made to reduce and refine animal use while ensuring the validity and reliability of research results. NHPs are intelligent, long-lived, non-domesticated and social animals that can experience pain and suffering, and the use of animals in research is a contentious issue (Tardif *et al.*, 2013). Moreover, insufficient welfare can significantly influence research outcomes, compromising their validity and reliability, thereby exacerbating the reproducibility crisis in the behavioural sciences (Garner, 2005; Richter, Garner and Würbel, 2009). To address these concerns, regulations have been established to ensure appropriate treatment of NHPs in research (Tardif *et al.*, 2013). For example, the Animal Welfare Act in the United States sets minimum standards for the care and use of animals in research, and the Institutional Animal Care and Use Committee (IACUC) is responsible for reviewing and approving all research protocols involving animals (Cohen, 2006; Tardif *et al.*, 2013). In addition, many efforts have been made to reduce and refine animal use while ensuring the validity and reliability of research results. These efforts are guided by the principles of the "3Rs": Replacement, Reduction, and Refinement (Prescott, 2023). Replacement refers to the use of alternative methods that do not involve animals, reduction refers to minimizing the number of animals used in research while maintaining the scientific validity of the research, and refinement refers to minimizing the potential for pain, suffering, and distress in animals used in research (Guhad, 2005). The implementation of the 3Rs has led to the development of guidelines and regulations governing the care and use of animals in research, including NHPs.

In conclusion, while the use of macaques in biomedical research is crucial for scientific and medical progress, it is also a contentious issue due to ethical concerns regarding the treatment of these animals. To address these concerns, regulations, and guidelines have been established to ensure their appropriate treatment in research while minimizing their potential for pain, suffering, and distress. These efforts have led to the development of guidelines governing the care and use of animals in research, including macaques, and will continue to be essential in ensuring that research is conducted ethically while advancing scientific and medical progress.

1.5 Automated Methodologies Applied to Animal Welfare

In this thesis, I will focus on automated methodologies based on machine learning (ML). Machine learning algorithms are designed to learn from data and identify patterns that enable them to make predictions or classify new, unseen data. These algorithms can be grouped into different types:

1. Supervised learning: Algorithms learn from labelled data, where each example has a known outcome or target variable. The goal is to learn a mapping between the input data and the corresponding output.

2. Unsupervised learning: Algorithms analyse unlabelled data to discover patterns, structures, or relationships within the data. They do not have predefined target variables but aim to uncover inherent patterns or groupings.

3. Reinforcement learning: Algorithms learn through trial and error and interaction with an environment. They receive feedback in the form of rewards or penalties based on their actions and use this feedback to optimize their decision-making processes.

In this thesis, all models are based on supervised learning algorithms.

ML can tackle challenging tasks that are otherwise difficult to handle. By leveraging ML, it is possible to effectively address significant questions across various domains, social structure, collective behaviour, communication, and animal welfare (Valletta *et al.*, 2017). Deep learning is a specific subset of machine learning that utilizes deep neural networks to automatically learn complex patterns and representations directly from raw data. It eliminates the need for extensive manual feature engineering by allowing the network to extract hierarchical features from the data itself.

Machine learning and deep learning technologies have emerged as powerful tools for evaluating and measuring animal welfare in diverse species and settings, spanning from laboratory animals to farm animals.

Particularly, laboratory mice and rats have received significant attention in this context, as they are extensively used in biomedical research and exhibit distinct postures and movements that are comparatively easier to discern than those of primates. By leveraging machine learning models, it becomes possible to uncover underlying patterns in mouse behaviour without any subjective biases from observers. These approaches open up new avenues for assessing even subtle changes in behaviour, particularly in the context of pain assessment in laboratory rodents (Tuttle *et al.*, 2018; Jirkof, Rudeck and Lewejohann, 2019; Andresen *et al.*, 2020; Fried *et al.*, 2020).

In addition, automated behaviour detection has been extensively explored in farm animals to enhance animal welfare. Notably, there have been several significant advancements in the context of pigs, specifically in early detection of tail biting, which is a significant concern for their welfare (Matthews *et al.*, 2016; D'Eath *et al.*, 2018; Chen, Zhu and Norton, 2021). Similarly, attention has been dedicated to improving the welfare of cows, with a primary focus on locomotion and addressing the prevalent issue of lameness (Chen, Zhu and Norton, 2021; Shahinfar *et al.*, 2021; Tassinari *et al.*, 2021). Comparable approaches have also been employed in other farm species, such as sheep and chickens, to monitor and enhance their welfare (i.e. sheep and chicken (Fogarty *et al.*, 2020; de Alencar Nääs *et al.*, 2021)).

The automated techniques utilized in this thesis will predominantly rely on computer vision methodologies. Computer vision, a branch of artificial intelligence, is dedicated to empowering machines to extract valuable information from images, videos, and various visual inputs. This decision and its application are influenced by several factors, including the widespread availability of video recordings documenting macaque behaviour and the non-intrusive nature of the approach. By relying on computer vision, it becomes possible to collect behavioural data without the need for additional devices such as tags, accelerometers, or trackers. This approach offers a cost-effective and less intrusive means of studying and analysing macaque behaviour in a captive setting.

1.5.1 Computer vision techniques

Computer vision is based on two specific technologies: deep learning and convolutional neural network (CNN). The former is a type of machine learning that uses training data to categorize image content without a priori specification of image features (LeCun, Bengio and Hinton, 2015; Weinstein, 2018). The latter helps the previous learning process, breaking images down into pixels and assigning them tags or labels (Fujiyoshi, Hirakawa and Yamashita, 2019). The CNN takes this name from a mathematical linear operation between matrixes called convolution that is used to make predictions from the previously extracted information (Albawi, Mohammed and Al-Zawi, 2017; Fujiyoshi, Hirakawa and Yamashita, 2019).

A computer vision system typically involves several key steps (Figure 1), including:

1. Image acquisition: To begin, the system acquires image or video data using a camera or other sensors.
2. Pre-processing: The raw data is then pre-processed to enhance the quality of the image, which includes noise reduction, contrast adjustment, and resizing.
3. Feature extraction: In this stage, using a CNN algorithm, the system identifies key features of the image or video, such as edges, shapes, and textures.
4. Object detection: Once the features are identified, the system can detect objects in the image or video, and classify them into different categories, such as people, cars, or animals.
5. Recognition and classification: The system then matches the detected objects to a pre-defined database of known objects or categories and classifies them accordingly.
6. Interpretation: Finally, the system interprets the data, which could involve recognizing patterns, identifying anomalies, or making predictions.

Computer vision has been used to solve several tasks; we will be focusing on the ones involving object recognition and pose estimation.

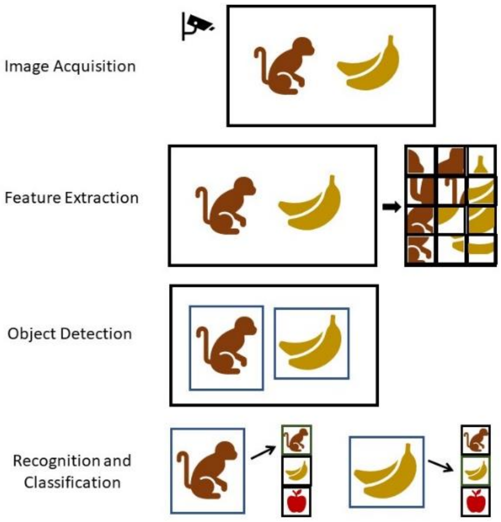


Figure 1: Example of key steps characterizing a computer vision system.

1.5.2 Object Recognition

Object recognition involves finding one or more objects in an image, recognizing them, and returning their position(s) within the image. This information can be obtained at the image-level, using an object detection algorithm that extract bounding boxes (rectangular regions delineating objects within images), or at the pixel-level, using a instance segmentation algorithm that produce instance masks (binary image where each pixel is assigned a value indicating whether it belongs to a specific object) (Matthews *et al.*, 2017; Wang *et al.*, 2020; Wu *et al.*, 2021) (Refer to Chapter 2 for more details).

Object recognition is a valuable tool in animal welfare research, where it can be used to detect and track animal behaviour and improve the quality of care provided to animals. For example, object recognition techniques can be used to monitor the health and welfare of livestock animals, such as cattle and pigs. By using deep learning methods to recognize behaviours such as feeding, drinking, aggressions, lameness and nursing, it is possible to detect changes in an animal's posture, gait, and activity levels, which can indicate potential health issues. (Chen, Zhu and Norton, 2021). This information can then be used to improve the welfare of the animals and reduce the risk of disease outbreaks. Object recognition can also be used to monitor the use of enrichment items in captive animals, such as toys or puzzles. By tracking the animals' interactions with these items, it is possible to gain insights into the animals' behaviour, preferences, and welfare (de Chaumont *et al.*, 2019; Chen *et al.*, 2020). Moreover, object recognition can be used in wildlife conservation efforts. For instance, camera traps and drones can gather information on the number and distribution of animals in their natural habitats (van Gemert *et al.*, 2015; Carl *et al.*, 2020; Chalmers *et al.*, 2021). This information can be used to track population trends, identify habitat hotspots, and develop conservation strategies to protect these animals and their ecosystems. In the case of macaques, object recognition has been employed to detect and track the animals in their natural habitat as well as monitor their behaviour during cognitive tasks, demonstrating its potential for furthering our understanding of their welfare (Chiverton, Micheletta and Waller, 2015; Bethell, Khan and Hussain, 2022; Pineda *et al.*, 2023). Overall, object recognition is a versatile tool that can be used in a variety of animal welfare applications, from monitoring the health of livestock animals to assess animal welfare in captivity.

For the purpose of this thesis, I used object recognition to detect both the macaques and various enrichment items within their pens. I have tested several object detection algorithms, including SIPEC (Marks *et al.*, 2020) and YOLO (Ma and Yang, 2022), but encountered implementation difficulties. However, better results were achieved using Argos (Ray and Stopfer, 2022), a toolkit that utilizes YOLACT (You Only Look At CoefficienTs) for CNN-based segmentation. Consequently, YOLACT (Bolya *et al.*, 2019) was chosen as the primary algorithm for tracking both macaques and the objects of interest. One of the advantages of YOLACT is its ease of installation compared to other algorithms, and it can deliver fast results using a single Graphics Processing Unit (GPU). This feature is particularly valuable in the field of animal behaviour research, as many animal facilities lack the resources to invest in multiple GPUs.

Towards the culmination of my PhD, a real-time object detection algorithm called YOLOv8 was introduced. This iteration of YOLO (You Only Look Once) represented a significant advancement over its predecessors, including YOLOv5, particularly in terms of ease of installation and usability (Jocher, G., Chaurasia, A., & Qiu, 2023; Terven and Cordova-Esparza, 2023). Built on the PyTorch framework and implemented in Python, this latest version of YOLO's single-shot detection approach maintains the core objective of optimizing the balance between speed and accuracy. It achieves this by keeping the model size small, making it convenient for users to work with on a single GPU, similar to YOLACT. Ultralytics, the developers of YOLOv8, released the algorithm in January 2023, offering multiple modes that support training, prediction, and validation for various tasks, such as detection, segmentation, and pose estimation. This versatility enhances the algorithm's utility across different applications.

1.5.3 Pose Estimation

Pose estimation is the process of determining the position and orientation of an object or objects within an image or video. It is largely used to track individuals and detect their geometrical configuration of multiple body parts (Andriluka *et al.*, 2018; Mathis *et al.*, 2018; Labuguen *et al.*, 2019). There are several approaches to pose estimation in computer vision, including deep learning-based methods. These approaches utilize advanced neural networks, such as convolutional neural networks (CNNs), to learn features and patterns, and then use this knowledge to estimate the poses of objects. Deep learning-based methods have several key advantages over traditional pose estimation techniques. For example, they can handle more complex and varied data inputs, learn from large datasets, and adapt to different scenarios without the need for extensive manual tuning.

Pose estimation has been increasingly used in animal welfare research to monitor the behaviour and welfare of animals. By accurately estimating the pose of animals, it is possible to gain valuable insights into their movements, posture, and activity levels, which can be indicative of their welfare state. Animal pose estimation and tracking has been applied to several species both in the wild and in captive settings with the aim of collecting behaviour and promoting welfare (de Chaumont *et al.*, 2019; Marks *et al.*, 2020, 2022; Blanco Negrete *et al.*, 2021; Joska *et al.*, 2021; Arnkærn *et al.*, 2022; Nasiri *et al.*, 2022; Yang *et al.*, 2022).

Over the past few years, numerous models have been developed to detect and recognize the poses and behaviours of non-human primates (NHPs). Some of these models use 2D cameras to estimate pose in NHPs that move freely within indoor enclosures (Bala *et al.*, 2020; Marks *et al.*, 2022), while others focus on outdoor NHPs (Desai *et al.*, 2022; Yao *et al.*, 2023) or primates in their natural habitats (Labuguen *et al.*, 2019, 2021). All of these models aim to address similar challenges: identifying NHPs' movements and behaviours in diverse environments, despite occlusions, unusual postures, and varying lighting conditions. Compared to other animals, primates are more challenging to track due to their highly flexible body joints, which allow them to assume complex positions and move in three-dimensional space. These characteristics present two major issues: substantial obstruction and extensive variability in appearance (Hayden, Park and Zimmermann, 2022). These problems are further intensified in a breeding colony environment due to the large population confined to a limited space, the presence of multiple structures and enrichment elements within the enclosure, and the constraints imposed by camera angles and positions. Consequently, the development of a dependable algorithm for accurately estimating NHPs poses remains an elusive goal.

For the purpose of this thesis, I have selected DeepLabCut (DLC) (Mathis *et al.*, 2018) as the pose estimation algorithm for detecting macaque body parts. One of the advantages of DLC is that it does not require any markers or sensors to be attached to the animals, which is impractical with macaques. Furthermore, DLC yields reliable results using only a single camera, which is crucial as the videos used in these studies were previously recorded for other purposes. DLC offers a user-friendly GUI interface that encompasses all the necessary functions. It has been utilized to train the MacaquePose open dataset (Labuguen *et al.*, 2021) and to develop a model capable of detecting face landmarks on macaques across a wide range of ages and genders. The model was trained using images of individuals from the breeding colony at the Centre for Macaques.

Initially, I trained DLC to identify and monitor individuals within various macaque groups. However, I encountered challenges, including overlapping animals and suboptimal video quality, which impeded my ability to achieve the desired outcomes. Consequently, I shifted my approach to employ DLC for the purpose of detecting specific body parts of individual macaques during temperament tests. This modified strategy enabled me to concentrate on a single macaque at a time, thereby mitigating the aforementioned issues.

1.6 Conclusion

Understanding the behaviour of NHPs is important in fields such as biology, primatology, and biomedicine. Despite this, accurately quantifying primate behaviour has been a longstanding challenge because of its complexity, cost, and low data availability. However, recent technological advancements have sparked a behavioural measurement revolution that offers affordable and scalable rigor (Hayden, Park and Zimmermann, 2022). These novel techniques have the ability to analyse vast quantities of data, enabling the detection of subtle and infrequent behaviours, as well as changes in behaviour over time. Additionally, these techniques can actively enhance welfare by focusing on evaluating animal outputs such as their physiological, health, or behavioural responses to these environmental inputs (Hewson, 2003; Truelove *et al.*, 2020; Knaebe *et al.*, 2022). Furthermore, behavioural tracking and imaging techniques in neuroscience experiments can impact the 3Rs by providing vast amounts of data for post hoc analyses, richer behaviour, and fewer individuals needed, ultimately improving their housing and husbandry conditions (Bethell, Khan and Hussain, 2022; Knaebe *et al.*, 2022).

In conclusion, my PhD project utilized these technologies to enhance the welfare of captive macaques, with a specific emphasis on enrichments, foraging patterns, and temperament assessment. These areas of investigation provide insights that can guide interventions aimed at enabling animals to engage in their natural behaviours. By leveraging machine learning based technologies, we can better understand and address the unique needs and preferences of captive primates, ultimately improving their quality of life and promoting their overall welfare.

1.7 Aims and Objectives

The primary aim of this research is to develop automated methodologies to collect and analyse rhesus macaques' behaviours that can be used to improve their welfare in a breeding colony setting. Automating the analysis of primate behaviour in videos poses a significant challenge that necessitates breaking it down into manageable objectives. For my thesis, I have specifically focused on the following goals:

1. Assess temperament in individual macaques by automating the video analyses of standardised temperament tests.
2. Evaluate enrichment usage in different groups of macaques by automating the video analyses of CCTV systems.
3. Analyse foraging behaviour in different groups of macaques by automating the video analyses of CCTV systems.

1.8 Chapter References

Albawi, S., Mohammed, T.A. and Al-Zawi, S. (2017) ‘Understanding of a convolutional neural network’, in *2017 International Conference on Engineering and Technology (ICET)*. IEEE, pp. 1–6. doi:10.1109/ICEngTechnol.2017.8308186.

Albert, R., Jeong, H. and Barabási, A.-L. (2000) ‘Error and attack tolerance of complex networks’, *Nature*, 406(6794), pp. 378–382. doi:10.1038/35019019.

de Alencar Nääs, I. *et al.* (2021) ‘Lameness prediction in broiler chicken using a machine learning technique’, *Information Processing in Agriculture*, 8(3), pp. 409–418. doi:10.1016/j.inpa.2020.10.003.

Andresen, N. *et al.* (2020) ‘Towards a fully automated surveillance of well-being status in laboratory mice using deep learning: Starting with facial expression analysis’, *PLOS ONE*. Edited by C.-Q. Gao, 15(4), p. e0228059. doi:10.1371/journal.pone.0228059.

Andriluka, M. *et al.* (2018) ‘Posetrack: A benchmark for human pose estimation and tracking’, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5167–5176.

Arnkærn, B. *et al.* (2022) ‘Deep Learning Based Multiple Animal Pose Estimation’, *IS and T International Symposium on Electronic Imaging Science and Technology*, 34(6), pp. 1–6. doi:10.2352/EI.2022.34.6.IRIACV-276.

Aroeira, R.M.C. *et al.* (2016) ‘Non-invasive methods of computer vision in the posture evaluation of adolescent idiopathic scoliosis’, *Journal of Bodywork and Movement Therapies*, 20(4), pp. 832–843. doi:10.1016/j.jbmt.2016.02.004.

Bala, P.C. *et al.* (2020) ‘Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio’, *Nature Communications*, 11(1), p. 4560. doi:10.1038/s41467-020-18441-5.

Balasubramaniam, K.N. *et al.* (2014) ‘Group size, contest competition, and social structure in Cayo Santiago rhesus macaques’, *Behaviour*, 151(12–13), pp. 1759–1798.

doi:10.1163/1568539X-00003216.

Barnard, S. *et al.* (2016) 'Quick, Accurate, Smart: 3D Computer Vision Technology Helps Assessing Confined Animals' Behaviour', *PLOS ONE*. Edited by C. Wade, 11(7), p. e0158748. doi:10.1371/journal.pone.0158748.

Bateson, M. and Martin, P. (2021) *Measuring behaviour: an introductory guide*. Cambridge university press.

Bayne, K. *et al.* (1992) 'The Use of Artificial Turf as a Foraging Substrate for Individually Housed Rhesus Monkeys (Macaca Mulatta)', *Animal Welfare*, 1(1), pp. 39–53. doi:10.1017/S0962728600014706.

Beisner, B.A. *et al.* (2015) 'Detection of social group instability among captive rhesus macaques using joint network modeling', *Current Zoology*, 61(1), pp. 70–84. doi:10.1093/czoolo/61.1.70.

Beisner, B.A. *et al.* (2023) 'Sociality, Health, and Welfare in Nonhuman Primates', in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 413–443. doi:10.1007/978-3-030-82708-3_18.

Beisner, B.A. and Isbell, L.A. (2011) 'Factors affecting aggression among females in captive groups of rhesus macaques (Macaca mulatta)', *American Journal of Primatology*, 73(11), pp. 1152–1159. doi:10.1002/ajp.20982.

Bellanca, R.U. and Crockett, C.M. (2002) 'Factors predicting increased incidence of abnormal behavior in male pigtailed macaques', *American Journal of Primatology*, 58(2), pp. 57–69. doi:10.1002/ajp.10052.

Bethell, E.J., Khan, W. and Hussain, A. (2022) 'A deep transfer learning model for head pose estimation in rhesus macaques during cognitive tasks: Towards a nonrestraint noninvasive 3Rs approach', *Applied Animal Behaviour Science*, 255, p. 105708. doi:10.1016/j.applanim.2022.105708.

Blanco Negrete, S. *et al.* (2021) 'Multiple Monkey Pose Estimation Using OpenPose', *bioRxiv*, p. 2021.01.28.428726. doi:10.1101/2021.01.28.428726.

Blois-Heulin, C. and Jubin, R. (2004) 'Influence of the presence of seeds and litter on the behaviour of captive red-capped mangabeys Cercocebus torquatus torquatus', *Applied Animal Behaviour Science*, 85(3–4), pp. 349–362. doi:10.1016/j.applanim.2003.10.005.

Blumrosen, G., Hawellek, D. and Pesaran, B. (2017) 'Towards automated recognition of facial expressions in animal models', *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2018-Janua, pp. 2810–2819. doi:10.1109/ICCVW.2017.332.

Bolya, D. *et al.* (2019) 'Yolact: Real-time instance segmentation', in *Proceedings of the*

IEEE/CVF international conference on computer vision, pp. 9157–9166.

Bolya, D. *et al.* (no date) *YOLACT Real-time Instance Segmentation*. Available at: https://github.com/dbolya/yolact.

Brambell, F.W. (1965) ‘Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems’, *(No Title)* [Preprint].

Buchanan-Smith, H.M. (2011) ‘Environmental enrichment for primates in laboratories’, *Advances in Science and Research*, 5(1), pp. 41–56. doi:10.5194/asr-5-41-2010.

Buchanan-Smith, H.M. *et al.* (2023) ‘Welfare of primates in laboratories: opportunities for refinement’, in *Nonhuman Primate Welfare: From History, Science, and Ethics to Practice*. Springer, pp. 97–120.

Capitanio, J.P. and Emborg, M.E. (2008) ‘Contributions of non-human primates to neuroscience research’, *The Lancet*, 371(9618), pp. 1126–1135. doi:10.1016/S0140-6736(08)60489-4.

Carenzi, C. and Verga, M. (2009) ‘Animal welfare: review of the scientific concept and definition’, *Italian Journal of Animal Science*, 8(sup1), pp. 21–30. doi:10.4081/ijas.2009.s1.21.

Carl, C. *et al.* (2020) ‘Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model’, *European Journal of Wildlife Research*, 66(4), p. 62. doi:10.1007/s10344-020-01404-y.

Caselli, L. and Chelazzi, L. (2011) ‘Does the Macaque Monkey Provide a Good Model for Studying Human Executive Control? A Comparative Behavioral Study of Task Switching’, *PLoS ONE*. Edited by G. Chapouthier, 6(6), p. e21489. doi:10.1371/journal.pone.0021489.

Chalmers, C. *et al.* (2021) ‘Video analysis for the detection of animals using convolutional neural networks and consumer-grade drones’, *Journal of Unmanned Vehicle Systems*, 9(2), pp. 112–127. doi:10.1139/juvs-2020-0018.

de Chaumont, F. *et al.* (2019) ‘Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning’, *Nature Biomedical Engineering*, 3(11), pp. 930–942. doi:10.1038/s41551-019-0396-1.

Chen, C. *et al.* (2020) ‘A computer vision approach for recognition of the engagement of pigs with different enrichment objects’, *Computers and Electronics in Agriculture*, 175, p. 105580. doi:10.1016/j.compag.2020.105580.

Chen, C., Zhu, W. and Norton, T. (2021) ‘Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning’, *Computers and Electronics in Agriculture*, 187, p. 106255. doi:10.1016/j.compag.2021.106255.

Chiou, K.L. *et al.* (2020) ‘Rhesus macaques as a tractable physiological model of human ageing’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1811), p. 20190612. doi:10.1098/rstb.2019.0612.

Chiverton, J., Micheletta, J. and Waller, B. (2015) ‘Detecting and Tracking Bottoms and Faces of the Crested Black Macaque in the Wild’, in *Procedings of the Machine Vision of Animals and their Behaviour Workshop 2015*. British Machine Vision Association, pp. 9.1-9.8. doi:10.5244/C.29.MVAB.9.

Cohen, H. (2006) ‘The animal welfare act’, *J. animal l.*, 2, p. 13.

Coleman, K. *et al.* (2023) ‘Common Husbandry, Housing, and Animal Care Practices’, in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 323–354. doi:10.1007/978-3-030-82708-3_14.

Coleman, K. and Novak, M.A. (2017) ‘Environmental Enrichment in the 21st Century’, *ILAR Journal*, 58(2), pp. 295–307. doi:10.1093/ilar/ilx008.

Crast, J., Bloomsmith, M.A. and Jonesteller, T. (2015) ‘Effects of changing housing conditions on mangabey behavior (Cercocebus atys): Spatial density, housing quality, and novelty effects’, *American Journal of Primatology*, 77(9), pp. 1001–1014. doi:10.1002/ajp.22430.

D’Eath, R.B. *et al.* (2018) ‘Automatic early warning of tail biting in pigs: 3D cameras can detect lowered tail posture before an outbreak’, *PLOS ONE*. Edited by A. Yildirim, 13(4), p. e0194524. doi:10.1371/journal.pone.0194524.

Dawkins, M.S. (2021a) ‘Does Smart Farming Improve or Damage Animal Welfare? Technology and What Animals Want’, *Frontiers in Animal Science*, 2. doi:10.3389/fanim.2021.736536.

Dawkins, M.S. (2021b) *The science of animal welfare: Understanding what animals want*. Oxford University Press, USA.

Dawkins, M.S., Cain, R. and Roberts, S.J. (2012) ‘Optical flow, flock behaviour and chicken welfare’, *Animal Behaviour*, 84(1), pp. 219–223. doi:10.1016/j.anbehav.2012.04.036.

Desai, N. *et al.* (2022) ‘OpenApePose: a database of annotated ape photographs for pose estimation’, *arXiv preprint arXiv:2212.00741* [Preprint].

Duncan, I.J.H. (2006) ‘The changing concept of animal sentience’, *Applied Animal Behaviour Science*, 100(1–2), pp. 11–19. doi:10.1016/j.applanim.2006.04.011.

Fogarty, E.S. *et al.* (2020) ‘Behaviour classification of extensively grazed sheep using machine learning’, *Computers and Electronics in Agriculture*, 169, p. 105175. doi:10.1016/j.compag.2019.105175.

Fried, N.T. *et al.* (2020) ‘Improving pain assessment in mice and rats with advanced videography and computational approaches’, *Pain*, 161(7), pp. 1420–1424. doi:10.1097/j.pain.0000000000001843.

Fujiyoshi, H., Hirakawa, T. and Yamashita, T. (2019) ‘Deep learning-based image recognition for autonomous driving’, *IATSS Research*, 43(4), pp. 244–252. doi:10.1016/j.iatssr.2019.11.008.

Gardner, M.B. and Luciw, P.A. (2008) ‘Macaque Models of Human Infectious Disease’, *ILAR Journal*, 49(2), pp. 220–255. doi:10.1093/ilar.49.2.220.

Garner, J.P. (2005) ‘Stereotypies and Other Abnormal Repetitive Behaviors: Potential Impact on Validity, Reliability, and Replicability of Scientific Outcomes’, *ILAR Journal*, 46(2), pp. 106–117. doi:10.1093/ilar.46.2.106.

van Gemert, J.C. *et al.* (2015) ‘Nature Conservation Drones for Automatic Localization and Counting of Animals’, in, pp. 255–270. doi:10.1007/978-3-319-16178-5_17.

Gottlieb, D.H., Capitanio, J.P. and McCowan, B. (2013) ‘Risk factors for stereotypic behavior and self-biting in rhesus macaques (Macaca mulatta): Animal’s history, current environment, and personality’, *American Journal of Primatology*, 75(10), pp. 995–1008. doi:10.1002/ajp.22161.

Gottlieb, D.H., Coleman, K. and McCowan, B. (2013) ‘The effects of predictability in daily husbandry routines on captive rhesus macaques (Macaca mulatta)’, *Applied Animal Behaviour Science*, 143(2–4), pp. 117–127. doi:10.1016/j.applanim.2012.10.010.

Green, J. (2018) ‘Investigating the relationship between nocturnal activity and psychological wellbeing in captive Rhesus macaques (Macaca mulatta)’.

Gronskyte, R. *et al.* (2016) ‘Monitoring pig movement at the slaughterhouse using optical flow and modified angular histograms’, *Biosystems Engineering*, 141, pp. 19–30. doi:10.1016/j.biosystemseng.2015.10.002.

Guhad, F. (2005) ‘Introduction to the 3Rs (refinement, reduction and replacement)’, *Journal of the American Association for Laboratory Animal Science*, 44(2), pp. 58–59.

Ha, J.C. and Sussman, A.F. (2023) ‘Primate Breeding Colonies: Colony Management and Welfare’, in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 307–321. doi:10.1007/978-3-030-82708-3_13.

Hannibal, D.L. *et al.* (2017) ‘Laboratory rhesus macaque social housing and social changes: Implications for research’, *American Journal of Primatology*, 79(1), p. e22528. doi:10.1002/ajp.22528.

Hanya, G. (2004) ‘Diet of a Japanese Macaque Troop in the Coniferous Forest of Yakushima’, *International Journal of Primatology*, 25(1), pp. 55–71. doi:10.1023/B:IJOP.0000014645.78610.32.

Hasan, M.K. *et al.* (2013) ‘Distribution of Rhesus Macaques (Macaca mulatta) in Bangladesh: Inter-population Variation in Group Size and Composition’, *Primate Conservation*, 26(1), pp. 125–132. doi:10.1896/052.026.0103.

Hayden, B.Y., Park, H.S. and Zimmermann, J. (2022) ‘Automated pose estimation in primates’, *American Journal of Primatology*, 84(10). doi:10.1002/ajp.23348.

Hewson, C.J. (2003) ‘What is animal welfare? Common definitions and their practical consequences.’, *The Canadian veterinary journal = La revue veterinaire canadienne*, 44(6), pp. 496–499.

Hill, D.A. (1997) ‘Seasonal variation in the feeding behavior and diet of Japanese macaques (Macaca fuscata yakui) in lowland forest of Yakushima’, *American Journal of Primatology*, 43(4), pp. 305–320. doi:10.1002/(SICI)1098-2345(1997)43:4<305::AID-AJP2>3.0.CO;2-0.

Hoy, J.M., Murray, P.J. and Tribe, A. (2009) ‘Thirty years later: enrichment practices for captive mammals’, *Zoo Biology*, 29(3), pp. 303–316. doi:10.1002/zoo.20254.

Isoda, M., Atsushi, N. and Ninomiya, T. (2018) ‘Development of social systems neuroscience using macaques’, *Proceedings of the Japan Academy, Series B*, 94(7), pp. 305–323. doi:10.2183/pjab.94.020.

Jaman, M.F. and Huffman, M.A. (2013) ‘The effect of urban and rural habitats and resource type on activity budgets of commensal rhesus macaques (Macaca mulatta) in Bangladesh’, *Primates*, 54(1), pp. 49–59. doi:10.1007/s10329-012-0330-6.

Jirkof, P., Rudeck, J. and Lewejohann, L. (2019) ‘Assessing Affective State in Laboratory Rodents to Promote Animal Welfare—What Is the Progress in Applied Refinement Research?’, *Animals*, 9(12), p. 1026. doi:10.3390/ani9121026.

Jocher, G., Chaurasia, A., & Qiu, J. (2023) ‘YOLO by Ultralytics (Version 8.0.0)’.

Joska, D. *et al.* (2021) ‘AcinoSet: a 3D pose estimation dataset and baseline models for Cheetahs in the wild’, in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 13901–13908.

Kanthaswamy, S. *et al.* (2006) ‘Microsatellite markers for standardized genetic management of captive colonies of rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 68(1), pp. 73–95. doi:10.1002/ajp.20207.

Kemp, C. (2023) ‘Enrichment’, in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 463–500. doi:10.1007/978-3-030-82708-3_20.

Kennedy, A. (2022) ‘The what, how, and why of naturalistic behavior’, *Current Opinion in Neurobiology*, 74, p. 102549. doi:10.1016/j.conb.2022.102549.

Knaebe, B. *et al.* (2022) ‘The Promise of Behavioral Tracking Systems for Advancing Primate

Animal Welfare', *Animals*, 12(13), p. 1648. doi:10.3390/ani12131648.

Labuguen, R. *et al.* (2019) 'Primate Markerless Pose Estimation and Movement Analysis Using DeepLabCut', in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pp. 297–300. doi:10.1109/ICIEV.2019.8858533.

Labuguen, R. *et al.* (2021) 'MacaquePose: A Novel "In the Wild" Macaque Monkey Pose Dataset for Markerless Motion Capture', *Frontiers in Behavioral Neuroscience*, 14. doi:10.3389/fnbeh.2020.581154.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436–444. doi:10.1038/nature14539.

Li, J. *et al.* (2022) 'Barriers to computer vision applications in pig production facilities', *Computers and Electronics in Agriculture*, 200, p. 107227. doi:10.1016/j.compag.2022.107227.

Line, S.W. *et al.* (1990) 'Responses of female rhesus macaques to an environmental enrichment apparatus', *Laboratory Animals*, 24(3), pp. 213–220. doi:10.1258/002367790780866245.

Lutz, C., Well, A. and Novak, M. (2003) 'Stereotypic and self-injurious behavior in rhesus macaques: A survey and retrospective analysis of environment and early experience', *American Journal of Primatology*, 60(1), pp. 1–15. doi:10.1002/ajp.10075.

Lutz, C.K. *et al.* (2007) 'Early predictors of self-biting in socially-housed rhesus macaques (Macaca mulatta)', *American Journal of Primatology*, 69(5), pp. 584–590. doi:10.1002/ajp.20370.

Lutz, C.K. and Baker, K.C. (2023) 'Using Behavior to Assess Primate Welfare', in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 171–205. doi:10.1007/978-3-030-82708-3_8.

Lutz, C. K. and Novak, M.A. (2005) 'Environmental Enrichment for Nonhuman Primates: Theory and Application', *ILAR Journal*, 46(2), pp. 178–191. doi:10.1093/ilar.46.2.178.

Lutz, Corrine K. and Novak, M.A. (2005) *Primate Natural History and Social Behavior. Implications for Laboratory Housing*, *The Laboratory Primate*. Elsevier Limited. doi:10.1016/B978-012080261-6/50009-X.

Ma, D. and Yang, J. (2022) 'YOLO-Animal: An efficient wildlife detection network based on improved YOLOv5', in *2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*. IEEE, pp. 464–468. doi:10.1109/ICICML57342.2022.10009855.

Maestripieri, D. and Wallen, K. (1997) 'Affiliative and submissive communication in rhesus macaques', *Primates*, 38(2), pp. 127–138. doi:10.1007/BF02382003.

Marks, M. *et al.* (2020) 'SIPEC: the deep-learning Swiss knife for behavioral data analysis', *bioRxiv*, p. 2020.10.26.355115. doi:10.1101/2020.10.26.355115.

Marks, M. *et al.* (2022) 'Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments', *Nature Machine Intelligence*, 4(4), pp. 331–340. doi:10.1038/s42256-022-00477-5.

Mathis, A. *et al.* (2018) 'DeepLabCut: markerless pose estimation of user-defined body parts with deep learning', *Nature Neuroscience*, 21(9), pp. 1281–1289. doi:10.1038/s41593-018-0209-y.

Matthews, S.G. *et al.* (2016) 'Early detection of health and welfare compromises through automated detection of behavioural changes in pigs', *The Veterinary Journal*, 217, pp. 43–51. doi:10.1016/j.tvjl.2016.09.005.

Matthews, S.G. *et al.* (2017) 'Automated tracking to measure behavioural changes in pigs for health and welfare monitoring', *Scientific Reports*, 7(1), p. 17582. doi:10.1038/s41598-017-17451-6.

McCowan, B., Beisner, B. and Hannibal, D. (2018) 'Social management of laboratory rhesus macaques housed in large groups using a network approach: A review', *Behavioural Processes*, 156, pp. 77–82. doi:10.1016/j.beproc.2017.11.014.

Mellor, D. (2016) 'Updating Animal Welfare Thinking: Moving beyond the "Five Freedoms" towards "A Life Worth Living"', *Animals*, 6(3), p. 21. doi:10.3390/ani6030021.

Mellor, D.J. (2017) 'Operational details of the five domains model and its key applications to the assessment and management of animal welfare', *Animals*, 7(8), p. 60.

Mellor, D.J. *et al.* (2020) 'The 2020 Five Domains Model: Including Human–Animal Interactions in Assessments of Animal Welfare', *Animals*, 10(10), p. 1870. doi:10.3390/ani10101870.

Millman, S.T. *et al.* (2004) 'The impact of applied ethologists and the International Society for Applied Ethology in improving animal welfare', *Applied Animal Behaviour Science*, 86(3–4), pp. 299–311. doi:10.1016/j.applanim.2004.02.008.

Mitchell, A.S. *et al.* (2021) 'International primate neuroscience research regulation, public engagement and transparency opportunities', *NeuroImage*, 229, p. 117700. doi:10.1016/j.neuroimage.2020.117700.

Nakamura, T. *et al.* (2016) 'A Markerless 3D Computerized Motion Capture System Incorporating a Skeleton Model for Monkeys', *PLOS ONE*. Edited by M.H.E. de Lussanet, 11(11), p. e0166154. doi:10.1371/journal.pone.0166154.

Nasiri, A. *et al.* (2022) ‘Pose estimation-based lameness recognition in broiler using CNN-LSTM network’, *Computers and Electronics in Agriculture*, 197, p. 106931. doi:10.1016/j.compag.2022.106931.

Norrby, E. and Prusiner, S.B. (2007) ‘Polio and nobel prizes: Looking back 50 years’, *Annals of Neurology*, 61(5), pp. 385–395. doi:10.1002/ana.21153.

Novak, M.A. and Suomi, S.J. (1988) ‘Psychological well-being of primates in captivity.’, *American Psychologist*, 43(10), pp. 765–773. doi:10.1037/0003-066X.43.10.765.

Okinda, Cedric *et al.* (2020) ‘A review on computer vision systems in monitoring of poultry: A welfare perspective’, *Artificial Intelligence in Agriculture*, 4, pp. 184–208. doi:10.1016/j.aiia.2020.09.002.

Passingham, R. (2009) ‘How good is the macaque monkey model of the human brain?’, *Current Opinion in Neurobiology*, 19(1), pp. 6–11. doi:10.1016/j.conb.2009.01.002.

Phillips, K.A. *et al.* (2014) ‘Why primate models matter’, *American Journal of Primatology*, 76(9), pp. 801–827. doi:10.1002/ajp.22281.

Pimm, S.L. *et al.* (2015) ‘Emerging Technologies to Conserve Biodiversity’, *Trends in Ecology & Evolution*, 30(11), pp. 685–696. doi:10.1016/j.tree.2015.08.008.

Pineda, R.R. *et al.* (2023) ‘Deep MAnTra: deep learning-based multi-animal tracking for Japanese macaques’, *Artificial Life and Robotics*, 28(1), pp. 127–138.

Porto, S.M.C. *et al.* (2013) ‘A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns’, *Biosystems Engineering*, 115(2), pp. 184–194. doi:10.1016/j.biosystemseng.2013.03.002.

Prescott, M.J. (2023) ‘Using primates in captivity: research, conservation, and education’, in *Nonhuman primate welfare: from history, science, and ethics to practice*. Springer, pp. 57–78.

Pruetz, J.D. and Isbell, L.A. (2000) ‘Correlations of food distribution and patch size with agonistic interactions in female vervets (Chlorocebus aethiops) and patas monkeys (Erythrocebus patas) living in simple habitats’, *Behavioral Ecology and Sociobiology*, 49(1), pp. 38–47. doi:10.1007/s002650000272.

Ray, S. and Stopfer, M.A. (2022) ‘Argos: A toolkit for tracking multiple animals in complex visual environments’, *Methods in Ecology and Evolution*, 13(3), pp. 585–595. doi:10.1111/2041-210X.13776.

Richter, S.H., Garner, J.P. and Würbel, H. (2009) ‘Environmental standardization: cure or cause of poor reproducibility in animal experiments?’, *Nature Methods*, 6(4), pp. 257–261. doi:10.1038/nmeth.1312.

Robinson, L.M. and Weiss, A. (2023) ‘Primate Personality and Welfare BT - Nonhuman Primate Welfare: From History, Science, and Ethics to Practice’, in Robinson, L.M. and Weiss, A. (eds). Cham: Springer International Publishing, pp. 395–411. doi:10.1007/978-3-030-82708-3_17.

Roelfsema, P.R. and Treue, S. (2014) ‘Basic Neuroscience Research with Nonhuman Primates: A Small but Indispensable Component of Biomedical Research’, *Neuron*, 82(6), pp. 1200–1204. doi:10.1016/j.neuron.2014.06.003.

Rushen, J., Chapinal, N. and de Passillé, A. (2012) ‘Automated monitoring of behavioural-based animal welfare indicators’, *Animal Welfare*, 21(3), pp. 339–350. doi:10.7120/09627286.21.3.339.

Rushen, J. and Mason, G. (2006) ‘A decade-or-more’s progress in understanding stereotypic behaviour.’, in *Stereotypic animal behaviour: fundamentals and applications to welfare*. UK: CABI, pp. 1–18. doi:10.1079/9780851990040.0001.

Salguero, F.J. *et al.* (2021) ‘Comparison of rhesus and cynomolgus macaques as an infection model for COVID-19’, *Nature Communications*, 12(1), p. 1260. doi:10.1038/s41467-021-21389-9.

Shahinfar, S. *et al.* (2021) ‘Machine learning approaches for the prediction of lameness in dairy cows’, *Animal*, 15(11), p. 100391. doi:10.1016/j.animal.2021.100391.

Tardif, S.D. *et al.* (2013) ‘IACUC Review of Nonhuman Primate Research’, *ILAR Journal*, 54(2), pp. 234–245. doi:10.1093/ilar/ilt040.

Tassinari, P. *et al.* (2021) ‘A computer vision approach based on deep learning for the detection of dairy cows in free stall barn’, *Computers and Electronics in Agriculture*, 182, p. 106030. doi:10.1016/j.compag.2021.106030.

Terven, J. and Cordova-Esparza, D. (2023) ‘A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond’, *arXiv preprint arXiv:2304.00501* [Preprint].

Thierry, B., Singh, M. and Kaumanns, W. (2004) *Macaque societies: a model for the study of social organization*. Cambridge University Press.

Truelove, M.A. *et al.* (2020) ‘The identification of effective welfare indicators for laboratory-housed macaques using a Delphi consultation process’, *Scientific Reports*, 10(1), p. 20402. doi:10.1038/s41598-020-77437-9.

Turner, P. V (2023) ‘The history of chimpanzees in biomedical research’, in *Nonhuman Primate Welfare: From History, Science, and Ethics to Practice*. Springer, pp. 31–55.

Tuttle, A.H. *et al.* (2018) ‘A deep neural network to assess spontaneous pain from mouse facial expressions’, *Molecular Pain*, 14, p. 174480691876365. doi:10.1177/1744806918763658.

Tuyttens, F.A.M. *et al.* (2014) ‘Observer bias in animal behaviour research: can we believe what we score, if we score what we believe?’, *Animal Behaviour*, 90, pp. 273–280. doi:10.1016/j.anbehav.2014.02.007.

Tuyttens, F.A.M., Molento, C.F.M. and Benaissa, S. (2022) ‘Twelve Threats of Precision Livestock Farming (PLF) for Animal Welfare’, *Frontiers in Veterinary Science*, 9. doi:10.3389/fvets.2022.889623.

Uno, Y., Uehara, S. and Yamazaki, H. (2016) ‘Utility of non-human primates in drug development: Comparison of non-human primate and human drug-metabolizing cytochrome P450 enzymes’, *Biochemical Pharmacology*, 121, pp. 1–7. doi:10.1016/j.bcp.2016.06.008.

Urano, E. *et al.* (2021) ‘COVID-19 cynomolgus macaque model reflecting human COVID-19 pathological conditions’, *Proceedings of the National Academy of Sciences*, 118(43). doi:10.1073/pnas.2104847118.

Valletta, J.J. *et al.* (2017) ‘Applications of machine learning in animal behaviour studies’, *Animal Behaviour*, 124, pp. 203–220. doi:10.1016/j.anbehav.2016.12.005.

Vidal, M. *et al.* (2021) ‘Perspectives on Individual Animal Identification from Biology and Computer Vision’, *Integrative and Comparative Biology*, 61(3), pp. 900–916. doi:10.1093/icb/icab107.

Waitt, C. and Buchanan-Smith, H.M. (2001) ‘What time is feeding?’, *Applied Animal Behaviour Science*, 75(1), pp. 75–85. doi:10.1016/S0168-1591(01)00174-5.

Wang, H. (2019) ‘Macaque monkeys as a non-human primate circadian model’, *National Science Review*, 6(2), pp. 302–303. doi:10.1093/nsr/nwz020.

Wang, S. *et al.* (2020) ‘RDSNet: A New Deep Architecture forReciprocal Object Detection and Instance Segmentation’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), pp. 12208–12215. doi:10.1609/aaai.v34i07.6902.

Weinstein, B.G. (2018) ‘A computer vision for animal ecology’, *Journal of Animal Ecology*. Edited by L. Prugh, 87(3), pp. 533–545. doi:10.1111/1365-2656.12780.

Weinstein, T.A.R. and Capitanio, J.P. (2008) ‘Individual differences in infant temperament predict social relationships of yearling rhesus monkeys, Macaca mulatta’, *Animal Behaviour*, 76(2), pp. 455–465. doi:10.1016/j.anbehav.2008.01.024.

Weiss, A. *et al.* (2011) ‘Rhesus macaques (Macaca mulatta) as living fossils of hominoid personality and subjective well-being.’, *Journal of Comparative Psychology*, 125(1), pp. 72–83. doi:10.1037/a0021187.

Wiesel, T.N. (1982) ‘Postnatal development of the visual cortex and the influence of environment’, *Nature*, 299(5884), pp. 583–591. doi:10.1038/299583a0.

Wu, Z. *et al.* (2021) ‘A Deep Detection Network Based on Interaction of Instance Segmentation and Object Detection for SAR Images’, *Remote Sensing*, 13(13), p. 2582. doi:10.3390/rs13132582.

Wurtz, K. *et al.* (2019) ‘Recording behaviour of indoor-housed farm animals automatically using machine vision technology: A systematic review’, *PLOS ONE*. Edited by D. Raboisson, 14(12), p. e0226669. doi:10.1371/journal.pone.0226669.

Yang, Y. *et al.* (2022) ‘APT-36K: A Large-scale Benchmark for Animal Pose Estimation and Tracking’, (NeurIPS). Available at: http://arxiv.org/abs/2206.05683.

Yao, Y. *et al.* (2023) ‘OpenMonkeyChallenge: Dataset and Benchmark Challenges for Pose Estimation of Non-human Primates’, *International Journal of Computer Vision*, 131(1), pp. 243–258. doi:10.1007/s11263-022-01698-2.

Yeager, C.P. (1996) ‘Feeding ecology of the long-tailed macaque(Macaca fascicularis) in Kalimantan Tengah, Indonesia’, *International Journal of Primatology*, 17(1), pp. 51–62. doi:10.1007/BF02696158.

Yeates, J.W. (2011) ‘Is ‘a life worth living’a concept worth having?’, *Animal Welfare*, 20(3), pp. 397–406.

Young, M. *et al.* (1979) ‘Inter-Observer Reliability’, *Behaviour*, 69(3–4), pp. 303–315. doi:10.1163/156853979X00520.

Chapter 2. Materials and Methods

2.1 Introduction

The studies presented in this thesis are based on data collected from the Medical Research Council Centre for Macaques (CFM) in Porton Down, Salisbury, UK. The establishment of CFM dates back to 2003 when it was initiated as a breeding colony of rhesus macaques. Its primary purpose is to provide non-human primates to UK universities for academic research.

All data obtained for the studies described in this thesis is derived from camera recordings. Videos of group-housed macaques, were automatically captured and stored using a CCTV system (See sub-section The data encompass a range of information, including weaning dates, location transitions, births (pertaining to breeding females), health assessments, medication records, treatments, veterinary checks, reports of injuries, and individual weight measurements. In the context of the Foraging Project, the data on injuries rate and individuals' weight, have been used in conjunction with the information derived from the automated methodology.

Footage of the Macaques). Videos of temperament tests carried out on individual animals were collected using camcorders by the student or member of staff responsible for conducting the tests.

In this chapter, I will provide detailed information about the animals involved in the studies, including their environment and husbandry practices. I will also elucidate various deep learning models employed for the extraction of information from video data. In addition, I will explain the software used to prepare the data for training and validation of deep learning models used in extracting information from videos. Finally, I will detail the methodologies utilized for cleaning and analysing the output data generated by the sequential arrangement of automated steps (i.e. models, R scripts, Python scripts), known as pipelines, designed to efficiently execute a series of tasks or processes.

For a thorough understanding of the number of animals, group composition, and model specifics in the studies, please refer to the dedicated chapters, providing extensive insights into the research, methodologies, and outcomes of each project.

2.2 The Centre for Macaques

2.2.1 The Animals

The non-human primates bred at the CFM are Indian origin rhesus macaques (*Macaca mulatta*). They are medium-sized primates with an average body length of about 50 to 70 centimetres and a tail length of approximately 20 to 30 centimetres. They have a robust build with a pinkish-brown or greyish-brown fur colour. The face is pale pink, and the cheeks are adorned with characteristic whisker-like tufts. They live in groups that typically consist of several adult males, multiple females, and their offspring. Within the group, there is a dominance hierarchy that determines access to resources and mating opportunities. The macaques at the CFM tend to breed between October to January and to give birth from April to July, after a gestation period between 160 to 175 days.

Due to their anatomy and behaviour, rhesus macaques make it challenging to securely attach and maintain any wearable sensors (such as collars, or bracelets). These animals are agile climbers and have dexterous hands, which could result in them removing or damaging any attached identification tags on their bodies. Therefore, to distinguish individual monkeys within the colony, a unique identification system is implemented. Each monkey is assigned initials that are tattooed on their chest by a trained technician while the monkey is under sedation. This process takes place during their first annual health screening, typically when the monkey reaches approximately one year of age.

2.2.2 The Facility

CFM is a self-sustaining breeding colony, meaning there is no need to acquire animals from outside the facility. Currently, it produces an average of 25 infants per year, with some infants being retained for use in breeding. Macaques are typically supplied to universities between the ages of 3 and 5 years and are most frequently used in neuroscience research, with smaller numbers used in ophthalmology and immunology research.

The colony comprises male and female macaques of various ages, which are housed in either breeding groups, typically consisting of one adult male, multiple females, and their offspring, or single-sex juvenile groups. The group size can vary from 2 to 27 individuals and each group is housed in a separate enclosure within the facility.

As part of the breeding program, a subset of the infants born within the breeding groups, typically between 12 to 30 months of age, are transitioned into single-sex groups. These groups are specifically established to raise and prepare the macaques for their future role in biomedical research. Subsequently, these individuals are provided to universities to support various research studies in the field of biomedicine.

The CFM facility provides appropriate habitats and living conditions that emulate the natural environment of macaques, ensuring their well-being and reproductive health. At the centre, each enclosure is divided into two distinct areas: the play pen and the cage room.

The play pen has dimensions of 8.04 meters in length, 3.35 meters in width, and 2.8 meters in height (Figure 2). This area is equipped with various enrichment items, suspended structures, visual barriers, and shelves to facilitate natural behaviours such as climbing, jumping, and resting at different heights.

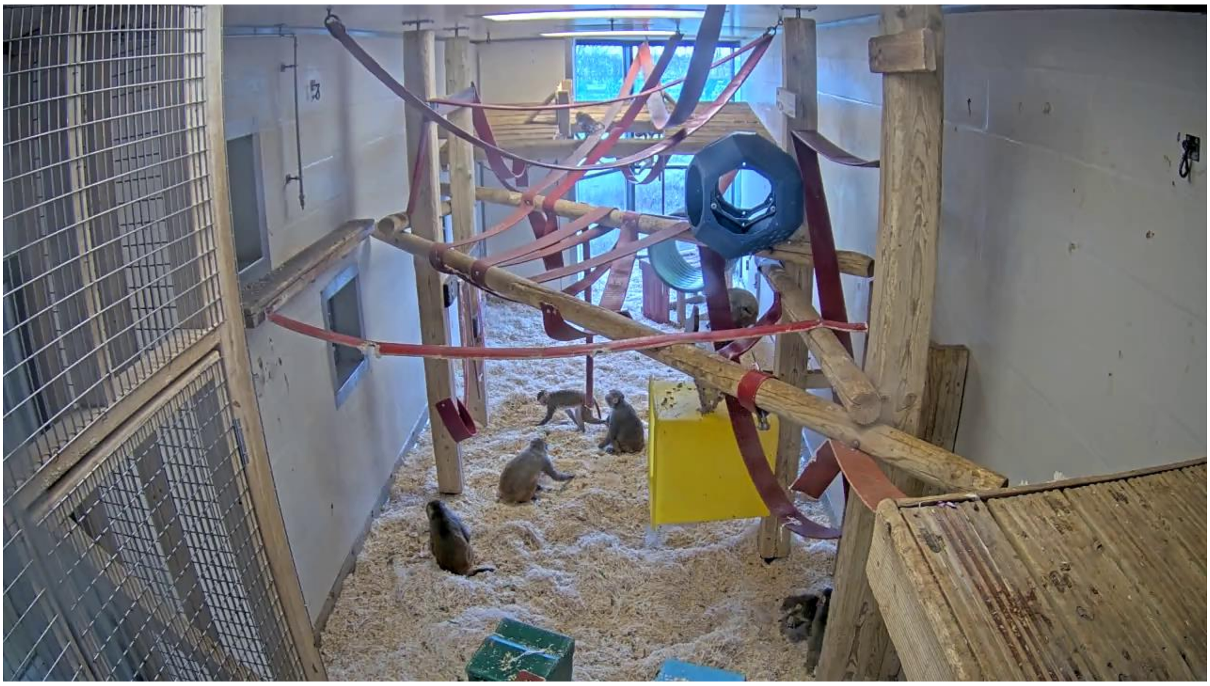


Figure 2: Images extracted from the CCTV system recording a play pen at CFM.

The cage room measures 6.12 meters in length, 1.5 meters in width, and 2.8 meters in height. This area is divided into three levels and can be further sectioned off as needed (Figure 3).

The macaques have unrestricted access between the play pen and the cage room through four hatches located at different heights, except when separation is required for cleaning, veterinary treatment, or testing. Both areas are illuminated by artificial lighting following a 12-hour light and 12-hour dark cycle, while the play pen also receives natural light through a large bay window. The enclosures are maintained at a consistent temperature between 18 to 20 degrees Celsius and a constant humidity level ranging from 55% to 65%.

Figure 3: Example of one of the cage rooms at CFM.

2.2.3 The Husbandry

At the CFM, a team of 10 animal technicians is dedicated to the care and well-being of the macaques. These technicians work from 8:00 to 16:00, providing care for the macaques seven days a week. However, there is a reduced staff presence during weekends, with only 2 members of staff at the facility. The primary responsibilities of the technicians include feeding the animals, maintaining cleanliness in the enclosures, and regularly introducing new enrichments to promote the macaques' mental and physical stimulation.

Rhesus macaques display an omnivorous diet, encompassing a wide range of food sources such as fruits, seeds, leaves, insects, and small vertebrates. At the colony, the macaques are provided with one feeding session each morning. Their diet primarily consists of specially formulated complete diet pellets, designed to meet all their nutritional requirements. A diverse array of fruit and vegetable-based forage mixes are incorporated with the diet pellets, totalling seven unique mixes assigned to specific days of the week. This approach serves multiple purposes, encouraging the macaques to engage in natural foraging behaviours while enhancing the variety in their diet. To facilitate feeding, staff enter the play pen through a door. They proceed approximately 1.5 meters into the room and evenly distribute the food by tossing buckets containing the food items onto the ground. Water dispensers are placed throughout both the play pen and the cage room, always allowing the macaques continuous access to fresh water.

Cleaning takes place on a biweekly schedule, specifically allocated for two days per group. The staff follows a rotational pattern, attending to either the play pen or the cage room during each cleaning session. The process entails removing all existing wood shavings bedding, thoroughly washing the entire enclosure, and subsequently replenishing it with fresh straw bedding. Cleaning a single room typically requires approximately 4 to 6 hours. During this period, the monkeys are temporarily relocated to the adjacent room, a measure taken to safeguard both their safety and that of the staff. It is important to note that the enclosures are exclusively cleaned on weekdays.

There are few events linked to the necessary husbandry practice that can be stressful for the macaques at CFM. The annual health screen is one of these. It occurs once per annum for each group. On the day of the screen every animal in the group apart from the infants are sedated with ketamine (10mg/kg) and transported to the on-site surgery room. The entire procedure takes approximately 1 hour. Juveniles born in the previous calendar year are also administered with an identification tattoo on their chest, if they are at least 12 months old.

Another stressful event is the removal or introduction of a male from or into a breeding group. To prevent inbreeding and enhance mating outcomes, new breeding groups are formed or disbanded annually. However, these transitions are carefully managed to prioritize the group's welfare and stability. The process involves pre-introduction assessment, quarantine, gradual exposure through sight and smell, controlled physical introductions, establishing group dynamics, and continuous monitoring.

As the primary focus of this thesis centres on examining the impact of management strategies, including feeding procedures and enrichment planning, on macaque welfare rather than quantifying the effects of stressful procedures, attention has been given to the selection of observation periods at the group level. This approach is aimed at mitigating potential sources of interference and upholding the integrity of the data collected. This includes avoiding cleaning days, weeks with health screenings, and other stressful events such as group disruptions or breakdowns. By minimizing these factors, the data collected can provide a clearer and more accurate representation of the natural behaviour and dynamics within the macaque groups.

2.3 Ethical Statement

The data collection took place at the Medical Research Council's 'Centre for Macaques', which adheres to the regulatory standards set by the UK Home Office for housing captive non-human primates. Since all the studies conducted in this thesis were observational in nature, no additional licensing was necessary. However, for the Enrichment Item Usage Study, which involved introducing objects into the enclosures, approval was obtained during the Centre for Macaques AWERB meeting held on March 13, 2023 (approval number: CFM2023E001). The PhD project itself was ethically approved by the Newcastle University Animal Welfare and Ethical Review Body.

2.4 Data Collection

The thesis contained three different studies: Temperament Test, Enrichment Item Usage, Foraging Project. The data used for the different studies are categorized into two main sources: information pertaining to the colony and recorded footage of the animals.

2.4.1 Information about the Colony

At the CFM, daily records for health and husbandry-related information are maintained in paper health records for each animal and day books for each group. Data from these records, are entered into an electronic database (Oracle based system provided by ENOS; https://poweredbyenos.com/). The data encompass a range of information, including weaning dates, location transitions, births (pertaining to breeding females), health assessments, medication records, treatments, veterinary checks, reports of injuries, and individual weight measurements. In the context of the Foraging Project, the data on injuries rate and individuals' weight, have been used in conjunction with the information derived from the automated methodology.

2.4.2 Footage of the Macaques

The videos utilized for the Temperament Test study were captured using two Sony HDR-CX625 Full HD Compact Camcorders cameras that were temporarily positioned outside the cage room specifically for the duration of the tests. In contrast, the videos employed for the other two studies were obtained through a CCTV system that continuously recorded the play pens round the clock. Each of these rooms has an Axis P1455-LE camera, these are record via Power-over-Ethernet cables to an Axis Camera Station S1148.

All the collected videos, regardless of the source, were subjected to analysis using deep learning-based models. These models enabled the automated processing and extraction of relevant behavioural information from the footage, facilitating the study and exploration of various aspects related to macaque welfare and behaviour.

2.5 Deep Learning Models

Deep learning is a subfield of machine learning that utilizes artificial neural networks, also known as deep neural networks, to learn and make predictions or decisions.

In this thesis, I will present three distinct projects, each employing various deep learning models:

1- Temperament Test Project
   This investigation encompasses three separate models:
   - Two models built on the foundation of DeepLabCut, designed to detect and distinguish various macaque body parts.
   - One model, based on YOLACT, dedicated to detecting and identifying six different objects.

2- Foraging Project
   This research revolves around a YOLACT-based model, focused on the task of detecting macaques within different playpens.

3- Enrichment Project
   This study comprises two models:
   - One built upon YOLACT, aimed at detecting and tracking a specific object.
   - Another model based on YOLOV8, intended to identify macaques within various playpens, parts of an enrichment puzzle, and the front platform within the playpens.

In order to enable a deep learning model to learn from data and make accurate predictions or decisions on new data, it is necessary to train it. Training a deep learning model involves two primary phases: training and validation. During the training phase, the model learns to make label predictions or segment objects in the image. Label predictions involves classifying objects in the image, while perform segmentation involves identifying and outlining object boundaries using techniques like bounding boxes and masks (Matthews *et al.*, 2017; Wang *et al.*, 2020; Wu *et al.*, 2021). Bounding boxes are rectangular regions that tightly enclose objects of interest in an image, whereas masks are pixel-level representations that precisely delineate the shape and extent of objects, providing a more fine-grained segmentation. accurate predictions or decisions on new data, it is necessary to train it. Training a deep learning model involves two primary phases: training and testing.

2.5.1 Training

In supervised learning (see 0Chapter 2 for more details), during the training phase, the model learns from a labelled dataset, also known as the training set, which consists of input data and corresponding target labels. The model iteratively adjusts its internal parameters, also known as weights and biases, to minimize the difference between its predicted labels and the true labels in the training set (Kavakiotis *et al.*, 2017).

The training process involves passing the training data through the neural network, calculating the predictions, comparing them to the true labels, and updating the weights using optimization algorithms such as gradient descent and backpropagation (Figure 4).

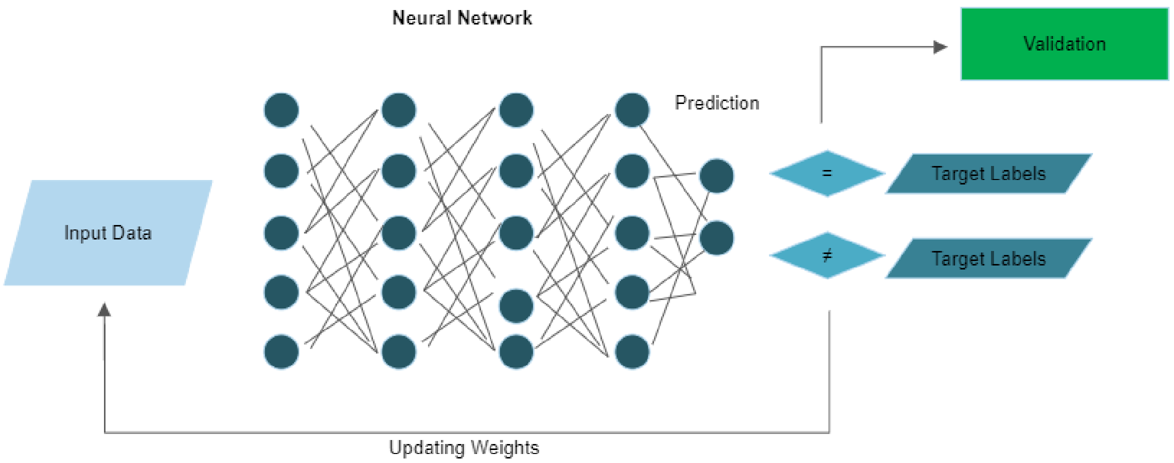


Figure 4: Diagram showing the main steps to train a deep learning-based model.

This iterative process continues until the model reaches a point where the predictions align closely with the true labels in the training set. The goal is to minimize the training loss, which quantifies the discrepancy between the predicted and true labels. After the model has completed training, it proceeds to the validation phase, where the hyperparameters are assessed. These hyperparameters are configuration settings predetermined before the commencement of training in a machine learning or deep learning model. They wield significant influence in shaping the model's architecture and its behaviour throughout the training process.

2.5.2 Testing and Evaluation

During testing, the model is evaluated using a separate dataset called the test set (Kavakiotis *et al.*, 2017). This dataset contains examples that the model has not seen during training. The purpose of the test set is to assess the generalization performance of the model. By making predictions on the test set and comparing them to the true labels, metrics such as accuracy, precision, recall (for categorical labels), or mean squared error and mean average precision (for continuous predictions) are calculated to measure the model's performance. The ground truth, which represents the true, accurate labels or annotations, is used for this comparison. This evaluation helps in understanding how well the model is likely to perform on new, unseen data.

In each study, I established a pipeline to extract the necessary information. These pipelines consisted of a sequence of interconnected and ordered processing steps. These steps included running deep learning models on videos, refining the model's output, and extracting essential information, such as the count of detections and the positions of detected objects. To evaluate these pipelines, I used the accuracy, recall and precision. Accuracy is a common evaluation metric used in machine learning to measure the performance of a classification model. It represents the proportion of correctly predicted instances out of the total number of predictions made. In other words, accuracy indicates how often the model's predictions align with the ground truth.

Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances that were attempted. It focuses on the ability of the model to identify all positive instances, avoiding false negatives. Recall is calculated by dividing the number of true positive predictions by the sum of true positive and false negative predictions.

Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It focuses on the accuracy of positive predictions and helps evaluate the model's ability to minimize false positives. Precision is calculated by dividing the number of true positive predictions by the sum of true positive and false positive predictions.

To perform a comprehensive evaluation of the individual models within each pipeline, I utilized three distinct evaluation metrics, each selected based on the inherent evaluation methods of the underlying algorithm used for the model. For instance, classification models, such as those based on YOLACT and YOLO (Bolya *et al.*, 2019; Jocher, G., Chaurasia, A., & Qiu, 2023), which yield class labels as outputs, were assessed using metrics that focus on classification accuracy, precision, and recall. Conversely, regression models like DeepLabCut (Mathis *et al.*, 2018), which produce numerical values within a range of real numbers, were evaluated using metrics designed to gauge the model's proficiency in making accurate numerical predictions.

I used the mean average precision (mAP) to evaluate the YOLCAT models' performance (Bolya *et al.*, no date; Liu, 2009). The mAP compares the ground-truth bounding box (and mask) to the detected box (and mask) and returns a score. It measures how accurately the model identifies and localizes objects within an image.

$$mAP = \frac{AP_1 + AP_2 + \ldots + AP_n}{n}$$

Where $AP_1, AP_2, \ldots, AP_n$ are the average precision values (AP) for each object category. AP_i is the area under the Precision- Recall Curve for the object i. This curve is obtained by plotting the model's precision and recall values as a function of the model's confidence score threshold. n is the total number of object categories.

The higher the score, the more accurate the model is in its detections. In addition, the mAP, can be calculated for different thresholds of Intersection over Union (IoU) (Rahman and Wang, 2016). IoU is a metric that evaluates the accuracy of bounding box predictions. It measures the overlap between the predicted bounding box and the ground truth bounding box of an object in an image (

Figure 5).

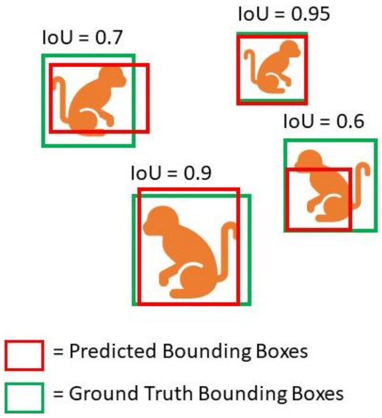
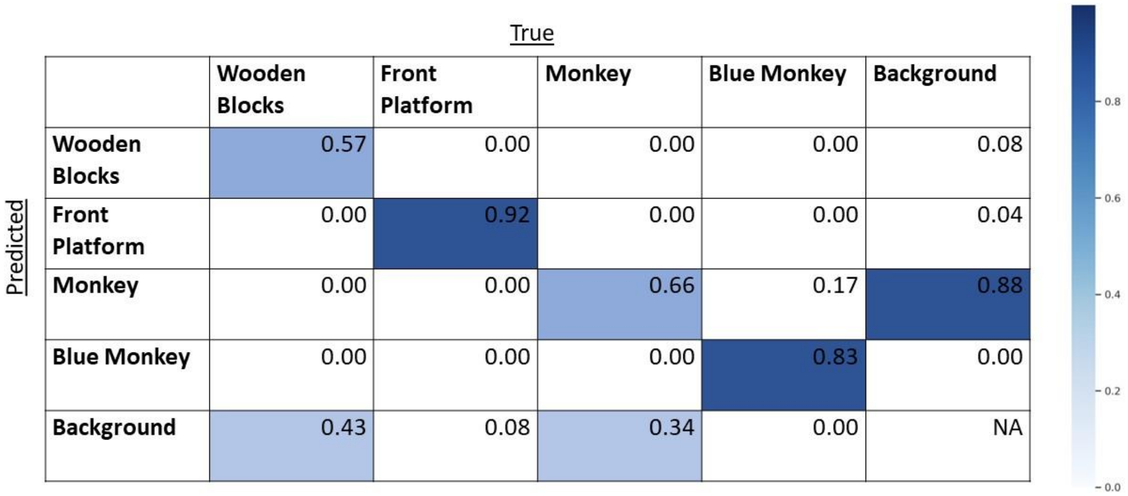


Figure 5: Example of Intersection of Union (IoU).

The confusion matrix was used as tabular representation to evaluate the YOLO model's performance in identifying objects within the videos (Li and Deng, 2019; Jocher, G., Chaurasia, A., & Qiu, 2023). In object detection, the matrix provides a detailed breakdown of the model's predictions for each class of objects, taking into account both classification and localization aspects.

The model using YOLO was trained to discern macaques, two enrichment items, and a wooden platform. Consequently, its associated confusion matrix featured five columns and five rows. These rows correspond to the individual classes (macaques, two enrichment items, wooden platform, and background), with an extra row allocated for the "background" class. This class assumes significance in delineating regions in an image devoid of any objects of interest. This inclusion empowers the model to discern that not every portion of an image necessarily hosts a detectable object. This proves particularly beneficial in scenarios where objects of interest are relatively sparse within video frames.

In the context of a confusion matrix, the rows signify the Predicted Classes, where each row encapsulates a distinct class of objects anticipated by the model. Conversely, the columns represent the True Classes, encapsulating the classes designated for detection by the model. The diagonal elements (top-left to bottom-right) of the matrix represent the cases where the model made correct predictions. While the off-diagonal elements represent cases where the model made wrong predictions (Figure 6).



	Wooden Blocks	Front Platform	Monkey	Blue Monkey	Background
Wooden Blocks	0.57	0.00	0.00	0.00	0.08
Front Platform	0.00	0.92	0.00	0.00	0.04
Monkey	0.00	0.00	0.66	0.17	0.88
Blue Monkey	0.00	0.00	0.00	0.83	0.00
Background	0.43	0.08	0.34	0.00	NA

Figure 6: Confusion matrix from the YOLO model used in the Enrichment Project.

Finally, I used the Mean Squared Error (MSE) to evaluate the DeepLabCut models' performance (Mathis *et al.*, 2018). The MSE loss function is calculated as the squared difference between the model's prediction and the ground truth, averaged across the whole dataset.

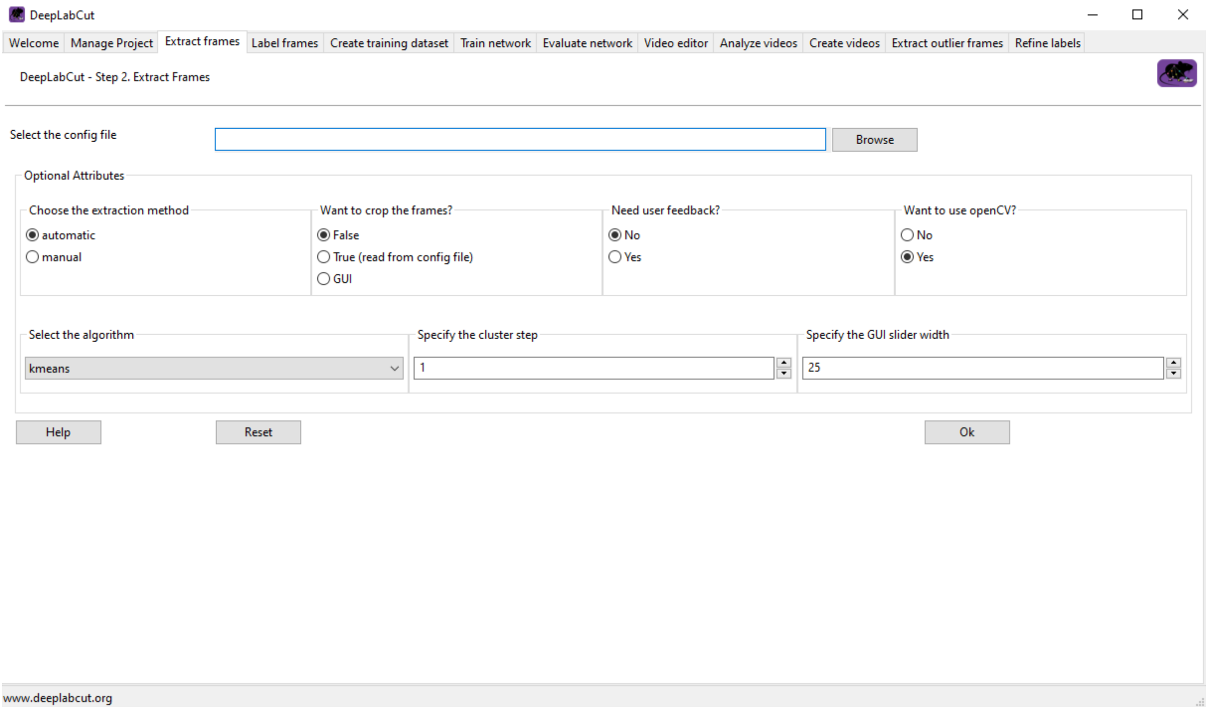
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Where y_i are the ground truth values, $\hat{y}_i$ are the model's prediction values, and N is the total number of objects.

2.5.3 *Preparing the Databases*

DeepLabCut, YOLACT (You Only Look At CoefficienTs) and YOLO (You Only Look Once) require supervised learning, where both the input data and the corresponding target labels are provided (Mathis *et al.*, 2018; Bolya *et al.*, 2019; Jocher, Chaurasia and Qiu, 2023). To create these datasets, I used three graphical user interfaces (GUI).

DeepLabCut provides a GUI (Figure 7) that allows the user to extract the frames from videos, label them, train the model, and use it on new videos (Mathis *et al.*, 2018).



DeepLabCut - Step 2. Extract Frames

Select the config file

Optional Attributes

Choose the extraction method
● automatic
○ manual

Want to crop the frames?
● False
○ True (read from config file)
○ GUI

Need user feedback?
● No
○ Yes

Want to use openCV?
○ No
● Yes

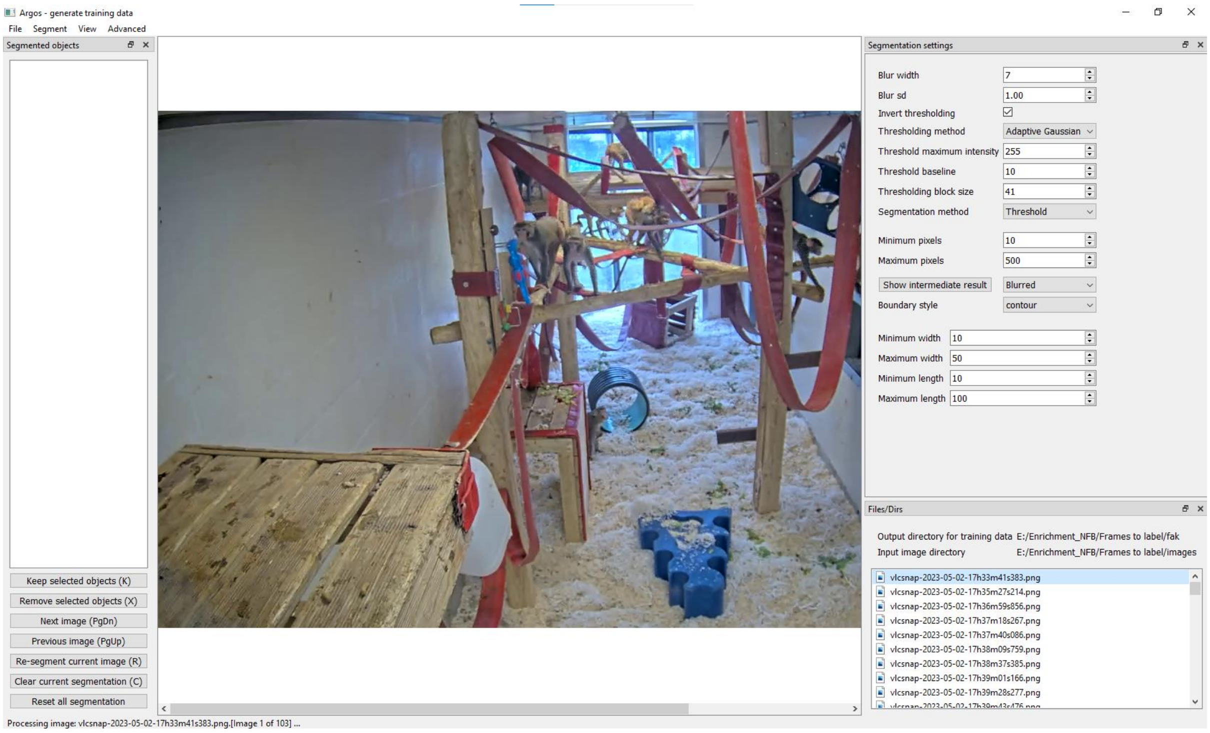
Select the algorithm: kmeans

Specify the cluster step: 1

Specify the GUI slider width: 25

Figure 7: DeepLabCut GUI. Extract Frames Tab.

To create the datasets for training and testing of the YOLACT models, I used the annotation tool provided by Argos (Ray and Stopfer, 2022). This facilitated the process of labelling images and exporting them in a format compatible with YOLACT for training purposes (Figure 8).



Blur width 7
Blur sd 1.00
Invert thresholding ☑
Thresholding method Adaptive Gaussian
Threshold maximum intensity 255
Threshold baseline 10
Thresholding block size 41
Segmentation method Threshold
Minimum pixels 10
Maximum pixels 500
Show intermediate result Blurred
Boundary style contour
Minimum width 10
Maximum width 50
Minimum length 10
Maximum length 100

Output directory for training data E:/Enrichment_NFB/Frames to label/fak
Input image directory E:/Enrichment_NFB/Frames to label/images

vlcsnap-2023-05-02-17h33m41s383.png
vlcsnap-2023-05-02-17h35m27s214.png
vlcsnap-2023-05-02-17h36m59s856.png
vlcsnap-2023-05-02-17h37m18s267.png
vlcsnap-2023-05-02-17h37m40s086.png
vlcsnap-2023-05-02-17h38m09s759.png
vlcsnap-2023-05-02-17h38m37s385.png
vlcsnap-2023-05-02-17h39m01s166.png
vlcsnap-2023-05-02-17h39m28s277.png

Processing image: vlcsnap-2023-05-02-17h33m41s383.png.[Image 1 of 103] ...

Figure 8: Argos GUI.

To generate the necessary data for training and validating the YOLO model I employed a segment-anything-annotator that incorporates features from both the Segment Anything Model (SAM) and labelme (Wada, no date; Kirillov *et al.*, 2023) (Figure 9). This user interface (UI) was released in 2023, towards the conclusion of my PhD, limiting its usage to my latest project. The methodology boasts a significant advantage: it can produce high-quality object masks from simple input prompts, like points, resulting in a faster labelling process compared to conventional methods.

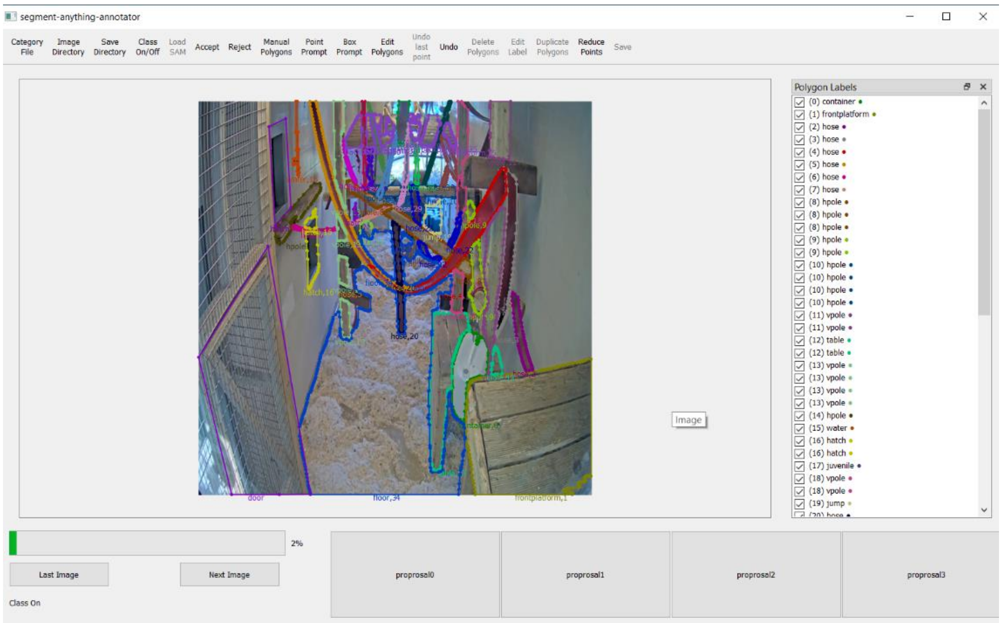


Figure 9: Segment Anything Annotator UI.

To attain satisfactory accuracy for each model, I needed to label fewer than 500 images for training and testing, this has been possible thanks to a machine learning technique called transfer learning. This involves leveraging knowledge gained from one task or domain to improve performance on another related task or domain. In transfer learning, a pre-trained model, which has been trained on a large dataset for a specific task, is utilized as a starting point or a feature extractor for a different but related task.

The idea behind transfer learning is that the features learned by the pre-trained model on the source task can be generalized and transferred to the target task, even if the target task has a different dataset or slightly different requirements. Rather than training a new model from scratch on the target task, transfer learning allows us to benefit from the knowledge already captured by the pre-trained model.

To pre-train a model, there are several available large-scale image databases. DeepLabCut, for instance, is pre-trained on ImageNet, a dataset comprising over 14 million images categorized into more than 20,000 classes or categories (Mathis *et al.*, 2018). On the other hand, both YOLACT and YOLO are pre-trained on COCO (Common Objects in Context), which contains 330,000 images covering 80 common object categories (Bolya *et al.*, 2019). Notably, the latter database provides both labels and segmentation data for each image, while ImageNet solely offers labels.

2.5.4 Data Processing

Once the models were trained and utilized to extract information from unseen videos, the outputs of these algorithms, DeepLabCut (DLC), YOLACT and YOLO, consist of CSV (Comma-Separated Values) files containing thousands of entries, each representing specific details captured during the analysis process.

The CSV files typically include the following types of information:

1- Body Part Coordinates: For DLC, the CSV file contains the x and y coordinates of each identified body part in the video frames. These coordinates represent the positions of body parts such as joints, limbs, or other points of interest. These values allow for tracking and analysing the movement or position of specific body parts over time.
2- Bounding Box Coordinates: For YOLACT and YOLO, the CSV file includes the bounding box coordinates for detected objects or classes in the video frames. The bounding box specifies the position and size of an object in the image. It consists of the top-left corner coordinates (x, y) and the width and height of the bounding box. This information helps in identifying and localizing objects of interest within the video frames.
3- Likelihood or Confidence Scores: Both DLC, YOLACT and YOLO provide likelihood or confidence scores associated with each data entry. These scores represent the model's level of certainty or confidence in its predictions or detections. Higher scores

indicate a higher degree of confidence in the accuracy of the predictions, while lower scores may indicate uncertainty or lower confidence.

4- Classes or Labels: For YOLACT and YOLO, the CSV file contains information about the detected object classes or labels. Each entry specifies the class or category to which an object belongs, such as "person", "macaques", "raisins" etc. This information enables the identification and classification of different objects present in the video frames.

Due to the large volume of data generated by the models, it was essential to undergo a data cleaning process to ensure the data align with the desired outcomes and objectives. Data cleaning involves a series of tasks such as removing unneeded data, handling missing values, standardizing formats, removing values associated with low confidence scores, correcting errors, and resolving inconsistencies in the data. This helps mitigate the impact of noisy or unreliable data on the interpretation of results, facilitating the data analyses.

2.6 Other Software

To create and compile the training dataset for the various models, I utilized the Python programming language (version 3.9.13). Python offers a wide range of libraries and frameworks that are well-suited for data manipulation, pre-processing, and model training.

For data processing and statistical analyses, I employed R, specifically version 4.0.4. R provides a comprehensive set of tools and packages for data manipulation, visualization, and statistical modelling, making it suitable for conducting rigorous analyses on the obtained data.

2.7 Chapter References

Bolya, D. *et al.* (2019) ‘Yolact: Real-time instance segmentation’, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166.

Bolya, D. *et al.* (no date) *YOLACT Real-time Instance Segmentation*. Available at: https://github.com/dbolya/yolact.

Jocher, G., Chaurasia, A., & Qiu, J. (2023) ‘YOLO by Ultralytics (Version 8.0.0)’.

Jocher, G., Chaurasia, A. and Qiu, J. (2023) ‘YOLO by Ultralytics’. Available at: https://github.com/ultralytics/ultralytics.

Kavakiotis, I. *et al.* (2017) ‘Machine learning and data mining methods in diabetes research’, *Computational and structural biotechnology journal*, 15, pp. 104–116.

Kirillov, A. *et al.* (2023) ‘Segment Anything’, *arXiv:2304.02643* [Preprint].

Li, S. and Deng, W. (2019) ‘Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning’, *International Journal of Computer Vision*, 127(6), pp. 884–906.

Liu, T.-Y. (2009) ‘Learning to rank for information retrieval’, *Foundations and Trends® in Information Retrieval*, 3(3), pp. 225–331.

Mathis, A. *et al.* (2018) ‘DeepLabCut: markerless pose estimation of user-defined body parts with deep learning’, *Nature Neuroscience*, 21(9), pp. 1281–1289. doi:10.1038/s41593-018-0209-y.

Matthews, S.G. *et al.* (2017) ‘Automated tracking to measure behavioural changes in pigs for health and welfare monitoring’, *Scientific Reports*, 7(1), p. 17582. doi:10.1038/s41598-017-17451-6.

Rahman, M.A. and Wang, Y. (2016) ‘Optimizing intersection-over-union in deep neural networks for image segmentation’, in *International symposium on visual computing*. Springer, pp. 234–244.

Ray, S. and Stopfer, M.A. (2022) ‘Argos: A toolkit for tracking multiple animals in complex visual environments’, *Methods in Ecology and Evolution*, 13(3), pp. 585–595. doi:10.1111/2041-210X.13776.

Wada, K. (no date) ‘Labelme: Image Polygonal Annotation with Python’. doi:10.5281/zenodo.5711226.

Wang, S. *et al.* (2020) ‘RDSNet: A New Deep Architecture forReciprocal Object Detection and Instance Segmentation’, *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), pp. 12208–12215. doi:10.1609/aaai.v34i07.6902.

Wu, Z. *et al.* (2021) ‘A Deep Detection Network Based on Interaction of Instance Segmentation and Object Detection for SAR Images’, *Remote Sensing*, 13(13), p. 2582. doi:10.3390/rs13132582.

Chapter 3. Exploring Macaque Temperament Assessment through Deep Learning: A Study on Behaviour Recognition in Temperament Tests

3.1 Abstract

Knowing temperament in captive non-human primates by gauging individual responses to novel objects holds significance in their management. Typically, temperament tests are evaluated through focal observation and manual behavioural coding, a method demanding trained personnel and significant time investment. In this study, I explored whether computer vision techniques could be used to automate collection of behavioural data from rhesus macaques (*Macaca mulatta*) during temperament tests.

A total of 130 individual temperament tests were recorded using two cameras, one positioned at the front of the cage and another on the side. Three distinct deep learning-based models were trained and validated using images extracted from a subset of temperament test videos. These models were employed to quantify relevant behavioural parameters crucial for temperament assessment, including movement patterns, exploratory behaviour, and latency to approach novel stimuli.

The results obtained from the pipeline constructed upon these models were compared with manually coded data collected by a human observer. Both datasets were then applied to a case study aimed at investigating the impact of the protocols in place during Covid-19 pandemic on macaque behaviour. It is worth noting that due to video limitations, I encountered challenges in obtaining reliable data regarding latency to approach to the novelties.

Nevertheless, my automated method successfully extracted previously unrecorded behaviours crucial for temperament assessment in macaques, enabling the demonstration that macaques born and raised during the Covid-19 pandemic exhibited a higher level of neophobia during temperament tests compared to the cohort tested before the pandemic.

3.2 Introduction

Annually, approximately 100,000 primates are employed in global research endeavors. Macaques, due to their close genetic resemblance and physiological-cognitive parallels to humans, are favored for studies with translational implications (Kalin and Shelton, 1989). Ensuring optimal welfare for these valuable research subjects necessitates informed management practices, with temperament assessment emerging as a key strategy in captive non-human primate care (Coleman *et al.*, 2012; Coleman and Schapiro, 2021). Temperament, sometimes synonymous with personality (Capitanio, Mendoza and Cole, 2011), reflects a biologically inherent disposition in an individual's consistent response to stimuli over time (Stamps and Groothuis, 2010; Schmidt and Poole, 2020). Quantitative temperament evaluation encompasses standardized tests measuring behavior in reaction to set stimuli or trait ratings by experienced caretakers, employing Likert scale ratings of descriptive adjectives to gauge distinct temperament facets (refer Freeman and Gosling, 2010; Coleman and Pierre, 2014).

Assessing temperament in rhesus macaques held for research offers insights into various challenges, including injury susceptibility, abnormal behaviour development, health status, group stability, training success, husbandry response, social bond formation, and pair housing compatibility (Coleman, Tully and McMillan, 2005; Weinstein and Capitanio, 2008, 2012; Vandeleest, McCowan and Capitanio, 2011; Button *et al.*, 2013; Gottlieb and Capitanio, 2013; Gottlieb, Coleman and McCowan, 2013; Gottlieb *et al.*, 2019; Doelling *et al.*, 2021; Fox *et al.*, 2021). This assessment helps identify individuals at risk and informs strategies for their management and research outcomes.

Numerous factors play a role in shaping temperament, including prenatal exposure to stress affecting infant emotional responsiveness and postnatal experiences (Timmermans and Vossen, 1996; Herrington, Del Rosso and Capitanio, 2016). Notably, assessing temperament in captive populations requires attention to habituation to human presence, which has been shown to influence responses to novel stimuli. Recent research on vervet monkeys (*Chlorocebus pygerythrus*) and orangutans (*Pongo sp.*) underscores how prior exposure to humans fosters curiosity and exploration of novel objects (Damerius *et al.*, 2017; Forss *et al.*, 2022). In addition, in captivity, habituation is pivotal for animal well-being, minimizing undue stress from human interaction. However, the COVID-19 pandemic disrupted habitual exposure for primates at the Medial Research Council's Centre for Macaques (CFM), potentially impacting their welfare and data quality for biomedical research (Schapiro and Hau, 2023). Ensuring consistent and documented levels of habituation is essential for maintaining reliable research outcomes involving primates.

Traditionally, temperament assessments entail laborious manual observations and behavioral data collection, reliant on skilled personnel (Porto *et al.*, 2013; Barnard *et al.*, 2016; Mathis *et al.*, 2018; Weinstein, 2018). The human observer needs to watch videos recording the test and code behaviours of interest throughout the entire test, typically spanning a duration of 15 to 20 minutes. However, strides in machine learning and computer vision offer promising avenues to address these challenges (Blumrosen, Hawellek and Pesaran, 2017; Kennedy, 2022).

This study focuses on analysing temperament tests to extract valuable insights for evaluating macaques' neophobia at the CFM. Specifically, the objective was to uncover the stimulus-monkey interaction by employing object detection, localization techniques, and pose estimation. These models were applied to examine the influence of reduced habituation due to COVID-19 procedures on macaques' responses to various stimuli in temperament tests.

3.3 Materials and Methods

3.3.1 Facility and Subjects

Temperament tests were conducted at the Medical Research Council's Centre for Macaques (CFM) involving 130 rhesus macaques, with 59 of them being females, spanning across four birth cohorts. These assessments took place around their 3rd birthday, with ages ranging from 2.97 to 3.51 years (mean ± SD = 3.07 ± 0.1), spanning the period from March 2019 to August 2022.

Staff at the CFM typically engages in routine habituation practices, which can encompass training sessions or providing treats like raisins and peanuts to the monkeys. However, due to the COVID-19 pandemic, staffing levels were halved, leading to a substantial reduction in habituation efforts. This proactive measure aimed to mitigate the risk of SARS-CoV-2 transmission to the monkeys (see detailed habituation hours in the colony from 2019 to 2022 in Table 1).

Year	Recorded hours of habituation
2019	135.08
2020	48.33
2021	18.17
2022	71.58

Table 1: Total recorded hours of habituation/training for all the macaques in the colony by year.

The CFM provides rhesus macaques with access to a playpen and an adjoining cage area (see Chapter 2). Within the cage room, two cage sections are positioned across from each other, separated by about 1.5 meters. This layout enables two distinct groups to have visual exposure to each other. During testing, neither the opposite group nor the individual's group mates had visual contact with the subject being tested. For the actual testing process, the individual was isolated from the group in the adjacent cage room, running parallel to their playpen. The testing area's dimensions were 6.0 meters in length, 1.5 meters in width, and 0.9 meters in height. To provide a visual barrier, a black sheet was placed on the front of the cage that is farthest from the door, covering approximately one-third of the cage area where observers entered.

The temperament tests initiated with a five-minute acclimatization phase, during which the monkey interacted with familiar food (raisins). Following this, three sequential two-minute phases were conducted: one with a novel food and two featuring novel objects. Each monkey was assigned randomly to one of two stimulus sets, and the presentation sequence consistently followed 'novel food' first, 'novel object 1' second, and 'novel object 2' third (as illustrated in Figure 10). The stimuli were positioned on a small wooden block placed at the centre of the cage exterior (Figure 11). Throughout the test, the monkey was solitary in the testing area, except when a researcher entered to replace the stimulus and introduce the next one at the phase's conclusion.

The temperament tests were recorded on two Sony HDR-CX625 Full HD Compact Camcorders: one situated on the outside of a clear panel on the side of the cage, the other in front of the cage where the stimulus was presented (Figure 11).

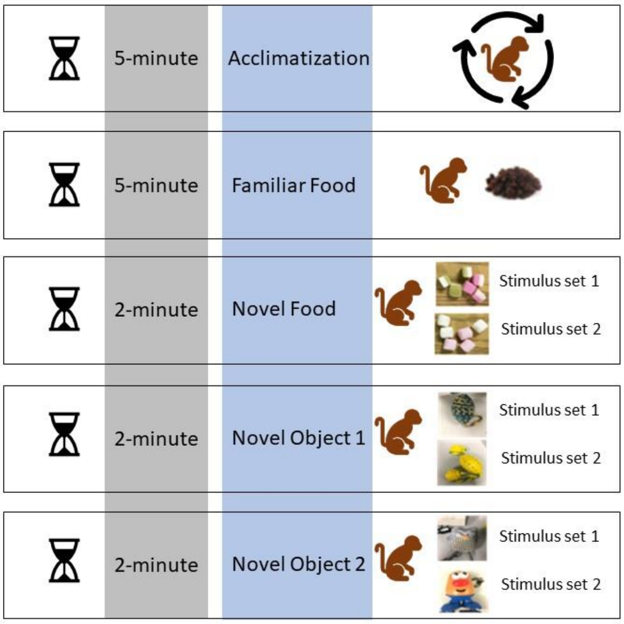


Figure 10: Different phases of a temperament test. The novelties were randomly assigned from two sets of stimuli: Stimulus set 1 and Stimulus set 2).

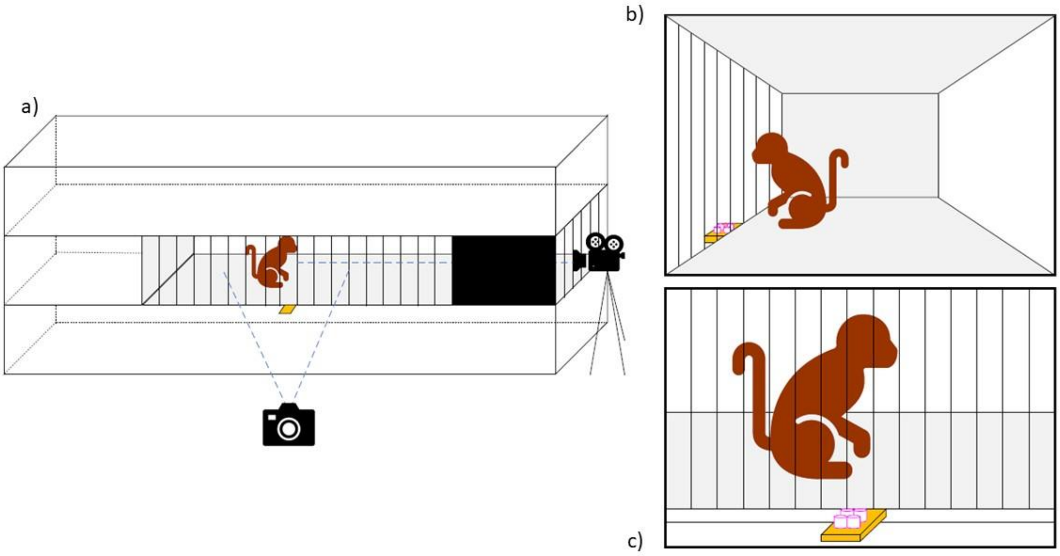


Figure 11: Cage setting during temperament tests. a) Cage room was set up with a wooden shelf outside of the cage where the different items used for the test were positioned. The tests were recorded using two cameras, one positioned on the side of the cage and the other on the front of the cage. b) Camera view from the side of the cage. c) Camera view from the front of the cage.

3.3.2 Automated Methodologies

To address the primary objective of assessing stimulus-monkey interaction, an approach encompassing object detection, localization, and pose estimation was selected to elucidate the intricacies of monkey-stimulus dynamics (further details about these approaches are available in Chapter 2). This methodology enables the determination of the monkey's spatial position, or relevant body parts, in relation to the stimulus.

To achieve this goal, two distinct computer vision models have been employed to analyse the video recording the temperament tests. Computer vision is a process by which information is extracted from visual inputs like images and videos using computational techniques. The models capitalize on the utilization of the real-time instance segmentation tool, You Only Look At CoefficienTs (YOLACT) (Bolya *et al.*, 2019), a deep learning technique rooted in Convolutional Neural Network (CNN) principles and renowned for its efficacy in localizing objects within complex visual scenes (LeCun, Bengio and Hinton, 2015; Weinstein, 2018). Together with YOLACT, DeepLabCut (Mathis *et al.*, 2018), a foundation stone in contemporary marker-less pose estimation, has been integrated into the analytical framework. This was essential to accurately track specific body parts on individual monkeys, enhancing the ability to analyse how the individuals interact with the stimuli.

Object Detection Model

The aim of the YOLACT-based was to detect the ID and the location (bbox x, y coordinate) of 6 objects used during the tests. This model processes video input sourced from the forward-positioned cage camera (Figure 11b) and yields outputs comprising object IDs along with their respective regions of interest (ROIs) (Figure 12). YOLACT, having undergone training, has acquired the capacity to discern six distinct object classes: Familiar Food (FF), Novel Food (NF), Novel Objects 1 (NO1), and Novel Objects 2 (NO2) (Figure 10). Notably, the model has been trained to conflate the familiar food items from both stimulus sets into a unified class, while concurrently trained to discriminate among all four novel objects. This stratagem was adopted due to the similar visual attributes of the two familiar food categories.

The model demonstrates reasonable performance in detecting Novel Objects, manifesting a nominal false-positive rate of 7.9% and a satisfactory true-positive rate of 96% (Table 2). A nuanced assessment of the familiar food items in terms of false-positive and true-positive rates was unfeasible. This is attributed to the likelihood of these food items being consumed during the testing phase, thereby potentially disrupting the frames where their detection was expected but remained unattainable due to consumption. For a comprehensive evaluation encompassing all six classes, refer to Table 3.

Rope Toy		**Owl Toy**		**Yellow Toy**		**Mr. Potato Toy**	
TP	FP	TP	FP	TP	FP	TP	FP
94%	7%	94%	1.5%	99%	16%	97%	7%

Table 2: Average percentage of true positive (TP) and false positive (FP) for the different Novel Objects.

	All	.50	.60	.70	.80	.90
Box	74	94	92	86	79	53
Mask	68	91	90	84	76	37

Table 3: Mean average precision (mAP) for the Object Detection Model. The mAP compares the ground-truth bounding box (and mask) to the detected box (and mask) and returns a score. These comparisons are reported for different IoU thresholds, which represent the proportion of overlap between the predicted bounding box (or mask) and the ground truth bounding box (or mask) in relation to their combined area. For instance, an IoU of .50 signifies a 50% overlap.

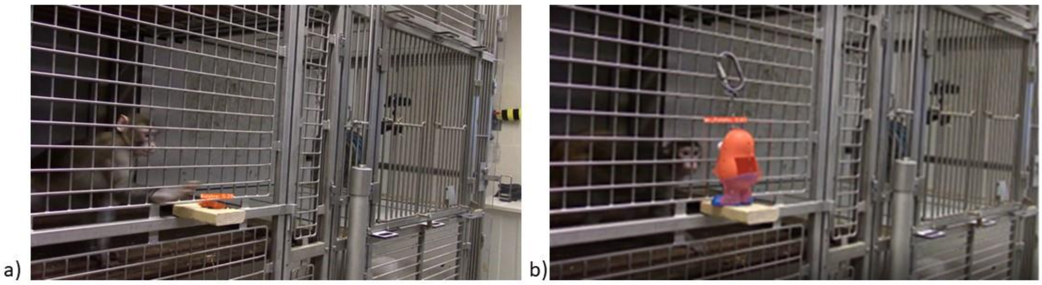


Figure 12: Examples of Object Detection model's visual outputs. a) showing the Familiar Food and b) showing the Novel Object 2 from Stimulus Set 2.

<u>Object Interaction Model</u>

The Object Interaction Model was designed to identify the locations of 8 specific body parts utilized for interacting with the food and objects provided during the tests. The model was employed to determine the positional attributes of specific anatomical elements, namely the monkey's hands, feet, eyes, mouth, and nose, within the video recordings (Table 5). Based on DeepLabCut, this model takes video inputs sourced from the forward-positioned cage camera (Figure 11b), adeptly identifying up to eight distinct body parts normally used by the monkeys to interact with novelties. Demonstrating sound performance, the model showcases a negligible overall error during both training and testing phases, as detailed in

	Training Error (pixels)	**Testing Error (pixels)**

Object Interaction Model	3.12 (3.0)	4.85 (3.67)
Tracking Model	1.91 (1.91)	8.34 (5.24)

Table 4.

	Training Error (pixels)	**Testing Error (pixels)**
Object Interaction Model	3.12 (3.0)	4.85 (3.67)
Tracking Model	1.91 (1.91)	8.34 (5.24)

Table 4: Mean Squared Error for the Tracking and Object Interaction Pipeline (from DeepLabCut evaluation). Numbers are given as overall error (error if threshold >0.9). The error is calculated as the difference in pixels between the ground truth and the model output.

The model's output, spotlighting the detected animal body parts, was used to capture interactions with the items of interest (Figure 13). Specifically, the coordinates (x and y) of these body parts, in conjunction with the output from the Object Detection Model, are harnessed to discern instances when the animal engages with the items. This determination is founded on the premise that the x, y coordinates derived from the Object Interaction Model, which identified the position of the body part within the frame, fall within the region of interest (ROI) outlined by the Object Detection Model. This model identifies the location (bounding box) of the object. If the x, y coordinates of the body parts are within the area of the ROI, then an approach is detected. (As illustrated in Figure 16).

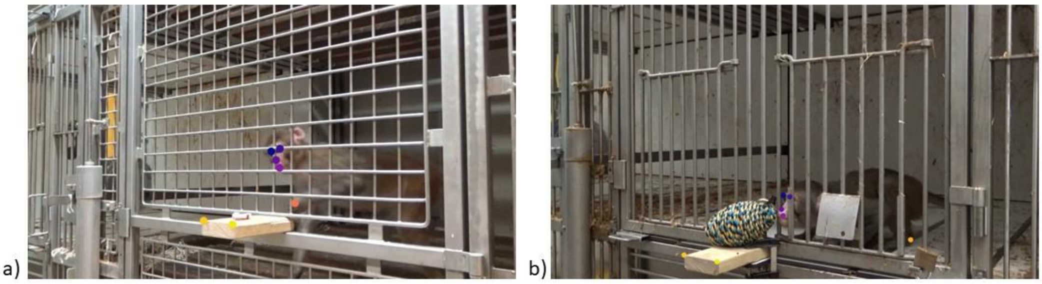


Figure 13: Examples of Object Interaction model's visual outputs. a) showing an approach from the Novel Food phase and b) showing an approach from the Novel Object 1 phase.

Utilizing both the Object Detection Model and the Object Interaction Model, it was possible to gain valuable insights into the monkeys' interactions with the stimuli, focusing on extracting behavioural metrics such as approach latency and the frequency of approaches to the stimuli (behaviours defined in 3.3.3).

In order to delve into the monkeys' responses to the diverse stimuli, a third model was employed. The Tracking Model had the aim of meticulously tracking alterations in individual relative positioning over time and it was used to extract the monkey movement patterns (Figure 16).

Tracking Model

Tracking Model body parts	Object Interaction Model body parts
Right and Left Eye	Right and Left Eye
Nose	Nose
Mouth	Mouth
Right and Left Foot	Right and Left Ear
Right and Left Hand	Right and Left Shoulder
	Top of the Head
	Back of the Neck
	Right and Left Elbow
	Right and Left Hand
	Right and Left Hip
	Right and Left Knee

	Right and Left Ankle
	Right and Left Foot
	Tail Base
	Tail Tip

The Tracking Model, based on DeepLabCut, was developed to utilize the coordinates of 24 body parts for tracking macaques around the cage, enabling the extraction of their movement patterns. The model utilizes video inputs sourced from the laterally positioned cage camera (Figure 11a). This model demonstrates proficiency in detecting and localizing up to 26 distinct body parts (Table 5), yielding commendable performance with an appreciably modest overall error rate during both training and testing stages (

	Training Error (pixels)	**Testing Error (pixels)**
Object Interaction Model	3.12 (3.0)	4.85 (3.67)
Tracking Model	1.91 (1.91)	8.34 (5.24)

Table 4).

Table 5: List of body parts detected by the Tracking Model and the Object Interaction Model.

The model's output has been employed to enable tracking of animal movement within the cage (Figure 14). This is achieved by extracting the central coordinate of the animal's head (Figure 16). To provide illustrative context, Figure 15 presents two distinct examples of cage exploration undertaken by separate individuals during their temperament tests.

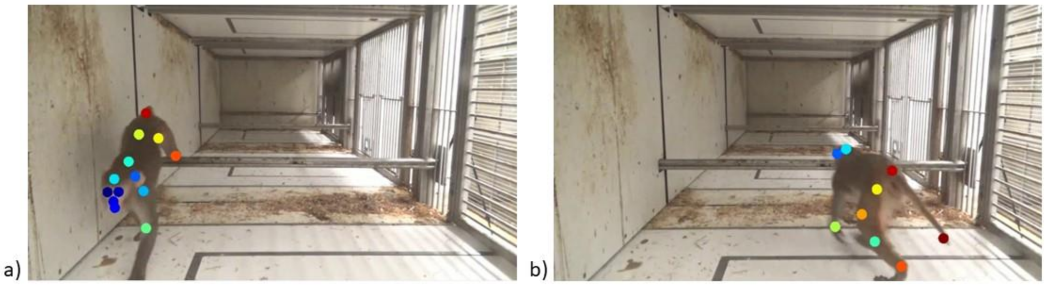


Figure 14: Examples of Tracking model's visual outputs. a) showing a monkey walking towards the side camera and b) showing a monkey walking away from the side camera.

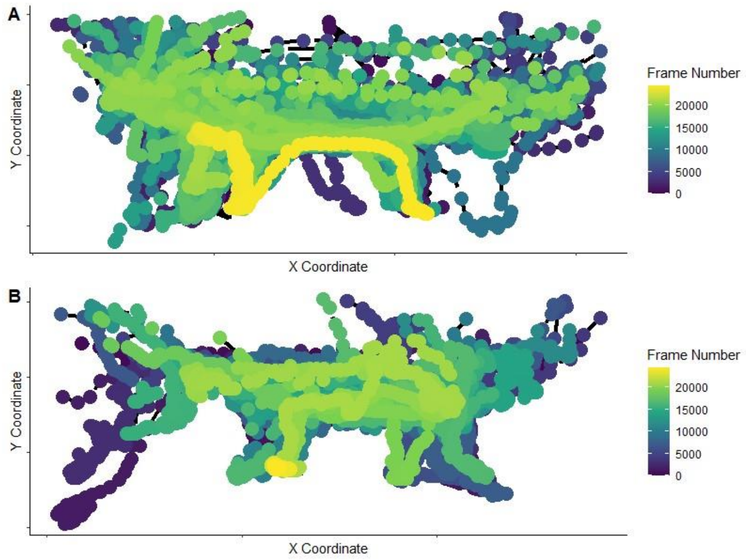


Figure 15: Example of two movement patterns extracted from the Tracking Model for two different individuals. The x and y axes represent the head coordinate extracted from the model. The darkest points represent the monkey position at the beginning of the video while the lightest ones represent its position towards the end of the recording.

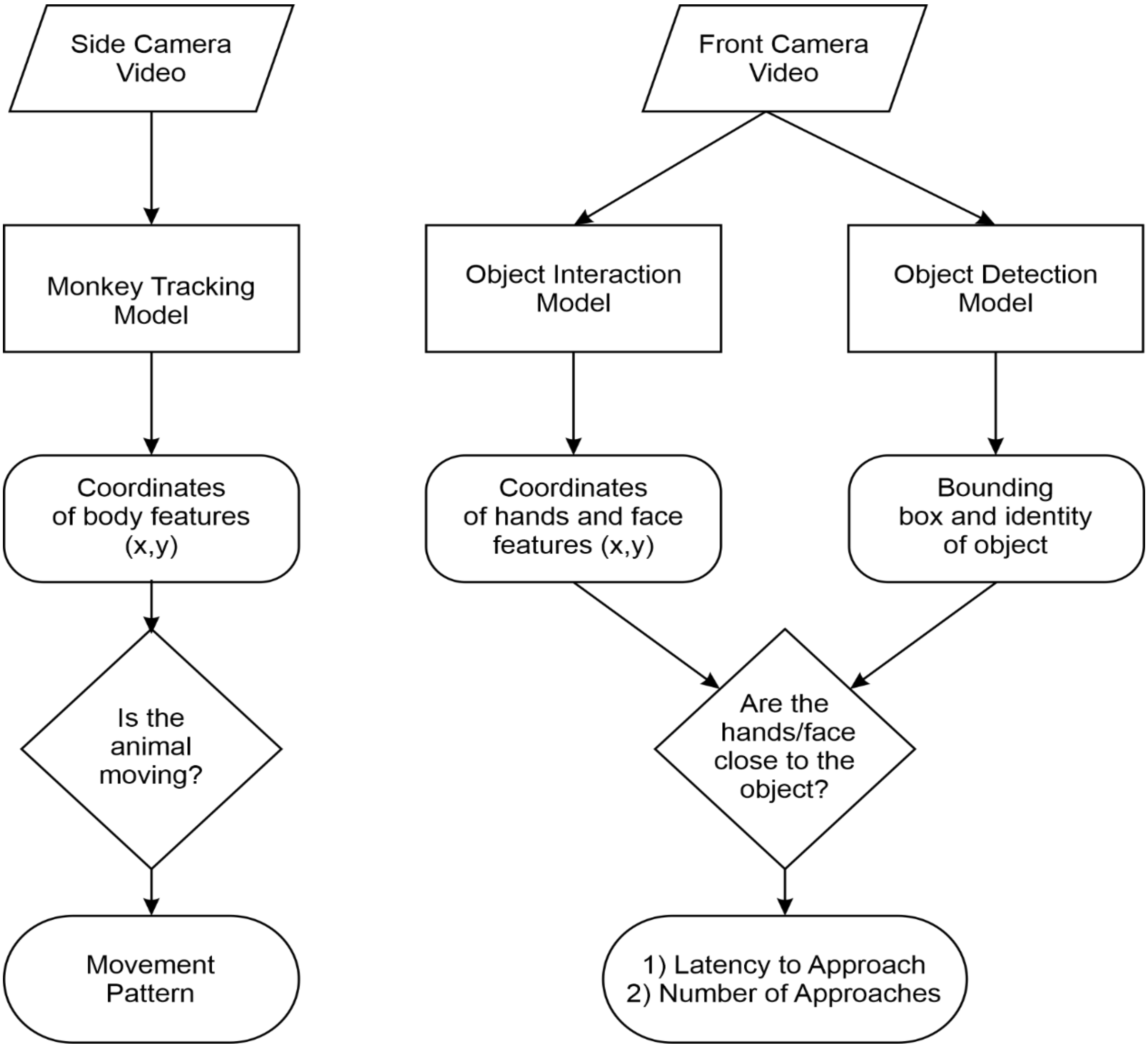


Figure 16: Diagram illustrating the pipeline for the analysis. The figure shows how each of the models were used to obtain the final behavioural variables coded.

3.3.3 Behaviour Extraction

Using our three models it has been possible to extract behavioural information used to assess temperament in rhesus macaques. In total, three behaviours were coded: number of approaches, latency to approach, amount of movement (Figure 16).

<u>Exploratory Behaviour</u>

This behaviour is defined as the amount of time that the individual spent close (approaching) to the objects. The quantification of approaches to distinct stimuli (FF, NF, NO1, and NO2) was achieved through the incorporation of the region of interest (ROI) derived from the Object Detection Model and the coordinates corresponding to body parts engaged in approaching novelty, as extracted from the Object Interaction Model. To ensure comprehensive coverage of all interaction types, the dimensions of the ROI were amplified tenfold for food-related stimuli and merely doubled for novel objects. This adjustment accounted for the divergent dimensions of the stimuli, facilitating the detection of approaches encompassing various interactions beyond mere physical contact, such as sniffing and close observation.

An approach event was registered when at least one body part, with a confidence score of 0.8 or higher, was detected within the expanded ROI. For characterization as an approach, the animal's body parts—nose, mouth, hands, and feet—needed to be situated within the ROI of the novelty stimulus. This encompassed a spectrum of behaviours including touching, inspecting, and licking. This behaviour offers a window into various dimensions of macaque temperament and neophobia, reflecting aspects such as exploratory tendencies, risk attitude, and impulsivity (Barr *et al.*, 2008; Santillán-Doherty *et al.*, 2010; Amici *et al.*, 2020).

The procedural framework employed for extracting the number of approaches is rooted in Python (version 3.9.13) and serves a multi-faceted purpose: (1) to refine the output from both the Object Detection Model and the Object Interaction Model, extracting pertinent information, like coordinates with confidence scores exceeding 0.8, (2) to determine instances where Object Interaction Model coordinates align with ROIs delineated by the Object Detection Model, and (3) to quantify the number of approach occurrences for each test phase.

Latency to First Approach

The computation of latency entailed measuring the time elapsed from the moment the item was introduced onto the wooden shelf to the instant when the animal's initial approach to it occurred. This data holds considerable significance within the realm of analyses focused on neophobia, novel object interest, and stress indicators studies (Kinnally *et al.*, 2008; Laudenslager *et al.*, 2011; Simpson *et al.*, 2019; Amici *et al.*, 2020). Leveraging the previously outlined framework, it becomes feasible to collect latency-related insights for each test condition. Notably, the calculation of latency to approach was confined to the first approach within each test phase, and it was computed by deducting the commencement of the respective phase from the time of interaction within that phase.

Amount of Movement

The Tracking Model was utilized to capture the time period during which macaques actively explored the cage environment. The evaluation of this behaviour requires a labour-intensive process when executed manually; however, comprehending the extent of a monkey's environmental exploration among novel circumstances or unfamiliar objects stands as imperative information for gauging neophobia, anxiety levels, and the broader response to novel stimuli (Pomeransky and Khriplovich, 1999; Rogers *et al.*, 2008; Gottlieb and Capitanio, 2013). To capture this behaviour, a Python-based pipeline was employed to analyse changes in coordinates over consecutive frames, allowing for the extraction and examination of movement patterns. This pipeline identified two distinct behaviours: movement and freeze. It accomplished this by tracking the average position of a macaque's face. In instances where the face was not visible (indicating that the monkey was not facing the camera), it used the average coordinates of other body parts. The pipeline then calculated the difference in these coordinates between two subsequent frames. Movements were recorded if the difference exceeded 4 pixels; otherwise, a freezing state was noted.

3.3.4 Behaviours Coded by the Human Observer

A manual coding of all the videos was performed by D. Massey using BORIS version 7.9.8 (Friard and Gamba, 2016). Behaviours were categorized as either states, indicating durations, or events, encompassing frequencies and latencies (Altmann, 1974). For a list and definition of the behaviour manually coded by the human observer see Table 6.

Behaviour	**Operational Definition**	**Outcome Variable**
Behind Visual Barrier	Individual is sat, moving, or climbing in the area of the cage behind the visual barrier. Freeze duration to be recorded separately.	Duration
Freeze	Individual is in a tense posture with ventrum pressed down or towards floor for more than 2 seconds.	Duration
Escape Attempts	Individual tries to squeeze one or more limb through the wooden slats; each attempt is separated by 2 seconds.	Frequency
Latency to Approach	The first time the individual initiates movement toward the direction of the novel object/food whilst looking in that direction.	Latency

Table 6: Operational definitions and outcome measures for behaviours coded by human observer.

The Latency to Approach served as a benchmark for validating the accuracy of the pipeline and the models. Given the time-intensive nature of quantifying both movement extent and the overall count of approaches through manual observation, these behaviours were not subjected to manual coding. On the other hand, behaviours like Escape Attempts and Behind Visual Barrier could not be reliably detected by the automated pipeline due to their intricate characteristics.

3.3.5 Statistical Analyses

Assessing reliability

To evaluate the consistency between two human coders, a subset of 30% of all videos was randomly selected and subjected to coding by an independent coder to gauge inter-observer reliability. The additional coder received comprehensive training on each behaviour outlined in the ethogram, alongside visual demonstrations from relevant videos. Importantly, the secondary coder was kept blind to both the identity of the monkeys and their weaning age. The assessment of reliability entailed the computation of Intraclass Correlation Coefficients (ICCs) using the ICC function within the 'psych' package in R (Revelle, 2020), employing an ICC (2,1) model that gauges the level of absolute agreement between coders. This model assumes that the measurement from a solitary coder serves as the foundational basis for measurement (Shrout and Fleiss, 1979; McGraw and Wong, 1996; Koo and Li, 2016). Notably, this reliability analysis was restricted to variables encompassed in the analytical framework.

Similarly, ICC (2,1) was computed to determine absolute agreement between a single human coder (D. Massey) and the automated pipeline for (1) latency to approach in each phase across all videos and (2) latency to approach in each phase specifically for videos where both the human and the pipeline identified an approach. Furthermore, to gauge the strength of the correlation between latency to approach as coded by the human and the pipeline, Pearson's correlation coefficient was calculated.

Principal Component Analysis

Principal Component Analysis (PCA) was employed to reduce the complexity of two datasets: (1) variables derived from human coding and (2) variables exclusively from the pipeline. For human-derived variables, 'behind visual barrier', 'escape attempts', and 'freeze' were aggregated across the three novel phases to create an overarching measure of neophobia. Meanwhile, latency variables for each phase were retained individually to account for potential variances in monkeys' responses to different food and objects. In the pipeline-derived dataset, 'movement' and 'freeze' were combined across all phases, akin to the human PCA approach, with phase-specific latency variables also included.

All variables slated for PCA underwent standardization, ensuring a mean of 0 and a standard deviation of 1. The optimal number of components was determined through parallel analysis utilizing the 'paran' package in R (Horn, 1965; Dinno and Dinno, 2018).

For human-derived data, parallel analysis disclosed the retention of a single component, rendering PCA rotation unnecessary. Conversely, pipeline-derived data revealed the need to preserve two components, necessitating the application of an orthogonal (varimax) rotation to uphold the independence of these components. PCA was carried out using the 'principal' function within the psych package (Revelle, 2020), followed by the extraction of differentially weighted component scores for subsequent analyses.

Model Fitting

Two distinct models were applied to the dataset, each having an individual principal component as its outcome. To assess the combined impact, a multiple regression approach was employed using the 'lm' function in R. The independent factors considered for analysis included Sex (categorical: female or male), Cohort (categorical: 2019, 2020, 2021, or 2022), and the duration of macaque habituation treatment (continuous). While the dependent variables where the two individual principal components. Subsequently, a range of assessments was conducted to ensure model robustness and credibility, encompassing diagnostics such as Cook's distance, DFBetas, DFFits, leverage, and Variance Inflation Factors. Additionally, checks were made for the distribution of residuals and the plotting of residuals against fitted values. Notably, none of these assessments revealed any conspicuous instances of influential cases or significant deviations from the assumptions of normality and homogeneity of residual (Quinn and Keough, 2002; Field, 2005).

3.4 Results

3.4.1 Agreement between Manual Observations and Model Estimations for Latency to First Approach

The pipeline demonstrated an overall detection accuracy of 88% for identifying initial approaches in all experimental phases compared to manual coding.

In the context of the Familiar Food phase (FF), the machine learning (ML)-based pipeline exhibited instances in which it inferred that certain monkeys had approached the stimuli, a conclusion that was not corroborated by the human observer's analyses. For the remaining experimental phases, the pipeline exhibited an underestimation of the number of monkey approaches when compared to human observations. In addition, it is noteworthy that there were occasions of concurrence between the human observer and the ML model (refer to Table 7 for more details).

Phase	**Human Observer**	**Pipeline**	**Agreement**
Familiar Food	100	118	93%
Novel Food	83	96	87%
Novel Object 1	75	56	64%
Novel Object 2	73	72	83%

Table 7: Agreement between the human and ML pipeline for the count of monkeys approaching, categorized by phase. The "Agreement" column within the table represents the percentage of individual macaques approaching concurrently detected by both the human observer and the pipeline.

Yet, when encompassing only all the monkeys that were independently recognized as approaching by both the human observer and the pipeline in the dataset, the level of complete agreement between the pipeline's output and the human observer fell under 50% for all phases except for Novel Object 2. Furthermore, the correlation coefficients exhibited a lower level of agreement for the familiar and novel food phases, while they demonstrated a moderate level of agreement for the novel object phases (*Table 8*). As a contrasting point of consideration, the inter-observer reliability between two human coders, when evaluating the latency to approach, demonstrated a high level of absolute agreement (ICC (2,1) = 0.9; 95% CI (lower, upper) = 0.87, 0.93).

	ICC			**Pearson's correlation**			
Phase	ICC (2,1)	CI (lower, upper)	*p*	*r*	df	CI (lower, upper)	*p*
Familiar Food	0.21	0.04,0.37	0.020	0.26	127	0.09,0.41	0.003
Novel Food	0.44	0.29,0.57	<0.0001	0.45	127	0.3,0.6	<0.0001
Novel Object 1	0.47	0.3,0.58	<0.0001	0.47	127	0.32,0.6	<0.0001
Novel Object 2	0.65	0.54,0.74	<0.0001	0.65	127	0.54,0.74	<0.0001

Table 8: Intraclass Correlation Coefficients (ICC) and reliability (r) for latency to first approach for full dataset.

In contrast, when analysing only the individuals identified as approaching by both the human observer and the pipeline in the different phases, the level of agreement was lower for the familiar and novel food phases, while it reached a moderate level for the novel object phases. Notably, the correlations between these subsets of data displayed a stronger connection compared to the entire dataset (Table 9 and Figure 17).

	ICC			**Pearson's correlation**			
Phase	ICC (2,1)	CI (lower, upper)	*p*	*r*	df	CI (lower, upper)	*p*
Familiar Food	0.11	-0.6,0.3	0.09	0.16	91	-0.04,0.35	0.12
Novel Food	0.52	0.33,0.67	<0.0001	0.53	71	0.34,0.68	<0.0001
Novel Object 1	0.78	0.64,0.87	<0.0001	0.78	46	0.65,0.87	<0.0001
Novel Object 2	0.66	0.5,0.78	<0.0001	0.67	59	0.51,0.79	<0.0001

Table 9: Intraclass Correlation Coefficients (ICC) and reliability (r) for latency to first approach within the subset of data where concordance exists between human observer and ML pipeline in identifying approaching monkeys.

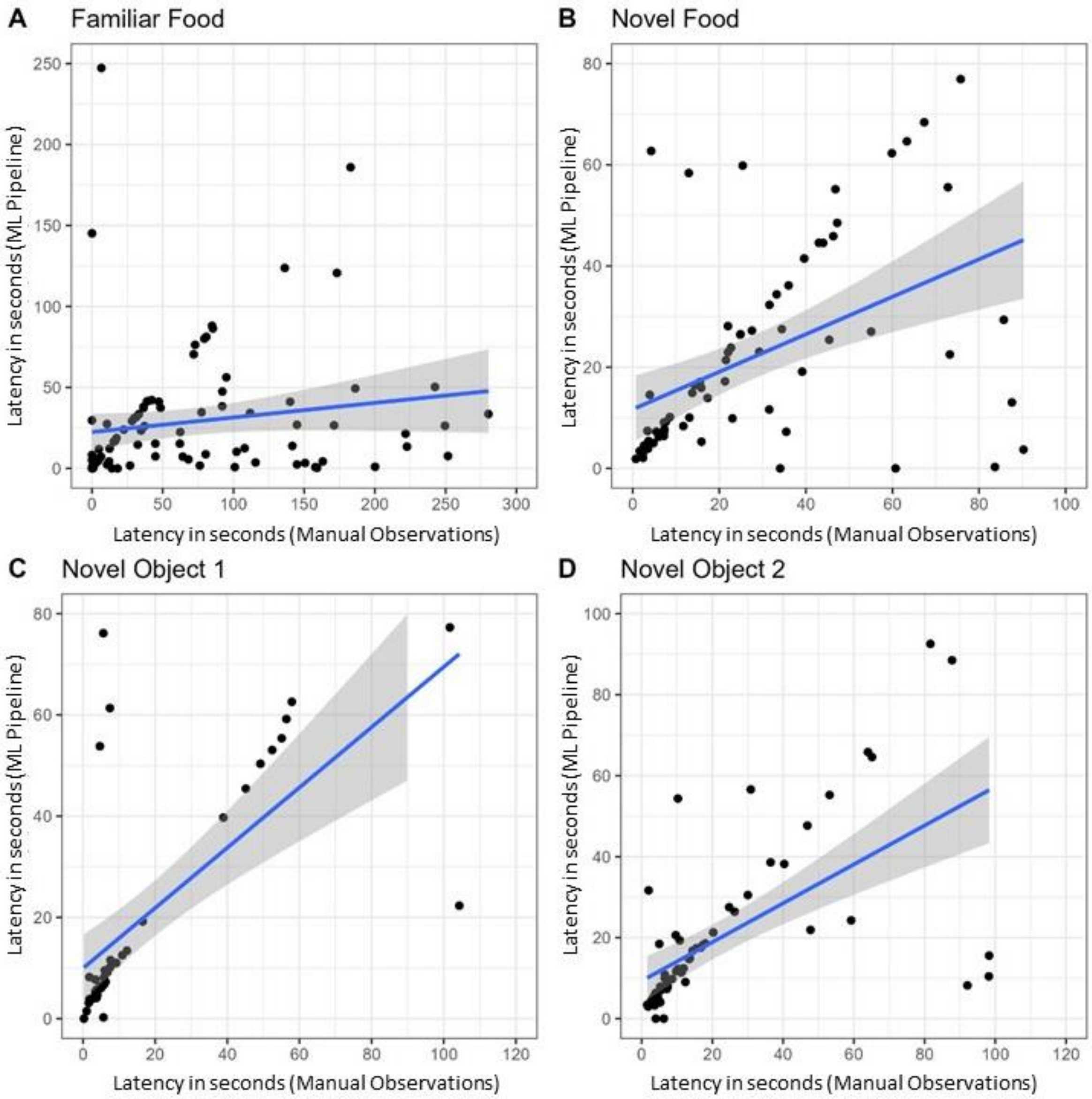


Figure 17: Scatter plots showing the correlation between human and ML pipeline latency to first approach in each phase: A) Familiar Food; B) Novel Food; C) Novel Object 1; D) Novel Object 2. All latency variables are measured in seconds.

3.4.2 Principal Component Analysis

Machine Learning Pipeline

Three components with eigenvalues exceeding 1 were identified, and parallel analysis for component retention indicated that one component exhibited a magnitude greater than what would typically be anticipated from random data (as determined by 300 iterations) at the 95th percentile. This specific component demonstrated positive salient loadings for latency to the first approach within each phase of the temperament test, indicating longer latency periods. In addition, it had negative loadings for exploratory behaviour and time spent moving, along with a positive salient loading for time spent still. As a result, this component was labelled as 'Neophobia_ML' (Table 10). Notably, 'Neophobia_ML' accounted for 38% of the total variance observed across the individual variables.

Variable	**Loading 'Neophobia_ML'**	***h2***
Latency (FF)	0.63	0.4
Latency (NF)	0.64	0.42
Latency (NO1)	0.67	0.45
Latency (NO2)	0.63	0.4
Movement	-0.69	0.47
Time Spent Still	0.61	0.37
Exploratory Behaviour	-0.42	0.18

Variable	**Loading 'Neophobia_ML'**	***h2***
Latency (FF)	0.63	0.4
Latency (NF)	0.64	0.42
Latency (NO1)	0.67	0.45
Latency (NO2)	0.63	0.4
Movement	-0.69	0.47

Time Spent Still	0.61	0.37
Exploratory Behaviour	-0.42	0.18

Table 10: Loadings and Communality (denoted by h2) of the first principal component from the ML pipeline output.

Human Observer Coding

Without the variables extracted from the pipeline, the analysis unveiled the presence of three components with eigenvalues exceeding the value of 1. Subsequently, conducting parallel analysis for component retention revealed a solitary component that surpassed the expected magnitude derived from random data (based on 210 iterations) at the 95th percentile. This particular component, defined as Neophobia_HO, exhibited positive salient loadings for latency to approach across all phases, as well as freeze behaviour (as presented in Table 11), thereby elucidating 30% of the total variance encompassing individual variables. However, including the data obtained from the ML pipeline, it was possible to explicate a more substantial proportion of the variance within the principal component.

Variable	**Loading Neophobia_HO**	***h2***
Latency (FF)	0.63	0.39
Latency (NF)	0.68	0.46
Latency (NO1)	0.62	0.39
Latency (NO2)	0.71	0.49
Behind Visual Barrier	0.15	0.02
Freezing Behaviour	0.51	0.27
Escape Attempts	-0.28	0.08

Table 11: Loadings of the variables into the Principal Component Analyses and Communality (denoted by h2). Variables from human coding.

3.4.3 Cohort Differences Due to Human Exposure

Two linear models were fit to each principal component: one derived from the machine learning pipeline (Neophobia_ML) and one from the manual human observations (Neophobia_HO), see **Error! Reference source not found.**.

Differences arose in the models assessing the effects of sex, cohort, and routine habituation practices days on neophobia. The Neophobia_ML model, which relied on data from the automated pipeline, indicated that none of the predictor variables had a discernible influence on the variation in neophobia scores (as illustrated in Table 12 and Figure 18).

In contrast, the Neophobia_HO model, derived from human observations, found significant effects of cohort and a marginally non-significant effect of routine habituation practices (treatment) days on neophobia scores. Specifically, the Neophobia_HO model estimated that cohorts from 2020 and 2022 exhibited approximately half a standard deviation higher levels of neophobia compared to the 2019 cohort, while no significant differences were observed among other cohorts (refer to Table 13 and Figure 18**Error! Reference source not found.**). Additionally, it suggested a decrease in neophobia by 0.02 standard deviations for each additional day of routine habituation practices (treatment).

Despite the absence of statistical significance in the results obtained from the ML pipeline, it is noteworthy that the estimated trends aligned with those derived from human observations (as detailed in**Error! Reference source not found.** and Table 13).

Neophobia_ML						
	Intercept	**Sex**	**Cohort (2020)**	**Cohort (2021)**	**Cohort (2022)**	**Treatment**
CI(Upper)	0.28	0.6	0.84	0.62	0.71	0.004
CI(Lower)	-0.61	-0.13	-0.13	-0.41	-0.35	-0.04
p	0.46	0.2	0.15	0.7	0.5	0.1
t	-0.73	1.23	0.14	0.41	0.7	-1.6
SE	0.22	0.179	0.24	0.26	0.3	0.01
Est	-0.16	0.22	0.35	0.1	0.18	-0.02

Table 12: Coefficients for the ML pipeline model fits to each principal component.

Neophobia_HO						
	Intercept	**Sex**	**Cohort (2020)**	**Cohort (2021)**	**Cohort (2022)**	**Treatment**
CI(Upper)	0.273	0.344	0.985	0.678	1.027	0.0003
CI(Lower)	-0.572	-0.346	0.059	-0.313	0.005	-0.041
p	0.485	0.996	**0.027**	0.468	**0.048**	**0.053**
t	-0.701	-0.005	2.231	0.728	1.998	-1.951
SE	0.213	0.174	0.234	0.250	0.258	0.010
Est	-0.150	-0.001	0.522	0.182	0.516	-0.020

Table 13: Coefficients for the Human Observer model fits to each principal component.

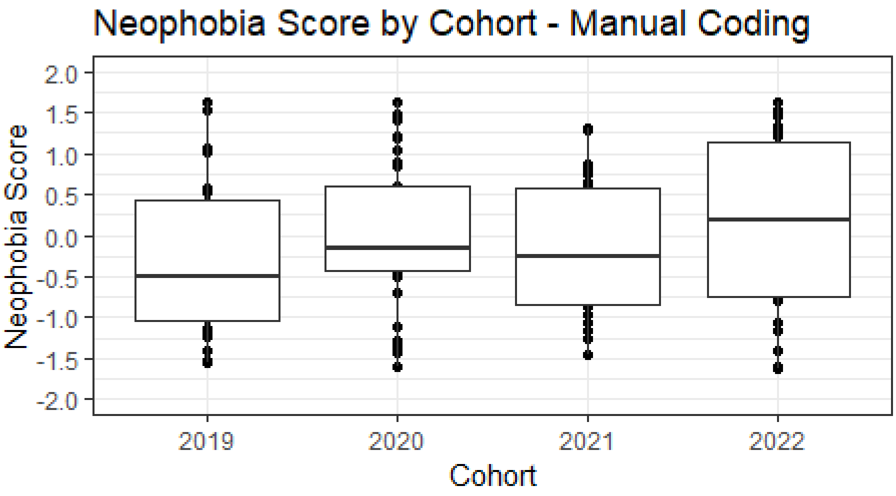


Figure 18: Boxplots displaying neophobia scores by cohort for the model deriving from human observations. Dots represent individual data points. Boxes represent the lower and upper quartiles. Solid horizontal lines represent the median. Solid vertical lines represent the range of the data. Dots that are disconnected from the solid vertical line represent outliers that are 1.5 times the inter-quartile-range larger or smaller than the upper or lower quartiles, respectively. Dashed horizontal line represents mean neophobia score at zero.

.

3.5 Discussion

Using three computer vision models, I was able to show how it could be possible to collect behaviours useful to assess neophobia and temperament using automated methodologies.

The machine learning based pipeline exhibited an overall accuracy of 88%. Nevertheless, there were notable challenges, particularly in the familiar food phase, where a considerable number of false positives were detected. Across the other phases, the pipeline performed relatively better but still not within an acceptable range when assessed using the Interclass Correlation Coefficient. The potential for enhancement in accuracy is evident, primarily through the acquisition of higher-quality video data (Aqqa, Mantini and Shah, 2019; Bergstrom and Messinger, 2023). Notably, the accuracy of detecting the latency to approach in this study is compromised due to the pre-recorded nature of the videos, which were not initially intended for automating temperament tests. Variations in camera resolution, position, and angles across different tests may have impacted the accuracy of the models trained on the front camera videos, leading to the discrepancies from human observer identifications. To enhance accuracy, future studies should focus on synchronizing the two cameras to provide a precise 3D layout of the cage room, minimizing false positives, and ensuring the actual presence of the monkey near the novelty, rather than just passing by in the background. Additionally, refining bounding box overlap assessments from both camera angles would offer a more comprehensive understanding of the approach. These goals are attainable with adjusted camera settings for studies specifically dedicated to automating approach detection.

Using videos from the side camera, it was possible to train the Tracking Model and extract data on movement and freezing behaviours that would have been impractical for a human observer to collect. Specifically, I tracked the animals' movements around the cage and extracted movement patterns for each individual, in every different test condition. While manual analysis during the temperament test would have been too time-consuming, both freezing and movement are important indicators in evaluating temperament, fear responses, and neophobia in primates. Movement tends to be linked with a gentle temperament, indicating a greater willingness to engage with novelty (Gottlieb, Capitanio and McCowan, 2013b). Conversely, freezing is a recognized response associated with fear and anxiety (Kalin and Shelton, 1989; Bethell *et al.*, 2019). Furthermore, the automated pipeline allowed me to extract the exploratory behaviour, which, too, would have been impractical to manually code. This behaviour is commonly used in the context of evaluating temperament in macaques, where individuals who allocate more time to exploring novel objects are typically classified as more exploratory and less fearful than their counterparts (Coleman, Tully and McMillan, 2005).

These findings shed light on the behaviour of individuals with heightened levels of neophobia, indicating that they tend to engage in less movement and exploration while exhibiting a greater propensity for freezing. These outcomes align with existing literature on neophobia and fearfulness in macaques and other species.

In the case of the study investigating neophobia differences among various cohorts of macaques before, during, and after the COVID-19 pandemic, both the automated pipeline and manual coding identified an increase in neophobia among cohorts born and raised during and after the pandemic, as compared to the 2019 cohort. However, the data extracted from the pipeline, while demonstrating a similar trend to that observed by the human coder, failed to yield statistically significant results. This discrepancy can be attributed to the pipeline's low accuracy during the food phases, which was exacerbated by video data not being optimally suited for training machine learning models.

This study, however, lays a foundational framework, demonstrating the viability of automatically detecting macaques approaching novelty in various phases, providing valuable insights into the steps needed to achieve this. In addition, it highlights the impact of human interaction on reducing neophobia in macaques and show the significance of exposing these animals at the CFM, as well as in similar facilities, to human contact.

3.6 Chapter References

Altmann, J. (1974) ‘Observational Study of Behavior: Sampling Methods’, *Behaviour*, 49(3–4), pp. 227–266. doi:10.1163/156853974X00534.

Amici, F. *et al.* (2020) ‘Innovation in wild Barbary macaques (Macaca sylvanus)’, *Scientific Reports*, 10(1), p. 4597. doi:10.1038/s41598-020-61558-2.

Aqqa, M., Mantini, P. and Shah, S.K. (2019) ‘Understanding How Video Quality Affects Object Detection Algorithms.’, in *VISIGRAPP (5: VISAPP)*, pp. 96–104.

Barnard, S. *et al.* (2016) ‘Quick, Accurate, Smart: 3D Computer Vision Technology Helps Assessing Confined Animals’ Behaviour’, *PLOS ONE*. Edited by C. Wade, 11(7), p. e0158748. doi:10.1371/journal.pone.0158748.

Barr, C.S. *et al.* (2008) ‘CRH Haplotype as a Factor Influencing Cerebrospinal Fluid Levels of Corticotropin-Releasing Hormone, Hypothalamic-Pituitary-Adrenal Axis Activity, Temperament, and Alcohol Consumption in Rhesus Macaques’, *Archives of General Psychiatry*, 65(8), p. 934. doi:10.1001/archpsyc.65.8.934.

Bergstrom, A.C. and Messinger, D.W. (2023) ‘Image quality and object detection performance of convolutional neural networks’, in *Pattern Recognition and Tracking XXXIV*. SPIE, pp. 159–177.

Bethell, E.J. *et al.* (2019) ‘Toward a Standardized Test of Fearful Temperament in Primates: A Sensitive Alternative to the Human Intruder Task for Laboratory-Housed Rhesus Macaques (Macaca mulatta)’, *Frontiers in Psychology*, 10. doi:10.3389/fpsyg.2019.01051.

Blumrosen, G., Hawellek, D. and Pesaran, B. (2017) ‘Towards automated recognition of facial expressions in animal models’, *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2018-Janua, pp. 2810–2819. doi:10.1109/ICCVW.2017.332.

Bolya, D. *et al.* (2019) 'Yolact: Real-time instance segmentation', in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166.

Button, K.S. *et al.* (2013) 'Power failure: why small sample size undermines the reliability of neuroscience', *Nature Reviews Neuroscience*, 14(5), pp. 365–376. doi:10.1038/nrn3475.

Capitanio, J.P., Mendoza, S.P. and Cole, S.W. (2011) 'Nervous temperament in infant monkeys is associated with reduced sensitivity of leukocytes to cortisol's influence on trafficking', *Brain, Behavior, and Immunity*, 25(1), pp. 151–159. doi:10.1016/j.bbi.2010.09.008.

Coleman, K. *et al.* (2012) 'Behavioral Management, Enrichment, and Psychological Well-being of Laboratory Nonhuman Primates', in *Nonhuman Primates in Biomedical Research*. Elsevier, pp. 149–176. doi:10.1016/B978-0-12-381365-7.00006-6.

Coleman, K. and Pierre, P.J. (2014) 'Assessing Anxiety in Nonhuman Primates', *ILAR Journal*, 55(2), pp. 333–346. doi:10.1093/ilar/ilu019.

Coleman, K. and Schapiro, S.J. (2021) *Behavioral biology of laboratory animals*. CRC Press.

Coleman, K., Tully, L.A. and McMillan, J.L. (2005) 'Temperament correlates with training success in adult rhesus macaques', *American Journal of Primatology*, 65(1), pp. 63–71. doi:10.1002/ajp.20097.

Damerius, L.A. *et al.* (2017) 'Curiosity boosts orang-utan problem-solving ability', *Animal Behaviour*, 134, pp. 57–70.

Dinno, A. and Dinno, M.A. (2018) 'Package "paran"', *Dortmund, Germany: R package version*, 1(2).

Doelling, C.R. *et al.* (2021) 'The relationship between personality, season, and wounding receipt in zoo-housed Japanese macaques (Macaca fuscata): A multi-institutional study', *American Journal of Primatology*, 83(12), pp. 1–10. doi:10.1002/ajp.23332.

Field, A.P. (2005) 'Is the meta-analysis of correlation coefficients accurate when population correlations vary?', *Psychological methods*, 10(4), p. 444.

Forss, S.I.F. *et al.* (2022) 'Captivity and habituation to humans raise curiosity in vervet monkeys', *Animal Cognition*, 25(3), pp. 671–682. doi:10.1007/s10071-021-01589-y.

Fox, A.S. *et al.* (2021) 'Infant inhibited temperament in primates predicts adult behavior, is heritable, and is associated with anxiety-relevant genetic variation', *Molecular Psychiatry*, 26(11), pp. 6609–6618. doi:10.1038/s41380-021-01156-4.

Freeman, H.D. and Gosling, S.D. (2010) 'Personality in nonhuman primates: A review and evaluation of past research', *American Journal of Primatology*, 72(8), pp. 653–671. doi:10.1002/ajp.20833.

Friard, O. and Gamba, M. (2016) ‘BORIS : a free, versatile open-source event-logging software for video/audio coding and live observations’, *Methods in Ecology and Evolution*. Edited by R. Fitzjohn, 7(11), pp. 1325–1330. doi:10.1111/2041-210X.12584.

Gottlieb, D.H. *et al.* (2019) ‘Personality, environmental stressors, and diarrhea in rhesus macaques: an interactionist perspective’, *American Journal of Primatology*, 80(12), pp. 1–27. doi:10.1002/ajp.22908.Personality.

Gottlieb, D.H. and Capitanio, J.P. (2013) ‘Latent Variables Affecting Behavioral Response to the Human Intruder Test in Infant Rhesus Macaques (Macaca mulatta)’, *American Journal of Primatology*, 75(4), pp. 314–323. doi:10.1002/ajp.22107.

Gottlieb, D.H., Capitanio, J.P. and McCowan, B. (2013) ‘Risk factors for stereotypic behavior and self-biting in rhesus macaques (Macaca mulatta): Animal’s history, current environment, and personality’, *American Journal of Primatology*, 75(10), pp. 995–1008. doi:10.1002/ajp.22161.

Gottlieb, D.H., Coleman, K. and McCowan, B. (2013) ‘The effects of predictability in daily husbandry routines on captive rhesus macaques (Macaca mulatta)’, *Applied Animal Behaviour Science*, 143(2–4), pp. 117–127. doi:10.1016/j.applanim.2012.10.010.

Herrington, J.A., Del Rosso, L.A. and Capitanio, J.P. (2016) ‘Biobehavioral consequences of prenatal exposure to a matrilineal overthrow and relocation in captive infant rhesus (Macaca mulatta) monkeys’, *American journal of primatology*, 78(9), pp. 895–903. doi:10.1002/ajp.22557.

Horn, J.L. (1965) ‘A rationale and test for the number of factors in factor analysis’, *Psychometrika*, 30, pp. 179–185.

Kalin, N.H. and Shelton, S.E. (1989) ‘Defensive behaviors in infant rhesus monkeys: environmental cues and neurochemical regulation’, *Science*, 243(4899), pp. 1718–1721.

Kennedy, A. (2022) ‘The what, how, and why of naturalistic behavior’, *Current Opinion in Neurobiology*, 74, p. 102549. doi:10.1016/j.conb.2022.102549.

Kinnally, E.L. *et al.* (2008) ‘Dimensions of response to novelty are associated with social engagement and aggression in adult male rhesus macaques (Macaca mulatta).’, *Journal of Comparative Psychology*, 122(2), pp. 195–203. doi:10.1037/0735-7036.122.2.195.

Koo, T.K. and Li, M.Y. (2016) ‘A guideline of selecting and reporting intraclass correlation coefficients for reliability research’, *Journal of chiropractic medicine*, 15(2), pp. 155–163.

Laudenslager, M.L. *et al.* (2011) ‘A novelty seeking phenotype is related to chronic hypothalamic-pituitary-adrenal activity reflected by hair cortisol’, *Physiology & Behavior*, 104(2), pp. 291–295. doi:10.1016/j.physbeh.2011.03.003.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) ‘Deep learning’, *Nature*, 521(7553), pp. 436–444.

doi:10.1038/nature14539.

Mathis, A. *et al.* (2018) ‘DeepLabCut: markerless pose estimation of user-defined body parts with deep learning’, *Nature Neuroscience*, 21(9), pp. 1281–1289. doi:10.1038/s41593-018-0209-y.

McGraw, K.O. and Wong, S.P. (1996) ‘Forming inferences about some intraclass correlation coefficients.’, *Psychological methods*, 1(1), p. 30.

Pomeransky, A.A. and Khriplovich, I.B. (1999) ‘Equations of motion of spinning relativistic particle in external fields’, *Surveys in High Energy Physics*, pp. 145–173. doi:10.1080/01422419908228843.

Porto, S.M.C. *et al.* (2013) ‘A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns’, *Biosystems Engineering*, 115(2), pp. 184–194. doi:10.1016/j.biosystemseng.2013.03.002.

Quinn, G.P. and Keough, M.J. (2002) *Experimental design and data analysis for biologists*. Cambridge university press.

Revelle, W. (2020) ‘How to: Use the psych package for factor analysis and data reduction’, *Northwestern University, Department of Psychology: Evanston, IL, USA* [Preprint].

Rogers, J. *et al.* (2008) ‘Genetic influences on behavioral inhibition and anxiety in juvenile rhesus macaques’, *Genes, Brain and Behavior*, 7(4), pp. 463–469. doi:10.1111/j.1601-183X.2007.00381.x.

Santillán-Doherty, A.M. *et al.* (2010) ‘Novelty-seeking temperament in captive stumptail macaques (macaca arctoides) and spider monkeys (ateles geoffroyi).’, *Journal of Comparative Psychology*, 124(2), pp. 211–218. doi:10.1037/a0018267.

Schapiro, S.J. and Hau, J. (2023) ‘Research Benefits of Improving Welfare in Captive Primates’, in *Nonhuman Primate Welfare: From History, Science, and Ethics to Practice*. Springer, pp. 445–462.

Schmidt, L.A. and Poole, K.L. (2020) *Adaptive Shyness: Multiple Perspectives on Behavior and Development*, *Adaptive Shyness: Multiple Perspectives on Behavior and Development*. doi:10.1007/978-3-030-38877-5.

Shrout, P.E. and Fleiss, J.L. (1979) ‘Intraclass correlations: uses in assessing rater reliability.’, *Psychological bulletin*, 86(2), p. 420.

Simpson, E.A. *et al.* (2019) ‘Handling newborn monkeys alters later exploratory, cognitive, and social behaviors’, *Developmental Cognitive Neuroscience*, 35, pp. 12–19. doi:10.1016/j.dcn.2017.07.010.

Stamps, J. and Groothuis, T.G.G. (2010) ‘The development of animal personality: Relevance, concepts and perspectives', *Biological Reviews*, 85(2), pp. 301–325. doi:10.1111/j.1469-185X.2009.00103.x.

Timmermans, P.J.A. and Vossen, J.M.H. (1996) 'The influence of rearing conditions on maternal behavior in cynomolgus macaques (Macaca fascicularis)', *International journal of primatology*, 17, pp. 259–276.

Vandeleest, J.J., McCowan, B. and Capitanio, J.P. (2011) 'Early rearing interacts with temperament and housing to influence the risk for motor stereotypy in rhesus monkeys (Macaca mulatta)', *Applied Animal Behaviour Science*, 132(1–2), pp. 81–89. doi:10.1016/j.applanim.2011.02.010.

Weinstein, B.G. (2018) 'A computer vision for animal ecology', *Journal of Animal Ecology*. Edited by L. Prugh, 87(3), pp. 533–545. doi:10.1111/1365-2656.12780.

Weinstein, T.A.R. and Capitanio, J.P. (2008) 'Individual differences in infant temperament predict social relationships of yearling rhesus monkeys, Macaca mulatta', *Animal Behaviour*, 76(2), pp. 455–465. doi:10.1016/j.anbehav.2008.01.024.

Weinstein, T.A.R. and Capitanio, J.P. (2012) 'Longitudinal stability of friendships in Rhesus monkeys (Macaca mulatta): Individual-and relationship-level effects', *Journal of Comparative Psychology*, 126(1), pp. 97–108. doi:10.1037/a0025607.

Chapter 4. Unravelling Macaque Foraging Behaviour: An Object Detection Approach

4.1 Abstract

In the context of captive facilities hosting animals, ensuring optimal nutrient intake while mitigating issues such as obesity is paramount. Notably, obesity poses a prevalent concern among captive Non-Human Primates (NHP), often surpassing their wild counterparts in weight. The nature of the diet and its method of delivery also significantly impact aggression and competitive behaviour, particularly among rhesus macaques, recognized for their competition for resources in both captivity and the wild.

Furthermore, foraging in natural habitats exhibits profound seasonal variations that can influence food choices and the effort required for food procurement. Therefore, gathering comprehensive information on foraging behaviour in different situations becomes essential for maintaining high welfare standards in captivity. However, collecting such data is time-consuming, given that foraging behaviours can span hours and involve the entire group simultaneously.

This study explores the effects of foraging mixtures, pellet sizes, and seasonal fluctuations on captive rhesus macaques at the Centre for Macaques (CFM). To streamline data collection and alleviate the human workload, automated techniques were employed to extract information from video recordings of macaques foraging at CFM. Subsequently, this data was analysed using a computer vision-based pipeline. The YOLACT object detection algorithm was utilized to identify foraging macaques across 15 distinct enclosures, successfully detecting their activity without requiring human observation. Thanks to the application of these automated methodologies, grounded in deep learning, it was possible to yield valuable insights that not only enhance animal management but also inform dietary decisions effectively. Through this approach, I discovered that macaques exhibit increased foraging behaviour when presented with chopped food, observed changes in foraging levels in response to alterations in food size, and discerned distinct foraging patterns across varying seasons.

4.2 Introduction

Macaques typically allocate over 30% of their time to foraging activities (Menon and Poirier, 1996; Hanya, 2004b; Hoshino *et al.*, 2022). Foraging not only provides essential nutrients and sustains a healthy body condition but also mitigates issues like boredom and abnormal behaviours commonly observed in captive animals (Novak *et al.*, 1998; Gottlieb, Maier and Coleman, 2015). Engaging in foraging activities involves exploring the surroundings, moving around, and manipulating food and objects, all of which are intrinsic to macaque species-typical behaviour (Pyke, Pulliam and Charnov, 1977; Agetsuma, 1995). In their natural habitat, macaques predominantly consume leaves, fruits, and buds, adhering to a high fibre, low-fat diet (Yeager, 1996; Hill, 1997; Hanya, 2004a). Replicating such a diet in captivity is crucial to prevent overweight-related health problems (Zijlmans *et al.*, 2021). While many primate facilities provide various foraging mixes and a diverse diet, different foods require different foraging strategies, influencing the animals' behaviour, time allocation, and overall well-being (Leigh, 1994; Agetsuma, 1995; Schwitzer and Kaumanns, 2001).

At the Centre for Macaques (CFM), approximately 200 rhesus macaques (*Macaca mulatta*) are housed and bred. This macaque species exhibits more frequent and severe aggression compared to other macaque species, both in the wild and in captivity (Lutz, Well and Novak, 2003; Thierry, Singh and Kaumanns, 2004; Corrine K. Lutz and Novak, 2005; Lutz *et al.*, 2007a; Beisner and Isbell, 2011b; Gottlieb, Capitanio and McCowan, 2013a). In a captive setting, aggression is more prevalent and escalates more rapidly (Thierry, Singh and Kaumanns, 2004; McCowan *et al.*, 2008; Beisner and Isbell, 2011b). Resource competition is a common cause of aggression, as it limits the access of lower-ranking individuals to food (Southwick, 1967; Mathy and Isbell, 2002; Chancellor and Isbell, 2008). In order to address this issue, the CFM has implemented an intervention by increasing the size of food pellets. This modification allows lower-ranking individuals to obtain larger chunks of food, enabling them to eat in quieter areas of the enclosure away from more dominant animals. Although this change in food size benefits subordinate and younger individuals, it may also influence the time spent foraging, as larger food is easier to locate.

Rhesus macaques at CFM exhibit seasonal breeding patterns, typically breeding between October and January, with births occurring from April to July after a gestation period of 160 to 175 days. During the breeding season, male macaques compete for mating opportunities, and energetically demanding behaviours like consortship can enhance reproductive success (Higham, Heistermann and Maestripieri, 2011). In addition, high-ranking males allocate more time to feeding prior the mating season, compared to low-ranking males (Higham, Heistermann and Maestripieri, 2011). Consequently, by the end of the mating season, high-ranking males are often in poorer physical condition. In female rhesus macaques, body size and higher physical condition during mating season influence the likelihood of first conception, and smaller females generally have lower reproductive success (Bercovitch *et al.*, 1999). Moreover, during the lactation period, mothers of other macaque species adopt an energy-conserving strategy, trading increased resting time for reduced feeding time (Marlies Heesen, 2014). Many studies report as, in both males and females, the time spent foraging is strongly influenced by the quantity and quality of available food in the wild (Agetsuma, 1995; Hill, 1997; Jaman, Huffman and Takemoto, 2010). However, the impact of seasons, such as the breeding and birth seasons, on the foraging patterns of captive rhesus macaques remains understudied due to the constant and reliable availability of appropriate food throughout the year. Furthermore, collecting foraging data is a time-intensive process, primarily because this behaviour can persist for hours each day and typically involves the majority of animals within the group simultaneously.

This study aims to investigate how various foraging mixtures and varying pellet sizes impact the foraging behaviour of captive rhesus macaques. Specifically, I seek to determine which foraging mixtures elicit more foraging activity, such as those containing smaller food portions or more enticing food items. Additionally, I aim to assess if the presence of larger pellets significantly prolongs foraging time within the group. Furthermore, I aim to examine the potential influence of different seasons on the foraging patterns of rhesus macaques at CFM, shedding light on this relatively understudied topic.

In addition to foraging data, information on the occurrence of injuries in the group and the weight of the animals were collected. The injury rates were used to explore whether increased foraging behaviour correlates with a lower occurrence of aggression. It is well-established that providing more opportunities for foraging can decrease competition and aggression in macaque groups (van Schaik *et al.*, 1983; Saito Chiemi, 1996; Steenbeek and

Sterck, 1997; Beisner and Isbell, 2011a; Grove, 2012). Weight data were employed to examine the potential correlation between increased foraging behaviour and higher body weight, with the potential of assessing the issue of obesity, a prevalent concern among captive NHP populations and closely tied to individual dietary habits (Schwitzer and Kaumanns, 2001; Bauer *et al.*, 2011; Pontzer and Pontzer, 2023).

To minimize time and cost associated with data collection, automated methodologies were employed in this study. Recent advancements in machine learning have facilitated the development of tools capable of detecting and identifying objects in videos. Leveraging this technology, we were able to collect consistent and reliable data in a faster and non-invasive manner (Rushen, Chapinal and de Passillé, 2012; Porto *et al.*, 2013; Aroeira *et al.*, 2016; Okinda *et al.*, 2020; Chen, Zhu and Norton, 2021).

In this project, the YOLACT object detection algorithm was employed to identify macaques across 15 diverse enclosures. Due to the intricacies of enclosure layouts and camera positioning, directly detecting foraging behaviour proved challenging. Nevertheless, since at the CFM macaques are fed with food dispersed onto the ground, a region of interest encompassing the enclosure floor was designated to recognize a macaque's foraging activity (refer to Materials and Methods for comprehensive information).

4.3 Materials and Methods

4.3.1 Facility and subjects

At CFM, rhesus macaques were socially housed in breeding groups consisting of one adult male, multiple females, and their offspring. After weaning, the young monkeys were relocated to same-sex peer groups of similar age. The study groups, comprising up to 27 individuals, resided in enclosures with a playpen area covered in sawdust, cleaned every two weeks. The monkeys were provided with a daily diet consisting of commercial food pellets, seed and lentil mix, occasional raisins and nuts during training sessions, and 7 different forage mixes (

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Mushroom and Radish	Leek and Tomato	Red Cabbage and Courgette	Eggs and Sweet Potato	Parsnip and Pepper	Spinach/ Watercress	Orange

Table 14). The staff prepared the vegetables by chopping them, while oranges and tomatoes were provided to the monkeys in their whole form. On the 29th of September 2022, the CFM started providing larger pellets to the macaques. The food was distributed on the ground by technicians from the front to the middle of the playpen each morning (Figure 19).

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Mushroom and Radish 	Leek and Tomato 	Red Cabbage and Courgette 	Eggs and Sweet Potato 	Parsnip and Pepper 	Spinach/ Watercress 	Orange 

Table 14: List of foraging mixes for each day of the week.

4.3.2 Data Collected

For this study, video recordings were captured at a frame rate of 15 frames per second using Axis P1435-LE CCTV cameras, which were installed individually in each play pen. The recorded videos were then analysed on a standard commercial laptop (XPS-15 with a Nvidia Geforce GTX 1650 Ti card). I utilized three different datasets to investigate (1) the effect of different foraging mixes on foraging behaviour, (2) the effect of pellet size on foraging behaviour, and (3) the effect of seasons on foraging behaviour.

1. Foraging Mixes study

The first dataset included videos of 15 macaque groups, comprising 9 breeding groups (BG) and 6 juvenile groups (JG), with a mean group size of 10 (SD = 2.7). Each group was recorded for 3 weeks after feeding time, for 2 hours per day. The total data collected for this study amounted to 588 hours, and the analysis was conducted over a period of 5 days using an automated pipeline.

2. *Pellet Size study*

The second dataset consisted of videos of 14 macaque groups, comprising of 9 breeding groups (BG) and 5 juvenile groups (JG), with a mean group size of 10 (SD = 2.8). For each group, I collected 4 weeks of videos (2 weeks before and 2 weeks after a pellet size change) after feeding time, for 2 hours per day. The total data collected for this study amounted to 784 hours, and the analysis was completed within a week using the automated pipeline.

In these datasets, weeks involving room cleaning, health screenings, or other management procedures were excluded. The process of room cleaning necessitates relocating the animals from the playpen to the cage room, which subtracts time that could otherwise be spent on foraging. In addition, it is important to note that health and management procedures are recognized stressors for the animals, and this stress can influence their behaviour, including their foraging activities.

3. *Seasons study*

The third dataset comprised videos of 14 macaque breeding groups, with a mean group size of 11 (SD = 2.8). For each group, I collected videos for 13 weekends from October to December 2022 (breeding season) and 13 weekends from March to May 2023 (birth season), capturing 2 hours after feeding time. To limit disturbance from technicians and ensure consistent data collection, while excluding days with exceptional proceedings such as the removal of an individual, health checks, or injured animals, I specifically focused on weekends. The total data collected for this study amounted to 1456 hours, and the analysis was conducted over a period of 2 weeks using the automated pipeline.

For this study, data on injuries and weights were extracted from the CFM database (see Chapter 2: Materials and Methods). The objective was to explore the impact of foraging behaviour on aggression and obesity, as previous studies have demonstrated its potential influence on both these factors (Schwitzer and Kaumanns, 2001; Mathy and Isbell, 2002; Vogel, 2005; Farmer *et al.*, 2010; Bauer *et al.*, 2011; Brereton, 2022; Pontzer and Pontzer, 2023). As this study was conducted on different macaque groups, a conversion was necessary to transform the data from individual-level to group-level analysis.

In the study, only injuries that required veterinary attention, such as medicine administration or suturing, were included. Specifically, injuries resulting from aggressions and fighting were considered, while those with unknown causes, which were more likely to be accidental, were excluded.

To convert the data from individual-level (injuries per animal in the group) to group-level (average number of injuries per group), the following calculations were performed:

1- The number of injuries per month for each individual.
2- The average number of injuries recorded per month per individual.
3- The average number of injuries recorded per month for each group.

Weights data at the CFM are collected during the annual health screenings for each individual. Utilizing a total of 11,921 data points gathered over the past 15 years, weight values for each age class and sex were determined by calculating the median of individual weights. Age class are defined as inTable 15.

Age class	**Age range in years**
Infant	0-1
Juvenile	1-2.5
Adolescent	2.5-4
Young Adult	4-7
Adult	7+

Table 15: Summary of the age class.

To streamline the dataset and convert the weights information from the individual-level to the group-level, the following calculations were carried out:

1- Each individual was assigned a score representing the percentage deviation of its weight from the median weight for its corresponding sex/age class.
2- The average score was calculated for each group.
3- The number of individuals above the median weight was recorded for each group.

These calculations allowed for a more concise representation of the weight data, facilitating the analysis and comparison of weight trends at the group level.

4.3.3 Automated Methodology

The methodology used to collect the data for this study is based on computer vision: a process that enable computers to extract information from visual inputs. In this case, the algorithm YOLACT (Bolya *et al.*, 2019) was used to build a model able to detect the macaques in the CCTV videos (Chapter 2, for more details). The model was trained on 674 labelled images of macaques foraging, with a split of 70% for training and 30% for testing (See Section 2.5.3 for more details). These images were randomly extracted from a subset of the whole dataset of videos recording the macaques foraging. The model precision, showing the intersection over union (IoU) for different thresholds is shown in

	All	**.50**	**.60**	**.70**	**.80**	**.90**
Box	25.5	56.3	27.7	6.9	1.2	0.2
Mask	21.5	55.4	43.0	21.3	1.6	0

Table 16. IOU is a measure of the overlap between the predicted bounding box (and mask) and the ground truth bounding box (and mask) (see Chapter 2 for details).

	All	**.50**	**.60**	**.70**	**.80**	**.90**

Box	25.5	56.3	27.7	6.9	1.2	0.2
Mask	21.5	55.4	43.0	21.3	1.6	0

Table 16: Mean average precision (mAP) for the Foraging Model. The mAP compares the ground-truth bounding box (and mask) to the detected box (and mask) and returns a score. These comparisons are reported for different IoU thresholds, which represent the proportion of overlap between the predicted bounding box (or mask) and the ground truth bounding box (or mask) in relation to their combined area. For instance, an IoU of .50 signifies a 50% overlap.

Because of the complexity of the enclosures (i.e., many structures and enrichment present in the play pen) and the position of the cameras (i.e., cameras recording from above), I was not able to use existing computer vision models to properly identify macaques' behaviour. This is a common problem in automated detection of animal behaviour, in fact most of the time the studies using such technologies focused on animal detection, tracking and position, rather than actual behaviour identification (Pons, Jaen and Catala, 2017; Valletta *et al.*, 2017; Kleanthous *et al.*, 2022). Therefore, to address these limitations, I built a pipeline able to identify the macaques foraging following this definition: a monkey foraging is any individual detected by the model in a specific region of interest (ROI) that is the part of the play pen where the monkeys spend more time foraging and where the food is located (Figure 19). This assumption has been validated by comparing manually coded estimates of the number of monkeys foraging with the pipeline output (

Number of Videos	15
Number of Frames	165
True Positive (TP): Instances where the model correctly detects the presence of a macaque	307
True Negative (TN): Instances where the model correctly detects the absence of a macaque	27
False Positive (FP): Instances where the model incorrectly detects a macaque when there is none present	34
False Negative (FN): Instances where the model fails to detect a macaque when one is present	73
Accuracy in detecting the macaques in the play pen $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$	76%
Accuracy in detecting the macaques foraging $\left(\frac{\text{Detected monkey in ROI}}{Foraging\ monkey\ in\ ROI+Foraging\ monkey\ outside\ ROI}\right)$	97%
Recall in detecting macaques foraging $\left(\frac{\text{Detected monkey in ROI}}{Foraging\ monkey\ in\ ROI+Non-detected\ monkey\ foraging\ in\ ROI}\right)$	81%
Precision in detecting macaques foraging $\left(\frac{\text{Detected monkey in ROI}}{Foraging\ monkey\ in\ ROI+Non-foraging\ detected\ monkey\ in\ ROI}\right)$	90%

Table 17). As indicated in the table below, the accuracy in detecting macaques within the enclosure is lower than that for detecting macaques while foraging. This discrepancy arises

from the challenges in accurately identifying macaques positioned behind structures or at the rear of the enclosures. Notably, the region of interest (ROI) employed for detecting foraging macaques excludes these areas, resulting in a higher accuracy in foraging detection compared to the overall accuracy across the entire playpen.

The manual data extraction process involved selecting one video at random for each of the 16 groups. From these videos, a total of 11 frames were extracted, spaced 10 minutes apart. Within these frames, manual counts were performed to determine the number of macaques within ROI (MPR), the number of macaques engaged in foraging within the ROI (MFIR), and the number of macaques foraging outside the ROI (MFOR). Subsequently, these manually collected data were compared to the number of detected monkeys within the ROI (MO) as identified by the automated process (Table 18).

To extract the percentage of macaques foraging I built a pipeline using YOLACT to detect the monkeys, a Python script to extract the ROI and R to extract the percentage of macaques foraging (Figure 20). This pipeline is able to detect monkeys foraging in 15 different enclosures.

Number of Videos	15
Number of Frames	165
True Positive (TP): Instances where the model correctly detects the presence of a macaque	307
True Negative (TN): Instances where the model correctly detects the absence of a macaque	27
False Positive (FP): Instances where the model incorrectly detects a macaque when there is none present	34
False Negative (FN): Instances where the model fails to detect a macaque when one is present	73
Accuracy in detecting the macaques in the play pen $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$	76%
Accuracy in detecting the macaques foraging $\left(\frac{\text{Detected monkey in ROI}}{Foraging\ monkey\ in\ ROI+Foraging\ monkey\ outside\ ROI}\right)$	97%
Recall in detecting macaques foraging $\left(\frac{\text{Detected monkey in ROI}}{Foraging\ monkey\ in\ ROI+Non-detected\ monkey\ foraging\ in\ ROI}\right)$	81%
Precision in detecting macaques foraging $\left(\frac{\text{Detected monkey in ROI}}{Foraging\ monkey\ in\ ROI+Non-foraging\ detected\ monkey\ in\ ROI}\right)$	90%

Table 17: The table shows the evaluation of the pipeline. It reports the accuracy of the model in detecting the macaques, the accuracy, recall and precision of the pipeline in detecting the macaques foraging in the region of interest (ROI).

Group	Minutes	MO	MPR	MFIR	MFOR	TP	FP	TN	FN	Accuracy Detection	Accuracy Foraging
G03	0	0	0	0	0	0	0	1	0	1	1
G03	10	7	9	8	0	7	0	0	2	0.77	0.87
G03	20	7	9	8	0	7	0	0	2	0.77	0.87
G03	30	3	6	4	0	3	0	0	3	0.5	0.75
G03	40	4	5	5	0	4	0	0	1	0.8	0.8
G03	50	7	8	7	0	7	0	0	1	0.87	1
G03	60	4	7	6	0	4	0	0	3	0.57	0.66
G03	70	4	4	4	0	4	0	0	0	1	1
G03	80	0	0	0	0	0	0	1	0	1	1
G03	90	0	0	0	0	0	0	1	0	1	1
G03	100	0	0	0	0	0	0	1	0	1	1

Table 18: Comparison of Model Detection and Manual Coding Data for G03 Group Video: MO (Model Output), MPR (Manual Presence in ROI), MFIR (Manual Foraging in ROI), MFOR (Manual Foraging outside ROI). TP (True Positive) obtains as (min(MO,MFIR)), FP (False Positive) as (MO-MFIR), True Negative as (no monkeys in both manual and model output), False Negative (MFIR-MO) were calculated for model accuracy, with model accuracy in detecting foraging macaques as MO / (MFIR + MFOR).

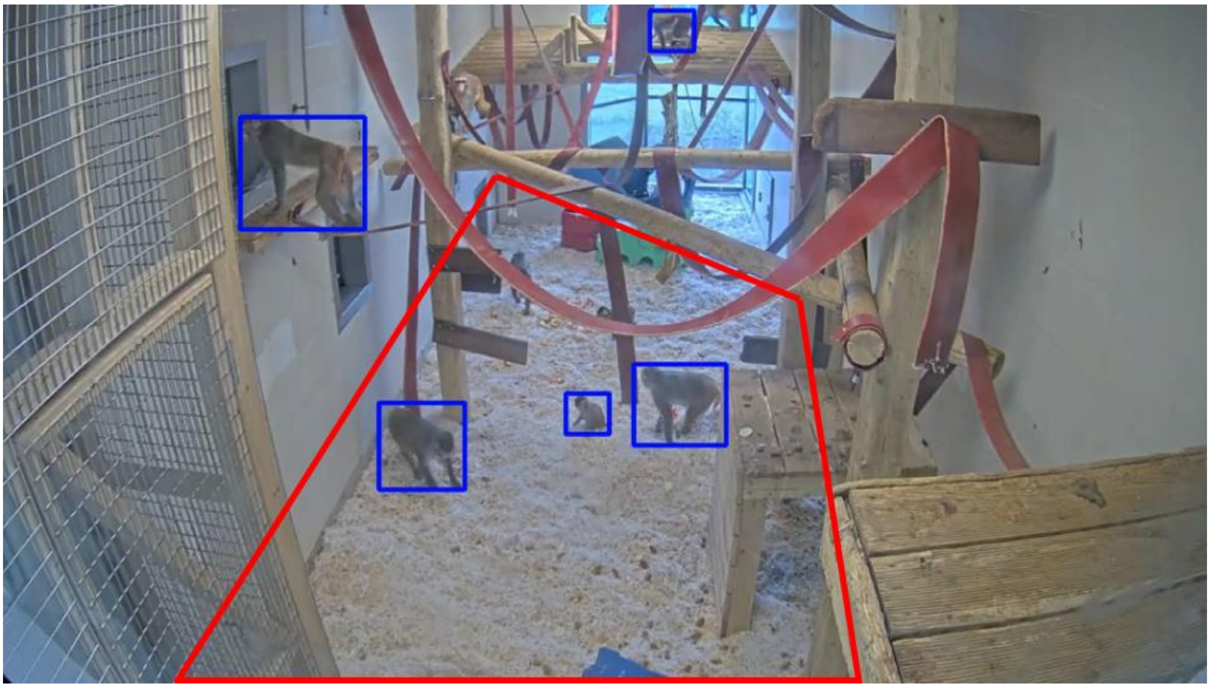


Figure 19: Model output. The individuals in the blue square (bbox) are the one detected by the model. The red polygon is the ROI where most of the food is located. The blue bbox located within the red ROI are the individuals identified as macaques foraging by the pipeline.

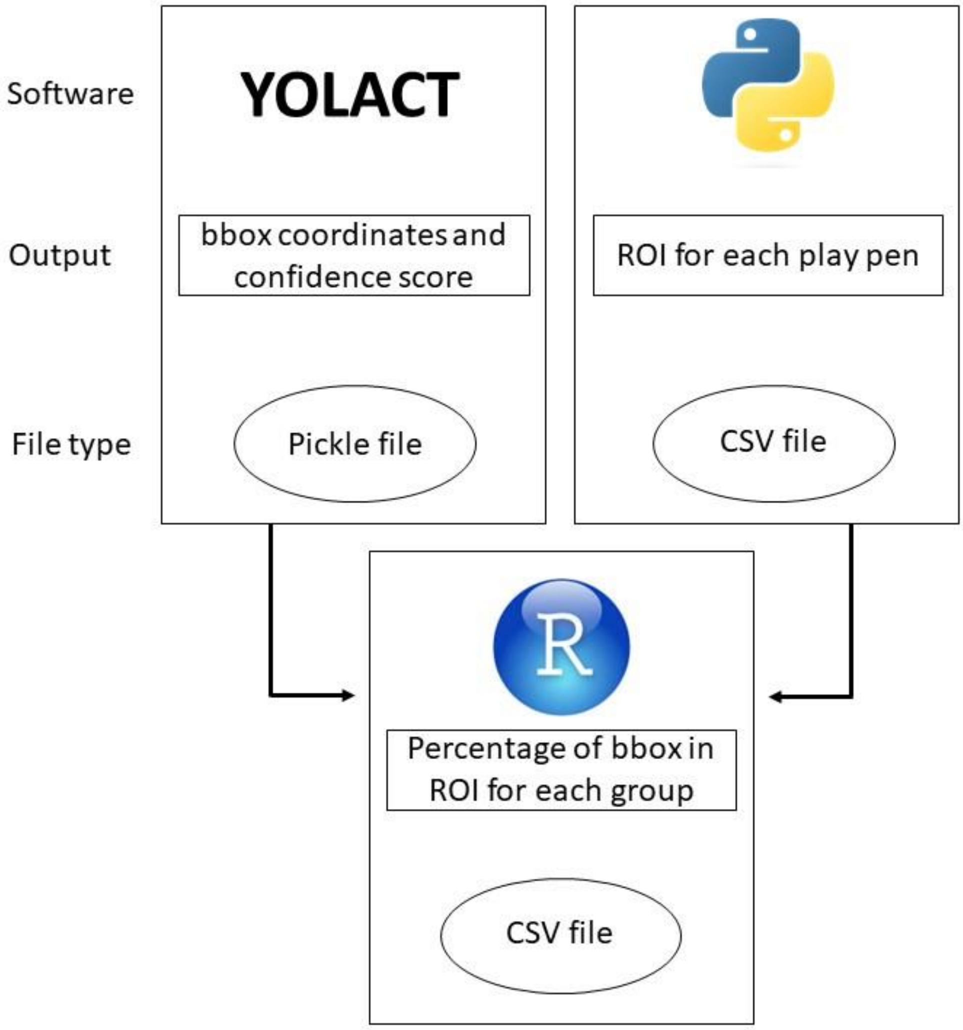


Figure 20: The pipeline begins with the Foraging model extracting bounding boxes of the monkeys in action. Next, a Python script defines the Region of Interest (ROI) for each enclosure. Finally, leveraging these outputs, R calculates the percentage of monkeys engaged in foraging activities.

4.3.4 Statistical Analyses

To analyse the output data from the pipeline, I utilized R Studio (version 4.0.4; R Core Team 2021) on Windows 10. All statistical analyses were two-tailed, and an alpha level of 0.05 was set. A linear mixed model analysis was employed to examine the data, allowing for the incorporation of both fixed and random effects to account for potential sources of variation within the dataset. To test the model assumptions histograms, normal probability plots of residuals, and quantile-quantile plots were employed (Hartig, 2018). The lmer function from the lme4 package was used to model the foraging data (Bates, D and Maechler, M and Bolker, BM and Walker, 2014). The percentage of monkeys engaged in foraging served as the dependent variable in all studies, while the different foraging mixes, pellet size (small or big), and seasons were independent variables for their respective studies. Additionally, each model included group size and group type (breeding group or juvenile group) as fixed factors, and group identity as a random factor. To control for group size, the percentage of monkeys foraging was calculated by dividing the number of detected animals in the ROI in each frame by the number of macaques in the group.

Furthermore, a Pearson's correlation coefficient was computed to examine the relationship between the percentage of monkeys foraging, the group size, the average number of injuries per group, and the number of individuals above the median weight of the group.

4.4 Results

1. *Foraging Mixes study*

Using the pipeline, I was able to demonstrate a significant effect of different foraging mixes on the average percentage of monkeys engaged in foraging activities (F (6,84) = 6.35, p < 0.001). The results indicate that foraging mixes consisting of smaller food items, such as mushroom and radish on Mondays, and spinach and watercress on Saturdays, are associated with a higher percentage of macaques foraging (*Figure 21*). The linear mixed-effects model (lmm), also showed a significant negative effect of the group size on the percentage of monkey foraging (F (1, 12) = 8.07, p = 0.01).

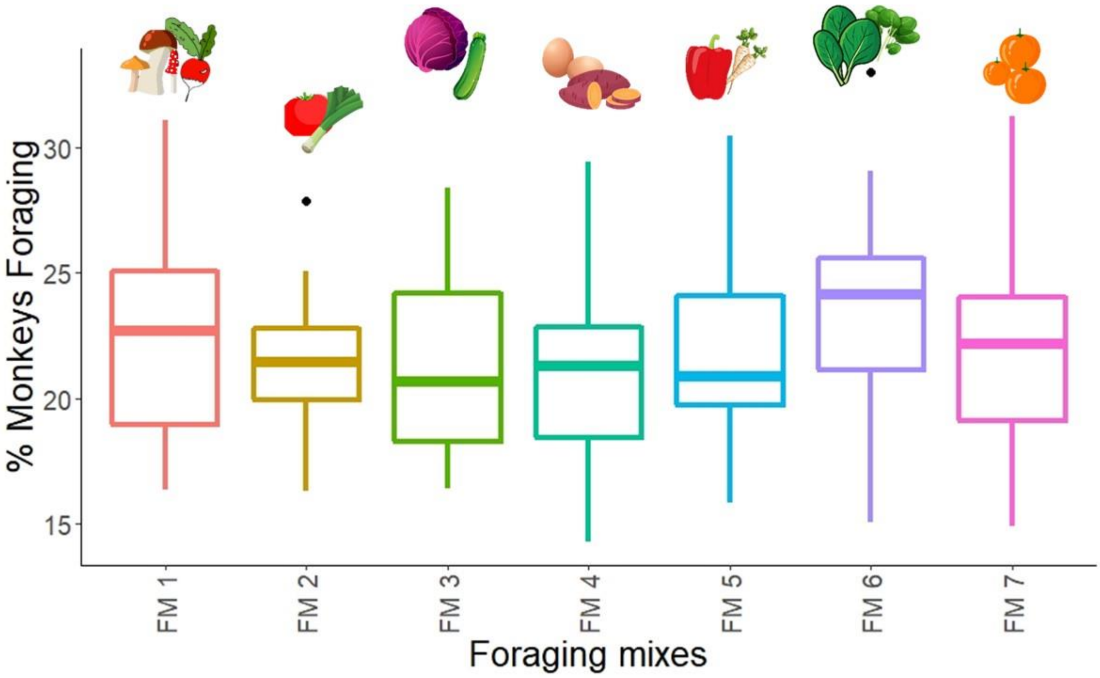


Figure 21: Boxplot illustrating the distribution of mean percentages of monkeys foraging across different foraging mixes. Each boxplot represents a unique foraging mix (FM), with the central line indicating the median percentage of monkeys foraging. The box extends from the lower to upper quartiles, while the whiskers show the data range. Outliers are denoted by black dots.

2. Pellet Size study

Using the pipeline, I observed an increase in the percentage of monkeys engaged in foraging after the change in pellet size (Figure 22). However, this change was not statistically significant (F (1, 13) = 0.02, p = 0.8) with a mean difference in percentage of macaques foraging of 0.127, 95% CI [-1.23, 1.56]. Similar to the previous study, the linear mixed-effects model (lmm) revealed a significant negative effect of group size on the percentage of monkeys engaged in foraging (F (1, 11) = 39.1, p < 0.001).

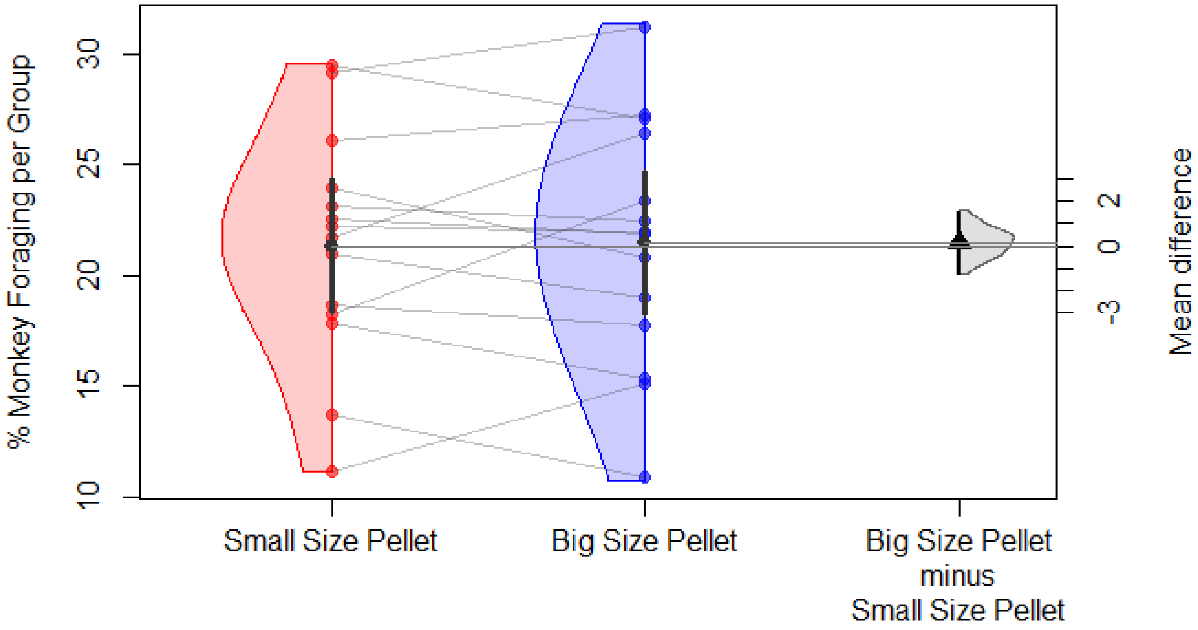


Figure 22: Percentage of monkeys engaged in foraging calculated over a period of 4 weeks before and after the change in pellet size, from small to big.

3. Season study

The data obtained through the automated methodology indicates a significant effect of the seasons (breeding season and birth season) on the percentage of monkeys engaged in foraging (F (5, 351) = 6.7, p < 0.001). The statistical analysis revealed a significant increase in the percentage of macaques foraging during the final months of the breeding season (Figure 23, see

Months	**Mean Difference**	**95% CI**
December - November	-5.16	[-6.84, -3.32]
December - October	-1.60	[-3.45, 0.36]
December - May	-0.33	[-2.12, 1.76]
December - April	-2.30	[-3.83, -0.78]
December - March	-1.69	[-3.32 -0.043]
November - October	0.14	[-1.97, 2.0]
November - May	1.41	[-0.64, 3.86]
November - April	-0.55	[-2.19, 1.43]
November - March	0.05	[-1.51, 1.84]
October - May	4.31	[2.14, 6.81]
October - April	2.34	[0.18, 4.96]
October - March	2.96	[0.44, 6.01]
May - April	-3.96	[-5.43, -2.71]

May - March	-3.35	[-4.53, -2.06]
April - March	3.01	[1.77, 4.55]

Table 19 for mean effect size information). Similar to the previous studies, the linear mixed-effects model (lmm) revealed a significant effect of group size on the percentage of monkeys engaged in foraging, with larger group showing a lower percentage of macaques foraging (F (1, 277) = 48.8, p < 0.001).

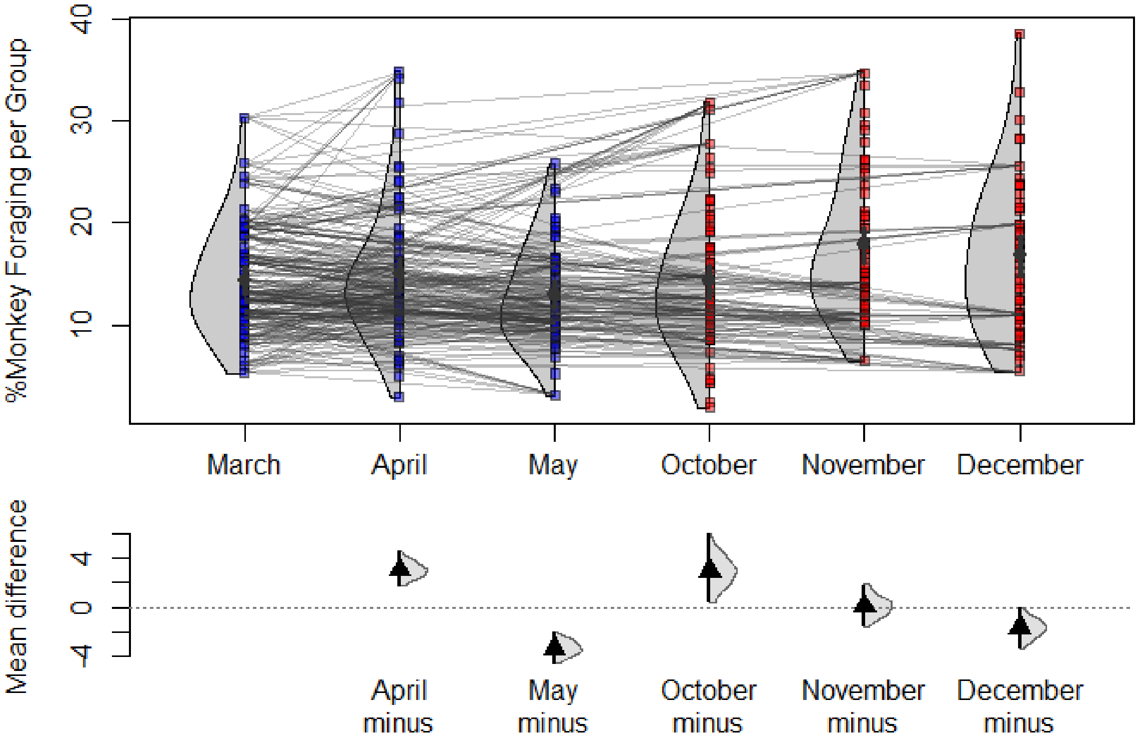


Figure 23: Percentage of monkeys engaged in foraging calculated over a period of 6 months, specifically during the breeding season (October to December) and the birth season (March to May). The blue dots represent the percentage of macaques foraging for each group during the birth season, while the red dots represent the percentage of macaques foraging for each group during the breeding season.

Months	**Mean Difference**	**95% CI**
December - November	-5.16	[-6.84, -3.32]
December - October	-1.60	[-3.45, 0.36]
December - May	-0.33	[-2.12, 1.76]
December - April	-2.30	[-3.83, -0.78]
December - March	-1.69	[-3.32 -0.043]
November - October	0.14	[-1.97, 2.0]
November - May	1.41	[-0.64, 3.86]
November - April	-0.55	[-2.19, 1.43]
November - March	0.05	[-1.51, 1.84]
October - May	4.31	[2.14, 6.81]
October - April	2.34	[0.18, 4.96]
October - March	2.96	[0.44, 6.01]
May - April	-3.96	[-5.43, -2.71]
May - March	-3.35	[-4.53, -2.06]
April - March	3.01	[1.77, 4.55]

Table 19: Mean Difference and Confidence Intervals for the different months.

Following the previous results, I observed a significant negative correlation between the percentage of macaques foraging and group size (r (96) = -0.31, p = 0.001), indicating that smaller groups exhibit a higher percentage of monkeys engaged in foraging (Figure 24).

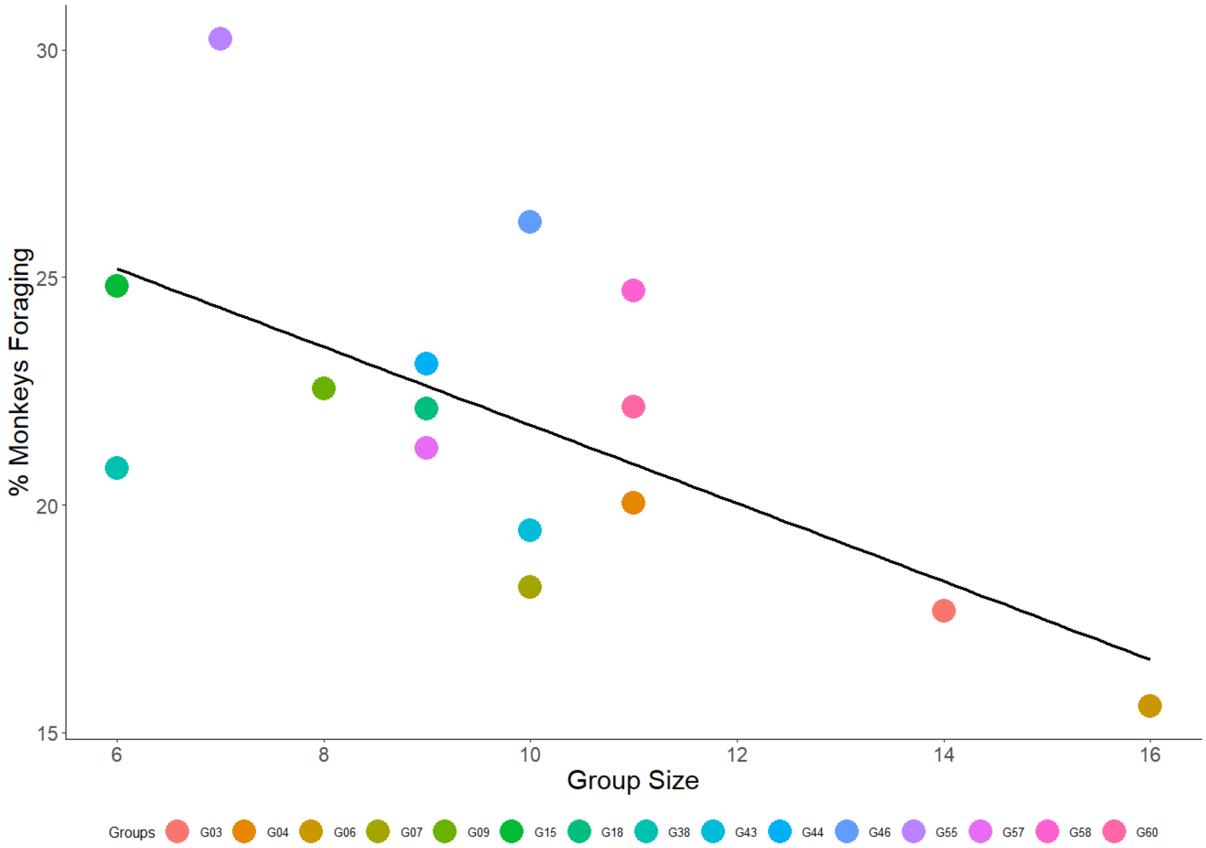


Figure 24: Relationship between the percentage of monkeys foraging and the group size, examined for the 15 groups of macaques that take part in the Foraging Mixes study.

Additionally, I found a significant negative correlation between the percentage of macaques foraging and the average number of injuries recorded per group over the study period (r (96) = -0.63, p = 0.002). This suggests that groups with a higher number of injuries tend to have a lower percentage of monkeys participating in foraging activities (Figure 25).

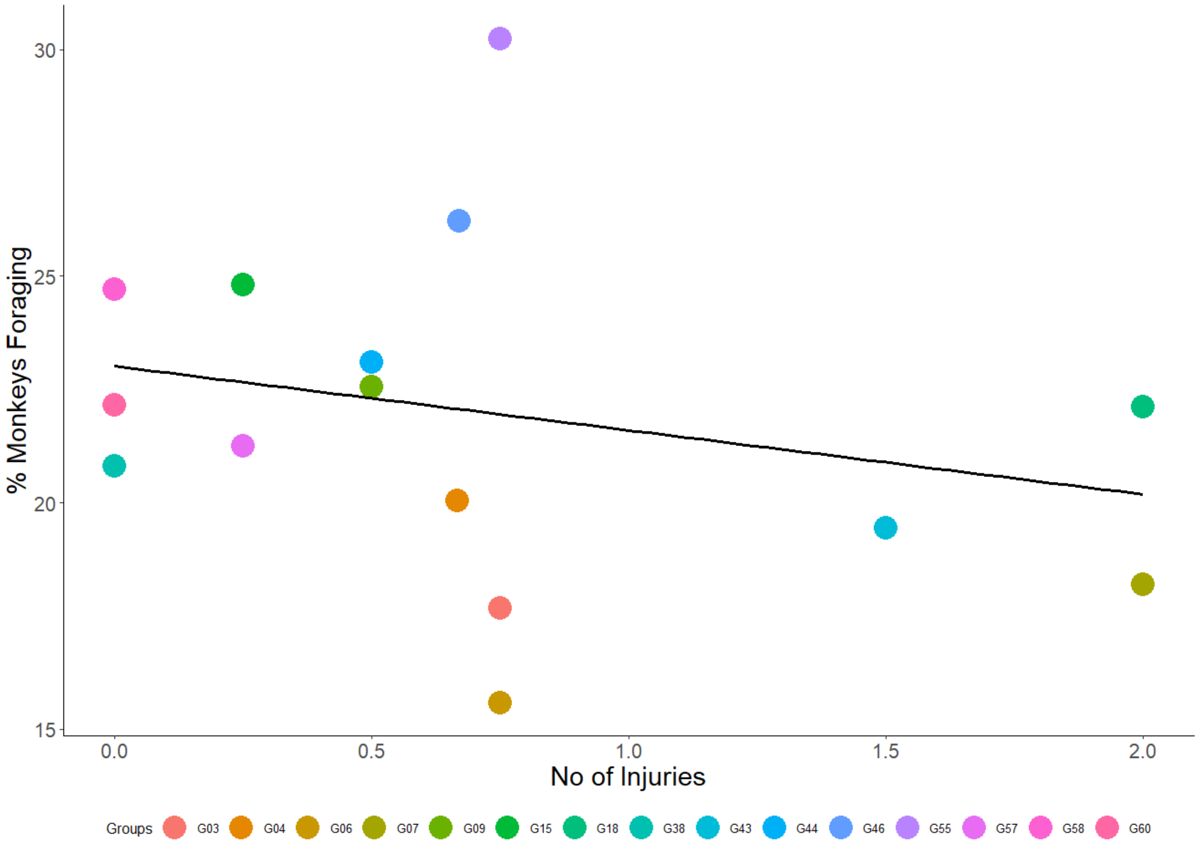


Figure 25: Relationship between the percentage of monkeys foraging and the number of injuries recorded, examined for the 15 groups of macaques that take part in the Foraging Mixes study.

Moreover, I explored the relationship between the percentage of macaques foraging and the number of individuals with weights above the median weight of the group. However, this relationship was not found to be statistically significant (r (96) = 0.17, p = 0.09), indicating that I was not able to show how the number of heavier individuals in a group influence the percentage of monkeys engaged in foraging (Figure 26).

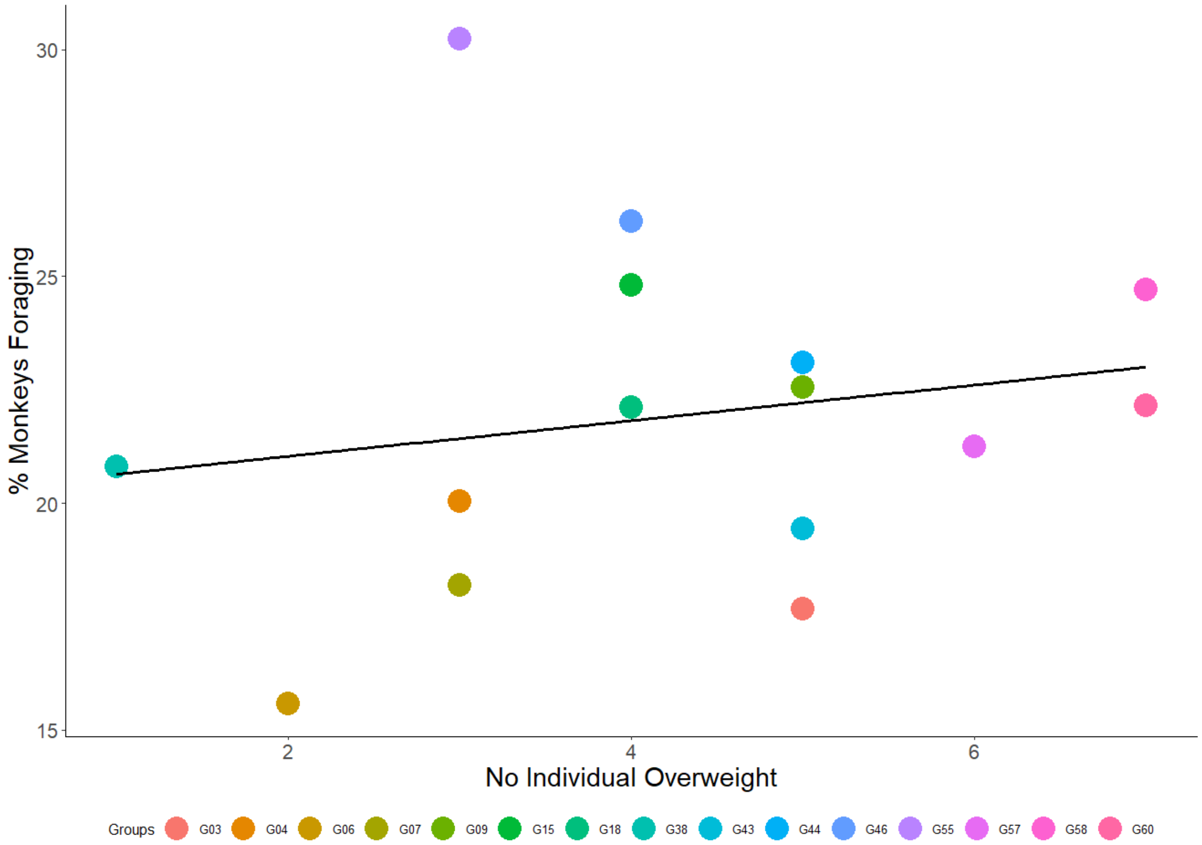


Figure 26: Relationship between the percentage of monkeys foraging and the number of individuals above the median weight of the group, examined for 15 groups of macaques.

4.5 Discussion

Using a machine learning based pipeline, it was possible to explore feeding strategies across diverse situations, obtaining valuable insights into group foraging patterns and the influence of different foods on foraging behaviour. Non-human primates (NHPs), like many other species in captivity, often face challenges such as boredom, social disruption, and welfare issues. Facilities that care for wild animals invest considerable effort in understanding their behaviour and adjusting husbandry practices to enhance their welfare.

Chopping food is a commonly practiced feeding method in zoos and other facilities housing wild animals (Farmer *et al.*, 2010). Several studies have shown that chopped food results in reduced stress, increased activity, and more foraging behaviour in many primate species, whereas clumped food leads to decreased activity and less foraging (Waasdorp *et al.*, 2021). Chopped food has also been to decrease aggression within groups of rhesus macaques (Mathy and Isbell, 2002). However, there are conflicting results in other species, with some studies showing no significant effect on aggression and foraging behaviour, or even an increase in food manipulation when whole food is provided (Farmer *et al.*, 2010; Brereton, 2022). Using my model, I was able to demonstrate that providing chopped foods such as spinach, watercress, and mushrooms keeps the macaques more engaged in foraging compared to foraging mixes with higher carbohydrate, fructose, and glucose contents (e.g., oranges and tomatoes), which are preferred by other primate species (Laska, 2001; Jildmalm, Amundin and Laska, 2008; Clay *et al.*, 2009). Feeding strategies, gut morphology, and nutrient requirements vary greatly among primate species, posing challenges for providing appropriate diets in captivity (Crissey and Pribyl, 1997). This insight into macaques foraging behaviour will assist care staff and management at CFM in providing better food options for the animals. In addition, these findings were previously unavailable due to the time-consuming nature of observing long-lasting behaviours like foraging, which require constant monitoring by care staff.

By understanding the factors influencing foraging, CFM can now provide better food options for the animals, taking into account the intense competition for food resources often observed among NHPs (Janson, 1985; Robichaud, Lefebvre and Robidoux, 1996; Saito Chiemi, 1996; Koenig, 2000). Similar patterns have been observed in capuchin monkeys, where food intake and foraging activities are correlated with dominance status and the number of aggressions received (Janson, 1985; Vogel, 2005). In captive environments with concentrated and easy to monopolize resources, contest competition tends to be even higher (Vogel, 2005). Dominant individuals within a group have higher food intake and may exclude subordinates from feeding opportunities (Koenig, 2000; Vogel, 2005). Additionally, larger groups of monkeys show a decrease in per-capita food intake due to indirect food competition and aggressive food competition. At CFM, the change in pellet size was intended to help subordinate individuals access more food and prevent dominant individuals from monopolizing all the resources (Smith, Lindburg and Vehrencamp, 1989; Farmer *et al.*, 2010). The pipeline's findings regarding the percentage of individuals foraging with larger pellet sizes at CFM, although not statistically significant, show a modest increase. Importantly, this result supports the implementation of pellet size adjustments without negatively impacting foraging patterns. While the study did not observe changes in subordinate foraging behaviour, it provides valuable insights into the stability of foraging behaviour in macaques throughout the year.

Previous studies on wild macaques have consistently shown that foraging behaviour remains relatively stable over time, with variations primarily driven by food availability in the environment (Agetsuma, 1995; Jaman, Huffman and Takemoto, 2010). Seasonal variations in food availability can affect foraging behaviour and drive dietary shifts towards fruits or leaves based on the time of year (Garber, 1987; Agetsuma, 1995; Jaman, Huffman and Takemoto, 2010; Trapanese, Meunier and Masi, 2019). However, information about the feeding strategies and nutrient intake of free-ranging primates, particularly during breeding and birth seasons, remains incomplete (Crissey and Pribyl, 1997). In some NHPs, lactating females consume more food than males and non-reproductive females, both in captivity and in the wild (Garber, 1987). However, in other monkey species, lactating females exhibit a decrease in feeding time, which could be attributed to the birth season coinciding with the dry season, a period of food scarcity (Harrison, 1984). The foraging behaviour observed during the birth season, as detected by the pipeline, shares similarities with patterns observed in black howler monkeys where lactating females spending more time inactive compared to non-lactating females (Dias, Rangel-Negrín and Canales-Espinosa, 2011). Furthermore, in the black howler monkey species, females initially exhibit higher levels of inactivity during the lactation period but gradually increase their foraging activities towards the end of the season (Dias, Rangel-Negrín and Canales-Espinosa, 2011). Similar low foraging behaviour patterns can be observed in baboons, where heightened vigilance levels are observed during the beginning of the birth season, resulting in decrease overall time dedicated to feeding during lactation compared to non-lactating periods (Barrett, Halliday and Henzi, 2006).

In terms of primate reproduction, males typically invest in body size, sexual traits, and mate searching, while females focus their efforts on mate choice, pregnancy, and lactation (Wolfe, 1986; Gomendio, 1991; Crissey and Pribyl, 1997; Soltis *et al.*, 1999; Kappeler and Van Schaik, 2004; Thomsen *et al.*, 2006). For instance, male rhesus macaques tend to gain more weight and accumulate greater fat reserves prior to the breeding season to cope with the energetic demands of mating (Muehlenbein *et al.*, 2002). Similarly, female rhesus macaques exhibit higher body mass and fat levels at the beginning of the mating season, ensuring sufficient physical condition to cover the energy costs associated with mating activity (Garcia *et al.*, 2011). Interestingly, despite the constant availability of food, male rhesus macaques display a decrease in feeding time between mating and non-mating seasons, suggesting a potential prioritization of energy towards mating rather than feeding (Bercovitch, 1997; Garcia *et al.*, 2011). Similar patterns have been observed at CFM, where males' weights after the breeding season were recorded to be lower than in previous periods. However, my results indicate a significant increase in the percentage of macaques foraging during the breeding seasons. This is not in conflict with the weight loss in males since most individuals in the breeding groups are females, who may increase their time spent foraging to cope with the energy expenditure of mating and subsequent pregnancy. These findings shed light on an understudied topic concerning the effect of seasons on foraging behaviours in captive non-human primates, where food is consistently available throughout the year, but seasonal breeding still occurs.

Additionally, it is important to recognize that social dynamics play a significant role in the behaviour and well-being of non-human primates (NHPs). Different housing conditions and group settings in rhesus monkeys (*Macaca mulatta*) have been found to result in abnormal behaviours and increased aggression. Rhesus monkeys already exhibit a higher frequency and severity of aggression compared to other macaque species, making deleterious aggression a major concern in group-housed macaques (Thierry, Singh and Kaumanns, 2004; Corrine K. Lutz and Novak, 2005; Lutz *et al.*, 2007a; Beisner and Isbell, 2011a; Gottlieb, Capitanio and McCowan, 2013b; Lutz and Baker, 2023). Food monopolization is one of the causes of aggression in macaques, and numerous studies have demonstrated how they tend to monopolize food patches through aggressive behaviours. Larger and less crowded food patches can reduce competition within the group (van Schaik *et al.*, 1983; Saito Chiemi, 1996; Steenbeek and Sterck, 1997; Beisner and Isbell, 2011a; Grove, 2012). While this behaviour occurs both in the wild and in captivity, aggression rates are often higher in captive settings due to the limited options for animals to escape aggressive encounters. This can lead to escalated aggression and an increased risk of injuries (Thierry, Singh and Kaumanns, 2004; McCowan *et al.*, 2008; Beisner and Isbell, 2011a; K.N. Balasubramaniam *et al.*, 2014). The study results demonstrate how a deep learning-based pipeline can effectively collect, analyse, and extract information that aligns with existing knowledge about NHP behaviour in social groups. Specifically, the findings indicate that smaller groups tend to have a higher percentage of monkeys engaged in foraging, and a higher percentage of monkeys foraging correlates with lower injury rates within the group. This can be attributed to reduced interference and competition among group members, resulting in improved access to food resources (van Schaik *et al.*, 1983; Saito Chiemi, 1996; Steenbeek and Sterck, 1997; Beisner and Isbell, 2011a; Grove, 2012). These insights underscore the importance of considering group size and composition when establishing social groups and monitoring behaviors such as foraging. By managing these factors, it becomes possible to better care for the group, enhance their welfare, and mitigate aggression within the group (Bayne *et al.*, 1992).

The data collected by the pipeline shows a non-significant positive correlation between the percentage of macaques foraging and the number of overweight individuals in the group. This can be due to foraging not being the only factor influencing macaques body mass (Altmann *et al.*, 1993; Bauer *et al.*, 2011).

Overall, this study provides valuable insights into the feeding behaviours of NHPs, highlighting the importance of considering food presentation and group dynamics to enhance welfare and manage aggression within social groups.

4.6 Chapter References

Agetsuma, N. (1995) ‘Foraging strategies of yakushima macaques(Macaca fuscata yakui)’, *International Journal of Primatology*, 16(4), pp. 595–609. doi:10.1007/BF02735283.

Altmann, J. *et al.* (1993) ‘Body size and fatness of free-living baboons reflect food availability and activity levels’, *American journal of primatology*, 30(2), pp. 149–161.

Aroeira, R.M.C. *et al.* (2016) ‘Non-invasive methods of computer vision in the posture evaluation of adolescent idiopathic scoliosis’, *Journal of Bodywork and Movement Therapies*, 20(4), pp. 832–843. doi:10.1016/j.jbmt.2016.02.004.

Balasubramaniam, K.N. *et al.* (2014) ‘Group size, contest competition, and social structure in Cayo Santiago rhesus macaques’, *Behaviour*, 151(12–13), pp. 1759–1798. doi:10.1163/1568539X-00003216.

Barrett, L., Halliday, J. and Henzi, S.P. (2006) ‘The ecology of motherhood: the structuring of lactation costs by chacma baboons’, *Journal of Animal Ecology*, 75(4), pp. 875–886. doi:10.1111/j.1365-2656.2006.01105.x.

Bates, D. *et al.* (2015) ‘Fitting Linear Mixed-Effects Models Using lme4’, *Journal of Statistical Software*, 67(1 SE-Articles), pp. 1–48. doi:10.18637/jss.v067.i01.

Bauer, S.A. *et al.* (2011) ‘Obesity in rhesus and cynomolgus macaques: a comparative review of the condition and its implications for research’, *Comparative medicine*, 61(6), pp. 514–526.

Bayne, K. *et al.* (1992) ‘The Use of Artificial Turf as a Foraging Substrate for Individually Housed Rhesus Monkeys (Macaca Mulatta)’, *Animal Welfare*, 1(1), pp. 39–53. doi:10.1017/S0962728600014706.

Beisner, B.A. and Isbell, L.A. (2011a) ‘Factors affecting aggression among females in captive groups of rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 73(11), pp. 1152–1159. doi:10.1002/ajp.20982.

Beisner, B.A. and Isbell, L.A. (2011b) ‘Factors affecting aggression among females in captive groups of rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 73(11), pp. 1152–1159. doi:10.1002/ajp.20982.

Bercovitch, F.B. (1997) ‘Reproductive strategies of rhesus macaques’, *Primates*, 38(3), pp. 247–263. doi:10.1007/BF02381613.

Bercovitch, F.B. *et al.* (1999) ‘Primigravidity, body weight, and costs of rearing first offspring in rhesus macaques’, *American Journal of Primatology*, 46(2), pp. 135–144. doi:10.1002/(SICI)1098-2345(1998)46:2<135::AID-AJP3>3.0.CO;2-X.

Bolya, D. *et al.* (2019) ‘Yolact: Real-time instance segmentation’, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166.

Brereton, J.E. (2022) ‘Should Zoo Foods be Chopped or Should We ‘Lemur’them Whole’, *Medp Nutr Food Sci*, 1(1).

Chancellor, R.L. and Isbell, L.A. (2008) ‘Punishment and competition over food in captive rhesus macaques, Macaca mulatta’, *Animal Behaviour*, 75(6), pp. 1939–1947. doi:10.1016/j.anbehav.2007.11.007.

Chen, C., Zhu, W. and Norton, T. (2021) ‘Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning’, *Computers and Electronics in Agriculture*, 187, p. 106255. doi:10.1016/j.compag.2021.106255.

Clay, A.W. *et al.* (2009) ‘Systematic Investigation of the Stability of Food Preferences in Captive Orangutans: Implications for Positive Reinforcement Training’, *Journal of Applied Animal Welfare Science*, 12(4), pp. 306–313. doi:10.1080/10888700903163492.

Crissey, S.D. and Pribyl, L.S. (1997) ‘Utilizing wild foraging ecology information to provide captive primates with an appropriate diet’, *Proceedings of the Nutrition Society*, 56(3), pp. 1083–1094. doi:10.1079/PNS19970112.

Dias, P.A.D., Rangel-Negrín, A. and Canales-Espinosa, D. (2011) ‘Effects of lactation on the time-budgets and foraging patterns of female black howlers (Alouatta pigra)’, *American Journal of Physical Anthropology*, 145(1), pp. 137–146. doi:10.1002/ajpa.21481.

Farmer, H. *et al.* (2010) *Should Zoo Food be Chopped?*

Garber, P.A. (1987) ‘Foraging Strategies Among Living Primates’, *Annual Review of Anthropology*, 16(1), pp. 339–364. doi:10.1146/annurev.an.16.100187.002011.

Garcia, C. *et al.* (2011) ‘Energetic consequences of seasonal breeding in female Japanese macaques (Macaca fuscata)’, *American Journal of Physical Anthropology*, 146(2), pp. 161–170. doi:10.1002/ajpa.21553.

Gomendio, M. (1991) ‘Parent/offspring conflict and maternal investment in rhesus macaques’, *Animal Behaviour*, 42(6), pp. 993–1005. doi:10.1016/S0003-3472(05)80152-6.

Gottlieb, D.H., Capitanio, J.P. and McCowan, B. (2013a) ‘Risk factors for stereotypic behavior and self-biting in rhesus macaques (Macaca mulatta): Animal’s history, current environment, and personality’, *American Journal of Primatology*, 75(10), pp. 995–1008. doi:10.1002/ajp.22161.

Gottlieb, D.H., Capitanio, J.P. and McCowan, B. (2013b) ‘Risk factors for stereotypic behavior and self-biting in rhesus macaques (Macaca mulatta): Animal’s history, current environment, and personality’, *American Journal of Primatology*, 75(10), pp. 995–1008. doi:10.1002/ajp.22161.

Gottlieb, D.H., Maier, A. and Coleman, K. (2015) ‘Evaluation of environmental and intrinsic factors that contribute to stereotypic behavior in captive rhesus macaques (Macaca mulatta)’, *Applied Animal Behaviour Science*, 171, pp. 184–191. doi:10.1016/j.applanim.2015.08.005.

Grove, M. (2012) ‘Space, time, and group size: a model of constraints on primate social foraging’, *Animal Behaviour*, 83(2), pp. 411–419. doi:10.1016/j.anbehav.2011.11.011.

Hanya, G. (2004a) ‘Diet of a Japanese Macaque Troop in the Coniferous Forest of

Yakushima', *International Journal of Primatology*, 25(1), pp. 55–71. doi:10.1023/B:IJOP.0000014645.78610.32.

Hanya, G. (2004b) 'Seasonal variations in the activity budget of Japanese macaques in the coniferous forest of Yakushima: Effects of food and temperature', *American Journal of Primatology*, 63(3), pp. 165–177. doi:10.1002/ajp.20049.

Harrison, M.J.S. (1984) 'Optimal foraging strategies in the diet of the green monkey,Cercopithecus sabaeus, at Mt. Assirik, Senegal', *International Journal of Primatology*, 5(5), pp. 435–471. doi:10.1007/BF02692269.

Hartig, F. 2018. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. Version 0.2.0.

Higham, J.P., Heistermann, M. and Maestripieri, D. (2011) 'The energetics of male–male endurance rivalry in free-ranging rhesus macaques, Macaca mulatta', *Animal Behaviour*, 81(5), pp. 1001–1007. doi:10.1016/j.anbehav.2011.02.001.

Hill, D.A. (1997) 'Seasonal variation in the feeding behavior and diet of Japanese macaques (Macaca fuscata yakui) in lowland forest of Yakushima', *American Journal of Primatology*, 43(4), pp. 305–320. doi:10.1002/(SICI)1098-2345(1997)43:4<305::AID-AJP2>3.0.CO;2-0.

Hoshino, S. *et al.* (2022) 'Modifying the diets of captive proboscis monkeys in a temperate zoo to reduce weight loss and renal disease', *Primates*. doi:10.1007/s10329-022-01031-y.

Jaman, M.F., Huffman, M.A. and Takemoto, H. (2010) 'The foraging behavior of Japanese macaques Macaca fuscata in a forested enclosure: Effects of nutrient composition, energy and its seasonal variation on the consumption of natural plant foods', *Current Zoology*, 56(2), pp. 198–208. doi:10.1093/czoolo/56.2.198.

Janson, C. (1985) 'Aggresive competition and individual food consumption in wild brown capuchin monkeys (Cebus apella)', *Behavioral Ecology and Sociobiology*, 18(2), pp. 125–138. doi:10.1007/BF00299041.

Jildmalm, R., Amundin, M. and Laska, M. (2008) 'Food Preferences and Nutrient Composition in Captive White-handed Gibbons, Hylobates lar', *International Journal of Primatology*, 29(6), pp. 1535–1547. doi:10.1007/s10764-008-9314-1.

Kappeler, P.M. and Van Schaik, C.P. (2004) *Sexual selection in primates: new and comparative perspectives*. Cambridge University Press.

Kleanthous, N. *et al.* (2022) 'A survey of machine learning approaches in animal behaviour', *Neurocomputing*, 491, pp. 442–463. doi:10.1016/j.neucom.2021.10.126.

Koenig, A. (2000) 'Competitive regimes in forest-dwelling Hanuman langur females (Semnopithecus entellus)', *Behavioral Ecology and Sociobiology*, 48(2), pp. 93–109. doi:10.1007/s002650000198.

Laska, M. (2001) ‘A comparison of food preferences and nutrient composition in captive squirrel monkeys, Saimiri sciureus, and pigtail macaques, Macaca nemestrina’, *Physiology & Behavior*, 73(1–2), pp. 111–120. doi:10.1016/S0031-9384(01)00439-5.

Leigh, S.R. (1994) ‘Ontoaenetic correlates of diet in anthropoid primates’, *American Journal of Physical Anthropology*, 94(4), pp. 499–522. doi:10.1002/ajpa.1330940406.

Lutz, C., Well, A. and Novak, M. (2003) ‘Stereotypic and self-injurious behavior in rhesus macaques: A survey and retrospective analysis of environment and early experience’, *American Journal of Primatology*, 60(1), pp. 1–15. doi:10.1002/ajp.10075.

Lutz, C.K. *et al.* (2007) ‘Early predictors of self-biting in socially-housed rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 69(5), pp. 584–590. doi:10.1002/ajp.20370.

Lutz, C.K. and Baker, K.C. (2023) ‘Using Behavior to Assess Primate Welfare’, in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 171–205. doi:10.1007/978-3-030-82708-3_8.

Lutz, C.K. and Novak, M.A. (2005) *Primate Natural History and Social Behavior. Implications for Laboratory Housing*, *The Laboratory Primate*. Elsevier Limited. doi:10.1016/B978-012080261-6/50009-X.

Marlies Heesen (2014) *Feeding competition in wild female Assamese macaques (Macaca assamensis)*. Georg-August-University Göttingen. doi:10.53846/goediss-4698.

Mathy, J.W. and Isbell, L.A. (2002) ‘The Relative Importance of Size of Food and Interfood Distance in Eliciting Aggression in Captive Rhesus Macaques (Macaca mulatta)’, *Folia Primatologica*, 72(5), pp. 268–277. doi:10.1159/000049948.

McCowan, B. *et al.* (2008) ‘Utility of social network analysis for primate behavioral management and well-being’, *Applied Animal Behaviour Science*, 109(2–4), pp. 396–405. doi:10.1016/j.applanim.2007.02.009.

Menon, S. and Poirier, F.E. (1996) ‘Lion-tailed macaques (Macaca silenus) in a disturbed forest fragment: Activity patterns and time budget’, *International Journal of Primatology*, 17(6), pp. 969–985. doi:10.1007/BF02735297.

Muehlenbein, M.P. *et al.* (2002) ‘Morphological and hormonal parameters in two species of macaques: Impact of seasonal breeding’, *American Journal of Physical Anthropology*, 117(3), pp. 218–227. doi:10.1002/ajpa.10030.

Novak, M.A. *et al.* (1998) ‘Effects of puzzle feeders on pathological behavior in individually housed rhesus monkeys’, *American Journal of Primatology*, 46(3), pp. 213–227. doi:10.1002/(SICI)1098-2345(1998)46:3<213::AID-AJP3>3.0.CO;2-L.

Okinda, Cedric *et al.* (2020) ‘A review on computer vision systems in monitoring of poultry:

A welfare perspective', *Artificial Intelligence in Agriculture*, 4, pp. 184–208. doi:10.1016/j.aiia.2020.09.002.

Pons, P., Jaen, J. and Catala, A. (2017) 'Assessing machine learning classifiers for the detection of animals' behavior using depth-based tracking', *Expert Systems with Applications*, 86, pp. 235–246. doi:10.1016/j.eswa.2017.05.063.

Pontzer, H. and Pontzer, H. (2023) 'The provisioned primate : patterns of obesity across lemurs , monkeys , apes and humans'.

Porto, S.M.C. *et al.* (2013) 'A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns', *Biosystems Engineering*, 115(2), pp. 184–194. doi:10.1016/j.biosystemseng.2013.03.002.

Pyke, G.H., Pulliam, H.R. and Charnov, E.L. (1977) 'Optimal Foraging: A Selective Review of Theory and Tests', *The Quarterly Review of Biology*, 52(2), pp. 137–154. doi:10.1086/409852.

Robichaud, D., Lefebvre, L. and Robidoux, L. (1996) 'Dominance affects resource partitioning in pigeons, but pair bonds do not', *Canadian Journal of Zoology*, 74(5), pp. 833–840. doi:10.1139/z96-096.

Rushen, J., Chapinal, N. and de Passillé, A. (2012) 'Automated monitoring of behavioural-based animal welfare indicators', *Animal Welfare*, 21(3), pp. 339–350. doi:10.7120/09627286.21.3.339.

Saito Chiemi (1996) 'Dominance and feeding success in female Japanese macaques,Macaca fuscata: effects of food patch size and inter-patch distance', *Animal Behaviour*, 51(5), pp. 967–980. doi:10.1006/anbe.1996.0100.

van Schaik, C.P. *et al.* (1983) 'The effect of group size on time budgets and social behaviour in wild long-tailed macaques (Macaca fascicularis)', *Behavioral Ecology and Sociobiology*, 13(3), pp. 173–181. doi:10.1007/BF00299920.

Schwitzer, C. and Kaumanns, W. (2001) 'Body weights of ruffed lemurs (Varecia variegata) in European zoos with reference to the problem of obesity', *Zoo Biology*, 20(4), pp. 261–269. doi:10.1002/zoo.1026.

Smith, A., Lindburg, D.G. and Vehrencamp, S. (1989) 'Effect of food preparation on feeding behavior of lion-tailed macaques', *Zoo Biology*, 8(1), pp. 57–65. doi:10.1002/zoo.1430080108.

Soltis, J. *et al.* (1999) 'Female mating strategy in an enclosed group of Japanese macaques', *American Journal of Primatology*, 47(4), pp. 263–278. doi:10.1002/(SICI)1098-2345(1999)47:4<263::AID-AJP1>3.0.CO;2-F.

Southwick, C.H. (1967) 'An Experimental Study of Intragroup Agonistic Behavior in Rhesus

Monkeys (Macaca mulatta)', *Behaviour*, 28(1–2), pp. 182–209. doi:10.1163/156853967X00235.

Steenbeek, R. and Sterck, E.H.M. (1997) 'Female Dominance Relationships and Food Competition in the Sympatric Thomas Langur and Long-Tailed Macaque', *Behaviour*, 134(9–10), pp. 749–774. doi:10.1163/156853997X00052.

Thierry, B., Singh, M. and Kaumanns, W. (2004) *Macaque societies: a model for the study of social organization*. Cambridge University Press.

Thomsen, R. *et al.* (2006) 'How costly are ejaculates for Japanese macaques?', *Primates*, 47(3), pp. 272–274. doi:10.1007/s10329-005-0171-7.

Trapanese, C., Meunier, H. and Masi, S. (2019) 'What, where and when: spatial foraging decisions in primates', *Biological Reviews*, 94(2), pp. 483–502. doi:10.1111/brv.12462.

Valletta, J.J. *et al.* (2017) 'Applications of machine learning in animal behaviour studies', *Animal Behaviour*, 124, pp. 203–220. doi:10.1016/j.anbehav.2016.12.005.

Vogel, E.R. (2005) 'Rank differences in energy intake rates in white-faced capuchin monkeys, Cebus capucinus: the effects of contest competition', *Behavioral Ecology and Sociobiology*, 58(4), pp. 333–344. doi:10.1007/s00265-005-0960-4.

Waasdorp, S. *et al.* (2021) 'Chopped and dispersed food enhances foraging and reduces stress-related behaviours in captive white-naped mangabeys (Cercocebus lunulatus)', *Applied Animal Behaviour Science*, 241, p. 105392. doi:10.1016/j.applanim.2021.105392.

Wolfe, L.D. (1986) 'Sexual strategies of female Japanese macaques (Macaca fuscata)', *Human Evolution*, 1(3), pp. 267–275. doi:10.1007/BF02436584.

Yeager, C.P. (1996) 'Feeding ecology of the long-tailed macaque(Macaca fascicularis) in Kalimantan Tengah, Indonesia', *International Journal of Primatology*, 17(1), pp. 51–62. doi:10.1007/BF02696158.

Zijlmans, D.G.M. *et al.* (2021) 'Retrospective Evaluation of a Minor Dietary Change in Non-Diabetic Group-Housed Long-Tailed Macaques (Macaca fascicularis)', *Animals*, 11(9), p. 2749. doi:10.3390/ani11092749.

Chapter 5. Enhancing Enrichment Evaluation: An Object Detection Approach for Assessing Enrichment Usage in Macaques

5.1 Abstract

Enrichment programs are pivotal for improving the welfare of captive animals, especially non-human primates (NHPs), leading to enhanced physical well-being and more reliable research outcomes. It is crucial to not only establish a well-thought-out enrichment plan but also to regularly assess and fine-tune it. This is important because animals may habituate to constant stimuli, and diligent monitoring of their behavioural responses guarantees that resources are optimally allocated. Although it is crucial to monitor behavioural responses to enrichment, this task can be time intensive. Therefore, automating this process is a significant goal to pursue.

In this project, I employed two distinct computer vision-based pipelines to evaluate monkeys' interactions with different enrichment items: a white tank containing raisins and a non-food-based puzzle. The first pipeline effectively detects and quantifies the movement of the enrichment item, analysing its usage both when containing food and when empty. The second pipeline accurately counts the number of monkeys interacting with the puzzle over time, shedding light on how quickly interest wanes. Through the application of these automated methods, it is evident that the macaques maintain their interest in the white tank enrichment, even after several months since its introduction. Additionally, it is possible to observe a notable attraction towards a new non-food based enrichment, particularly during peak activity periods. This serves as a compelling illustration of how these methodologies can effectively and consistently monitor macaque engagement with enrichments in an automated manner.

5.2 Introduction

Enrichment programs are designed to provide stimulating and engaging experiences for animals in captivity, promoting their physical and psychological well-being. Enrichments serve multiple goals, ranging from reducing abnormal and stereotypical behaviours to promoting more naturalistic activities like exploration, foraging, and play, ultimately leading to improved reproduction and breeding success (Shepherdson, 1994; Newberry, 1995;

Swaisgood and Shepherdson, 2006; Kemp, 2023). As a result, facilities housing wild animals dedicate significant time and effort to plan, introduce, and evaluate enrichment programs. However, not all enrichments are created equal, and careful consideration of the animals' needs, biological nature, history, and the intended purpose of the enrichment is crucial (Kemp, 2023). Practicality, safety, and cost considerations also play a vital role in the decision-making process (Hare, Rich and Worley, 2007; Kemp, 2023).

Despite the extensive literature on enrichment usage in captive settings for non-human primates (NHPs) and other species, generalizations may not always be applicable. The success of an enrichment strategy can be influenced by factors such as enclosure structure, group dynamics, and the animals' past experiences, making individualized approaches essential (Corrine K. Lutz and Novak, 2005; Kemp, 2023).

To effectively implement enrichments, it becomes evident that enrichment is not simply the introduction of novel toys or structures into the enclosure. Rather, it requires careful planning and the establishment of a comprehensive program that considers all these aspects.

Once the enrichment program is designed, the type of enrichment to be utilized becomes a crucial decision. Among the various categories of enrichment, this study will focus on inanimate objects, which encompass structural enrichment (e.g., shelves and platforms), food-based enrichment, sensory enrichment, and cognitive enrichment (Keeling, Alford and Bloomsmith, 1991; Corrine K. Lutz and Novak, 2005; Buchanan-Smith, 2011b; Kemp, 2023).

Structural enrichments play a pivotal role for NHPs, as they enable them to exhibit naturalistic behaviours like climbing, jumping, and exploring, enriching their living environment (de França Santos *et al.*, 2022; Kemp, 2023). However, their continuous availability might make them less interesting or challenging for the animals over time.

On the other hand, food-based enrichment tends to retain the animals' interest and elicit an immediate response, keeping them engaged for longer periods. Moreover, it promotes foraging behaviour, encouraging natural activities (Sha *et al.*, 2012). However, food-based enrichment must be carefully managed, as providing additional food to the animals' diet can lead to nutritional problems such as obesity (Kemp, 2023). Obesity is a common issue in captive primates and is associated with health problems, decreased welfare, and adverse effects on reproduction outcomes (Leigh, 1994; Schwitzer and Kaumanns, 2001; Elwell and Vaglio, 2023).

Another type of enrichment is sensory enrichment, which involves the use of coloured items, objects with different textures, shapes, and smells. NHPs are known to use sensory cues to communicate and interact with their environment (Prescott, 2006; Passarelli, Gamberini and Fattori, 2021). This form of enrichment can stimulate naturalistic behaviours, such as play, visual exploration, and tactile exploration. However, animals may lose interest relatively quickly if a sensory item becomes familiar and lacks novelty, unlike food-based enrichment (Kemp, 2023).

To maintain the animals' engagement in non-food based enrichment, complexity can be added to the items. Cognitive enrichment has the potential to engage NHPs in various tasks, stimulating their problem-solving abilities, tool usage, and exploration of different ways to interact with the objects, and it can even promote cooperation between individuals (Meehan and Mench, 2007; Kemp, 2023).

The effectiveness of enrichment, regardless of its type, needs to be assessed. This is crucial since planning and implementing an enrichment program can be costly and time-consuming (Newberry, 1995). Monitoring the effectiveness of enrichment ensures that resources are being wisely invested and that the program aligns with its intended goals. Assessing the enrichment also provides insights into whether the animals are still interested and engaged with the stimuli, particularly when the enrichment remains a constant presence in their enclosure.

Repeated exposure to a constant object in an animal's environment can result in decreased interest compared to an intermittent object (Kemp, 2023). The phenomenon behind this reduced interest is habituation, a well-documented concept in animal behaviour (Gallistel, 1990; Kuczaj *et al.*, 2002). Thus, it is necessary not only to evaluate the enrichment's effectiveness initially but also to monitor it over time, considering both immediate and prolonged responses for a comprehensive assessment.

Behavioural responses are commonly used to assess the effectiveness of enrichment, particularly since the introduction of stimuli often aims to promote specific behaviours (Kemp, 2023). However, collecting behavioural data from groups of captive animals can be complex and time-consuming, leading to inadequately tested enrichments (Swaisgood and Shepherdson, 2005). Consequently, some stimuli may remain in the enclosures even when their effectiveness is no longer assured.

In this chapter, I will demonstrate how automated methodologies can be used in assessing enrichment use and changes in use over time. Specifically, I will evaluate the use of two different types of enrichment:

1- A food-based enrichment that is consistently present in the enclosure, aiming to assess its effectiveness and how usage changes when the food container is empty or full.
2- A cognitive and sensory enrichment, specifically introduced for this study, with the objective of detecting changes in macaques' interest in it over time.

5.3 Materials and Methods

5.3.1 Facility and subjects

At the Centre for Macaques (CFM), rhesus macaques were housed in socially structured breeding groups consisting of one adult male, multiple females, and their offspring. Once the young monkeys were weaned, they were moved to same-sex peer groups of similar age. Each study group comprised up to 18 individuals and resided in enclosures featuring a play pen area covered in sawdust, which was cleaned every two weeks (Chapter 2, for more details). The play pen was equipped with various enrichments, including structural objects like poles, shelves, and hoses, as well as food-based enrichment. For the purpose of this study, my focus was on two specific enrichments: a white tank filled with raisins and a cognitive puzzle.

The white tank was suspended with a carabiner from one of the horizontal poles in the playpen and was filled with raisins every Monday morning during the regular feeding time for the macaques (Figure 27). The tank featured an open top, prompting the macaques to employ various methods to extract the raisins. They may either shake and overturn the tank from above, from the hose to which it is affixed with a carabiner, or access it from below while positioned on the ground. Raisins were a favoured treat, reserved exclusively for training and human habituation exercises. The quantity of raisins distributed within the tank corresponded to the size of the macaque group. Typically, the macaques demonstrated their adeptness by emptying the tank within a few hours. This enrichment was always present in the playpen and was already part of the enrichment plan when the study commenced. Although it was not possible to determine the exact date of its introduction due to different timelines, schedule changes, and group variations, the white tank was in each group for more than two months before the start of the study, ensuring it was no longer a novelty for the animals.

Figure 27: Example of the white tank containing raisins.

The cognitive puzzle was attached to one of the vertical poles in the play pen. Initially, it consisted of three metal hooks, one with a metal circle attached, one with a blue toy depicting a monkey (blue monkey), and one with several pieces of coloured wood of various shapes (wooden blocks). However, during the study, the blue monkey became damaged during one of the trials and was subsequently replaced with several pieces of coloured wood of different shapes (Figure 28). The puzzle was intentionally designed to encourage macaques to manipulate the metal circle and wooden block around the metal hooks, providing them with exposure to varied materials, colours, and textures. In the case of the blue monkey, it afforded greater movement options because it was attached solely from its top side, and its material properties permitted the macaques to pull and twist it. This enrichment was introduced into the enclosure specifically for the purpose of the study and has not been previously presented to the macaques.

Figure 28:The two enrichment items. On the left is the version with the blue toy and on the right is the version with the wooden blocks.

5.3.2 Data Collected

For this study, video recordings were captured using Axis P1435-LE CCTV cameras, which were installed individually in each play pen. The recorded videos for the white tank enrichment were then analysed on a standard commercial laptop (XPS-15 with a Nvidia Geforce GTX 1650 Ti card). While the videos concerning the puzzle enrichment were analysed with a Scan Systems 3XS Deep Learning DBP G2-18C machines with two Nvidia Geforce RTX 3080 Turbo v2 cards, running Ubuntu 22.04.

White Tank Enrichment

This study involved a dataset comprising information from 9 different macaque groups, consisting of 5 breeding groups (BG) and 4 juvenile groups (JG), with an average group size of 10 (SD = 2.9). Each group was observed and recorded for a total of 3 Mondays (when the white tank contained raisins) and 3 Thursdays (when the tank was empty), all within the same 3 weeks, after the regular feeding time. Monday was selected due to enrichment being filled on that day, while Thursday was chosen to ensure the enrichment was emptied, being several days from Monday. This also ensured that Thursday was not too close to the weekend when staffing levels are reduced and routines differ from weekdays. Each observation period lasted for 6 hours per day.

Puzzle Enrichment

This study involved a dataset comprising information from 12 different macaque groups, consisting of 6 breeding groups (BG) and 6 juvenile groups (JG), with an average group size of 10 (SD = 2.8).

Each group was observed and recorded for a total of 27 days after the puzzle enrichment was introduced to the play pen. Each observation period lasted for 12 hours per day, starting at approximately 7:00. Days involving room cleaning were omitted from the datasets. This process requires temporarily relocating the animals from the play pen to the cage room, thus reducing the available time for interacting with the enrichment.

5.3.3 Automated Methodologies

The methodology used to collect the data for this study is based on computer vision: a process that enable computers to extract information from visual data.

White Tank Enrichment

A specialized model was trained to detect the white tank itself, rather than focusing on the macaques. The white tank’s distinctive bright colour and fixed shape facilitated a more streamlined training process, leading to considerable time savings during both the labelling and training stages. The model is based on the YOLACT algorithm (Bolya *et al.*, 2019) and is capable of detecting the white tank in the CCTV videos in each macaque group (Figure 27). The model was trained on 292 labelled images, with a split of 70% for training and 30% for validation. These images were randomly extracted from a subset of the whole dataset of videos recording the macaques interacting with the tank. However, to test the ability of the model to generalise, one group was excluded from the training set. These were randomly selected and did not include images from every group participating in the study. The model precision, showing the intersection over union (IoU) is shown on

	All	**.50**	**.60**	**.70**	**.80**	**.90**
Box	76.7	100	100	98	88.8	20.1
Mask	82.5	100	100	94.2	92.1	59.5

Table 20. IOU is a measure of the overlap between the predicted bounding box (and mask) and the ground truth bounding box (and mask) (see Chapter 2 for more details).

Figure 29: Model output. The blue square (bbox) contains the detect white tank.

	All	**.50**	**.60**	**.70**	**.80**	**.90**
Box	76.7	100	100	98	88.8	20.1
Mask	82.5	100	100	94.2	92.1	59.5

Table 20: Mean average precision (mAP) for the Enrichment Model. The mAP compares the ground-truth bounding box (and mask) to the detected box (and mask) and returns a score. These comparisons are reported for different IoU thresholds.

To gain insights into the macaques' interaction with the enrichment, the tank movement was used as a proxy for its usage. The object detection model provides coordinates of the bounding box (bbox) detecting the white tank (Figure 29). Extracting the tank's movement involved using an R script that took the x and y bbox coordinates as input and applied a Gaussian Kernel for smoothing. This process significantly enhanced the visualization of the model output in the CSV file.

Next, I computed the average variance difference between sequential x and y values, which helped determine an appropriate threshold for these coordinates. Subsequently, I conducted a convolution operation between the coordinate differences for each frame and the Gaussian kernel. By doing this, I could identify instances where the convolution results exceeded the calculated threshold, indicating significant object movement (Figure 30).

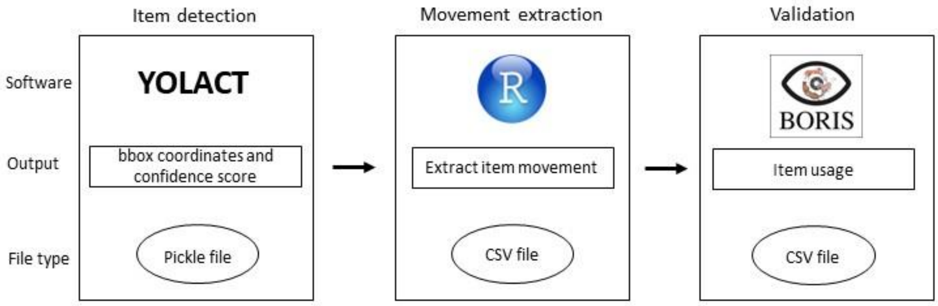


Figure 30: The pipeline begins with the Enrichment model extracting bounding boxes of the white tank. Next, a R script extracts the movement of the detected item. The output of the R script is validated using BORIS.

The validity of using item movement as a proxy for its usage was confirmed by comparing the pipeline's output with the ground truth. The ground truth data were obtained by analysing videos using BORIS (Friard and Gamba, 2016) which allowed for the collection of the actual time the macaques spent interacting with the item. For this validation, six videos were randomly selected from various macaque groups, and a manual analysis was performed, covering more than 111 thousand frames (Figure 31). These manual analyses recorded the time the animals spent interacting with the tank, both when it was moving and when it was not, collecting all the intervals of interaction (Table 21).

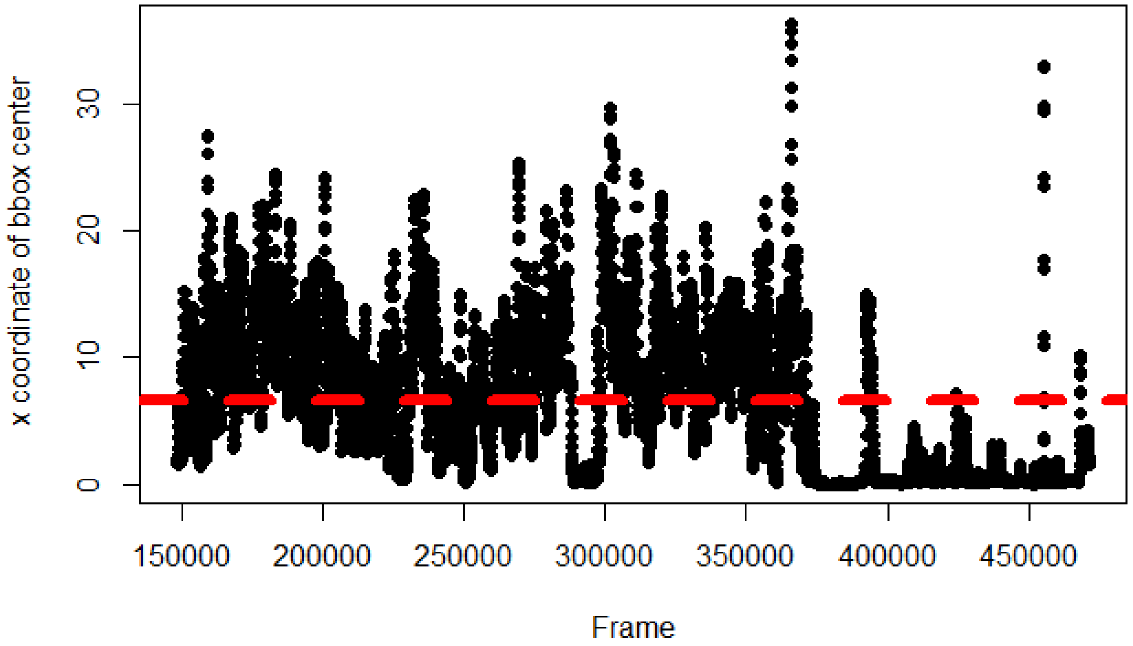


Figure 31: The figure displays the x coordinates of the centre of bounding boxes after being smoothed using the Gaussian kernel. The x-axis represents frame numbers, and the y-axis represents the smoothened x coordinates. The red horizontal line shows the x threshold providing a visual reference of the frames where the object is moving.

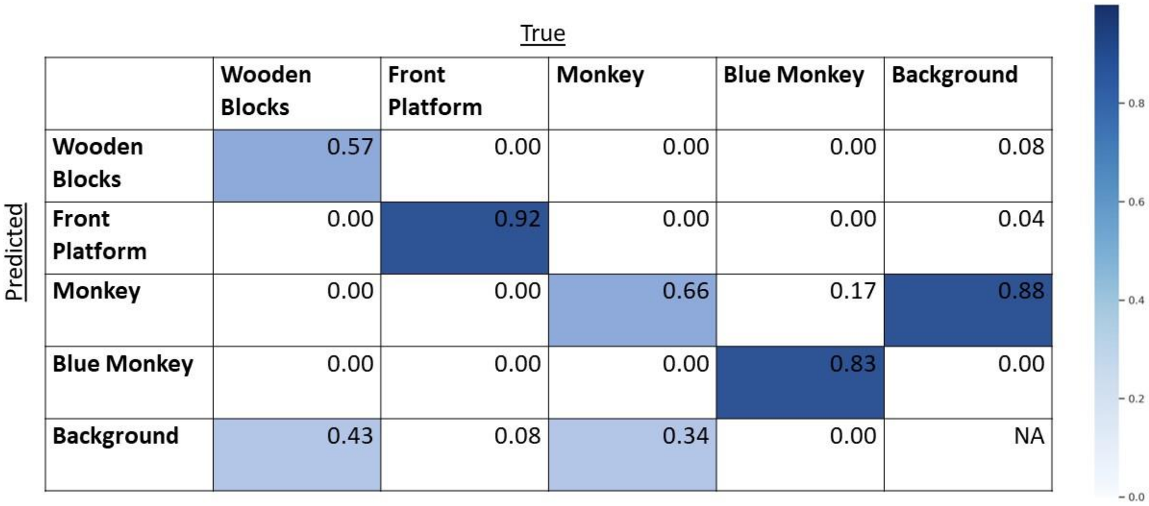
Number of Videos	6
Number of Frames	111,465
True Positive (TP): Instances where the model correctly detects the presence of a macaque	1150
True Negative (TN): Instances where the model correctly detects the absence of a macaque	110,309
False Positive (FP): Instances where the model incorrectly detects a macaque when there is none present	146
False Negative (FN): Instances where the model fails to detect a macaque when one is present	198
Accuracy in assessing enrichment usage $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$	99%
Recall in assessing enrichment usage $\left(\frac{TP}{TP+FN}\right)$	85%
Precision in assessing enrichment usage $\left(\frac{TP}{TP+FP}\right)$	88%

Table 21: The table shows the evaluation of the pipeline. It reports the accuracy, recall and precision of the pipeline in detecting the white tank usage.

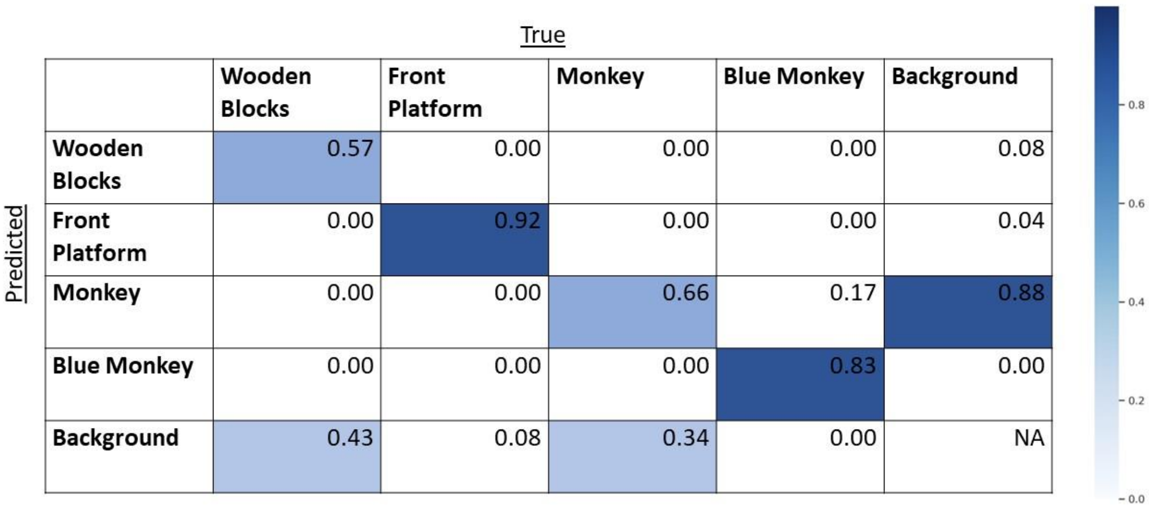
Puzzle Enrichment

In this case, the level of movement observed in the puzzle was relatively minimal compared to the white tank study. Consequently, a different approach was required to assess enrichment usage. To directly quantify individual interactions, a model capable of detecting both the macaques and the enrichment item was employed. The model utilized a combination of YOLOv8 (Terven and Cordova-Esparza, 2023) and Segment Anything Model (SAM) (Kirillov *et al.*, 2023) to label the data used for the training and validation phases (Refer to Chapter 2 for more details). Leveraging SAM for data annotation facilitated a faster and simpler labelling process, while YOLOv8 ensured more accurate detection of the macaques.

The YOLOv8 model underwent training to identify macaques, the wooden platform situated at the front of the enclosure, the blue monkey and the wooden blocks on the enrichment puzzle (Figure 28, Figure 32). The model was trained on a dataset of 258 labelled frames extracted from part of the CCTV system recording different groups of macaques, yielding an overall acceptable performance (Table 22). Notably, as demonstrated in the table below, objects closer to the cameras were detected with greater accuracy. This distinction proved valuable, as our focus was primarily on macaques interacting with the puzzle in the foreground. Conversely, those situated towards the rear of the enclosure, near the window, were not within the scope of this study, and were more prone to being mistaken for background elements (see

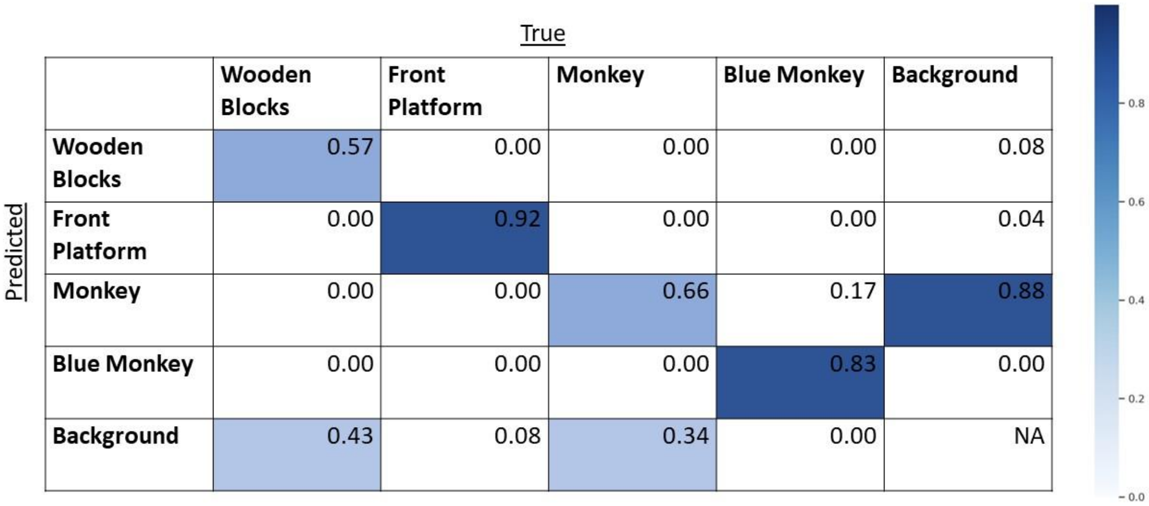


	Wooden Blocks	**Front Platform**	**Monkey**	**Blue Monkey**	**Background**
Wooden Blocks	0.57	0.00	0.00	0.00	0.08
Front Platform	0.00	0.92	0.00	0.00	0.04
Monkey	0.00	0.00	0.66	0.17	0.88
Blue Monkey	0.00	0.00	0.00	0.83	0.00
Background	0.43	0.08	0.34	0.00	NA



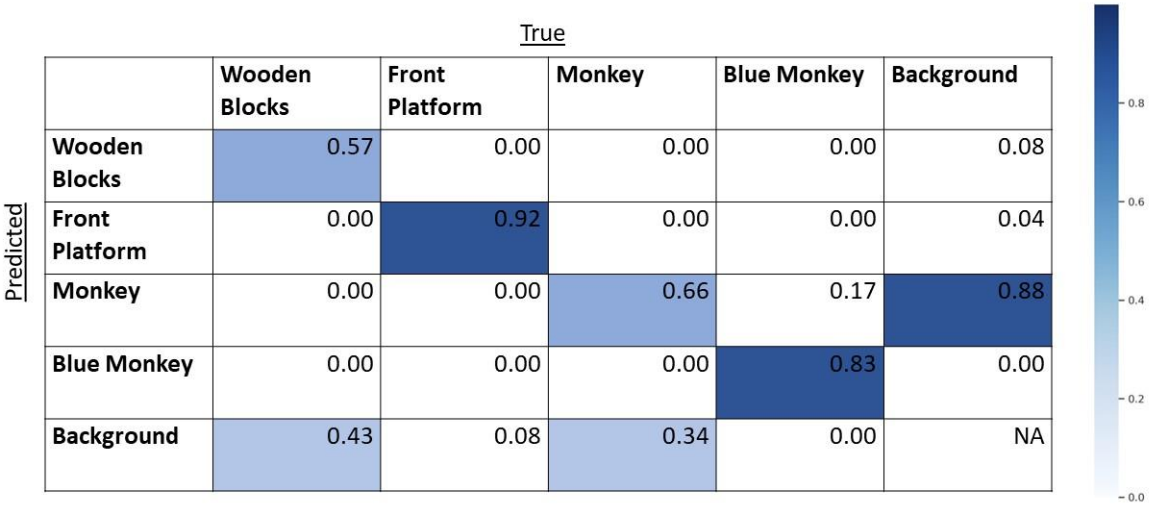
	Wooden Blocks	Front Platform	Monkey	Blue Monkey	Background
Wooden Blocks	0.57	0.00	0.00	0.00	0.08
Front Platform	0.00	0.92	0.00	0.00	0.04
Monkey	0.00	0.00	0.66	0.17	0.88
Blue Monkey	0.00	0.00	0.00	0.83	0.00
Background	0.43	0.08	0.34	0.00	NA

Table 22) due to their distance from the camera, presenting a challenge even for a human observer. Additionally, stable objects such as the blue monkey and the front platform exhibited the highest levels of accuracy (83% and 92%, see



	Wooden Blocks	Front Platform	Monkey	Blue Monkey	Background
Wooden Blocks	0.57	0.00	0.00	0.00	0.08
Front Platform	0.00	0.92	0.00	0.00	0.04
Monkey	0.00	0.00	0.66	0.17	0.88
Blue Monkey	0.00	0.00	0.00	0.83	0.00
Background	0.43	0.08	0.34	0.00	NA

Table 22).



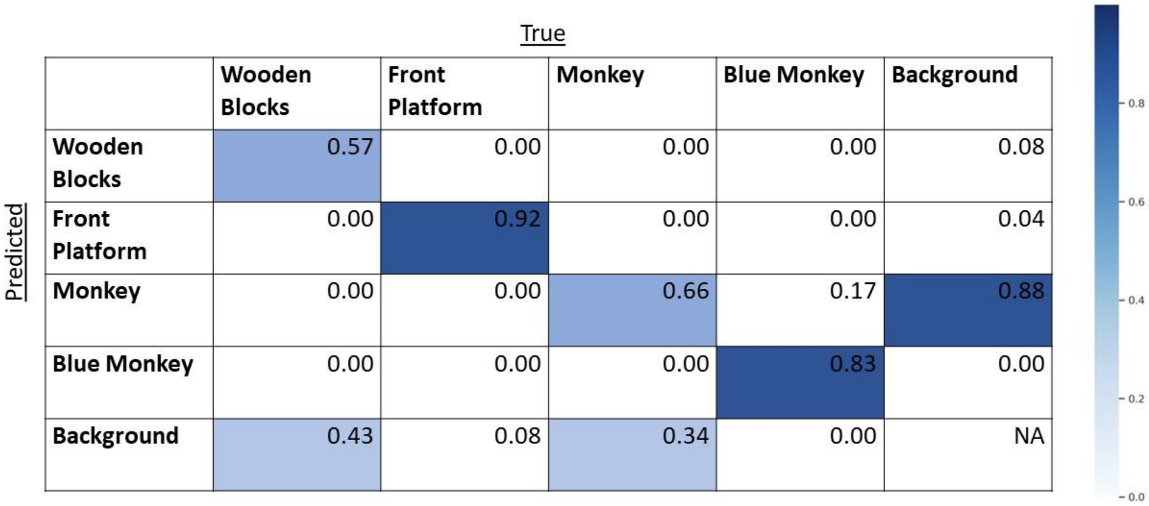
True

Predicted

	Wooden Blocks	Front Platform	Monkey	Blue Monkey	Background
Wooden Blocks	0.57	0.00	0.00	0.00	0.08
Front Platform	0.00	0.92	0.00	0.00	0.04
Monkey	0.00	0.00	0.66	0.17	0.88
Blue Monkey	0.00	0.00	0.00	0.83	0.00
Background	0.43	0.08	0.34	0.00	NA

Table 22: Confusion Matrix representing the accuracy of the model in detecting each object. Each row represents a specific class of objects that the model is trying to detect. Each column represents the classes that the model predicts. The diagonal elements of the matrix represent the cases where the model made correct predictions. The off-diagonal elements represent cases where the model made wrong prediction. On the axis there are the 4 classes of objects (Wooden Blocks from the enrichment puzzle, Front Platform, Monkey, Blue Monkey from the enrichment puzzle) and the background. The background are objects that do not belong to either of the classes but detected as one of them (false positive).

Due to unsatisfactory wooden block detection (57% accuracy, see Table 22), any macaque engaging with the enrichment was considered within the region of interest (ROI) surrounding the enrichment. Moreover, as the enrichment was positioned behind the front platform, to prevent counting macaques on the platform as individuals



True

Predicted

	Wooden Blocks	Front Platform	Monkey	Blue Monkey	Background
Wooden Blocks	0.57	0.00	0.00	0.00	0.08
Front Platform	0.00	0.92	0.00	0.00	0.04
Monkey	0.00	0.00	0.66	0.17	0.88
Blue Monkey	0.00	0.00	0.00	0.83	0.00
Background	0.43	0.08	0.34	0.00	NA

within the ROI, those whose area overlapped with the front platform were excluded from the count of individuals interacting with the enrichment (Figure 33).

The object detection model captures the count of interacting macaques with the enrichment in a single frame per second, recorded at a frame rate of 15 frames per second. This results in a substantial dataset, yielding a total of 3600 data points per hour of video. To streamline this dataset, an R script was employed to calculate the percentage of macaques engaging with the enrichment for each hour of observation.

The pipeline's accuracy was further assessed by comparing the number of macaques detected within the ROI with those identified by a human observer. In this analysis, one video was randomly selected from each of the 12 groups. For a total of 13 frames for each video, taken at hourly intervals, both the model and the human observer independently counted the number of individuals within the ROI. The model achieved an accuracy of 87.5% in detecting macaques within the ROI (

Number of Videos	12
Number of Frames	156
True Positive (TP)	45
True Negative (TN)	102
False Positive (FP)	10
False Negative (FN)	11
Accuracy in detecting the macaques in the ROI $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$	87.5%

Table 23).

Number of Videos	12
Number of Frames	156
True Positive (TP)	45
True Negative (TN)	102
False Positive (FP)	10
False Negative (FN)	11
Accuracy in detecting the macaques in the ROI $\left(\frac{TP+TN}{TP+TN+FP+FN} \right)$	87.5%

Table 23: The table shows the validation of the pipeline. It reports the accuracy of the model in detecting the macaques in the region of interest (ROI) around the enrichment puzzle.

Figure 32: Frame extracted from a YOLOv8 model output video. The objects enclosed in black boxes are the ones utilized to gather information regarding the enrichment interaction.

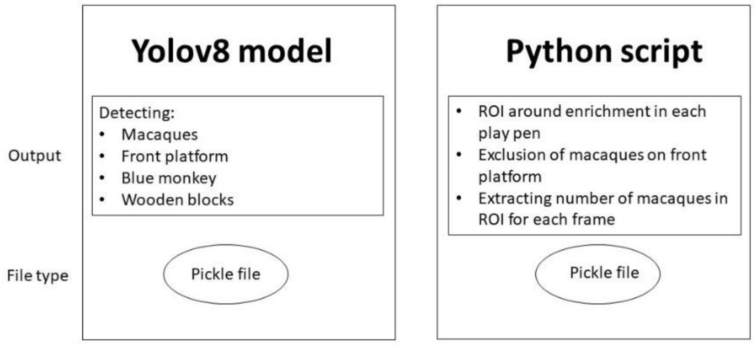


Figure 33: The pipeline begins with the Enrichment model based on Yolov8 detecting several objects. Next, a Python script extracts the number of macaques in the ROI, excluding the one positioned on the front platform.

5.3.4 Statistical Analyses

To analyse the output data from the two pipelines, I utilized R Studio (version 4.0.4; R Core Team 2021) on Windows 10. All statistical analyses were two-tailed, and an alpha level of

0.05 was set. A linear mixed model analysis was employed to examine the data, allowing for the incorporation of both fixed and random effects to account for potential sources of variation within the dataset. To test the model assumptions histograms, normal probability plots of residuals, and quantile-quantile plots were employed (Hartig, 2018). The lmer function from the lme4 package was used to model the interaction data (Bates, D and Maechler, M and Bolker, BM and Walker, 2014).

White Tank Enrichment

In this study, to account for potential instances of missed tank detection (false negative), movement was quantified as a percentage of the tank's motion when it was reliably detected by the model.

$$\frac{Number\ of\ frames\ where\ the\ tank\ is\ moving}{Number\ of\ frames\ where\ the\ tank\ is\ detected}$$

Consequently, the dependent variable was the percentage of tank movement defined with the above equation. While the independent variables focused on specific days of the week, specifically Mondays and Thursdays. Furthermore, the model incorporated fixed factors such as group size and group type (breeding group or juvenile group), while group identity was treated as a random factor.

Puzzle Enrichment

In this study, to control for group size, the number of monkeys interacting with the items was adjusted by dividing the number of detected animals in each frame by the number of macaques in the group. Therefore, the dependent variable was the percentage of monkeys interacting with the item, and the independent variables were the days since the enrichment was added in the group and the presence/absence of the blue monkey. Furthermore, the model incorporated fixed factors such as group size and group type (breeding group or juvenile group), weekday and hour of the day. Group identity was treated as a random factor.

In order to address a high number of zero values in the dataset—reflecting periods when the macaques were non-interactive with the enrichment—a logarithmic transformation was applied following the equation: $x = \log(y + 1)$

Where y were the percentage of macaques interacting with the enrichment.

This transformation was employed to mitigate the skewness caused by the presence of zeros and to ensure the data met the assumptions required for statistical analyses.

5.4 Results

5.4.1 White Tank Study

The object detection-based pipeline yielded significant findings in the percentage of item movement (ratio of frames in which the item was detected in motion to the total number of frames in which the item was detected) between Mondays, when the white tank was filled with raisins, and Thursdays, when the white tank remained empty. In total, the study accumulated 324 hours of data, which was subsequently analysed using an automated pipeline over a span of 3 days.

A significant difference (F (2,150) = 60, $p < 0.001$) was observed in the percentage of time where the item was in motion during the first hour after feeding time compared to the subsequent hours (Figure 34). However, no significant difference (F (1,7) = 0.1, p = 0.7) in item interaction was observed between Breeding groups and Juvenile groups (Figure 35).

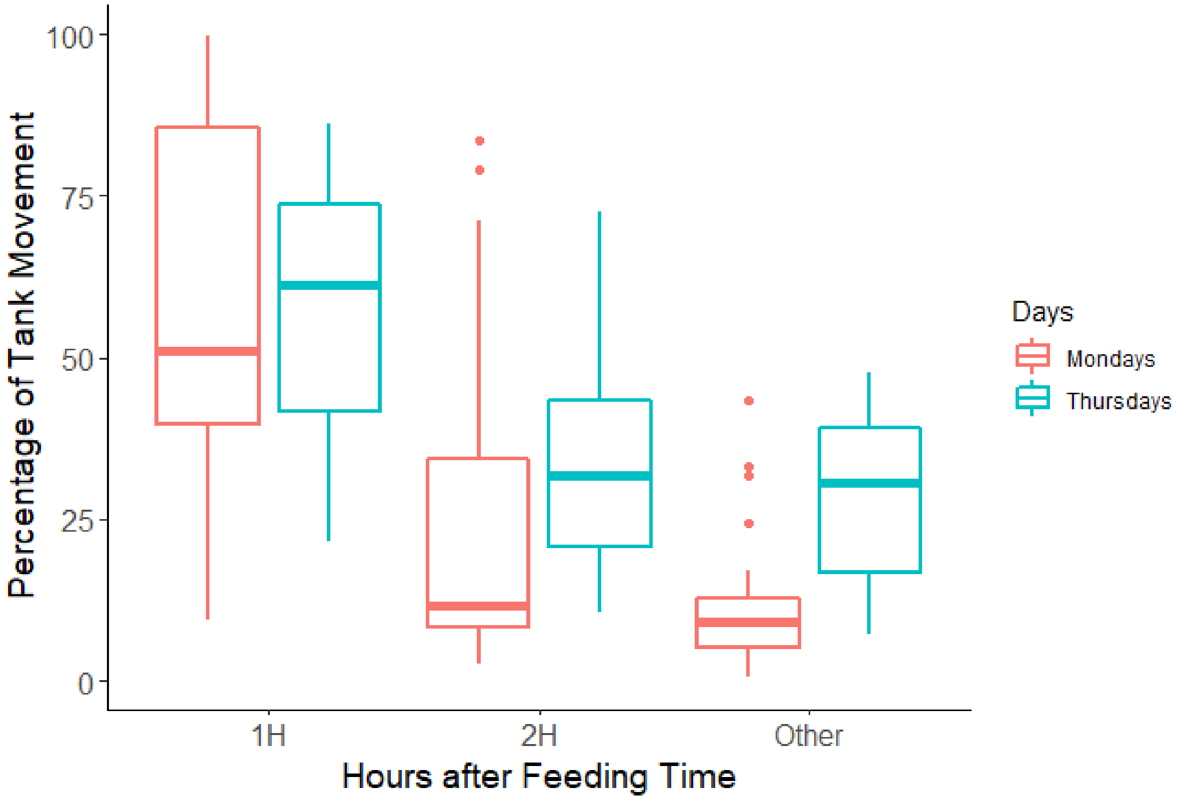


Figure 34: Percentage of frames where the tank was moving on Mondays (pink) and Thursdays (blue). The percentage are categorised into three-time intervals: first hour after feeding time (1H), two hours after feeding time (2H), last hours of observation (Other).

The results demonstrate that during the initial hour after feeding time on both days, there was a higher interaction with the item, which gradually decreased in the subsequent hours. Moreover, overall item movement was significant higher (F (2,150) = 8, p = 0.005) on Thursdays when the tank was empty, for both breeding groups and juvenile groups (Figure 35, see Table 24 for effect size information).

Group type	Mean Difference	Confidence Interval (95%)
Breeding Group	0.183	[0.038, 0.29]
Juvenile Group	0.34	[0.24, 0.43]

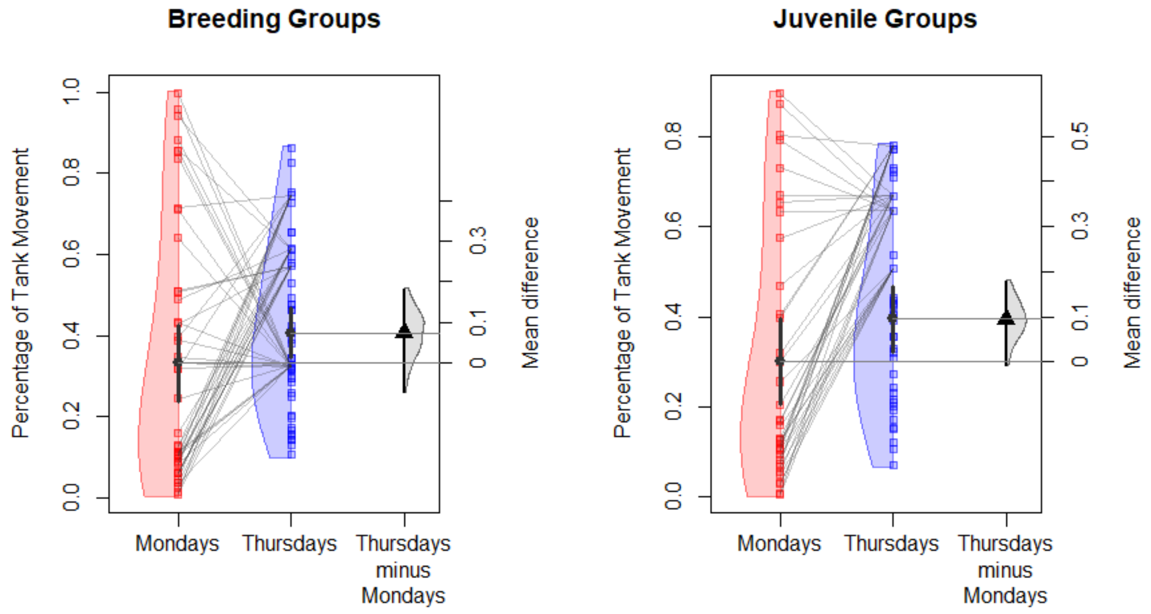


Figure 35: Percentage of frames where the white tank was in motion, categorized separately by group type: Breeding group (on the left) and Juvenile group (on the right).

Group type	Mean Difference	Confidence Interval (95%)
Breeding Group	0.183	[0.038, 0.29]

Juvenile Group	0.34	[0.24, 0.43]

Table 24: Mean Difference in white tank movement and Confidence Intervals for the different months.

5.4.2 Puzzle Enrichment

The object detection-based model analysed 303 days of observations for a total of 3636 hours of data and required just a single hour to analyse the content of a single day of video recordings. The pipeline identified a decreasing trend in the number of macaques interacting with the puzzle in the month after its introduction (F (1,3043) = 132, $p < 0.001$) (Figure 36).

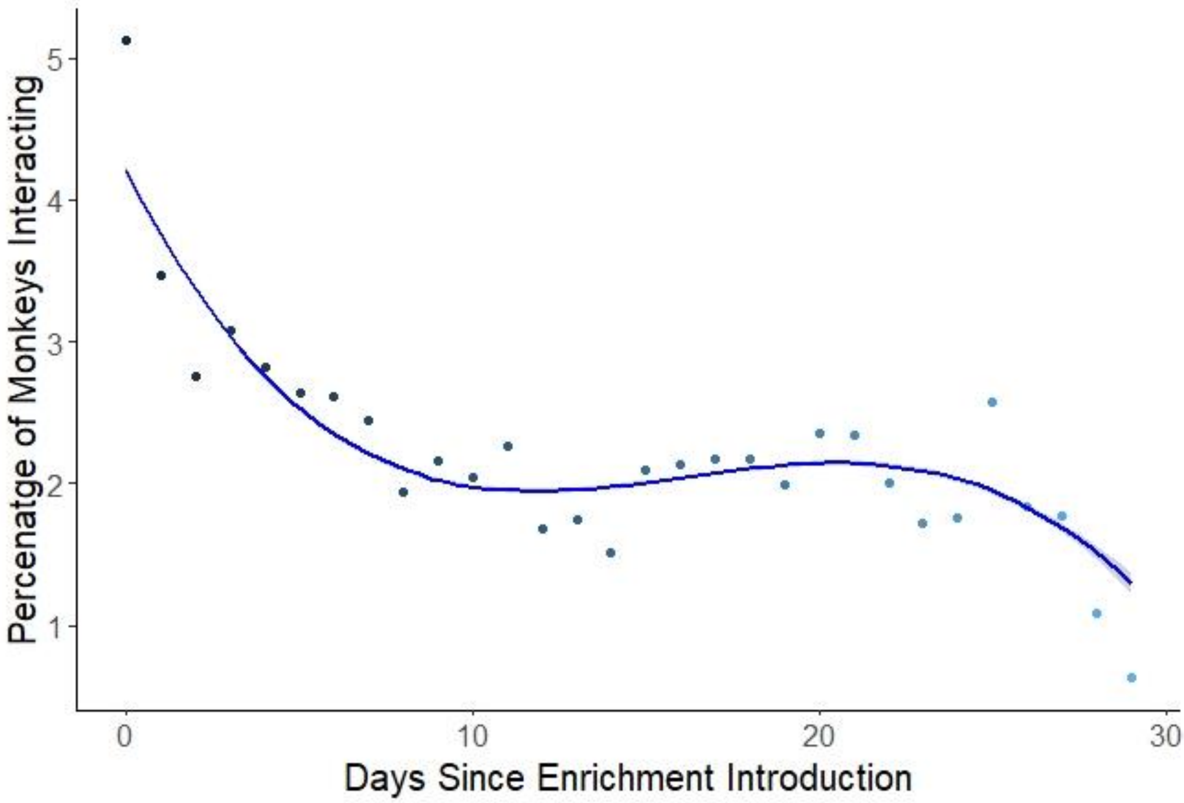


Figure 36: Percentage of monkeys interacting with the enrichment over the 27 days of observations.

Furthermore, the automated methodology identified a non-significant increase in puzzle engagement for the variant containing the blue monkey, compared to the one featuring only wooden blocks (F (1,5.1) = 5.3, p = 0.06) (Figure 37).

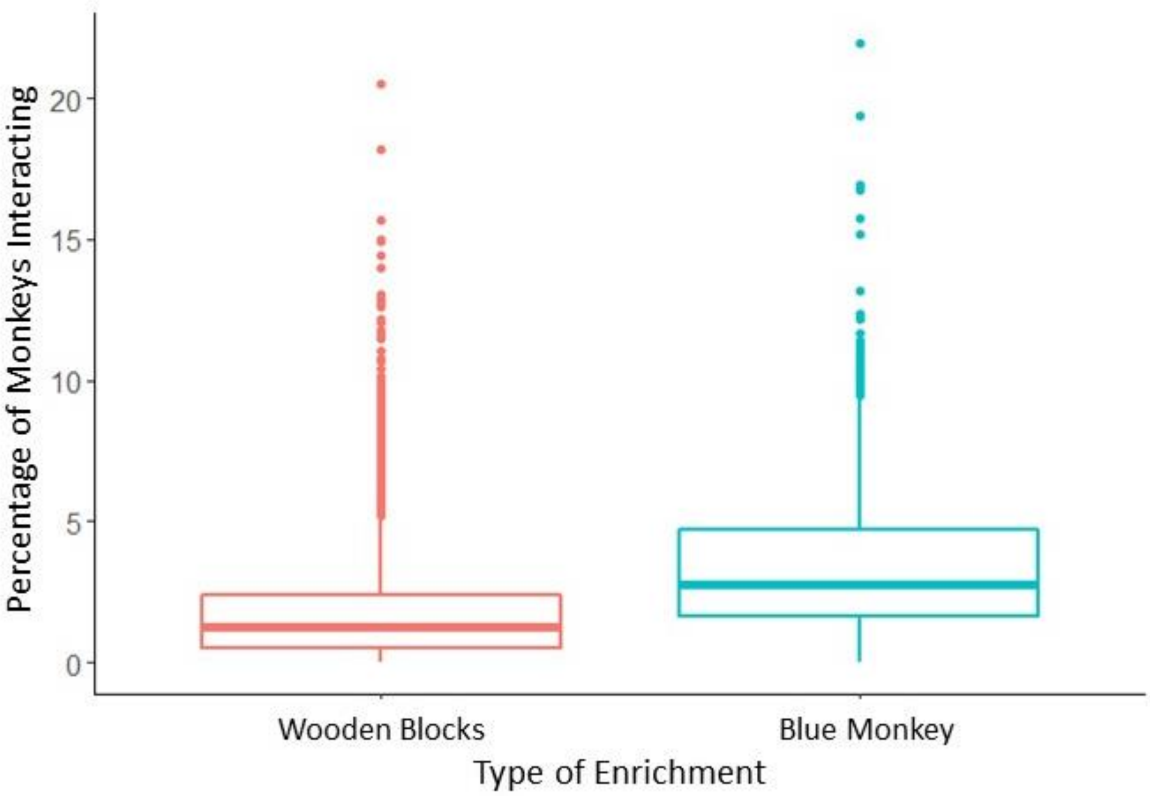


Figure 37: Percentage of monkeys interacting with the puzzle enrichment only featuring the wooden blocks and the one with the blue monkey.

The group size also had a significant negative effect on the percentage of macaques interacting with the enrichment, with bigger groups interacting less with the puzzle (F (1, 19) = 7.2, p = 0.01) (Figure 36).

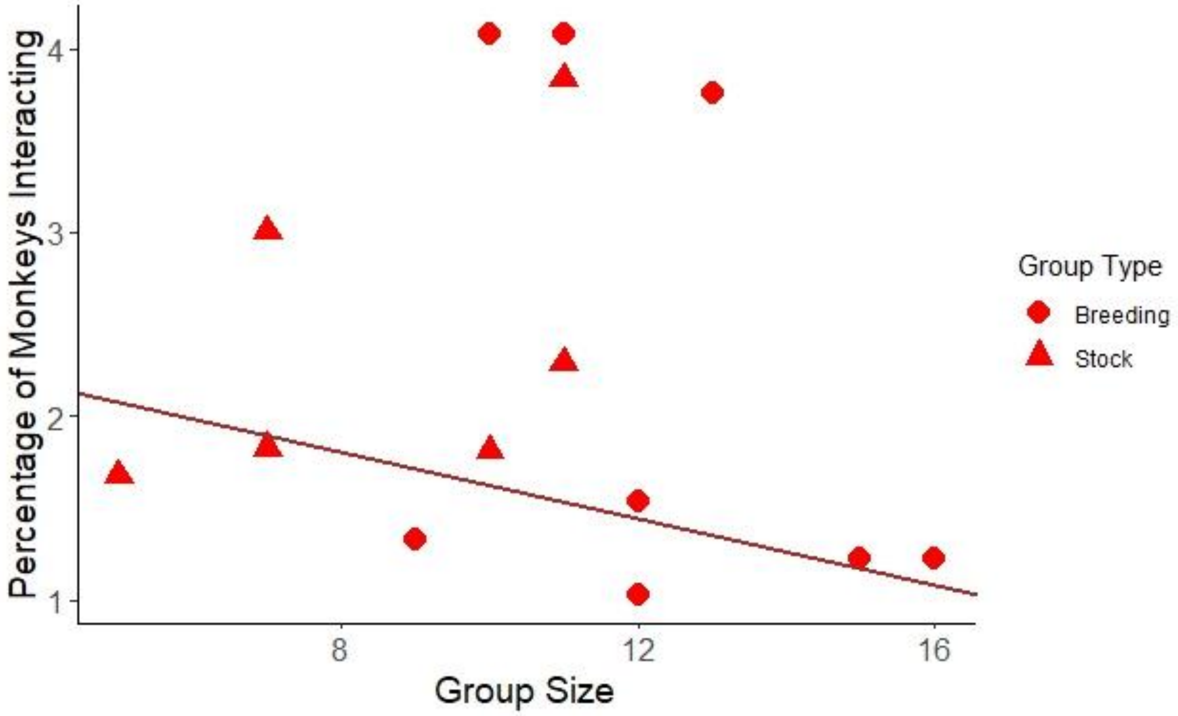


Figure 36: Percentage of Macaques Interacting with Puzzle Enrichment. Circles denote breeding groups; triangles represent stock groups.

The day of the week and the hour of the day also influenced the enrichment usage, with the puzzle being used less during the weekend and later in the evening (F (1, 4054) = 2.2, p = 0.03; F (1, 4050) = 273, p < 0.001) (Figure 37 and Table 25).

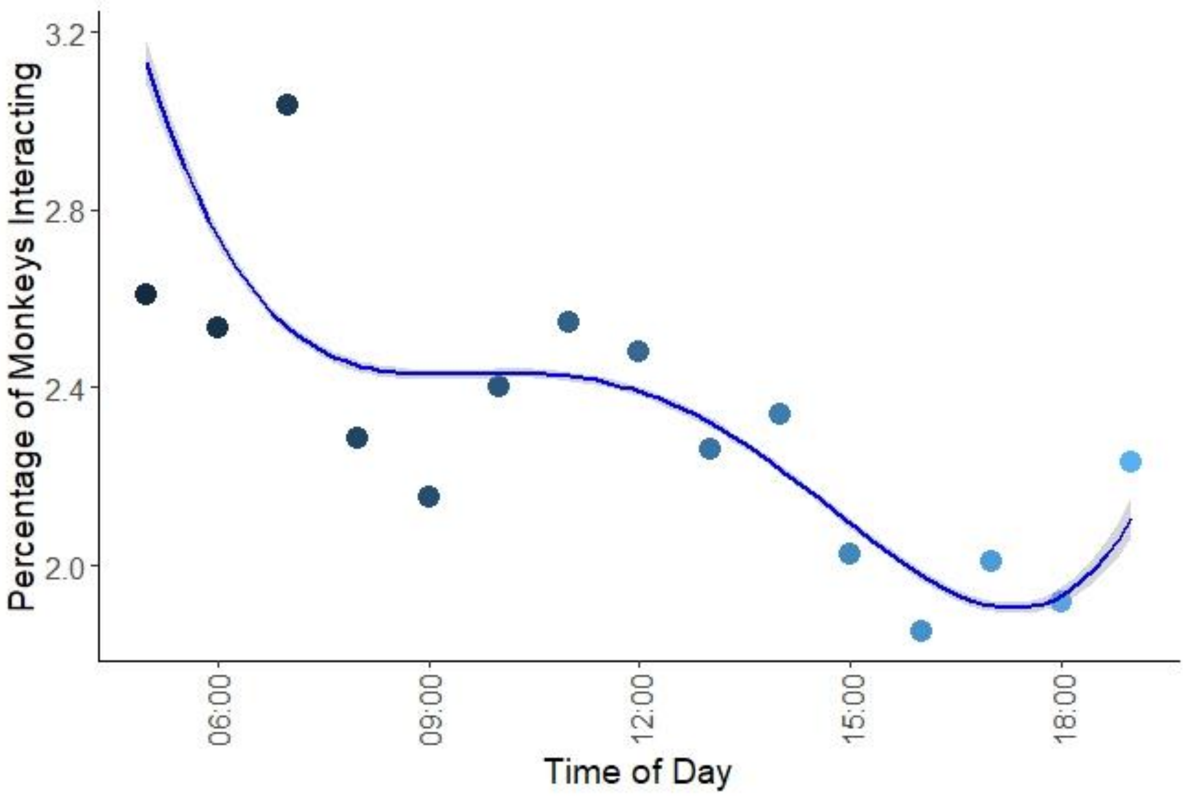


Figure 37: Percentage of macaques per each group interacting with the puzzle enrichment grouped by time of day (from 7:00 to 19:00).

Day of the Week	**Percentage of Monkeys Interacting**
Monday	2.20
Tuesday	2.36
Wednesday	2.43
Thursday	2.55
Friday	2.34
Saturday	2.10
Sunday	2.09

Table 25: Table showing percentage of monkeys interacting with the puzzle across days of the week.

5.5 Discussion

Both models demonstrated the capability to automatically detect macaque interactions with enrichments effectively. The approaches, one focusing on detecting the white tank and tracking its movements, and the other identifying macaques in close proximity to the puzzle enrichment, exhibited high accuracy and provided valuable insights. These insights will be used to enhance enrichment planning and identify when enrichments cease to be engaging, and macaques lose interest in them.

Even with the extended presence of the tank in the enclosures, the macaques continue to actively engage with and utilize the enrichment. Notably, the animals exhibited heightened interaction with the tank when it was devoid of contents. It is important to highlight that the tank is replenished subsequent to feeding sessions. On both days, irrespective of its contents, there was an elevated level of engagement shortly post-feeding compared to subsequent hours. However, in instances where the item was empty, macaque interaction surpassed that observed when it was filled. This heightened interaction during emptiness may be attributed to intensified efforts in extracting food from the tank, potentially stemming from the frustration induced by its empty state even after the designated feeding period. In fact, even though the tank is not removed when empty, this behaviour could be associated with a form of enrichment removal, which is known to be particularly frustrating and can reduce the satisfaction derived from low-reward enrichment (Amsel, 1958; Papini, 2003; Latham and Mason, 2010). However, since the animals at CFM consistently live in enriched enclosures, this change in how the white tank enrichment is presented is unlikely to impact their overall welfare (Latham and Mason, 2010).

While one may posit that if the tank consistently contains raisins only on Mondays, the animals at CFM should anticipate this routine, empirical evidence suggests that regular feeding times do not necessarily render routines reliably predictable (Waitt and Buchanan-Smith, 2001a). The anticipation of food appears to be elicited by cues preceding the feeding event, indicating that animals may form associations beyond strict temporal patterns (Waitt and Buchanan-Smith, 2001a). In essence, if the animals come to link the presence of raisins in the white tank with feeding time, an expectation of its replenishment on a daily basis may arise.

In summation, the white tank enrichment continues to captivate the macaques at CFM, eliciting species-specific behaviours, including manipulation and foraging. Importantly, these behaviours persist even in the absence of high-caloric food items like raisins, suggesting that the enrichment itself may serve as a stimulating factor independent of daily food provision.

Contrastingly, it is well-documented that straightforward toys experience a rapid decline in usage among NHPs (Bayne *et al.*, 1993; C. K. Lutz and Novak, 2005; Kemp, 2023). Therefore, when considering non-food-based enrichment, the ability to monitor animals' interest in these items is paramount. The employment of this automated pipeline enabled the detection of an initial decline in interest towards the puzzle in the majority of groups, followed by a subsequent surge of interaction towards the conclusion of the four-week experiment. Nevertheless, the results also evince variability in the percentage of macaques engaging with the puzzle across different groups, aligning with numerous other studies affirming that age, sex, and individual differences exert an influence on non-food based enrichment utilization (C. K. Lutz and Novak, 2005).

Additionally, factors such as destructibility, complexity, physical alterability, texture, manipulability, colour, and size of the objects may influence their utility (Pruetz and Isbell, 2000). The model successfully discerned disparities among various types of puzzles, illustrating how the presence of a brightly coloured rubber toy (i.e., the blue monkey) amplified interactions with the puzzle, even if not significantly. The blue monkey was brightly coloured and stood out distinctly from the neutral-toned wooden blocks, both in texture and appearance. Its design, featuring distinctive facial features like eyes, added an extra layer of visual interest. This observation suggests that puzzles with more complexity and a variety of toys might be more engaging for the macaques (Schapiro and Bloomsmith, 1995; Kemp, 2023).

The pipeline shows a decrease in percentage of macaques interacting with the puzzle during feeding time (8-9 am) and in the afternoon/evening (3-6 pm) while the macaques' group is known to be resting. In addition, the pipeline brought to light a notable decline in puzzle usage, particularly during weekends and late afternoons. This phenomenon can be attributed to reduced staffing levels and a generally quieter atmosphere at CFM during these times. Existing studies on non-human primates in zoo environments consistently demonstrate that their behaviour is influenced by human presence (Hosey and Druck, 1987; Chamove, Hosey and Schaetzel, 1988; Wells, 2005). Specifically, these animals tend to exhibit increased activity and spend more time near the front of their enclosures when there is a greater influx of visitors. Conversely, in times of low human presence, they tend to allocate more time to rest and relaxation (Wells, 2005). This dual effect likely contributes to both the weekend and late afternoon reductions in enrichment interaction observed at CFM, coinciding with lower staff presence. Using the YOLOv8-based pipeline will enable gathering of precise data on the positions of macaques within the play pen. This will facilitate an examination of whether the animals are indeed allocating more time to established resting zones, such as the front and back platforms.

In the case of the puzzle enrichment, it was observed that larger groups had a lower percentage of individuals interacting with the item. This phenomenon may be attributed to heightened competition and aggression within these larger macaque groups, potentially leading to a monopolization of the item (C. K. Lutz and Novak, 2005; Olsson and Westlund, 2007). Additionally, it is well-established that group size influences the time allocation of primates in social settings. They tend to invest more time in actions like grooming, social interactions, and maintaining vigilance, subtracting time to other activities (Lehmann, Korstjens and Dunbar, 2007). This suggests that in larger groups, providing multiple puzzle enrichments may better accommodate to their needs.

Despite prevailing studies suggesting that young animals tend to exhibit greater interest in novelty and enrichment (Schapiro and Bloomsmith, 1995; Schapiro *et al.*, 1996; C. K. Lutz and Novak, 2005), the current project did not yield any significant difference in enrichment interaction between group types. This could potentially be attributed to the presence of young individuals within the breeding group as well as in the juvenile groups.

In conclusion, the utilization of this automated pipelines not only enables the detection of macaque interactions with enrichments but also facilitates comparative analyses of various enrichments and their impact on different groups. Furthermore, the results can be harnessed to enhance the management of the enrichment program and formulate a more effective strategy tailored to the specific requirements and preferences of each group.

5.6 Chapter References

Amsel, A. (1958) ‘The role of frustrative nonreward in noncontinuous reward situations.’, *Psychological Bulletin*, 55(2), pp. 102–119. doi:10.1037/h0043125.

Bates, D. *et al.* (2015) ‘Fitting Linear Mixed-Effects Models Using lme4’, *Journal of Statistical Software*, 67(1 SE-Articles), pp. 1–48. doi:10.18637/jss.v067.i01.

Bayne, K.A.L. *et al.* (1993) ‘Kong [R] Toys for Laboratory Primates: Are They Really an Enrichment or Just Fomites?’, *Laboratory Animal Science*, 43, p. 78.

Bolya, D. *et al.* (2019) ‘Yolact: Real-time instance segmentation’, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166.

Buchanan-Smith, H.M. (2011) ‘Environmental enrichment for primates in laboratories’, *Advances in Science and Research*, 5(1), pp. 41–56. doi:10.5194/asr-5-41-2010.

Chamove, A.S., Hosey, G.R. and Schaetzel, P. (1988) ‘Visitors excite primates in zoos’, *Zoo Biology*, 7(4), pp. 359–369. doi:10.1002/zoo.1430070407.

Elwell, E.J. and Vaglio, S. (2023) ‘The Scent Enriched Primate’, *Animals*, 13(10), p. 1617. doi:10.3390/ani13101617.

de França Santos, M. *et al.* (2022) ‘Welfare Improvement by Enrichment Programs in Common Marmoset Females Under Social Isolation’, *Journal of Applied Animal Welfare Science*, 25(3), pp. 297–309. doi:10.1080/10888705.2021.1968863.

Friard, O. and Gamba, M. (2016) ‘BORIS : a free, versatile open-source event-logging software for video/audio coding and live observations’, *Methods in Ecology and Evolution*. Edited by R. Fitzjohn, 7(11), pp. 1325–1330. doi:10.1111/2041-210X.12584.

Gallistel, C.R. (1990) *The organization of learning.* The MIT Press.

Hare, V.J., Rich, B. and Worley, K.E. (2007) ‘Enrichment gone wrong’, in *Proceedings of the Eighth International Conference on Environmental Enrichment*, pp. 35–45.

Hartig, F. 2018. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. Version 0.2.0.

Hosey, G.R. and Druck, P.L. (1987) ‘The influence of zoo visitors on the behaviour of captive primates’, *Applied Animal Behaviour Science*, 18(1), pp. 19–29. doi:10.1016/0168-1591(87)90251-6.

Keeling, M.E., Alford, P.L. and Bloomsmith, M.A. (1991) ‘Decision analysis for developing programs of psychological well-being: A bias-for-action approach.’, in *Through the looking glass: Issues of psychological well-being in captive nonhuman primates.* Washington, DC, US: American Psychological Association, pp. 57–65. doi:10.1037/10080-007.

Kemp, C. (2023) ‘Enrichment’, in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 463–500. doi:10.1007/978-3-030-82708-3_20.

Kirillov, A. *et al.* (2023) ‘Segment Anything’, *arXiv:2304.02643* [Preprint].

Kuczaj, S. *et al.* (2002) ‘Keeping Environmental Enrichment Enriching’, *International Journal of Comparative Psychology*, 15(2). doi:10.46867/C4XK5N.

Latham, N. and Mason, G. (2010) ‘Frustration and perseveration in stereotypic captive animals: Is a taste of enrichment worse than none at all?’, *Behavioural Brain Research*, 211(1), pp. 96–104. doi:10.1016/j.bbr.2010.03.018.

Lehmann, J., Korstjens, A.H. and Dunbar, R.I.M. (2007) ‘Group size, grooming and social cohesion in primates’, *Animal Behaviour*, 74(6), pp. 1617–1629.

Leigh, S.R. (1994) ‘Ontoaenetic correlates of diet in anthropoid primates’, *American Journal* *of Physical Anthropology*, 94(4), pp. 499–522. doi:10.1002/ajpa.1330940406.

Lutz, C. K. and Novak, M.A. (2005) 'Environmental Enrichment for Nonhuman Primates: Theory and Application', *ILAR Journal*, 46(2), pp. 178–191. doi:10.1093/ilar.46.2.178.

Lutz, Corrine K. and Novak, M.A. (2005) *Primate Natural History and Social Behavior. Implications for Laboratory Housing*, *The Laboratory Primate*. Elsevier Limited.

Meehan, C.L. and Mench, J.A. (2007) 'The challenge of challenge: Can problem solving opportunities enhance animal welfare?', *Applied Animal Behaviour Science*, 102(3–4), pp. 246–261. doi:10.1016/j.applanim.2006.05.031.

Newberry, R.C. (1995) 'Environmental enrichment: Increasing the biological relevance of captive environments', *Applied Animal Behaviour Science*, 44(2–4), pp. 229–243. doi:10.1016/0168-1591(95)00616-Z.

Olsson, I.A.S. and Westlund, K. (2007) 'More than numbers matter: The effect of social factors on behaviour and welfare of laboratory rodents and non-human primates', *Applied Animal Behaviour Science*, 103(3–4), pp. 229–254.

Papini, M.R. (2003) 'Comparative Psychology of Surprising Nonreward', *Brain, Behavior and Evolution*, 62(2), pp. 83–95. doi:10.1159/000072439.

Passarelli, L., Gamberini, M. and Fattori, P. (2021) 'The superior parietal lobule of primates: a sensory-motor hub for interaction with the environment', *Journal of Integrative Neuroscience*, 20(1), p. 157. doi:10.31083/j.jin.2021.01.334.

Prescott, M. (2006) 'Primate sensory capabilities and communication signals: implications for care and use in the laboratory'.

Pruetz, J.D. and Isbell, L.A. (2000) 'Correlations of food distribution and patch size with agonistic interactions in female vervets (Chlorocebus aethiops) and patas monkeys (Erythrocebus patas) living in simple habitats', *Behavioral Ecology and Sociobiology*, 49(1), pp. 38–47. doi:10.1007/s002650000272.

Schapiro, S.J. *et al.* (1996) 'Enrichment effects on rhesus monkeys successively housed singly, in pairs, and in groups', *Applied Animal Behaviour Science*, 48(3–4), pp. 159–171.

Schapiro, S.J. and Bloomsmith, M.A. (1995) 'Behavioral effects of enrichment on singly-housed, yearling rhesus monkeys: An analysis including three enrichment conditions and a control group', *American Journal of Primatology*, 35(2), pp. 89–101. doi:10.1002/ajp.1350350202.

Schwitzer, C. and Kaumanns, W. (2001) 'Body weights of ruffed lemurs (Varecia variegata) in European zoos with reference to the problem of obesity', *Zoo Biology*, 20(4), pp. 261–269. doi:10.1002/zoo.1026.

Sha, J. *et al.* (2012) ‘Effects of single-use and group-use enrichment on stereotypy and intragroup aggressive and affiliative behaviors of a social group of squirrel monkeys (Saimiri sciureus) at the Singapore Zoo’, *Journal of Applied Animal Welfare Science*, 15(4), pp. 358–371.

Shepherdson, D. (1994) ‘The role of environmental enrichment in the captive breeding and reintroduction of endangered species’, in *Creative Conservation*. Dordrecht: Springer Netherlands, pp. 167–177. doi:10.1007/978-94-011-0721-1_8.

Swaisgood, R. and Shepherdson, D. (2006) ‘Environmental enrichment as a strategy for mitigating stereotypies in zoo animals: a literature review and meta-analysis.’, *Stereotypic animal behaviour: fundamentals and applications to welfare*, pp. 256–285.

Swaisgood, R.R. and Shepherdson, D.J. (2005) ‘Scientific approaches to enrichment and stereotypies in zoo animals: what’s been done and where should we go next?’, *Zoo Biology*, 24(6), pp. 499–518. doi:10.1002/zoo.20066.

Terven, J. and Cordova-Esparza, D. (2023) ‘A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond’, *arXiv preprint arXiv:2304.00501* [Preprint].

Waitt, C. and Buchanan-Smith, H.M. (2001) ‘What time is feeding?: How delays and anticipation of feeding schedules affect stump-tailed macaque behavior’, *Applied animal behaviour science*, 75(1), pp. 75–85.

Wells, D.L. (2005) ‘A note on the influence of visitors on the behaviour and welfare of zoo-housed gorillas’, *Applied Animal Behaviour Science*, 93(1–2), pp. 13–17. doi:10.1016/j.applanim.2005.06.019.

Chapter 6. Discussion and Future work

6.1 Thesis Objectives

The objective of this thesis was to leverage machine learning and computer vision to facilitate the acquisition and analysis of data essential for the assessment and enhancement of macaque welfare within a breeding colony. Three distinct projects were developed to investigate individual and group macaque behaviours.

Initially, pre-recorded videos of temperament tests were utilized to train three models designed to identify movement patterns, exploratory behaviour, and approaches to objects during the tests. The primary aim was to enable the automated assessment of neophobia in macaques while capturing both behaviours previously coded by human observers and new behaviours that had not been coded previously due to their labour-intensive nature.

Subsequently, a pipeline was constructed to detect macaques engaging in foraging activities, with the goal of investigating foraging patterns across different contexts. This permitted the collection of data pertaining to preferred foraging materials, and pellet sizes and seasonal foraging habits.

Additionally, the pipeline provided insights into one of the challenges within captive macaque populations: aggression (van Schaik *et al.*, 1983; Saito Chiemi, 1996; Steenbeek and Sterck, 1997; Thierry, Singh and Kaumanns, 2004; McCowan *et al.*, 2008; Beisner and Isbell, 2011c; Grove, 2012). It revealed how foraging could contribute to reducing injury rates within macaque groups.

The final project involved the automated evaluation of two enrichment items. The first item comprised a white tank periodically filled with raisins; a food-based enrichment that had been present in the enclosure for several months before the study commenced. The second enrichment item consisted of a cognitive puzzle introduced into the enclosure for the purpose of the study, representing a novelty for the monkeys. The primary objective of this project was to explore the association between boredom and non-food-based enrichment, with a focus on identifying the point at which these items would no longer captivate the interest of the macaques.

These findings underscore the significance of technological advancements in the field of animal welfare research and open avenues for further investigations into enhancing the welfare of captive macaque populations.

6.2 Achievements and Limitations

This thesis demonstrates the potential of computer vision to enhance the study of animal behaviour while reducing human involvement in data collection and analyses.

Through this methodology, previously unrecorded behaviours have been captured and integrated into a standardised test for macaque temperament, such as exploratory behaviour and movement patterns. The results derived from the automated pipeline revealed a correlation between reduced levels of neophobia and heightened exploratory behaviour and movement. Conversely, more fearful individuals exhibited increased freezing behaviour and decreased interaction with objects, consistent with findings in existing literature (Kalin and Shelton, 1989; Coleman, Tully and McMillan, 2005; Bethell *et al.*, 2019). These behaviours are known to be crucial for assessing macaques' temperament and can offer new insights into the study of neophobia (Barr *et al.*, 2008; Santillán-Doherty *et al.*, 2010; Gottlieb and Capitanio, 2013; Amici *et al.*, 2020). While this pipeline may not completely meet the standard for accurately detecting latency to approach when compared to manual coding, it serves as a proof of concept, illustrating the feasibility of such methodologies. It offers understandings into the requisites for video setup and study design aimed at training and establishing automated systems for behaviour collection in temperament assessments. This project paves the way for future studies explicitly tailored to exploit the potential of this automated approach.

Utilizing computer vision models based on YOLACT (Bolya *et al.*, 2019), a real-time object segmentation system, has enabled the collection of other previously unattainable data at the Centre for Macaques (CFM), including foraging patterns and food preferences. The findings demonstrated that offering chopped foods such as spinach, watercress, and mushrooms maintained macaques' heightened engagement in foraging when compared to foraging mixes rich in carbohydrates, fructose, and glucose (e.g., oranges and tomatoes), which are known to be favoured by other primate species (Laska, 2001; Jildmalm, Amundin and Laska, 2008; Clay *et al.*, 2009).

Moreover, the versatility of YOLACT extends to monitoring and assessing the utilization of enrichment. This approach provides insights into the dynamic patterns of enrichment usage, effectively minimizing the necessity for manual intervention. The automated pipeline revealed ongoing interest in the enrichment among the macaques, even when the item lacked food content. These findings assist in refining and adapting the enrichment strategy as necessary.

The most recent advances and applications in this study have been driven by YOLOv8, a real time object detection system, and the Segment Anything Model (SAM), a labelling and mask generator system (Jocher, G., Chaurasia, A., & Qiu, 2023; Kirillov *et al.*, 2023). The former offers faster information retrieval compared to YOLACT, as evidenced in the non-food-based enrichment evaluation project where the automated analysis indicated a decline in interest in item over the course of a one-month observation period, confirming that toys can lose their appeal quickly among macaques (Bayne *et al.*, 1993; Corrine K. Lutz and Novak, 2005; Kemp, 2023). SAM, on the other hand, has proven instrumental in streamlining the labelling aspect of the work, further reducing human involvement and increasing agreement between labellers (F (3, 300) = 10.996, $p < 0.001$) (Figure 38).

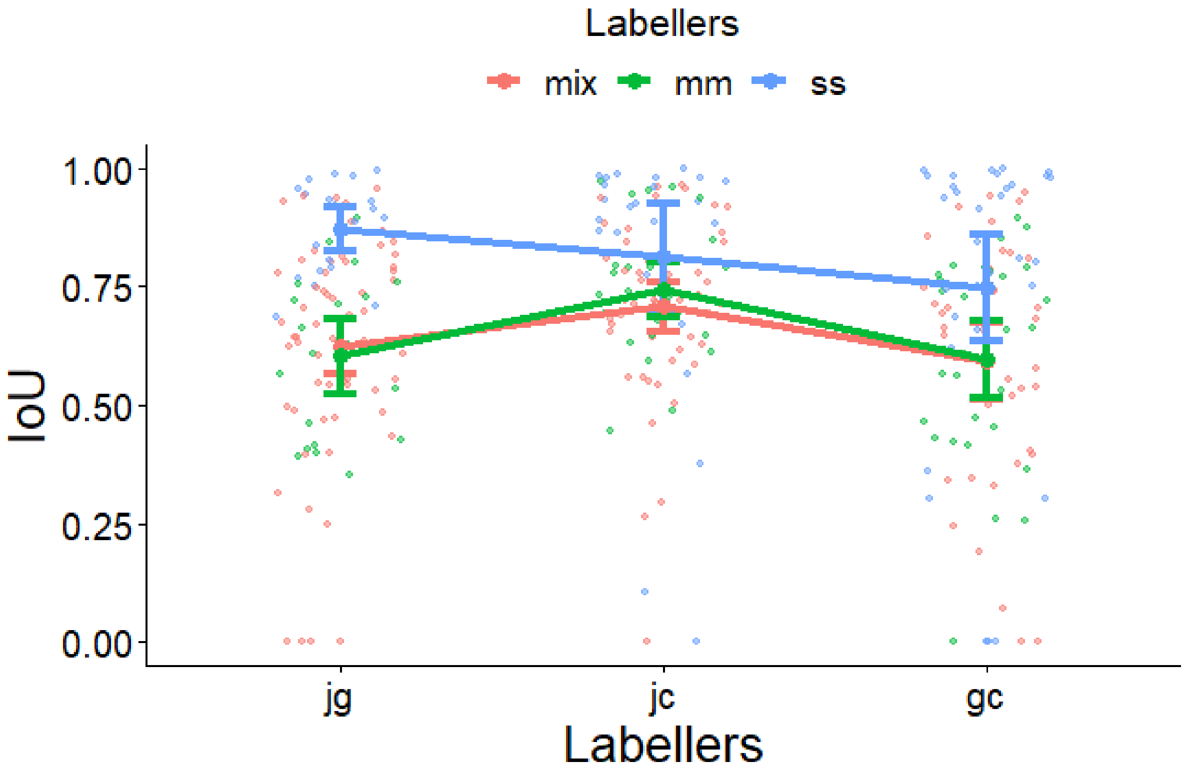


Figure 38: Agreement, indicated as Intersection over Union (IoU), between 3 labellers (j,g,c) labelling macaques. Blue line shows agreement between labellers using SAM, green line shows agreement between labellers using manual labelling, red line shows agreement between labellers using both SAM and the manual system.

The pipeline developed for the foraging patterns and food preferences is scheduled for integration into routine behaviour assessment and welfare monitoring at the breeding colony. This advance will empower staff to better monitor macaques' foraging behaviours and promptly detect any changes that might affect their welfare. In addition, the model based on YOLOv8 will serve to monitor the macaques, gathering essential data on enclosure usage and movement patterns (see sections below for more details).

6.3 Future Refinements

It is important to note that these methodologies are continuously evolving due to the rapid changes in the field of machine learning and computer vision. Therefore, the models proposed in this thesis hold the potential for refinement.

In particular, the temperament test pipeline requires further enhancement, mainly related to video quality. Emphasizing the need to standardize camera positions and achieve synchronization between the two cameras is pivotal. Such efforts will serve to increase the performance of the object detection model and enhance the overall accuracy of the pipeline, thereby diminishing the incidence of false positives.

Additionally, for both the foraging and enrichment pipelines, it would be advantageous to train a computer vision model capable of detecting when staff members enter the playpen to feed the animals. This would eliminate the need to manually identify feeding times from each video. This enhancement can be achieved using both YOLACT and YOLOv8 and would streamline the video selection and exportation process from the CFM's CCTV system. The dataset necessary to train the human detector model has already been prepared, although time constraints prevented me from training the model and applying it to the videos. Nevertheless, this enhancement is entirely feasible to integrate into the automated methodology outlined in this thesis.

6.4 Possible Applications

The adaptable nature of the pipeline employed for collecting and analysing data in the temperament test holds promise for similar assessments aimed at evaluating temperament and neophobia in Non-Human Primates (NHP). To facilitate this adaptation, it would entail retraining the DeepLabCut (DLC) (Mathis *et al.*, 2018) models using a limited set of new frames extracted from videos captured during similar tests conducted in different facilities. In cases where the tests involve macaques, minor retraining on a new dataset derived from videos of the novel tests would be necessary. Should the tests encompass other NHP species, the DLC models would require training on a dataset specific to the frames extracted from videos of these new tests. In parallel, the YOLACT model would necessitate training on a fresh dataset containing frames extracted from videos capturing the novelties utilized in these tests. A collaborative effort to tailor this pipeline for the temperament tests conducted at the Oregon National Primate Research Center is already underway, highlighting the potential for cross-facility applications and advances in temperament assessment methodologies.

As mentioned earlier, both the pipeline designed for identifying foraging patterns and the ones employed to assess enrichment usage are set to become integral components of the technology employed at the Centre of Macaques for enhancing macaques' welfare assessment. Furthermore, these pipelines hold the potential to offer benefits to larger macaque research centres, breeding colonies, as well as zoos and sanctuaries housing extensive primate populations. However, with the greater complexity of settings in these facilities, which include larger enclosures and a higher abundance of enrichment items, the need for a more diverse and complex dataset for thorough model training becomes vital. Fulfilling this requirement necessitates procuring a fresh set of labelled data. The implementation of SAM considerably streamlines the data labelling process, marking a noteworthy improvement over previous methods. This innovation implies that obtaining the necessary labelled data is now a more manageable and efficient task. Additionally, by training these models on the YOLOv8 architecture, the resulting pipeline stands to achieve heightened precision and reliability in detecting both the animals and the various structures present within their environment. This advancement is particularly valuable in navigating the complexities inherent in larger and more diverse primate habitats.

These insights are invaluable for management strategies, aiding in the informed planning of enrichment programs, food selection, and optimal presentation to the macaques. However, there exists the potential for further information extraction using analogous pipelines. Elements such as enclosure-wide movement, utilization of structures, spatial distribution, and deviations from established patterns hold promise for exploration. This realm of information could be harnessed to identify stress levels and social disruptions within macaque groups. Notably, in primates, patterns of spatial choice and movement are intrinsically linked to aggression and social hierarchy (Janson, 1990). For instance, chasing behaviours, involving multiple individuals, serve as indicators of aggression (Higley *et al.*, 1996; Muller and Wrangham, 2004). Thus, these methodologies possess the potential to delve deeper into the dynamics of macaque groups, equipping researchers with tools to pinpoint stressors and potential disruptions within their social fabric. This capacity holds special significance in the context of captive non-human primates, where the most challenging issue revolves around the identification (refer to Appendix A for more details) and resolution of aggressions and social disruptions (Thierry, Singh and Kaumanns, 2004;

Corrine K. Lutz and Novak, 2005; Lutz *et al.*, 2007b, 2007a; Gottlieb, Coleman and McCowan, 2013; Beisner *et al.*, 2015).

6.5 Conclusion

The methodologies outlined in this thesis have proven to be not only versatile but also dynamic, representing a multi-faceted approach that merges a spectrum of machine learning algorithms, ranging from pose estimation models to object detection techniques. These tailored pipelines have not only facilitated a comprehensive exploration but have also paved the way for an array of investigations into the intricate behaviour and social dynamics exhibited by macaques at the Centre for Macaques.

Through these pipelines, I have been able to unravel a diverse array of macaque behaviours, shedding light on individuals' interactions, exploration, and responses within their environment. Importantly, the data harnessed from these pipelines has emerged as an asset for colony management, empowering more informed decision-making when it comes to the macaques' welfare. The gathered insights have led to the refinement of feeding plans and enrichment strategies, both for optimizing existing protocols and charting out management activities. This not only reinforces the existing practices but also lays a foundation for future strategies and initiatives.

Furthermore, these pipelines are not confined to this stage alone. They are on the edge to transcend their current utility, being allocated for continual deployment at the Centre for Macaques in forthcoming research endeavours. Beyond these walls, there exists the potential for these pipelines to play a role in other macaque facilities, including the Oregon National Primate Research Center, thereby contributing to a broader understanding of primate behaviours.

As the realm of machine learning continues to evolve with each passing day, my thesis journey underscores the perpetual evolution of this field, unveiling new algorithms and models. While I have charted significant progress, the future indicates opportunities for enhancements. I have highlighted the avenues where adjustments and improvements are primed to refine accuracy and expedite data extraction from videos, further alleviating the burden on human resources. The use of these evolving technologies holds the promise of not only a more accurate understanding of animal behaviour but also a more efficient and humane approach to monitoring animal welfare.

6.6 Chapter References

Amici, F. *et al.* (2020) 'Innovation in wild Barbary macaques (Macaca sylvanus)', *Scientific Reports*, 10(1), p. 4597. doi:10.1038/s41598-020-61558-2.

Barr, C.S. *et al.* (2008) 'CRH Haplotype as a Factor Influencing Cerebrospinal Fluid Levels of Corticotropin-Releasing Hormone, Hypothalamic-Pituitary-Adrenal Axis Activity, Temperament, and Alcohol Consumption in Rhesus Macaques', *Archives of General Psychiatry*, 65(8), p. 934. doi:10.1001/archpsyc.65.8.934.

Bayne, K.A.L. *et al.* (1993) 'Kong [R] Toys for Laboratory Primates: Are They Really an Enrichment or Just Fomites?', *Laboratory Animal Science*, 43, p. 78.

Beisner, B.A. *et al.* (2015) 'Detection of social group instability among captive rhesus macaques using joint network modeling', *Current Zoology*, 61(1), pp. 70–84. doi:10.1093/czoolo/61.1.70.

Beisner, B.A. and Isbell, L.A. (2011) 'Factors affecting aggression among females in captive groups of rhesus macaques (Macaca mulatta)', *American Journal of Primatology*, 73(11), pp. 1152–1159. doi:10.1002/ajp.20982.

Bethell, E.J. *et al.* (2019) 'Toward a Standardized Test of Fearful Temperament in Primates: A Sensitive Alternative to the Human Intruder Task for Laboratory-Housed Rhesus Macaques (Macaca mulatta)', *Frontiers in Psychology*, 10. doi:10.3389/fpsyg.2019.01051.

Bolya, D. *et al.* (2019) 'Yolact: Real-time instance segmentation', in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166.

Clay, A.W. *et al.* (2009) 'Systematic Investigation of the Stability of Food Preferences in Captive Orangutans: Implications for Positive Reinforcement Training', *Journal of Applied Animal Welfare Science*, 12(4), pp. 306–313. doi:10.1080/10888700903163492.

Coleman, K., Tully, L.A. and McMillan, J.L. (2005) 'Temperament correlates with training success in adult rhesus macaques', *American Journal of Primatology*, 65(1), pp. 63–71. doi:10.1002/ajp.20097.

Gottlieb, D.H. and Capitanio, J.P. (2013) 'Latent Variables Affecting Behavioral Response to the Human Intruder Test in Infant Rhesus Macaques (Macaca mulatta)', *American Journal of Primatology*, 75(4), pp. 314–323. doi:10.1002/ajp.22107.

Gottlieb, D.H., Coleman, K. and McCowan, B. (2013) 'The effects of predictability in daily husbandry routines on captive rhesus macaques (Macaca mulatta)', *Applied Animal Behaviour Science*, 143(2–4), pp. 117–127. doi:10.1016/j.applanim.2012.10.010.

Grove, M. (2012) ‘Space, time, and group size: a model of constraints on primate social foraging’, *Animal Behaviour*, 83(2), pp. 411–419. doi:10.1016/j.anbehav.2011.11.011.

Higley, J.D. *et al.* (1996) ‘Stability of interindividual differences in serotonin function and its relationship to severe aggression and competent social behavior in rhesus macaque females’, *Neuropsychopharmacology*, 14(1), pp. 67–76.

Janson, C.H. (1990) ‘Social correlates of individual spatial choice in foraging groups of brown capuchin monkeys, Cebus apella’, *Animal Behaviour*, 40(5), pp. 910–921. doi:10.1016/S0003-3472(05)80993-5.

Jildmalm, R., Amundin, M. and Laska, M. (2008) ‘Food Preferences and Nutrient Composition in Captive White-handed Gibbons, Hylobates lar’, *International Journal of Primatology*, 29(6), pp. 1535–1547. doi:10.1007/s10764-008-9314-1.

Jocher, G., Chaurasia, A., & Qiu, J. (2023) ‘YOLO by Ultralytics (Version 8.0.0)’.

Kalin, N.H. and Shelton, S.E. (1989) ‘Defensive behaviors in infant rhesus monkeys: environmental cues and neurochemical regulation’, *Science*, 243(4899), pp. 1718–1721.

Kemp, C. (2023) ‘Enrichment’, in *Nonhuman Primate Welfare*. Cham: Springer International Publishing, pp. 463–500. doi:10.1007/978-3-030-82708-3_20.

Kirillov, A. *et al.* (2023) ‘Segment Anything’, *arXiv:2304.02643* [Preprint].

Laska, M. (2001) ‘A comparison of food preferences and nutrient composition in captive squirrel monkeys, Saimiri sciureus, and pigtail macaques, Macaca nemestrina’, *Physiology & Behavior*, 73(1–2), pp. 111–120. doi:10.1016/S0031-9384(01)00439-5.

Lutz, C.K. *et al.* (2007a) ‘Early predictors of self-biting in socially-housed rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 69(5), pp. 584–590. doi:10.1002/ajp.20370.

Lutz, C.K. *et al.* (2007b) ‘Early predictors of self-biting in socially-housed rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 69(5), pp. 584–590. doi:10.1002/ajp.20370.

Lutz, C.K. and Novak, M.A. (2005) *Primate Natural History and Social Behavior. Implications for Laboratory Housing*, *The Laboratory Primate*. Elsevier Limited. doi:10.1016/B978-012080261-6/50009-X.

Mathis, A. *et al.* (2018) ‘DeepLabCut: markerless pose estimation of user-defined body parts with deep learning’, *Nature Neuroscience*, 21(9), pp. 1281–1289. doi:10.1038/s41593-018-0209-y.

McCowan, B. *et al.* (2008) ‘Utility of social network analysis for primate behavioral management and well-being’, *Applied Animal Behaviour Science*, 109(2–4), pp. 396–405.

doi:10.1016/j.applanim.2007.02.009.

Muller, M.N. and Wrangham, R.W. (2004) ‘Dominance, aggression and testosterone in wild chimpanzees: a test of the “challenge hypothesis”’, *Animal behaviour*, 67(1), pp. 113–123.

Saito Chiemi (1996) ‘Dominance and feeding success in female Japanese macaques,Macaca fuscata: effects of food patch size and inter-patch distance’, *Animal Behaviour*, 51(5), pp. 967–980. doi:10.1006/anbe.1996.0100.

Santillán-Doherty, A.M. *et al.* (2010) ‘Novelty-seeking temperament in captive stumptail macaques (Macaca arctoides) and spider monkeys (ateles geoffroyi).’, *Journal of Comparative Psychology*, 124(2), pp. 211–218. doi:10.1037/a0018267.

van Schaik, C.P. *et al.* (1983) ‘The effect of group size on time budgets and social behaviour in wild long-tailed macaques (Macaca fascicularis)’, *Behavioral Ecology and Sociobiology*, 13(3), pp. 173–181. doi:10.1007/BF00299920.

Steenbeek, R. and Sterck, E.H.M. (1997) ‘Female Dominance Relationships and Food Competition in the Sympatric Thomas Langur and Long-Tailed Macaque’, *Behaviour*, 134(9–10), pp. 749–774. doi:10.1163/156853997X00052.

Thierry, B., Singh, M. and Kaumanns, W. (2004) *Macaque societies: a model for the study of social organization*. Cambridge University Press.

Appendix A

Identifying the individuals

Regarding the identification of social structures within macaque groups and their dynamic changes over time, a significant enhancement would involve the capability to individually identify each macaque. This advance would greatly contribute to the accurate detection and tracking of social structures and their fluctuations.

Automated detection of social structure

In the initial stages of my PhD research, I embarked on creating an automated system capable of extracting proximity data from videos featuring macaque breeding groups. The primary aim was to discern their social structure and capture insights into alterations within this structure over time using social network analysis. The knowledge of the monkeys' positions and their proximity to one another holds great potential for understanding the dominance hierarchy present within these groups (Corradino, 1990; Singh, D'Souza and Singh, 1992; Zhang *et al.*, 2012). This understanding is pivotal for effective colony management, given that successful management hinges on comprehending these dynamics. Furthermore, this knowledge is particularly valuable in comprehending the effects of stress-inducing scenarios, such as health assessments, the removal or introduction of individuals from groups, on the broader social structure (Ferin *et al.*, 1976; Fuller *et al.*, 1984; Beisner and Isbell, 2011c; Blumrosen, Hawellek and Pesaran, 2017; Hannibal *et al.*, 2017; McCowan, Beisner and Hannibal, 2018). Particularly, special attention was given to the last scenario.

Despite efforts to reduce aggressive behaviours in captive settings, such as those employed in CFM and other non-human primate breeding facilities, certain factors still contribute to heightened aggression (Pruetz and Isbell, 2000; Beisner and Isbell, 2011b). These include the lack of escape options in threatening situations and, notably, the introduction of new male individuals into breeding groups. While introducing new males is a common practice in maintaining genetic diversity and improving productivity within a breeding colony, it can lead to complex challenges, including the potential for failure and multiple injuries (Massey *et al.*, 2022). Such changes in group dynamics can cause instability, disrupting social orders and leading to ambiguous relationships, accompanied by physiological changes indicative of stress (Wolfensohn and Honess, 2008; Hannibal *et al.*, 2017; McCowan, Beisner and Hannibal, 2018). To address these situations more effectively, a heightened level of observational vigilance by management staff is required to accurately ascertain hierarchy rankings, facilitating the well-informed removal of specific animals if necessary (Coleman *et al.*, 2012). However, this approach is often constrained, focusing on a select few individuals and for a limited duration. Moreover, it mainly allows for the detection of significant alterations in already compromised animals, making early intervention challenging (Weissbrod *et al.*, 2013; Matthews *et al.*, 2017).

Given these challenges, there is a need to develop an automated model capable of detecting the social structure within macaque breeding groups and recognizing changes in group hierarchy resulting from the introduction of new male members. This methodology aimed to highlight the repercussions of such introductions on existing groups and to compare resultant social structure changes with the success or failure of these introductions.

At CFM, due to variations in lighting, the presence of obstructions, and the differing sizes of the monkeys in frame based on their location, identifying individual macaques becomes challenging when a single camera covers the entire enclosure. Consequently, a different approach was taken, training the model on videos capturing specific key areas (as illustrated in Figure 39).

Figure 39: Key areas recorded by the camera. The red square identifies the shelf in front of the window, and the blue area identifies the area on the ground in front of the windows. Both areas are mainly used for affiliative behaviours and resting behaviours.

The primary goals in detecting the social structure of these groups encompassed the following:

1. Identifying animals within the video.
2. Distinguishing individual macaques.
3. Tracking these animals and extracting information about their proximity to one another.
4. Performing social network analysis to discern shifts in interactions between individuals.

Drawing inspiration from C. Witham's previous work (Witham, 2018), a multi-faceted methodology was created to extract the aforementioned information:

1. Employing two models for face and profile detection to locate animals and their facial features. These detectors were based on the cascade object detector function in Matlab (for more detailes refer to Witham, 2018).
2. Utilizing a face recognition model to accurately identify individual macaques (for more detailes refer to Witham, 2018).
3. Employing DeepLabCut (DLC) landmark detection to validate detected faces, ensuring a minimum of three landmarks from DLC.
4. Implementing KLT tracking (Kanade-Lucas-Tomasi points tracker) to record proximity information.

Initially, the approach utilized DLC for single-animal tracking, as the multi-animal DLC version was unavailable at that time.

This pipeline encountered notable limitations, particularly when macaques were not squarely facing the camera or when they were entangled in overlapping configurations or unconventional postures. Regrettably, these situations led to suboptimal accuracy in comparison to human observations (

Video	**No Frames Tracked**	**No Frames Identified**	**No Frames Present GT**	**No Frames Present GT**	**Correctly Identified**
DA051214	1997	1584	4264	46.83%	79.32%
DA180315	4070	2341	12276	33.15%	57.52%
JU300614	2600	2253	11658	22.30%	86.65%
JU290614	2914	2267	10230	28.48%	77.80%
SO1507141	12589	11873	17543	71.76%	94.31%
SO1507142	10688	10655	14374	74.36%	99.69%
ST1902141	8243	6922	18012	45.76%	83.97%
UTO80416	7754	7404	22504	34.45%	95.48%

Table 26 and

	No Frames Tracked	**No Frames Identified**	**No Frames Present GT**	**Tracked**	**Correctly Identified**
Total	10688	10655	14374	74.36%	99.69%
Lala	3475	3475	4756	73.07%	100%

Libby	1743	1743	2591	67.27%	100%
Love	474	463	806	58.81%	97.68%
Meg	4996	4974	6002	83.24%	99.56%
Leah	0	0	74	0	
Wispa	0	0	41	0	
Yuletide (young)	0	0	36	0	
Zahara (young)	0	0	68	0	

Table 27).

Video	No Frames Tracked	No Frames Identified	No Frames Present GT	No Frames Present GT	Correctly Identified
DA051214	1997	1584	4264	46.83%	79.32%
DA180315	4070	2341	12276	33.15%	57.52%
JU300614	2600	2253	11658	22.30%	86.65%
JU290614	2914	2267	10230	28.48%	77.80%
SO1507141	12589	11873	17543	71.76%	94.31%
SO1507142	10688	10655	14374	74.36%	99.69%
ST1902141	8243	6922	18012	45.76%	83.97%
UTO80416	7754	7404	22504	34.45%	95.48%

Table 26: This table presents accuracy metrics for the multi-animal tracking and identification system across nine videos. The metrics include the number of frames with correct tracking (No Frames Tracked), correct identification (No Frames Identified), and frames with monkeys present in ground truth (No Frames Present GT). The model's accuracy is reflected in terms of the percentage of correctly tracked instances compared to ground truth (% Tracked) and the number of correctly identified instances compared to tracked instances (Correctly Identified).

	No Frames Tracked	No Frames Identified	No Frames Present GT	Tracked	Correctly Identified
Total	10688	10655	14374	74.36%	99.69%
Lala	3475	3475	4756	73.07%	100%
Libby	1743	1743	2591	67.27%	100%
Love	474	463	806	58.81%	97.68%
Meg	4996	4974	6002	83.24%	99.56%
Leah	0	0	74	0	
Wispa	0	0	41	0	
Yuletide (young)	0	0	36	0	
Zahara (young)	0	0	68	0	

Table 27. The table illustrates tracking and identification outcomes for a group of 8 monkeys engaged in stable behaviours such as sitting, grooming, and resting. Despite the stationary nature of these behaviours, the pipeline encountered challenges in tracking half of the individuals within the group.

Subsequently, I explored an alternative approach based on YOLACT, aiming to enhance macaque detection. This involved using YOLACT for animal detection and subsequently applying face and profile detection on the detected individuals. While this approach led to improved macaque detection, challenges persisted in finding a dependable tracking algorithm to furnish reliable proximity data.

However, there is a potential avenue to address this limitation by leveraging a robust tracking mechanism, like the one offered by YOLOv8. This advance could provide a solution to the challenge posed by animals not facing the camera. In essence, a proficient tracker could rectify the issue of unidentifiable instances. With the aid of such a tracker, the pipeline could assign identities to previously unidentified tracks, particularly when the animal in question comes into view of the camera. It is crucial to emphasize the significance of maintaining extended tracking periods and deploying an algorithm that can effectively track an individual even when it partially disappears behind objects or overlaps with others. In doing so, this approach holds the potential to significantly enhance the accuracy and completeness of the collected data, ultimately enabling a more comprehensive analysis for social network assessment.

References

Beisner, B.A. and Isbell, L.A. (2011a) ‘Factors affecting aggression among females in captive groups of rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 73(11), pp. 1152–1159. doi:10.1002/ajp.20982.

Beisner, B.A. and Isbell, L.A. (2011b) ‘Factors affecting aggression among females in captive groups of rhesus macaques (Macaca mulatta)’, *American Journal of Primatology*, 73(11), pp. 1152–1159. doi:10.1002/ajp.20982.

Blumrosen, G., Hawellek, D. and Pesaran, B. (2017) ‘Towards automated recognition of facial expressions in animal models’, *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2018-Janua, pp. 2810–2819. doi:10.1109/ICCVW.2017.332.

Coleman, K. *et al.* (2012) ‘Behavioral Management, Enrichment, and Psychological Well-being of Laboratory Nonhuman Primates’, in *Nonhuman Primates in Biomedical Research*. Elsevier, pp. 149–176. doi:10.1016/B978-0-12-381365-7.00006-6.

Corradino, C. (1990) ‘Proximity structure in a captive colony of Japanese monkeys (Macaca fuscata fuscata): An application of multidimensional scaling’, *Primates*, 31(3), pp. 351–362.

Ferin, M. *et al.* (1976) ‘Phencyclidine Sedation as a Technique for Handling Rhesus Monkeys: Effects on LH, GH, and Prolactin Secretion’, *Experimental Biology and Medicine*, 151(2), pp. 428–433. doi:10.3181/00379727-151-39227.

Fuller, G.B. *et al.* (1984) ‘Influence of Restraint and Ketamine Anesthesia on Adrenal Steroids, Progesterone, and Gonadotropins in Rhesus Monkeys’, *Experimental Biology and Medicine*, 175(4), pp. 487–490. doi:10.3181/00379727-175-41825.

Hannibal, D.L. *et al.* (2017) ‘Laboratory rhesus macaque social housing and social changes: Implications for research’, *American Journal of Primatology*, 79(1), p. e22528. doi:10.1002/ajp.22528.

Massey, D.A. *et al.* (2022) ‘Factors Associated with Injury Rate and Pregnancy Success in Rhesus Macaques’, *Biology*, 11(7), p. 979.

Matthews, S.G. *et al.* (2017) ‘Automated tracking to measure behavioural changes in pigs for health and welfare monitoring’, *Scientific Reports*, 7(1), p. 17582. doi:10.1038/s41598-017-17451-6.

McCowan, B., Beisner, B. and Hannibal, D. (2018) ‘Social management of laboratory rhesus macaques housed in large groups using a network approach: A review’, *Behavioural Processes*, 156, pp. 77–82. doi:10.1016/j.beproc.2017.11.014.

Pruetz, J.D. and Isbell, L.A. (2000) ‘Correlations of food distribution and patch size with agonistic interactions in female vervets (Chlorocebus aethiops) and patas monkeys (Erythrocebus patas) living in simple habitats’, *Behavioral Ecology and Sociobiology*, 49(1), pp. 38–47. doi:10.1007/s002650000272.

Singh, Mewa, D’Souza, L. and Singh, Mridula (1992) ‘Hierarchy, kinship and social interaction among Japanese monkeys (Macaca fuscata)’, *Journal of Biosciences*, 17, pp. 15–27.

Weissbrod, A. *et al.* (2013) ‘Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment’, *Nature Communications*, 4(1), p. 2018. doi:10.1038/ncomms3018.

Witham, C.L. (2018) ‘Automated face recognition of rhesus macaques’, *Journal of Neuroscience Methods*, 300, pp. 157–165. doi:10.1016/j.jneumeth.2017.07.020.

Wolfensohn, S. and Honess, P. (2008) *Handbook of primate husbandry and welfare*. John Wiley & Sons.

Zhang, P. *et al.* (2012) ‘A proximity-based social network of a group of Sichuan snub-nosed monkeys (Rhinopithecus roxellana)’, *International Journal of Primatology*, 33, pp. 1081–1095.