Communication

Behavioral Coding of Captive African Elephants (*Loxodonta africana*): Utilizing DeepLabCut and Create ML for Nocturnal Activity Tracking

Silje Marquardsen Lund [1,*,†], Jonas Nielsen [1,†], Frej Gammelgård [1,*,†], Maria Gytkjær Nielsen [1,†], Trine Hammer Jensen [1,2] and Cino Pertoldi [1,2]

1   Department of Chemistry and Bioscience, Aalborg University, Frederik Bajers Vej 7H, 9220 Aalborg, Denmark; jnielc21@student.aau.dk (J.N.); maniek21@student.aau.dk (M.G.N.); cp@bio.aau.dk (C.P.)
2   Aalborg Zoo, Mølleparkvej 63, 9000 Aalborg, Denmark
*   Correspondence: siljemarquardsenlund@gmail.com (S.M.L.); fgamme21@student.aau.dk (F.G.)
†   These authors contributed equally to this work.

Simple Summary: This paper presents a way to automate computer vision processes applied to behavior recognition on closed-circuit television (CCTV) footage of two captive African elephants. Object detection software using both Create ML and DeepLabCut was used to control the accuracy of using such models, and those models were subsequently used to analyze seven days' worth of nighttime footage to assess the general behavioral patterns of the elephants, showcasing the possibility of using automated tools for behavioral analysis.

Abstract: This study investigates the possibility of using machine learning models created in DeepLabCut and Create ML to automate aspects of behavioral coding and aid in behavioral analysis. Two models with different capabilities and complexities were constructed and compared to a manually observed control period. The accuracy of the models was assessed by comparison with manually scoring, before being applied to seven nights of footage of the nocturnal behavior of two African elephants (*Loxodonta africana*). The resulting data were used to draw conclusions regarding behavioral differences between the two elephants and between individually observed nights, thus proving that such models can aid researchers in behavioral analysis. The models were capable of tracking simple behaviors with high accuracy, but had certain limitations regarding detection of complex behaviors, such as the stereotyped behavior sway, and displayed confusion when deciding between visually similar behaviors. Further expansion of such models may be desired to create a more capable aid with the possibility of automating behavioral coding.

Keywords: machine learning; nocturnal behavior; computer vision; captive elephants

1. Introduction

1.1. Objectives of Wildlife Conservation

The World Association of Zoos and Aquaria (WAZA) aims to conserve endangered species through breeding programs and exchange of captive animals. This association requires certain standards of its members and emphasizes the importance of welfare among captive animals [1]. Based on consistent and current research, good physical and psychological welfare among captive animals must be maintained and therefore associations such as this play an important role in conservation [1–4].

Because of the importance of animal welfare, it is essential to ensure that captive animals are consistently studied in relation to their behavioral reactions to different aspects of their captive lives, such as enclosure design, enrichment, and much more [5–8]. An example of an animal that may require such studies to properly conserve the species is the

African elephant (*Loxodonta* sp.), which has faced great declines in population size all over Africa [5,9].

1.2. Captive Elephant Behavior and Welfare

To accommodate the welfare needs of captive elephants, normal behaviors must first be monitored and understood [6,10,11]. Behaviors such as foraging, locomotion, social behavior, etc., likely influence the welfare of elephants and help understand undesired behaviors that may indicate stress [12–14]. It is important to also address the nocturnal behavior of elephants since night behavior can differ from behaviors observed during the day. An example of one such behavior is recumbent sleep, where the elephants lay down and sleep, which exclusively occurs during the night [14,15]. On average, captive African elephants lie down to sleep around two hours per night and tend to lie down more if their bedding is comfortable [14–17]. Besides rest, feeding and atypical behaviors are the behaviors most commonly observed in captive elephants, yet the activity level is lower at night compared to during the day [14,18].

Atypical behaviors observed in animals kept in captivity are usually those that deviate from the norm for their species and are not commonly observed in their natural habitat. These behaviors are frequently regarded as signs of compromised welfare [19–22]. A type of stereotypic behavior is characterized by the consistent and inappropriate repetition of specific movements or body postures. These actions seemingly lack any purpose or function and appear to be coping mechanisms to reduce stress, but the exact causes remain unclear [21–25]. Different forms of stereotypic behavior have been observed in elephants, including whole-body movements [10,13,22,26,27]. Among whole-body movements, 'swaying' is most common and is defined as a rhythmic side-to-side movement of the body, typically observed while standing [12,18,22]. Consistent observations of elephants may be helpful in handling these behaviors when they arise.

1.3. Machine Learning as a Tool for Behavioral Analysis

One widely used method for observation of animal behavior is videography [14,28], although this method can be highly time-consuming [29–32]. Manual scoring is limited by human capabilities, such as the observer not recognizing behavioral patterns or failing to spot new patterns. At the same time, it is difficult to standardize the scoring of behaviors by human observers due to subjectivity [33]. Inconsistency between different observers can therefore not be avoided completely [31]. Furthermore, it is challenging to track multiple animals and behaviors at the same time, despite the use of video material [29,31]. Some of these logistical problems with behavioral analysis may be aided with the use of machine learning [32].

Machine learning used in video and image analysis (computer vision) has been explored in recent years in application to a variety of purposes [34,35]. The use of object detection as a machine learning tool to find and recognize a given object has been investigated previously and used to recognize animals and their behaviors [32,36–38]. Tools such as this may prove useful as a way of automating behavioral analysis in the near future, which may reduce the workhours required of the researcher [30,32]. However, these uses and methods are still in their infancy which necessitates further investigation of different machine learning models, methods, and implementations [36,37,39].

DeepLabCut is a Machine Learning software that specializes in pose estimation in video material, by using points for tracking specific body parts [40,41]. This is utilized in behavior tracking by marking body parts of interest on a relatively small dataset of images showing a diverse range of behaviors by the subject of interest [32]. Constructing a DeepLabCut model capable of accurately tracking body parts of interest may allow for automatization of behavior coding in behavioral studies [42].

Create ML is a built-in application for iOS products with the ability to train custom machine learning models such as object detection models with no code. Create ML models can be trained to detect and recognize objects of interest, such as an elephant executing a specific behavior, by annotating a relatively small and diverse dataset. This annotation can be used in various ways, such as using the image annotation tool RectLabel for marking boxes, polygons, or skeletons on the subject, and categorizing the behavior. This may allow for construction of a simple model that can recognize simple behaviors with relatively user-friendly software [43].

1.4. Aim of This Paper

This paper aims to use DeepLabCut and Create ML to construct models capable of tracking selected body parts and classifying elephant behaviors, aiming to streamline and automate behavioral analysis processes, thus ultimately alleviating the workload of researchers and zookeepers, and standardizing behavioral coding. This study simultaneously examines nocturnal activity of two captive African elephants with the following hypotheses:

- This study expects that the machine learning models can predict selected behaviors on the same level as manual scoring;
- This study expects that behavioral differences between the elephants and behavioral differences between days can be demonstrated using selected computer vision models.

2. Materials and Methods

2.1. Subjects and Enclosure

The behavior of two captive female African elephants, exhibited in Aalborg Zoo, Denmark, was examined. Both elephants were born wild in South Africa around 1982 and relocated to Aalborg Zoo in 1985. In this study, the elephants are referred to as Subject A and B.

The elephant enclosure consisted of an indoor area and an outdoor area. The elephants did not have access to the outdoor enclosure at night during the examined period. The indoor enclosure consisted of concrete floors and walls with metal wires towards the visitor area (see Appendix A). The two elephants were able to have physical contact during the night through metal bars between enclosures E1 and E2. Subject A had access to enclosure E1, and a corridor attached to the enclosure (56 square meters). Subject B had access to enclosures E2 and E3 as well as a corridor measuring a total of 116 square meters.

The elephants' diet consisted of branches, seed grass hanging from nets in the inside and outside enclosures, and concentrate pellets that would periodically be released into the enclosure from a timer-automated mechanism, as a type of enrichment. Fresh fruits and vegetables were spread around their outdoor enclosure daily, which allowed for foraging behaviors. Foraging boxes, accessible with use of their trunks at the back wall of their enclosure, were also opened periodically throughout each night, controlled by a timer.

2.2. Data Collection

Prior to data collection, an ethogram with selected behaviors was made. The behaviors were determined based on similar behavioral studies of the same elephants in previous publications, namely Bertelsen et al. (2020) and Andersen et al. (2020), and were modified for the purpose of this study [14,44]. This ethogram was used to conduct a manual control, using researchers experienced in behavioral coding, where the researchers were assigned a subject each. The behaviors were coded continually on a second-by-second basis. These scorings were used to compare manual coding with the models; see Table 1.

Table 1. Ethogram used for behavioral coding of the two subjects, used for both manual and automatic coding.

Behavior	Description
Standing	The elephant is standing or walking. This behavior is the default if no other selected behavior is taking place.
Lying down	The elephant is lying down on the floor of the enclosure.
Drinking	The elephant is drinking from a water bowl.
Foraging	The elephant is using the foraging boxes, accessed using trunks in the holes at the back of the enclosure.
Hay-net	The elephant is using its trunk to reach the hay-net at the top of the enclosure.
Swaying	The elephant is swaying from side to side for at least 5 s.
Out of view	The elephant is out of view of the camera. This may also include falsely unlabeled frames by the machine learning models.

The data were collected from the 12th to 18th of March 2024 to use for model creation, and from the 16th to the 22nd of April 2024 for analysis purposes, using three cameras (ABUS, 25 FPS). The three cameras were placed at the visitor viewing side facing towards each enclosure (see Appendix A). The two elephants were observed during the night for seven hours from 22:00 to 05:00 (DST).

2.3. Data Analysis

The analyzed days in this experiment provided data which were compared with statistical tests using Excel (Version 2404 Build 16.0.17531.20120) and RStudio (R version 4.1.2 (1 November 2021)).

For body part tracking, DeepLabCut (version 2.3.9.) was used [32,42]. Specifically, 244 frames taken from 70 30-min videos were labelled (95% were used for training), and no preprocessing was performed. A ResNet-50-based neural network was used with the parameters set to 400,000 training iterations. Validation was carried out with a single shuffle, and the test error found was 17.44 pixels, train 2.4 pixels (image size for creating the model was 1920 by 1080). A p-cutoff of 0.5 was used to condition the *x*- and *y*-coordinates for future analysis. DeepLabCut does not provide annotations of behavior, and it is required of the user to manually define and interpret the coordinates of the results. In this study, the behaviors were classified in Excel using parameters set by comparing the coordinates of relevant body parts with the manually recorded data and videos from the control period (April 16). Each parameter consisted of distinct coordinate limits and requirements to fit the ethogram, resulting in frame-by-frame behavioral coding. The locations of the selected behaviors can be seen in Appendix B. The algorithm for classifying behaviors using DeepLabCut can be seen below in Algorithm 1.

Algorithm 1. DeepLabCut operation

Input: Video containing the subject to be tracked
1. Load dataset. Load the video into DeepLabCut;
2. Define keypoints. Specify keypoints of interest (body parts like head, tail, limbs);
3. Annotate frames. Annotate a subset of frames manually by marking the keypoints;
4. Model training. Use annotated frames to train the pose estimation model;
5. Pose estimation. Apply the trained model to new video material;
6. Refine model (optional). Correct predictions and retrain the model;
7. Process coordinates. Extract CSV file and filter the coordinates for desired body parts;
8. Classify behaviors. Set limits for each coordinate corresponding to a desired behavior and filter frames that fulfill the criteria.
Output: Subset of data points that can be classified as a specific behavior.

Difficulties arose with defining some of the parameters, such as 'Drinking', which proved undefinable for both subjects as the label on the trunk tip was unstable. Furthermore, no distinct parameter was definable for the behavior 'Hay-net' for Subject B as the foraging box and the hay net were located at approximately the same place in the video frame.

For object detection, Create ML (version 5.0 (121.1)) was used [43]. Specifically, 370 image frames were extracted from the model creation period and annotated using bounding boxes in RectLabel Pro (version 2023.11.19). Each frame was annotated with a selected behavior as seen in Table 1. 'Swaying' was not labelled for the object detection part. The behavior 'Standing' was labelled 232 times, 'Foraging' was labelled 42 times, 'Lying down' was labelled 39 times, 'Drinking' was labelled 39 times, and 'Hay-net' was labelled 18 times. The dataset was split into two sets, one containing images for training and one containing images for validation. The model was trained with 6000 iterations, and the training set had an accuracy of 95% whilst the validation set had an accuracy of 75%. Create ML automatically classifies the behavior using bounding boxes resulting in a frame-by-frame behavioral coding output. The algorithm for classifying behaviors using Create ML is seen below in Algorithm 2.

Algorithm 2. Create ML operation

Input: Video containing the subject to be tracked.
1. Load dataset. Load the video into RectLabel and extract images;
2. Define bounding boxes. Specify categories of each bounding box of interest (behaviors, such as lying down or standing);
3. Annotate frames. Annotate a subset of frames manually by drawing bounding boxes;
4. Model training. Use annotated frames to train the model;
5. Pose estimation. Apply the trained model to new video material;
6. Refine model (optional). Correct predictions and retrain the model;
7. Process coordinates. Extract CSV file containing frames annotated with behaviors.
Output: Dataset containing the predicted behaviors at all analyzed frames.

To appraise the accuracy of the models designed with Create ML and DeepLabCut, a control was analyzed manually from the video footage from the 16th of April and used to compare the models' results. To test the accuracy of the models compared to the control, a confusion matrix was conducted. In this case, a multiple class confusion matrix was produced and analyzed [45]. This will give insight into where the model performs well and has a high accuracy, as well as where mistakes occur, such as when the model mislabels a behavior. The columns of such a matrix represent the manually observed behavior of the individual while the rows represent the predicted behavior from the models. Time budgets and cumulative graphs were also used to further appraise the models [46]. Furthermore, Kendall's Coefficient of Concordance was used to measure the agreement between the models and the control [47].

The behavioral analysis consisted of time budgets, cumulative graphs, and Kendall's Coefficient of Concordance between days. Time budgets were made for each elephant each day and for the whole study period. This was carried out using the sums, transformed into percentages, of observed time spent on each behavior, where the out of view percentage made up the time where no behavior was observed or classified. Time budgets for each day were used to investigate daily differences in behavior. The time budgets for the whole study period were used to see how much time was spent on each behavior in total and to compare the different models with the control. Kendall's Coefficient of Concordances were used on the data from the time budgets to analyze the similarity between the observed behavior between different days.

Cumulative graphs were made for each behavior each day. The graphs were made with both the manually recorded data and the model data for the control day, while the rest of the study period only had cumulative graphs made for the Create ML model.

A Spearman rank correlation test was used to investigate correlations between the subject's 'Lying Down' behavior from night to night. Correlations were also used to test the similarities between the subjects and if they exhibited similar behavioral patterns throughout the night.

The possibility to observe the stereotypic behavior 'Swaying' was also examined. This was accomplished by calculating the Euclidian distance between a given point of the trunk root, labelled by the DeepLabCut model, and the succeeding point [48]. The distances between the points were calculated and plotted as a cumulative graph together with the actual cumulative time spent on the sway behavior, so that sway could be observed as steep increases in the cumulative sum.

3. Results

3.1. Comparability of Manual and Automatic Behavioural Observations

3.1.1. A General Overview

First, the capabilities of the two machine learning models, compared to a manually conducted analysis using an ethogram, have been displayed using time budgets, illustrated below in Figure 1.
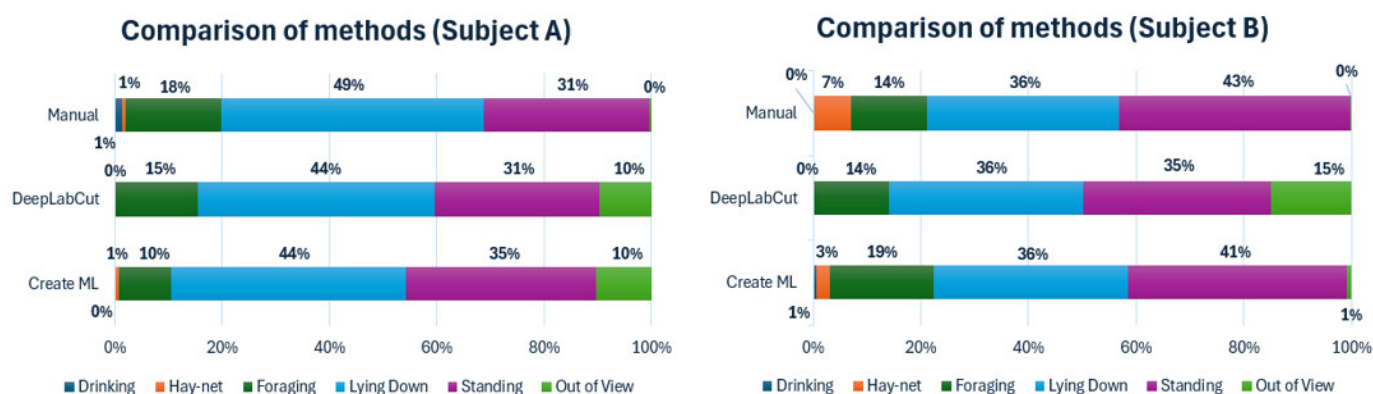
Figure 1. Time budgets for Subjects A and B of the 7 h control period, comparing manual observations with a DeepLabCut and a Create ML model. 'Drinking' is excluded for the DeepLabCut model.

The time budgets for both subjects, using both models, showed similar percentages to the manual observations regarding standing and lying down with relatively small differences. For Subject A, both models had a 10% out of view percentage. Since the manual observations showed an out of view percentage close to zero, this indicates that these were caused by uncertainty by the models, causing them to not label the subject. Both models for Subject A also displayed a lower percentage for 'Foraging' than the manual observations. For Subject B there were notable differences in 'Foraging' and 'Hay-net', which is likely to be caused by the hay-net being close to the foraging boxes, thus showing overlapping coordinates when observed by the DeepLabCut model. The Create ML model also had a lower 'Hay-net' percentage, but contrarily a higher 'Foraging' percentage than the manual observations. The out of view percentage for the DeepLabCut model was noticeably higher than the Create ML model, considering that the manual observations showed an out of view percentage close to zero. This was likely also caused by lack of labelling by the model. 'Drinking' was left out of the DeepLabCut model due to limitations in defining the parameters of this behavior, caused by a lack of consistent appropriate labels needed to properly categorize the behavior.

To further investigate the similarities between the machine learning models and the manually conducted observations, cumulative graphs for each subject, showing each behavior tracked with each method, have been constructed and displayed below in Figure 2.
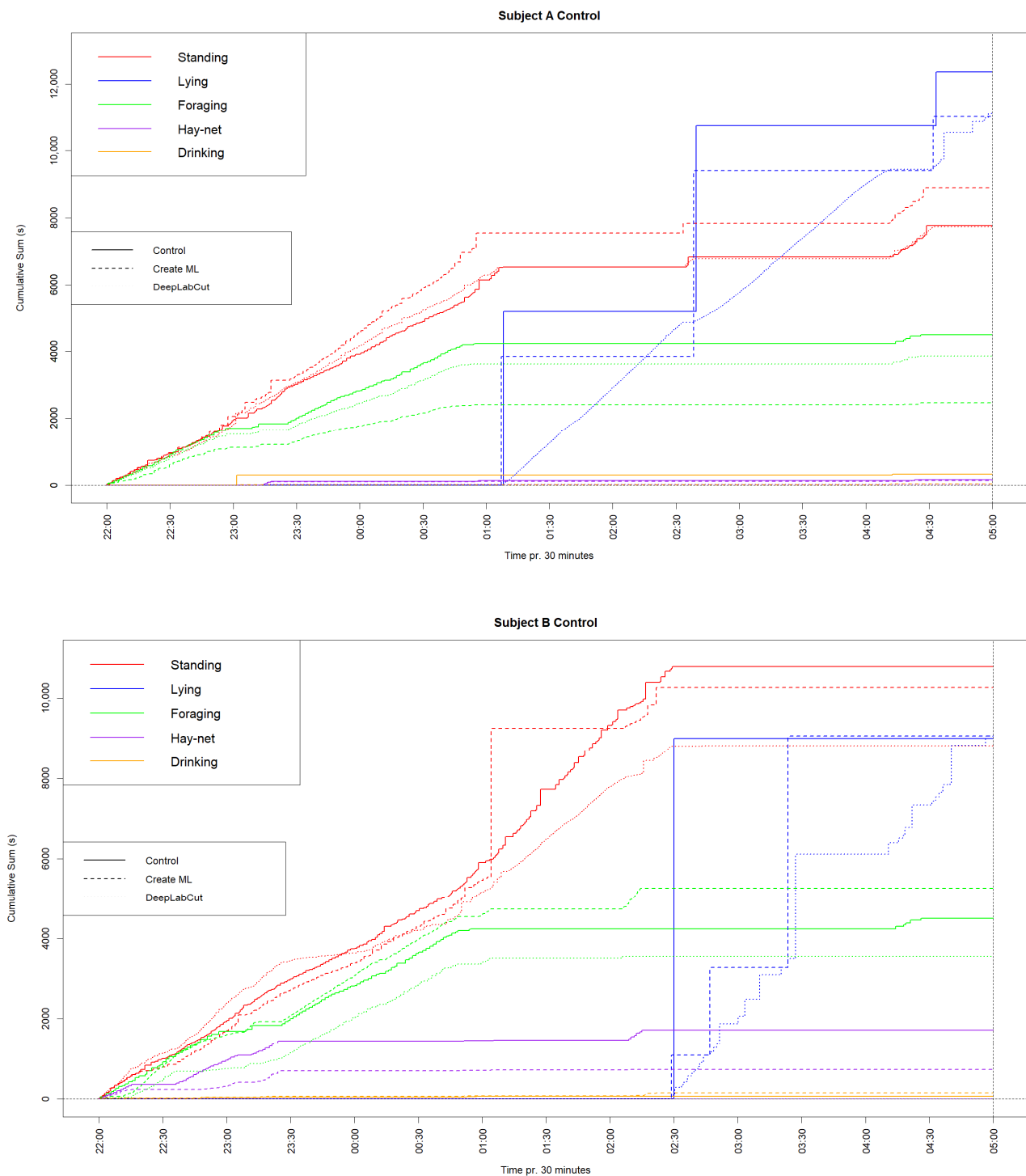
Figure 2. Graph showing the cumulative sums of each behavior for Subjects A and B during the 7 h control period, observed manually, using Create ML and DeepLabCut. Behaviors are distinguished by color and method.

The cumulative graph for Subject A showed a lower sum but similar shape for 'Lying down', which indicates that a period of observations of this behavior went unlabeled for both models. 'Standing' had a similar shape for all methods and a very similar sum for the DeepLabCut model, whereas the Create ML model had a higher sum. 'Foraging' had similar shapes for all methods; however, the sums were generally lower for the models. 'Hay-net' and 'Drinking' were difficult to distinguish clearly, due to low values.

The cumulative graph for Subject B had very similar values for 'Lying down' for all methods. All methods showed similar shapes for 'Standing', although the sums were lower for both models, most noticeably for the DeepLabCut model. 'Foraging' also had similar shapes but higher and lower sums for the Create ML model and DeepLabCut model, respectively. 'Hay-net' also showed a somewhat similar shape to the manual observations for the Create ML model, but this model had a lower sum. The DeepLabCut model showed a 'Hay-net' sum close to zero, due to difficulty in defining these parameters in the enclosure.

Similarly to the time budgets, 'Drinking' was left out of the cumulative graphs for the DeepLabCut model due to limitations in defining the parameters. Furthermore, the shapes in 'Lying down' for both subjects seem largely different; however, this is caused by the chosen type of cumulative graph.

3.1.2. Investigating the Reliability of Two Machine Learning Models

To investigate the reliability of the two models, Kendall's Coefficient of Concordance was utilized. The concordance (W-value) between the models and the control was found to be 0.85 with a p-value of 0.026 for Subject A, and 0.90 with a p-value of 0.019 for Subject B. This indicates a high concordance between the models that is not stochastic for both subjects. This high concordance means that the models and the manual scoring mostly agree on the observed behavior. However, this concordance is not perfect, so a slight disagreement is present.

To test the accuracy of the models, a confusion matrix for the control period was made and the models were compared to the manually observed values and normalized (Appendix C). As seen in Tables A1–A4, both models predicted highly similar values for the behavior 'Lying Down' for both subjects compared to the manually observed values. For the DeepLabCut model, the predicted values for 'Standing' and 'Foraging' were highly similar to the observed values for Subject A; but for Subject B, the model was more inaccurate. The opposite applies for Create ML, where the model was generally most accurate for Subject B. For the behavior 'Hay-net', the DeepLabCut model struggled to predict the correct behavior for Subject A, and for Subject B the classification parameters were not defined, and therefore no value was predicted for this behavior. The Create ML model for the behavior 'Hay-net' predicted highly similar values for Subject A, but for Subject B the behavior was often misclassified as 'Foraging'. For the behavior 'Drinking', the Create ML model often predicted the behavior as 'Standing' for both subjects and for the DeepLabCut model the classification parameters were not defined and therefore no values were predicted correctly. Finally, the predicted values by both models for out of view for Subject B were highly similar to the observed value but it is notable that the total manually observed value for this behavior is 17 s out of 7 h of observation time and is therefore arguably negligible.

3.2. Using Machine Learning Models for Behavioural Analysis

3.2.1. Assessing Behavioral Differences

To analyze behavioral differences between the two subjects, two time budgets for the total sums of each behavior for seven nights were constructed for both machine learning models, as is seen below in Figure 3.
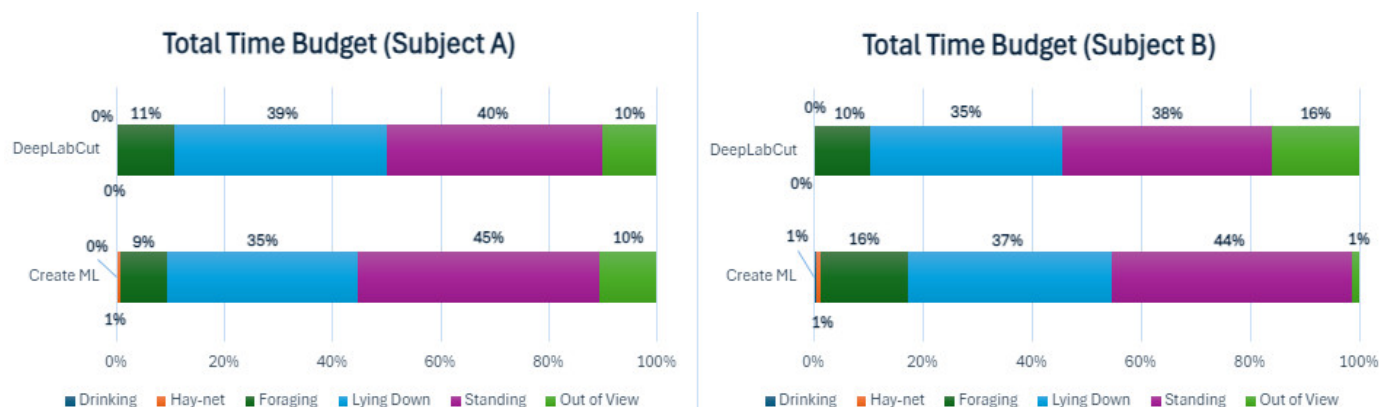
Figure 3. Time budgets for the total time spent on each behavior for Subjects A and B during all seven observed nights, for both ML models. The different colors show different behaviors.

The time budgets display some differences between the models, especially noticeable in the out of view percentages for Subject B. 'Standing' and 'Lying down' were similar for both models for both subjects. 'Foraging' was similar for Subject A in both models but was slightly higher for the Create ML model for Subject B, possibly due to the lower out of view percentage. The DeepLabCut model did not measure the 'Drinking' behavior for either subject. Comparing the total time budgets for the period between the two subjects only showed slight differences.

To further investigate the behaviors of both subjects during the observed period, the sums of behaviors for all individual days have been shown as time budgets in Appendix D. The time budgets for Subject A showed some variation in the behaviors, especially in 'Foraging'. 'Standing' and 'Lying down' also varied somewhat from night to night. Subject B also showed variation in 'Foraging', but generally less so than Subject A. 'Standing' and 'Lying down' varied somewhat for Subject B. There was a noticeable difference in out of view percentages for Subject B, depending on the model, with a consistently much lower percentage for Create ML. The behavioral differences were examined further using cumulative graphs (Appendix E).

Kendall's Coefficient of Concordance was used to examine if the amount of time spent on each behavior was the same each day. The analysis was conducted on the results of both the DeepLabCut and the Create ML model. Subject A showed a concordance of 0.935 with a *p*-value of 4.29×10^{-6} for the DeepLabCut model and 0.865 with a *p*-value of 1.31×10^{-5} for the Create ML model. Subject B showed a concordance of 0.951 with a *p*-value of 3.3×10^{-6} for the DeepLabCut model and 0.869 with a *p*-value of 1.21×10^{-5} for the Create ML model. All the concordance values were high with a significant *p*-value indicating that the high concordance is not stochastic. This high concordance indicates an agreement in the observed time a subject spends on different behaviors from day to day. This concordance is not perfect meaning some variations are still present in the subjects' nocturnal behavior.

Spearman's rank correlation for the behavior 'Lying down' was investigated for both models, and the analysis was split between days and individuals (Appendix F).

The analysis between days resulted in mainly positive correlations. Subject A had correlations between 0.965 and −0.053 for DeepLabCut and 0.970 and −0.135 for Create ML. Negative correlations were observed between the 20th and 22nd of April for DeepLabCut and the 19th and 20th of April for Create ML. Subject B had correlations between 0.999 and 0.019 for DeepLabCut and 1.000 and 0.311 for Create ML. No negative correlations were found for Subject B.

The analysis between the two subjects also resulted in mainly positive correlations (Appendix G). The results of the DeepLabCut model had correlations between 0.975 and −0.361. A single negative correlation was found between the 20th of April for Subject A and the 19th of April for Subject B. The results of the Create ML model had correlations between 0.977 and 0.052. No negative correlations were found for the Create ML model.

3.2.2. Investigating Further Applications of Automatic Behavioral Coding

Certain behaviors of a more complex character may potentially be assessed using machine learning methods for behavioral coding. One such behavior is the stereotyped behavior 'Swaying', which is largely relevant for elephants [14]. This behavior is difficult to categorize simply, as has been carried out for the previously mentioned behaviors, since swaying can happen anywhere in the frame and is primarily observable through a side-to-side motion of the elephant's trunk and head. To visualize this behavior using data from the DeepLabCut model, the cumulative distance moved by the point labelled at the trunk root of Subject B was plotted, along with the actual sway noted manually in the control period as a cumulative graph in Figure 4.
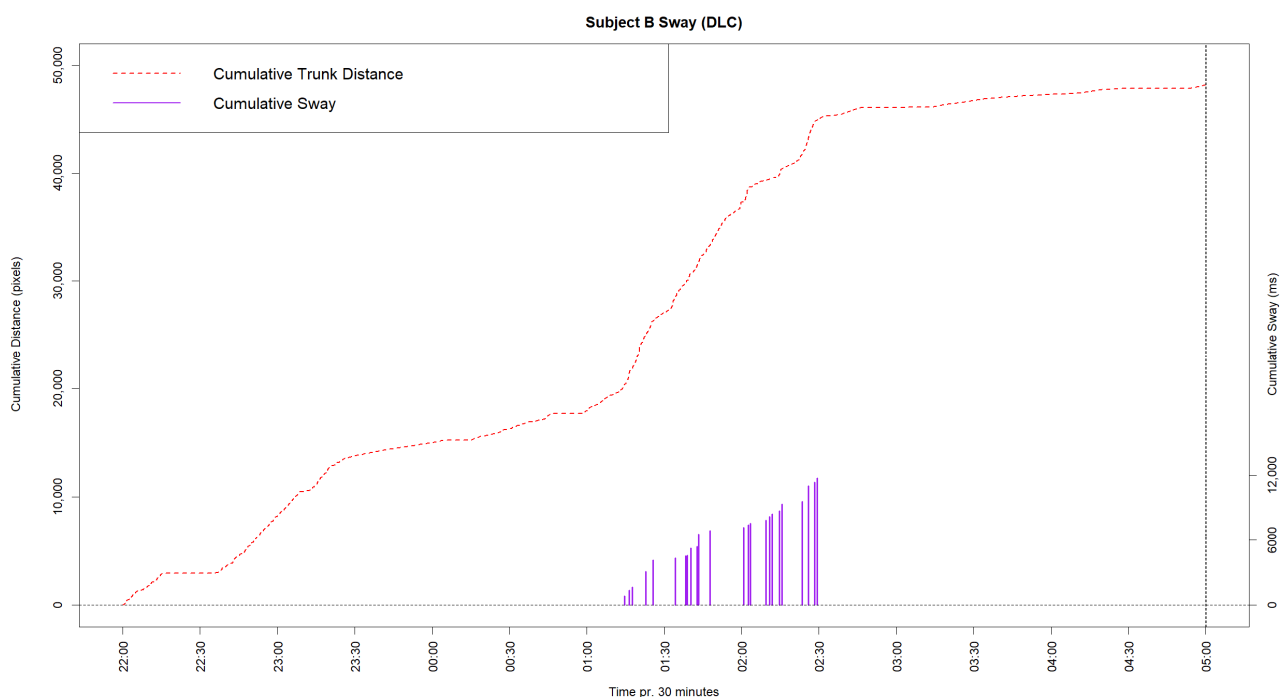
Figure 4. Cumulative graph of trunk root movement (red) and manually observed sway (purple) in the control period. The left *y*-axis shows cumulative pixel movement; the right *y*-axis shows the manually coded sway behavior in milliseconds.

As seen in the cumulative graph, a steep increase in the trunk distance appears around 01:15, which approximately matches with the manually observed sway behavior occurring at this time. This is because a steep increase in distance moved by the trunk root will occur as a result of the sway behavior.

4. Discussion

4.1. Performance and Limitations of the Two Machine Learning Models

Before using the two constructed machine learning models, it must first be investigated how accurate they are compared to manually recorded observations. The concordance test between the models showed a high agreement between the models and manual observations (W = 0.848), although there is seemingly room for improvement. The confusion matrixes for each model compared to the manual observations also displayed high accuracy in detecting some behaviors, such as 'Standing' and 'Lying down', although with certain challenges, such as confusing the 'Hay-net' behavior with 'Foraging' for Subject B. This suggests that there may be a need for improving the parameters of classifying each behavior or training the models with better or more material to account for different environments, footage qualities, and a broader range of behaviors. Mathis et al. (2018) discussed the

capabilities and limitations of DeepLabCut for various behaviors, noting similar challenges in behavior recognition and out of view instances [32].

It must, however, be noted that the accuracy of these models is based on comparison with manual observations, which itself has inherent problems. Manual observations are not entirely accurate, since they may lack precision in noting the exact time a behavior takes place, and there may be differences in how a sequence of behaviors is coded by different researchers, which usually necessitates inter-rater reliability tests [41,49,50]. These issues should not be present in machine learning models since a behavior is coded at the exact frame and can be standardized across studies. The capabilities of models constructed in both Create ML and DeepLabCut thus emphasize the potential of machine learning models to complement and enhance traditional manual observations in behavioral studies. Further optimization of such computer vision models may also include image processing such as exploring different color spaces and image augmentations [51,52].

4.2. Nocturnal Behavioral Differences of the Two Subjects

The two machine learning models were used to automatically observe the two subjects of the study, with the aim of assessing whether the nocturnal behaviors varied between each subject and individually across each observed night. This analysis was carried out using time budgets, cumulative graphs, and correlations.

Firstly, the total time budgets for each subject across all nights displayed only slight differences between the subjects, most notably in foraging behavior, which might be higher for Subject B. This slight difference is supported by the high concordance values between the days, meaning the observed behavior only differs slightly from day to day. It is, however, inconclusive whether Subject B generally carries out more foraging behavior, due to the differing out of view percentages caused by the lack of confidence by the models. With a closer look at the behavioral patterns using the cumulative graphs (Appendix E), it does, however, appear that the two subjects have some differences. Once again, it appears that 'Foraging' is generally higher for Subject B, along with 'Drinking'. From the 'Lying down' cumulative graph it also appears that sleeping patterns may be somewhat different, since Subject A appears to wake up and walk around more commonly throughout the night, whereas Subject B appears to be lying down for longer periods at a time. The correlations calculated regarding the sleeping patterns compared between the two subjects also showed some correlation, indicating that the subjects go to sleep at similar times, although this varies each night. This shows that the two elephants differ from each other in the nocturnal behavioral patterns; however, there do not appear to be large differences, and overall, the subjects typically carry out somewhat similar behavioral patterns throughout the night. This is in accordance with Bertelsen et al. (2020) which studied the same subjects and found some personality differences displayed through behavior, but similarly the differences were relatively small [14]. A study by Rees (2009) also found behavioral differences on an individual basis in captive Asian elephants (*Elephas maximus*) [10]. This is similar to Tobler (1992), Holdgate et al. (2016), and Schiffmann et al. (2023) who examined recumbent sleep behavior in zoo-housed Asian and African elephants, and also found differences on an individual basis [15,17,53].

It was investigated whether the individual subjects differed in their behavioral patterns from night to night, using time budgets for each night, along with cumulative graphs and correlations of their sleeping behavior. From the time budgets, both subjects appear to vary from night to night in all behaviors. 'Foraging' ranges from very low percentages (1–2%) to high percentages (18–22%), which is also apparent from the cumulative graph. This is in accordance with the study by Finch et al. (2021) who found varying feeding behavior in their nocturnal activity budgets for zoo-housed Asian elephants [54]. 'Standing' and 'Lying down' for both subjects also differed across nights with a range of approximately 20% difference for both behaviors. Further investigation of sleeping patterns using Spearman rank correlations showed that most days had very highly correlated values, indicating a general circadian rhythm; this is in accordance with a study by Casares et al. (2016)

that investigated cortisol levels to establish the circadian rhythm of African elephants [55]. However, some days showed much weaker correlation, suggesting differences in night-to-night sleeping patterns caused by the elephants going to sleep at different times. This confirms the hypothesis that computer vision models are capable of demonstrating that the nocturnal behavioral patterns differ from night to night for both individuals, although there is some uncertainty of exactly how much the behaviors differ, due to the out of view percentages. This result is, however, in accordance with the studies by Rees (2009) and Holdgate et al. (2016) who found considerable day-to-day variation in activity budgets for a group of captive Asian elephants and for African and Asian elephants, respectively [10,15].

This study showed that the subjects were lying down approximately 35–39% of the observed time, or just over 2.5 h until 5:00, at which point they would still be lying down, as is seen in the cumulative graphs. This is in accordance with studies such as Holdgate et al. (2016) and Schiffmann et al. (2023) who found that elephants in captivity generally tend to lie down for a similar amount of time during the night; although they also note that the elephants may not be sleeping throughout all of this time [15,17].

4.3. Other Applications of Machine Learning Models for Behavioral Coding

As was displayed in the results regarding the sway behavior, it may be difficult to address complex behaviors, even though it may be possible through different techniques. One such technique was displayed by plotting the distance moved by a point on the trunk root of Subject B. The steep incline on the graph largely matches with the manually observed sway during the night, which indicates that using such a measurement might be useful for observing 'Swaying'. Currently, the presentation of this behavior is, however, primarily visual since it may provide further challenges to precisely define the parameters capable of properly discerning when the gathered data should be categorized as sway behavior. Other challenges related to addressing complex behaviors, such as stereotypic and obsessive self-grooming in primates, face similar challenges since this behavior is also classified by a consistent repetition, which a computer vision model would have difficulties identifying on a frame-by-frame level. However, a study by Yin et al. (2024) is somewhat successful in showing distinct motion trajectories exhibited by a variety of different animals including tigers (*Panthera tigris*), bears (*Ursidae*), and wolves (*Canis lupus*), and classifying this as stereotypic behavior by assessing repetitive patterns [56]. Tackling issues that may arise using computer vision for behavior recognition is somewhat a case-by-case problem, where camera settings, image processing, and other issues should be considered, in order to fit the models appropriately to the research. However, developing such parameters and applying them to similar machine learning model data in the future would prove useful to quickly and accurately find stereotyped behavior to gain insight into the welfare of individual animals.

5. Conclusions

It is apparent from this study that using machine learning models from DeepLabCut and Create ML provides a capable tool for aiding or even replacing certain aspects of behavioral studies. The models could detect simple behaviors with high accuracy, although limitations were met when assessing repetitive behaviors such as 'Swaying'. Similar complex behaviors, such as certain stereotypic behaviors, in other animals may prove equally challenging and the detection of it may require further work to be adequate.

Applying the models to seven nights of footage of nocturnal behavior provided general insight into behavioral patterns and differences between the two studied subjects, as well as differences between individually observed days. This showed that the constructed computer vision models can effectively aid in behavioral analyses, and further expansion and adjustments may be desired. This could potentially be achieved through exploring image augmentation, classification of complex behavioral patterns, or implementing such models to be readily available as a tool for zoological gardens.

Appendix A. Floor Plan of Elephant Enclosure
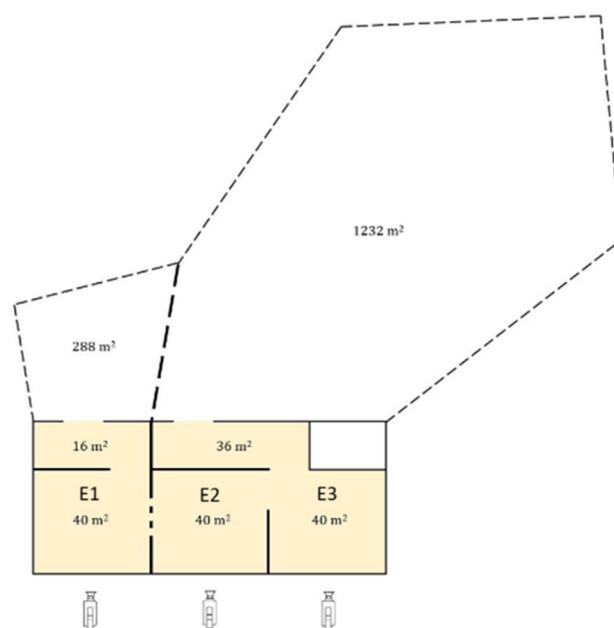
Figure A1. Illustration of the elephant enclosure in Aalborg Zoo. The indoor enclosures are colored yellow and have the enclosure names written with the dimensions. The dashed lines indicate the outdoor enclosure and the part of the wall between enclosures E1 and E2 where the subjects can have physical contact at night. The locations and positions of the cameras are indicated with the camera icons. The illustration is a modification from Bertelsen et al., 2020 [14].

Appendix B. Selected Behaviors Displayed in Subject Enclosures
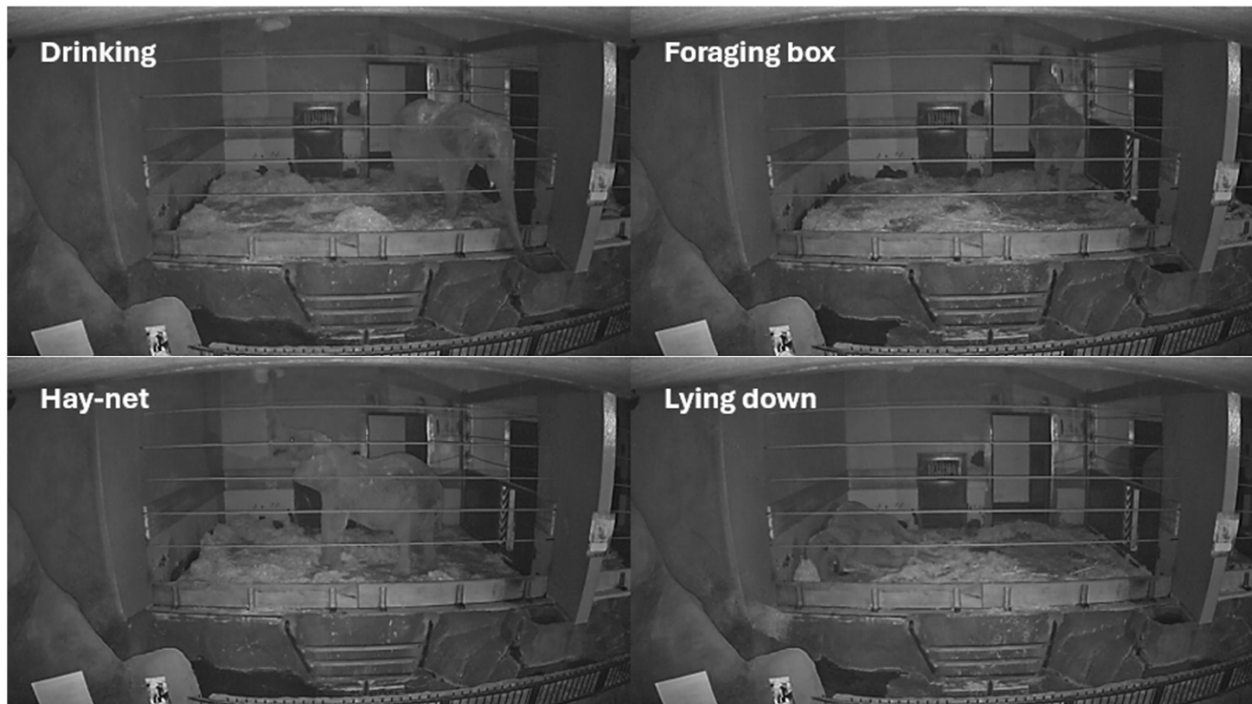
Figure A2. Selected behaviors displayed in the subject enclosure E1 as observed through video monitoring. The images showcase four distinct behaviors: drinking, foraging at the box, using the hay-net, and lying down.

Figure A3. Selected behaviors displayed in the subject enclosure E2 as observed through video monitoring. The images showcase four distinct behaviors: drinking, foraging at the box, using the hay-net, and lying down.

Appendix C. Confusion Matrix Comparison of Models

Table A1. Confusion matrix produced from the manually observed values compared to the predicted values from the DeepLabCut model for Subject A. Darker blue indicates better prediction.

Subject A	Behavior:	Manually Observed					
		Standing	Lying down	Foraging	Drinking	Hay-net	Out of view
Predicted by DeepLabCut model	Standing	0.86820	0.00016	0.13258	1.00000	0.39926	
	Lying down	0.00036	0.89418				
	Foraging	0.00701		0.84651			
	Drinking						
	Hay-net	0.00005				0.16883	
	Out of view	0.12437	0.10567	0.02091		0.43190	

Table A2. Confusion matrix produced from the manually observed values compared to the predicted values from the DeepLabCut model for Subject B. Darker blue indicates better prediction.

Subject B	Behavior:	Manually Observed					
		Standing	Lying down	Foraging	Drinking	Hay-net	Out of view
Predicted by DeepLabCut model	Standing	0.64320	0.00170	0.10185	0.76200	0.58339	0.18203
	Lying down	0.02517	0.99657		0.06867	0.00271	
	Foraging	0.02384		0.66607		0.14721	
	Drinking						
	Hay-net	0.00016					
	Out of view	0.30763	0.00173	0.23209	0.16933	0.26669	0.81797

Table A3. Confusion matrix produced from the manually observed values compared to the predicted values from the Create ML model for Subject A. Darker blue indicates better prediction.

Subject A	Behavior:	Manually Observed					
		Standing	Lying down	Foraging	Drinking	Hay-net	Out of view
Predicted by Create ML model	Standing	0.94253		0.29061	0.84503	0.07534	
	Lying down	0.00051	0.88842				
	Foraging	0.00718		0.53114			
	Drinking	0.00038			0.07602		
	Hay-net	0.00051				0.91096	
	Out of view	0.04887	0.11158	0.17825	0.07895	0.01370	

Table A4. Confusion matrix produced from the manually observed values compared to the predicted values from the Create ML model for Subject B. Darker blue indicates better prediction.

Subject B	Behavior:	Manually Observed					
		Standing	Lying down	Foraging	Drinking	Hay-net	Out of view
Predicted by Create ML model	Standing	0.91524	0.00121	0.04310	0.36667	0.06907	0.29412
	Lying down		0.99868				
	Foraging	0.05509		0.94939		0.50390	
	Drinking	0.00890	0.00011		0.61667		
	Hay-net	0.00130			0.01667	0.38559	
	Out of view	0.01948		0.00751		0.04144	0.70588

Appendix D. Time Budgets for Each Night

Figure A4. Time budgets for both models across each observed night for Subject A. Each color represents a behavior. Percentages were left out for 'Drinking' and 'Hay-net' due to low values.
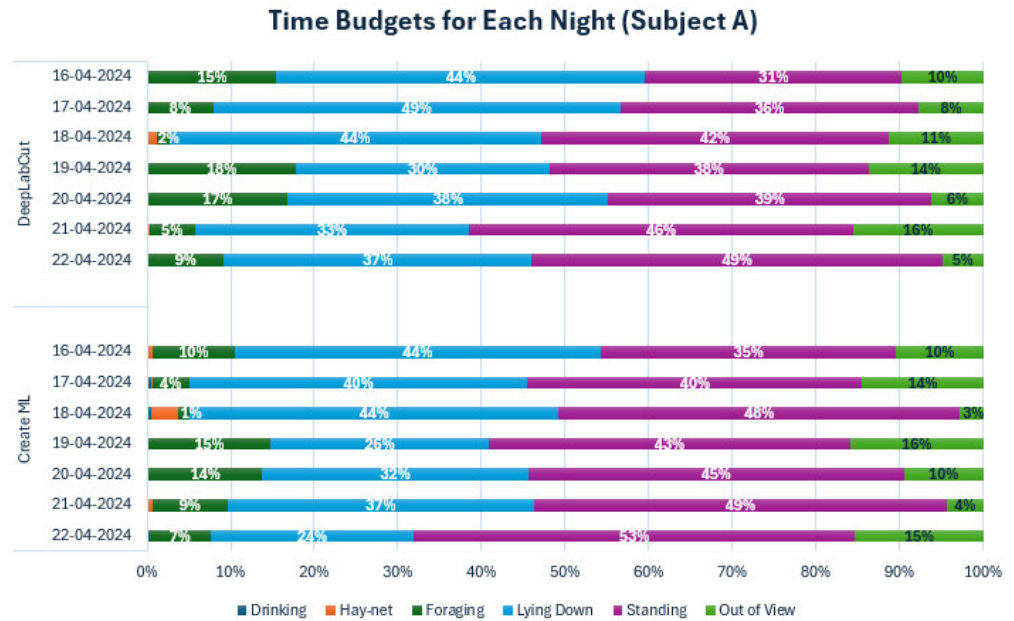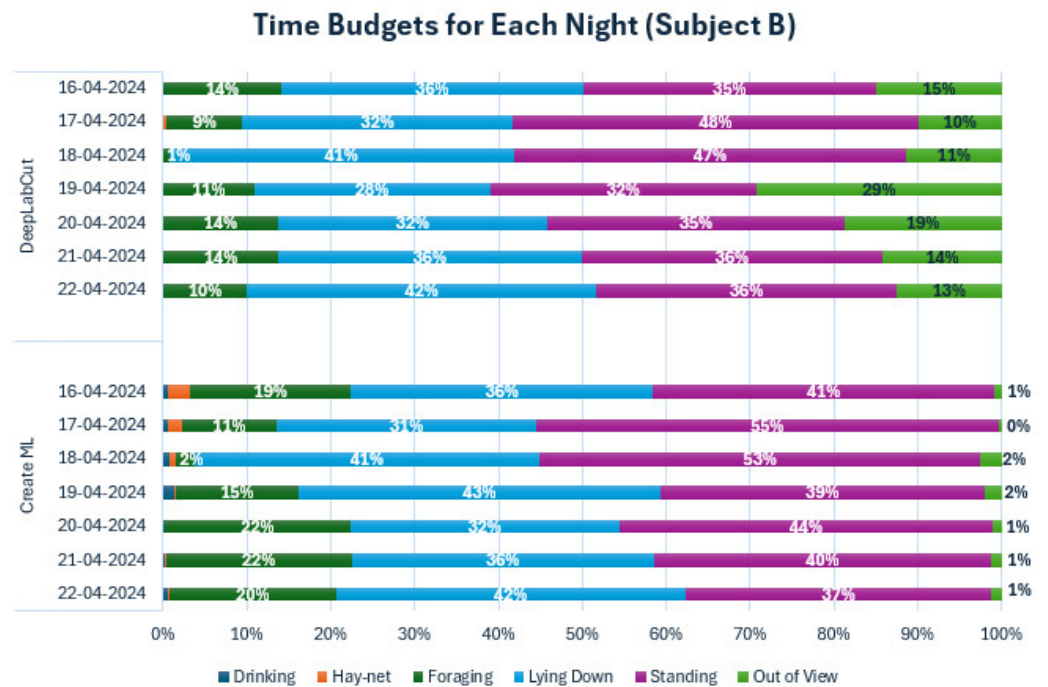

Figure A5. Time budgets for both models across each observed night for Subject B. Each color represents a behavior. Percentages were left out for 'Drinking' and 'Hay-net' due to low values.

Appendix E. Cumulative Graphs for Every Behavior Each Night
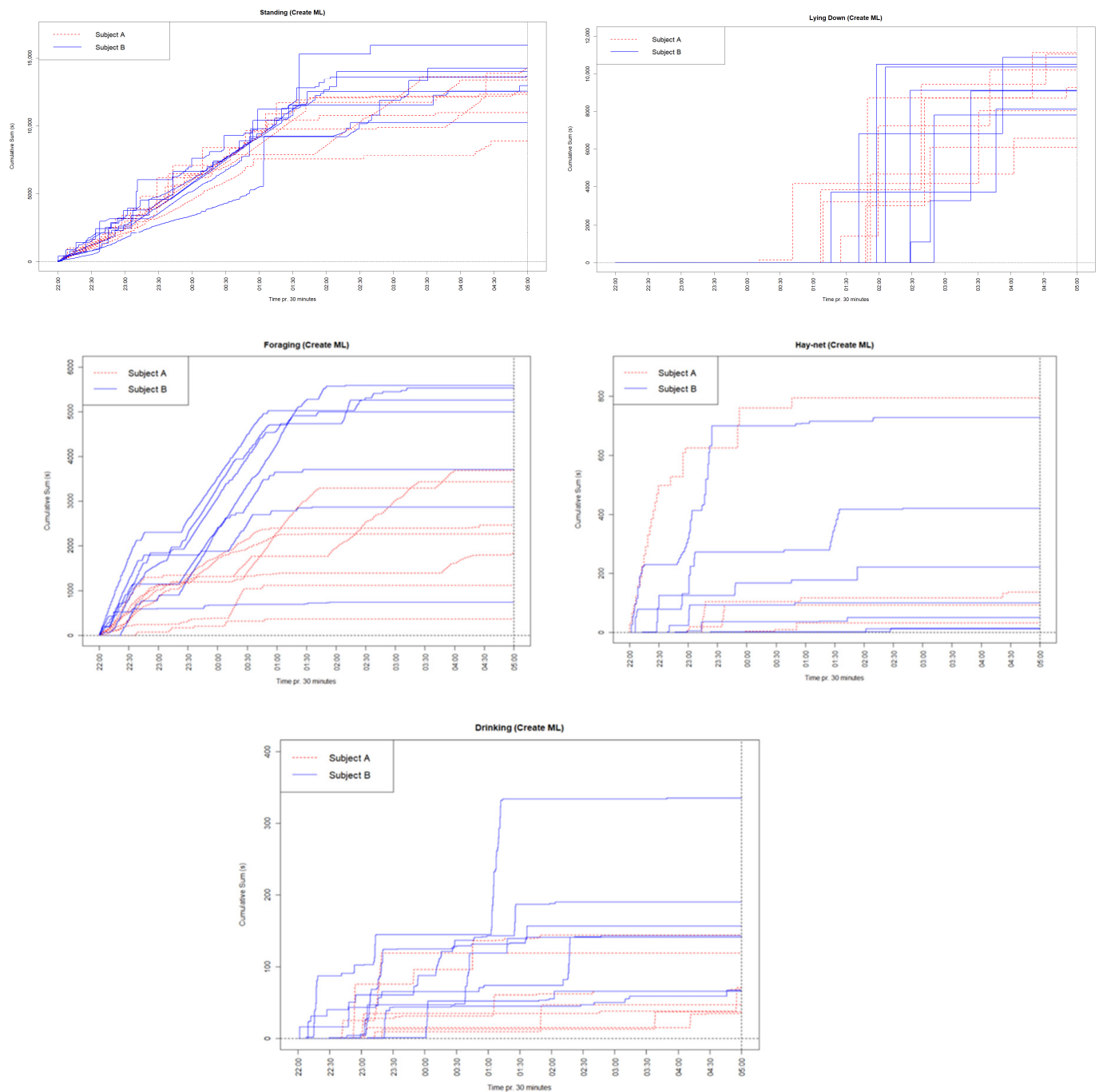
Figure A6. Cumulative graphs for every behavior for both subjects each observed night. Each color/line type signifies the subject, and every line signifies a different night.

The 'Standing' behavior was roughly similar in shape and sum for each subject and across nights, although Subject B had somewhat higher sums during several nights. 'Lying down' was also roughly similar for both subjects and between nights. 'Lying down' generally started occurring later at night but would then usually occur a lot for the rest of the night. 'Foraging' occurred mostly at the beginning of the night for both subjects, although Subject B tended to have higher sums than Subject A. 'Foraging' shapes and sums also varied from night to night. 'Hay-net' did not occur much for either individual but generally occurred most at the beginning of the night. 'Drinking' similarly did not

occur much for either individual but mostly occurred at the beginning of the night, and furthermore displayed higher sums for Subject B.

Appendix F. Spearman Rank Correlation between Days

Table A5. DeepLabCut—Subject A. Spearman rank correlation matrix for DeepLabCut model of Subject A, based on data collected between 16 April 2024 and 22 April 2024. The matrix displays the pairwise correlation coefficients between the data sets corresponding to each date. The color intensity represents the strength of the correlation, with darker green indicating higher positive correlation and darker red indicating higher negative correlation.

	16-04-2024	17-04-2024	18-04-2024	19-04-2024	20-04-2024	21-04-2024	22-04-2024
16-04-2024	1.000						
17-04-2024	0.820	1.000					
18-04-2024	0.756	0.965	1.000				
19-04-2024	0.465	0.779	0.791	1.000			
20-04-2024	0.367	0.307	0.237	0.207	1.000		
21-04-2024	0.716	0.938	0.924	0.700	0.276	1.000	
22-04-2024	0.786	0.612	0.520	0.250	−0.053	0.450	1.000

Table A6. DeepLabCut—Subject B. Spearman rank correlation matrix for DeepLabCut model of Subject B, based on data collected between 16 April 2024 and 22 April 2024. The matrix displays the pairwise correlation coefficients between the data sets corresponding to each date. The color intensity represents the strength of the correlation, with darker green indicating higher positive correlation and darker red indicating higher negative correlation.

	16-04-2024	17-04-2024	18-04-2024	19-04-2024	20-04-2024	21-04-2024	22-04-2024
16-04-2024	1.000						
17-04-2024	0.930	1.000					
18-04-2024	0.860	0.777	1.000				
19-04-2024	0.299	0.115	0.573	1.000			
20-04-2024	0.297	0.401	0.357	0.019	1.000		
21-04-2024	0.997	0.924	0.872	0.315	0.314	1.000	
22-04-2024	0.843	0.760	0.999	0.601	0.372	0.856	1.000

Table A7. Create ML—Subject A. Spearman rank correlation matrix for Create ML model of Subject A, based on data collected between 16 April 2024 and 22 April 2024. The matrix displays the pairwise correlation coefficients between the data sets corresponding to each date. The color intensity represents the strength of the correlation, with darker green indicating higher positive correlation and darker red indicating higher negative correlation.

	16-04-2024	17-04-2024	18-04-2024	19-04-2024	20-04-2024	21-04-2024	22-04-2024
16-04-2024	1.000						
17-04-2024	0.902	1.000					
18-04-2024	0.872	0.918	1.000				
19-04-2024	0.759	0.841	0.859	1.000			
20-04-2024	0.222	0.043	0.102	−0.135	1.000		
21-04-2024	0.896	0.934	0.970	0.855	0.112	1.000	
22-04-2024	0.737	0.741	0.501	0.473	0.048	0.516	1.000

Table A8. Create ML—Subject B. Spearman rank correlation matrix for Create ML model of Subject B, based on data collected between 16 April 2024 and 22 April 2024. The matrix displays the pairwise correlation coefficients between the data sets corresponding to each date. The color intensity represents the strength of the correlation, with darker green indicating higher positive correlation and darker red indicating higher negative correlation.

	16-04-2024	17-04-2024	18-04-2024	19-04-2024	20-04-2024	21-04-2024	22-04-2024
16-04-2024	1.000						
17-04-2024	0.933	1.000					
18-04-2024	0.913	0.842	1.000				
19-04-2024	0.747	0.662	0.882	1.000			
20-04-2024	0.311	0.414	0.379	0.558	1.000		
21-04-2024	1.000	0.933	0.917	0.752	0.313	1.000	
22-04-2024	0.852	0.775	0.988	0.892	0.373	0.857	1.000

Appendix G. Spearman Rank Correlation between Individuals

Table A9. Spearman rank correlation matrix comparing the data between Subject A and Subject B for the DeepLabCut model, from 16 April 2024 to 22 April 2024. The matrix presents the correlation coefficients between corresponding dates for the two subjects. The color intensity represents the strength of the correlation, with darker green indicating higher positive correlation and darker red indicating higher negative correlation.

		Subject B						
		16-04-2024	17-04-2024	18-04-2024	19-04-2024	20-04-2024	21-04-2024	22-04-2024
Subject A	**16-04-2024**	0.476						
	17-04-2024	0.772	0.700					
	18-04-2024	0.839	0.762	0.975				
	19-04-2024	0.884	0.790	0.783	0.407			
	20-04-2024	0.372	0.508	0.194	−0.361	0.717		
	21-04-2024	0.654	0.622	0.881	0.551	0.672	0.672	
	22-04-2024	0.203	0.068	0.441	0.879	0.197	0.224	0.474

Table A10. Spearman rank correlation matrix comparing the data between Subject A and Subject B for the Create ML model, from 16 April 2024 to 22 April 2024. The matrix presents the correlation coefficients between corresponding dates for the two subjects. The color intensity represents the strength of the correlation, with darker green indicating higher positive correlation and darker red indicating higher negative correlation.

		Subject B						
		16-04-2024	17-04-2024	18-04-2024	19-04-2024	20-04-2024	21-04-2024	22-04-2024
Subject A	**16-04-2024**	0.690						
	17-04-2024	0.673	0.563					
	18-04-2024	0.855	0.785	0.977				
	19-04-2024	0.532	0.470	0.790	0.920			
	20-04-2024	0.206	0.314	0.112	0.052	0.602		
	21-04-2024	0.743	0.711	0.918	0.874	0.491	0.748	
	22-04-2024	0.347	0.194	0.419	0.512	0.066	0.349	0.472

References

1. Veasey, J.S. Differing animal welfare conceptions and what they mean for the future of zoos and aquariums, insights from an animal welfare audit. *Zoo Biol.* **2022**, *41*, 292–307. [CrossRef] [PubMed]
2. Barongi, R.; Fisken, F.A.; Parker, M.; Gusset, M. *Committing to Conservation: The World Zoo and Aquarium Conservation Strategy*; World Association of Zoos and Aquariums (WAZA) Executive Office: Gland, Switzerland, 2015.
3. European Association of Zoos and Aquaria. *Standards for the Accommodation and Care of Animals in Zoos and Aquaria*; EAZA: Amsterdam, The Netherlands, 2014.

4.   Danish Association of Zoos and Aquaria. *DAZA Etiske Retningslinjer*; DAZA: Aalborg, Denmark, 2022. Available online: https://www.daza.dk/pdf/DAZA%20etiske%20retningslinjer%20-%20godkendt%20GF%2015.06.2022-underskrevet.pdf (accessed on 9 April 2023).

5.   Sutherland, W.J. The importance of behavioural studies in conservation biology. *Anim. Behav.* **1998**, *56*, 801–809. [CrossRef] [PubMed]

6.   Wolfensohn, S.; Shotton, J.; Bowley, H.; Davies, S.; Thompson, S.; Justice, W.S.M. Assessment of Welfare in Zoo Animals: Towards Optimum Quality of Life. *Animals* **2018**, *8*, 110. [CrossRef] [PubMed]

7.   Perdue, B.M.; Sherwen, S.L.; Maple, T.L. Editorial: The Science and Practice of Captive Animal Welfare. *Front. Psychol.* **2020**, *11*, 1851. [CrossRef] [PubMed]

8.   Sherwen, S.L.; Hemsworth, P.H. The Visitor Effect on Zoo Animals: Implications and Opportunities for Zoo Animal Welfare. *Animals* **2019**, *9*, 366. [CrossRef] [PubMed]

9.   Hauenstein, S.; Jassoy, N.; Mupepele, A.; Carroll, T.; Kshatriya, M.; Beale, C.M.; Dormann, C.F. A systematic map of demographic data from elephant populations throughout Africa: Implications for poaching and population analyses. *Mammal Rev.* **2022**, *52*, 438–453. [CrossRef]

10.   Rees, A.P. Activity budgets and the relationship between feeding and stereotypic behaviors in Asian elephants (*Elephas maximus*) in a Zoo. *Zoo Biol.* **2009**, *28*, 79–97. [CrossRef] [PubMed]

11.   Yon, L.; Williams, E.; Harvey, N.D.; Asher, L. Development of a behavioural welfare assessment tool for routine use with captive elephants. *PLoS ONE* **2019**, *14*, e0210783. [CrossRef] [PubMed]

12.   Greco, B.J.; Meehan, C.L.; Heinsius, J.L.; Mench, J.A. Why pace? The influence of social, housing, management, life history, and demographic characteristics on locomotor stereotypy in zoo elephants. *Appl. Anim. Behav. Sci.* **2017**, *194*, 104–111. [CrossRef]

13.   Bansiddhi, P.; Nganvongpanit, K.; Brown, J.L.; Punyapornwithaya, V.; Pongsopawijit, P.; Thitaram, C. Management factors affecting physical health and welfare of tourist camp elephants in Thailand. *Biodivers. Conserv.* **2019**, *7*, e6756. [CrossRef]

14.   Bertelsen, S.S.; Sørensen, A.S.; Pagh, S.; Pertoldi, C.; Jensen, T.H. Nocturnal Behaviour of Three Zoo Elephants (*Loxodonta africana*). *Genet. Biodivers. J.* **2020**, *4*, 92–113. [CrossRef]

15.   Holdgate, M.R.; Meehan, C.L.; Hogan, J.N.; Miller, L.J.; Rushen, J.; de Passillé, A.M.; Soltis, J.; Andrews, J.; Shepherdson, D.J. Recumbence Behavior in Zoo Elephants: Determination of Patterns and Frequency of Recumbent Rest and Associated Environmental and Social Factors. *PLoS ONE* **2016**, *11*, e0153301. [CrossRef] [PubMed]

16.   Boyle, S.A.; Roberts, B.; Pope, B.M.; Blake, M.R.; Leavelle, S.R.; Marshall, J.J.; Smith, A.; Hadicke, A.; Falcone, J.F.; Knott, K.; et al. Assessment of Flooring Renovations on African Elephant (*Loxodonta africana*) Behavior and Glucocorticoid Response. *PLoS ONE* **2015**, *10*, e0141009. [CrossRef]

17.   Schiffmann, C.; Hellriegel, L.; Clauss, M.; Stefan, B.; Knibbs, K.; Wenker, C.; Hård, T.; Galeffi, C. From left to right all through the night: Characteristics of lying rest in zoo elephants. *Zoo Biol.* **2022**, *42*, 17–25. [CrossRef]

18.   Greco, B.J.; Meehan, C.L.; Hogan, J.N.; Leighty, K.A.; Mellen, J.; Mason, G.J.; Mench, J.A. The Days and Nights of Zoo Elephants: Using Epidemiology to Better Understand Stereotypic Behvaior of African Elephants (*Loxodonta africana*) and Asian Elephants (*Elephas maximus*) in North American Zoos. *PLoS ONE* **2016**, *11*, e0144276. [CrossRef]

19.   Broom, D.M. Welfare of Animals: Behavior as a Basis for Decisions. *Encycl. Anim. Behav.* **2010**, *3*, 580–584.

20.   Jacobson, S.L.; Ross, S.R.; Bloomsmith, M.A. Characterizing abnormal behavior in a large population of zoo-housed chimpanzees: Prevalence and potential influencing factors. *PeerJ* **2016**, *4*, e2225. [CrossRef] [PubMed]

21.   Bacon, H. Behaviour-Based Husbandry—A Holistic Approach to the Management of Abnormal Repetitive Behaviors. *Animals* **2018**, *8*, 103. [CrossRef]

22.   Fuktong, S.; Yuttasaen, P.; Punyapornwithaya, V.; Brown, J.L.; Thitaram, C.; Luevitoonvechakij, N.; Bansiddhi, P. A survey of stereotypic behaviors in tourist camp elephants in Chiang Mai, Thailand. *Appl. Anim. Behav. Sci.* **2021**, *243*, 105456. [CrossRef]

23.   Mason, G.J. Stereotypies and suffering. *Behav. Process.* **1991**, *25*, 103–115. [CrossRef] [PubMed]

24.   Mason, G.J.; Rushen, J. *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare*; CABI Digital Library: Wallingford, UK, 2006.

25.   Mostard, K.E. *General Understanding, Neuro-Endocrinologic and (Epi)Genetic Factors of Stereotypy*; Radboud University of Nijmegen: Nijmegen, The Netherlands, 2011.

26.   Altmann, J. Observational Study of Behavior: Sampling Methods. *Behaviour* **1974**, *49*, 227–267. [CrossRef]

27.   Vanitha, V.; Thiyagesan, K.; Baskaran, N. Prevalence of stereotypies and its possible causes among captive Asian elephants (*Elephas maximus*) in Tamil Nadu, India. *Appl. Anim. Behav. Sci.* **2016**, *174*, 137–146.

28.   Adams, J.; Berg, J.K. Behavior of Female African Elephants (*Loxodonta africana*) in Captivity. *Appl. Anim. Ethol.* **1980**, *6*, 257–276. [CrossRef]

29.   Dell, A.I.; Bender, J.A.; Branson, K.; Couzin, I.D.; de Polavieja, G.G.; Noldus, L.P.; Pérez-Escudero, A.; Perona, P.; Straw, A.D.; Wikelski, M.; et al. Automated image-based tracking and its application in ecology. *Trends Ecol. Evol.* **2014**, *29*, 417–428. [CrossRef] [PubMed]

30.   Gomez-Marin, A.; Paton, J.J.; Kampff, A.R.; Costa, R.M.; Mainen, Z.F. Big Behavioral Data: Psychology, Ethology and the Foundations of Neuroscience. *Nat. Neurosci.* **2014**, *17*, 1455–1462. [CrossRef]

31.   Anderson, D.J.; Perona, P. Toward a Science of Computational Ethology. *Neuron* **2014**, *84*, 18–31. [CrossRef] [PubMed]

32. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289. [CrossRef]

33. Mirkó, E.; Dóka, A.; Miklósi, Á. Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (*Canis familiaris*). *Appl. Anim. Behav. Sci.* **2013**, *149*, 45–54. [CrossRef]

34. Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef] [PubMed]

35. Zhou, Z.-H.; Liu, S. *Machine Learning*; Springer Nature: Nanjing, China, 2021.

36. Lenzi, J.; Barnas, A.F.; ElSaid, A.A.; Desell, T.; Rockwell, R.F.; Ellis-Felege, S.N. Artificial intelligence for automated detection of large mammals creates path to upscale drone surveys. *Sci. Rep.* **2023**, *13*, 947. [CrossRef]

37. Bain, M.; Nagrani, A.; Schofield, D.; Berdugo, S.; Bessa, J.; Owen, J.; Hockings, K.J.; Matsuzawa, T.; Hayashi, M.; Biro, D.; et al. Automated audiovisual behavior recognition in wild primates. *Sci. Adv.* **2021**, *7*, eabi4883. [CrossRef] [PubMed]

38. Sakib, F.; Burghardt, T. Visual Recognition of Great Ape Behaviours in the Wild. *arXiv* **2020**, arXiv:2011.10759.

39. Hebbar, P.K.; Pullela, P.K. Deep Learning in Object Detection: Advancements in Machine Learning and AI. In Proceedings of the 2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM), Bangalore, India, 28–29 November 2023.

40. Hardin, A.; Schlupp, I. Using machine learning and DeepLabCut in animal behavior. *Acta Ethologica* **2022**, *25*, 125–133. [CrossRef]

41. Marks, M.; Jin, Q.; Sturman, O.; von Ziegler, L.; Kollmorgen, S.; von der Behrens, W.; Mante, V.; Bohacek, J.; Yanik, M.F. Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments. *Nat. Mach. Intell.* **2022**, *4*, 331–340. [CrossRef]

42. Nath, T.; Mathis, A.; Chen, A.C.; Patel, A.; Bethge, M.; Mathis, M.W. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* **2019**, *14*, 2152–2176. [CrossRef] [PubMed]

43. Marques, O. Machine Learning with Core ML. In *Image Processing and Computer Vision in IOS*; Springer International Publishing AG: Cham, Switzerland, 2020; pp. 29–40.

44. Andersen, T.A.; Herskind, C.; Maysfelt, J.; Rørbæk, R.W.; Schnoor, C.; Pertoldi, C. The nocturnal behaviour of African elephants (*Loxodonta africana*) in Aalborg Zoo and how changes in the environment affect them. *Genet. Biodivers. J.* **2020**, *4*, 114–130. [CrossRef]

45. Ruuska, S.; Hämäläinen, W.; Kajava, S.; Mughal, M.; Matilainen, P.; Mononen, J. Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behav. Process.* **2018**, *148*, 56–62. [CrossRef] [PubMed]

46. Larsen, J.F.; Andersen, K.K.D.; Cuprys, J.; Fosgaard, T.B.; Jacobsen, J.H.; Krysztofiak, D.; Lund, S.M.; Nielsen, B.; Pedersen, M.E.B.; Pedersen, M.J.; et al. Behavioral analysis of a captive male Bornean orangutan (*Pongo pygmaeus*) when. *Arch. Biol. Sci.* **2023**, *75*, 443–458. [CrossRef]

47. Field, A.P. Kendall's Coefficient of Concordance. *Encycl. Stat. Behav. Sci.* **2005**, *2*, 1010–1011.

48. Brownlee, J. Machine Learning Mastery. Guiding Tech Media. 19 August 2020. Available online: https://machinelearningmastery.com/distance-measures-for-machine-learning/ (accessed on 22 May 2024).

49. Malviya, M.; Buswell, N.T.; Berdanier, C.G.P. Visual and Statistical Methods to Calculate Intercoder Reliability for Time-Resolved Observational Research. *Int. J. Qual. Methods* **2021**, *20*, 16094069211002418. [CrossRef]

50. Whiteway, M.R.; Biderman, D.; Friedman, Y.; Dipoppa, M.; Buchanan, E.K.; Wu, A.; Zhou, J.; Bonacchi, N.; Miska, N.J.; Noel, J.-P.; et al. Partitioning variability in animal behavioral videos using semi-supervised variational autoencoders. *PLoS Comput. Biol.* **2021**, *17*, e1009439. [CrossRef] [PubMed]

51. García-Mateos, G.; Hernández-Hernández, J.; Escarabajal-Henarejos, D.; Jaén-Terrones, S.; Molina-Martínez, J. Study and comparison of color models for automatic image analysis in irrigation management applications. *Agric. Water Manag.* **2015**, *151*, 158–166. [CrossRef]

52. Hernández-Hernández, J.L.; García-Mateos, G.; González-Esquiva, J.M.; Escarabajal-Henarejos, D.; Ruiz-Canales, A.; Molina-Martínez, J.M. Optimal color space selection method for plant/soil segmentation in agriculture. *Comput. Electron. Agric.* **2016**, *122*, 124–132. [CrossRef]

53. Tobler, I. Fundamental Research Behavioral Sleep in the Asian Elephant in Captivity. *Sleep* **1992**, *15*, 1–12. [PubMed]

54. Finch, K.; Sach, F.; Fitzpatrick, M.; Rowden, L.J. Insights into Activity of Zoo Housed Asian Elephants (*Elephas maximus*) during Periods of Limited Staff and Visitor Presence, a Focus on Resting Behaviour. *Zool. Bot. Gard.* **2021**, *2*, 101–114. [CrossRef]

55. Casares, M.; Silván, G.; Carbonell, M.D.; Gerique, C.; Martinez-Fernandez, L.; Cáceres, S.; Illera, J.C. Circadian rhythm of salivary cortisol secretion in female zoo-kept African elephants (*Loxodonta africana*). *Zoo Biol.* **2016**, *35*, 65–69. [CrossRef] [PubMed]

56. Yin, Z.; Zhao, Y.; Xu, Z.; Yu, Q. Automatic detection of stereotypical behaviors of captive wild animals based on surveillance videos of zoos and animal reserves. *Ecol. Inform.* **2024**, *79*, 102450. [CrossRef]