COMP3850 Group 23 Requirements and Scoping Document

Revision history

Revision Number	Date	Person(s)	Changes
1.0	28/03/2024	Group 23	Initial version
1.1	27/04/2024	Michael Yee	Added paragraphs to Modelling section explaining how hyperparameter tuning would be performed and risks of algorithmic bias Added paragraphs to Evaluation section discussing specific measures for comparing model performance
	16/05/2024	N/A	Document reviewed and deemed reflective of the current state of the project, no changes made

1. Introduction

1.1 Purpose/Intended audience

This document will serve to outline the scope of Project APRA and how it will follow the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, a proven standard for data science projects that includes six phases: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment[1]. This document is intended for researchers and technicians who intend on using Social LEAP Estimates Animal Poses (SLEAP) as a means of posture recognition for animals and insects. Although users do not need to be familiar with data analysis or ant biology, it is assumed that users have basic computer knowledge and skills.

1.2 Overview

Project Automated Posture Recognition of Ants (APRA) will be utilising SLEAP as its main software solution for accurate multi-animal pose estimation of the weaver ants[2]. SLEAP is a computer vision framework designed specifically for identifying the postures of animals in video data and has been used in other research projects for posture analysis of insects such as ants and bees[3,4]. This unique software package will allow us to gain valuable insight into the ants' behaviour and how they interact with their environment and each other. More specifically, Project APRA will focus on how weaver ants are able to collaborate with one another to form living winches otherwise known as pulling chains, which are used to pull leaves together during the formation of their nests. Using SLEAP, we plan on analysing how the synchronised positioning of the legs of each ant is able to assist in creating a consistent pulling force.

Furthermore, by leveraging SLEAP's advanced machine learning technology, we hope that the newly discovered knowledge regarding the weaver ant's behaviours will be able to assist in future animal/insect projects as well as enhancements in other fields such as innovation in the robotics industry.

1.3 Scope

The scope of Project APRA includes the development and delivery of a data model used for analysis of the posture of ants whilst forming a pulling chain. The data model itself will contain, at least, an instance, skeleton and labels for a single ant's thorax, head and legs, and if time permits, an analysis with a conclusion as to the postures involved in the formation of a pulling chain.

Additionally, the scope of Project APRA includes the provision of a user guide for repeated application in future project iterations that users will be able to adopt in order to train relevant data and build new models.This guide will include how to carry out the analysis of rudimentary pulling chain scenarios, such as single ants at the beginning of a chain-forming activity.

The following list of software will be required to carry out Project APRA:

- SLEAP[5] for posture tracking and analysis (data mining) as well as model development.
- Google Docs[6] for developing required documentation.
- GitHub[7] for capturing and developing a user manual for ongoing APRA iterations.
- Trello[8] for navigating and organising required tasks.

Project APRA will deliver the following outcomes and resources to the client:

- A developed model to apply to ongoing APRA projects/iterations, including:
    - Skeletons and labels
    - Sample data
    - A recorded setup and training of data using the supplied skeletons and labels
- A document containing simplified and relevant instructions to set up and use SLEAP software.
- A GitHub repository for ongoing development of the models and user manual, including historical versioned instances of the manual and the capability to manage ongoing change, and versions of the resulting data model to use in ongoing analysis as an accompaniment to the user guide.
- A report detailing the outcomes and analysis produced as a result of carrying out Project APRA, which will include:
    - Details of the posture labels successfully captured and tracked.
    - Details of analysis we were able to carry out of the posture of at least one weaver ant during a pulling chain event.
    - The outcome of that analysis, if any, concerning the effect of various posture measures on force generation, in either a single ant or a chain of pulling ants.

1.3.1 Context and Flow Diagrams

Below is a context diagram (Figure 1), illustrating the external and internal entities of the APRA project and the transactions that occur between these entities in order to produce the final output of the project.

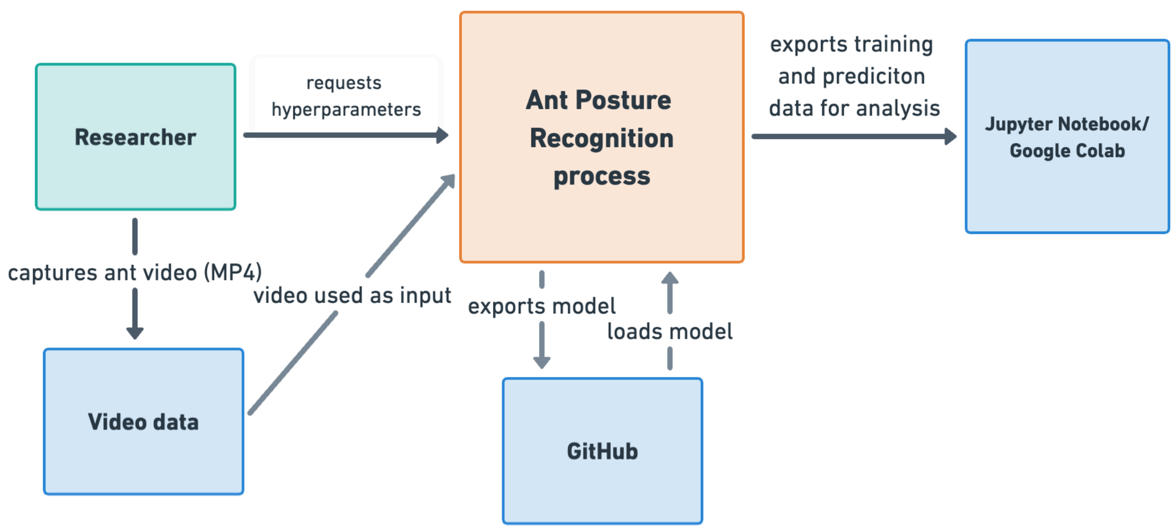

Figure 1: Context Diagram

Figure 2 below is a flow diagram capturing how the data resources, modelling, decision making mechanisms will work within the SLEAP software we will be using

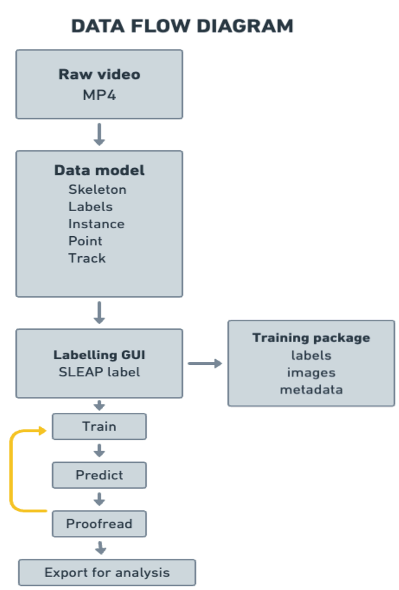

Figure 2: Data flow diagram

Figure 3 below illustrates the data model entities involved in creation of our final output.

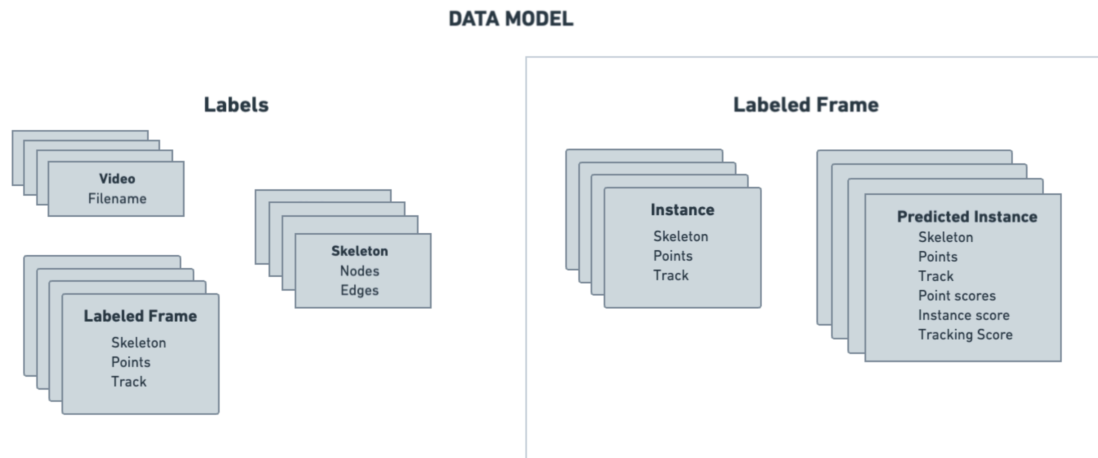

Figure 3: Data model

2. Data Understanding

2.1 Data description

The data is provided in the form of 23 video files recorded in MPEG-4 format, most of which are 3840 x 2160 pixels large but all are at least 1280 x 720 pixels in resolution. The file size of the videos range from 2.03GB to 19.70GB and the length of the videos range from 20 minutes to 45 minutes. The videos show weavers ants pulling each other in chains to fold a leaf shaped piece of paper to create a nest in. They also show individual ants pulling on the paper and many instances of ants walking over other ants.

2.2 Sources

The videos were recorded at the Macquarie University School of Natural Sciences by Masters student Madelyne Stewardson, working in the lab of Chris Reid and his team. The videos were recorded in June 2022 during the School of Natural Sciences experiments into measuring the pulling force of weaver ant chains when pulling an object.

2.3 Capture

The videos were filmed from above, perpendicular to the leaf substrate, with a Panasonic Lumix GH-5 digital camera equipped with a Macro 30mm lens at 24 frames per second. The videos were produced to explore how weaver ants produce force when pulling in isolation, and when pulling in teams of ants as pulling chains, specifically exploring whether ants in chains are able to produce more force per individual than the same number of ants pulling as single units.

2.4 Storage/Environment

The videos were provided to the team by Chris Reid through OneDrive where he uploaded all 23 of the videos. Team Member Michael Yee has cropped the videos to remove space where the ants weren't present to make it easier for other team members to download and so there is less labelling required in SLEAP. The cropped videos have been uploaded to Google Drive and made available to all team members.

2.5 Data quality

After watching the videos we can conclude that the quality of the data provided is very good for the following reasons:

- Data accuracy is high, as the client can provide a large amount of data with multiple hours of video.
- The data can be considered both relevant and reliable due to being used in previous analyses on the force produced by the ant chains.

- The data timeliness is high, as the initial results are ready to be published and the more detailed analysis we will produce will be perfectly positioned for publishing as a follow-up study.

Combined, these factors indicate the data quality is very high and will be very useful for our development of Project APRA. Therefore, it is highly likely that no further data will need to be collected to complete the project.

3. Data Preparation

As this project is a computer vision project, there are considerable differences in how the data are to be prepared to train a model. For example, a model built using a series of one-dimensional features to predict a particular label or continuous value may involve first selecting the appropriate data sources, cleaning the data to remove outliers and erroneous values and impute missing values, constructing new attributes based on calculations using existing fields from the raw data, integrating different data sources into a single dataset, and reformatting the fields to make analysis easier to perform. In terms of prediction labels, these can often be linked from other data sources, if they are not already part of one of the data sources selected initially.

A computer vision model, however, can often begin with a specific dataset instead of a business or research question alone. The data may be cleaned to remove images that were captured incorrectly (e.g., excessive blur, visual noise etc.) or to edit out irrelevant parts of the images, but data imputation typically is not a feasible option nor is constructing new data without being able to replicate the circumstances of the original dataset. Other data sources may be integrated if they are visually similar to the original dataset but this may not always be practical. Data formatting may be required to convert the images into a format that is more suitable to be passed as input into a computer vision model. The primary challenge with developing computer vision models is that the data often require significant amounts of manual labelling, particularly as the labels relevant for one project may not be relevant for another.

3.1 Data Cleaning

To reduce the amount of labelling that needs to be performed per frame, the videos will be cropped to focus only on the area of the videos that can contain pulling chains of ants. The experimental set-up used in recording each video is identical and consists of a top-down shot of paper cut into the teardrop-shape of a leaf similar to those used by weaver ants to construct their nests. The “leaf” is positioned with the leaf tip on the left side of the shot, and weaver ants will usually only pull on the tip of the leaf when they form pulling chains. Therefore, the videos will be cropped to only include the part of the frame in the general vicinity of the leaf tip.

3.2 Data Construction and Integration

Due to the large number of video frames that must be labelled to train an accurate multi-animal model in SLEAP, the work will be divided among all team members with each person assigned a subset of videos to label according to mutually agreed definitions of each animal part (which may differ from the anatomical definition). After discussing with the supervisor, the initial anatomical labelling points and predicted skeleton have been agreed upon and are shown in Figure 4. The number of anatomical points and the resulting skeleton can be altered after assessing the first round of training, to determine if a simpler or more detailed model is sufficient/necessary.

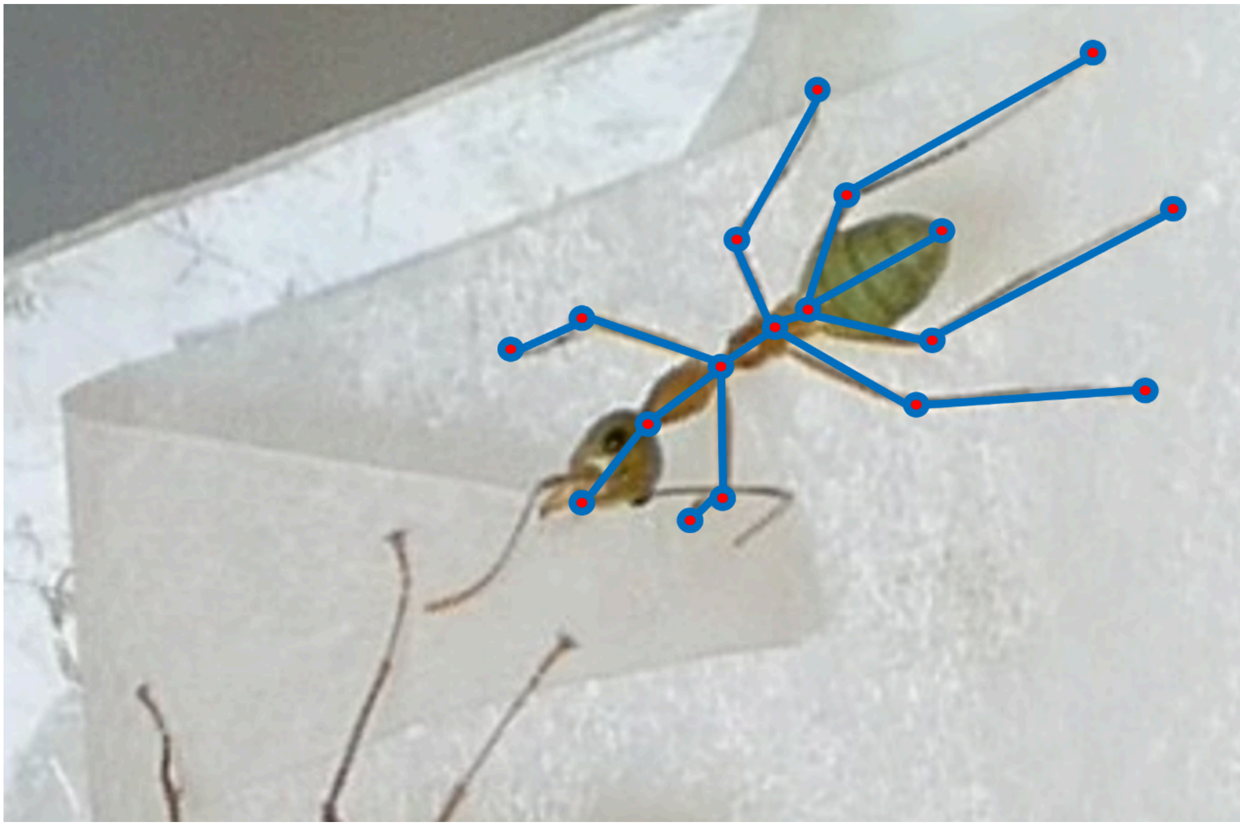

Figure 4: Predicted skeleton model of a pulling weaver ant, based on 18 initial anatomical points of interest (nodes) and their connecting edges.

Once a suitable number of video frames from each video have been labelled, the individual labelled datasets will be merged into a single integrated dataset using merge functionality built into SLEAP. This dataset will then form the base dataset from which potential models will be trained, validated and tested. It would not be suitable to include other videos of ant behaviour as the lab is specifically focused on weaver ants and the model will only be used for the specific experimental set-up used in the training videos.

The volume of training data will be further expanded by using image augmentation to produce new training data without having to acquire or label additional video frames. Image augmentation in computer vision works by applying various image transformations such as rotation, scaling, colour and brightness adjustment to labelled training data, which improves the model's ability to produce accurate inferences in a broader range of situations. In the case of the ant videos that our team will be using for training and inference, the videos were filmed exclusively from a top-down perspective, therefore applying a random rotation of ±180° to video frames during the training phase would be an appropriate as the transformed images would be similar to other images that would potentially be used as input. SLEAP provides a straight-forward way of configuring the image augmentation parameters via the included graphical user interface, so further adjustments can be made in future if deemed appropriate.

3.3 Data Formatting

As per the official SLEAP documentation, it is recommended that the video files used as input be encoded in a "reliably seekable" video format[9]. This means that the image data at a particular frame index in a video should be the same each time it is retrieved, and this is an essential step for producing consistent results. Many modern video codecs compress video data by taking advantage of redundancy in pixel data between frames and saving space by ignoring pixels in a frame that don't change significantly, instead using pixel data from the frames around it to reconstruct a good approximation of the original frame[10]. However, this can cause problems in a computer vision context as frames are often used as input in random order instead of sequentially, therefore having to reconstruct the frame by examining surrounding frames may result in different frame data being produced, depending on the video codec being used.

We will use the video editing software 'ffmpeg' to re-encode the videos into a format that is reliably seekable and hence make them more suitable for use in training a computer vision model. ffmpeg is a command-line multimedia editing application suite that can perform a broad range of operations on video and audio files including encoding and decoding, scaling, cropping and colour correction to name a few. Since the composition of each video is similar, the command-line interface will be leveraged to efficiently re-encode all of the videos using a batch script, which will be provided to the client for future use.

3.3.1 Data labelling

The data labelling stage is predicted to be the most time-consuming phase of this project as all of the videos provided by the client must be manually labelled to some degree. While it has been shown that SLEAP can produce reasonably accurate tracking results for a single animal with as few as 10-20 labelled video frames, tracking multiple animals reliably often requires hundreds of labelled frames to produce a similarly accurate result. By splitting this workload among all team members, this task should be made considerably easier.

One of the primary challenges will be ensuring that all data labels are consistent regardless of who performs the labelling, as labels must be applied to a single point on the ant and different people may have different views of exactly where that point should be placed on the ant's body part. As previously mentioned, the team's solution to this is to agree on precise definitions of the points on the ants' bodies that correspond to the model skeleton that will be used for training and inference, which will be approved by the client as being suitable for subsequent analysis before any labelling takes place.

In other cases, some ant body parts may be mistaken for other body parts on the same ant or the same body part on a different ant. This is particularly an issue for the limbs of the ants and especially in the case of ants engaged in a pulling chain, as there is often a considerable overlap in the position of ants' limbs in this situation. As the video data consists of sequential frames, correctly identifying the nature and owner of each limb will likely require examining video frames before and after the frame being labelled to provide an accurate label.

Finally, there is a potential for a labeller to identify the correct body part but apply the incorrect label due to a lapse in attention. This is also a problem that is most common with the limbs as weaver ants, being insects, have six limbs which are very similar in appearance. Aside from maintaining a heightened degree of attention to detail during the labelling process, the trained models will be used to make inferences on the training data, and instances where the manually labelled data and predictions consistently differ will be checked in case an incorrect label has been applied.

4. Modelling

4.1 Modelling technique selection

As this project relates to an image landmark localisation problem, the conventional approach towards such problems is to use a convolutional neural network, and we will be using one as the basis of our model for tracking ant postures. Convolutional neural networks (CNN) are an advancement on the traditional feed-forward neural network structure and they are particularly suited for making inferences with image inputs as they avoid many of the problems that arise when trying to convert images into a format that is suitable for use with a feed-forward neural network.

To explain why feed-forward networks are not suitable, it is necessary to first explain how they work. Feed-forward neural networks are among the simplest and earliest developed types of neural networks and consist of an input layer, an output layer, and one or more hidden layers. Within each layer is a set of nodes or ‘neurons’, with each individual neuron in one layer being connected to each of the nodes in the layers ahead of and behind it, with the exceptions of the input layer (which can only be connected to the layer ahead of it) and the output layer (which can only be connected to the layer behind it). A ‘weight’ value is associated with each of these connections, which are then adjusted as the model “learns” to produce an output consistent with the known correct output (the ‘ground truth’) for a given set of inputs.

However, the nature of image data makes it difficult to efficiently convert them into a format usable with such models. In order to use these feed-forward models with image data, it is necessary to first convert them into a one-dimensional format, usually by converting each pixel into a series of continuous values representing the pixel intensity, column-wise then row-wise or vice versa. The converted data are then used as input into the feed-forward model and training takes place as usual. There are two problems with this approach. Firstly, information relating a pixel to the pixels around it in two-dimensional space is lost when this transformation takes place. Secondly, since each pixel is considered a feature in its own right, the number of input features becomes equal to the number of pixels in the image, multiplied by the number of colour channels in the image. For a 1920 x 1080 pixel input with three colour channels (red, green and blue) this would result in an input of 6,220,800 features. This can lead to a problem known as the ‘curse of dimensionality’ where an enormous amount of training data is required to train an accurate model.

Convolutional neural networks avoid these problems by instead identifying features within an image by applying a ‘filter’ over the entire image and locating regions where different features exist, with deeper layers in the model using this information to identify the presence of objects in the image associated with the features. This minimises the number of neurons required in each convolutional layer of the network as each neuron only needs to track a particular value in the filter, which is often less than ten pixels in each dimension, as opposed to having one neuron for each pixel in the entire image. This reduces the number of neurons that need to be trained to generate an accurate model, which speeds up training time and reduces the likelihood of overfitting. This approach also retains the 2-D positional information from the image.

4.2 Model design constraints

In terms of designing a new convolutional neural network structure, it is highly improbable that we would be able to design a new deep learning architecture from scratch that is capable of capturing the complexity of the problem to be solved. Many of the well-known and well-used CNN models such as VGGNet and Inception are the result of significant in-depth research and often form the basis of an entire research project in their own right. Given that the team does not have the same level of experience, it would be more appropriate and time-efficient to leverage an existing architecture and train it to track the ant postures instead. SLEAP has several pre-constructed models built into the software, UNet, LEAP, hourglass, and ResNet, and the parameters for these models are readily configurable via the SLEAP interface. Therefore, we will be constrained to the available parameters for whichever model we decide to use, however this still provides a significant degree of customisation.

It is also necessary to consider the constraints within SLEAP itself, most significantly is that SLEAP only allows for a single skeleton to be used per model. Therefore, videos that contain multiple different animals cannot be used to train a single model that can track the postures of all of them. Since the training data that has been provided by the client only consist of videos of a single type of animal, namely weaver ants, this is not a significant issue. If we need to pivot to using other video data sourced from the client in the event that tracking the ants in the videos that we have been provided proves to be an insurmountable problem, these videos may contain other insects (e.g., videos of ants pulling collectively on crickets) which also need to be tracked in relation to the ants. In this case, it may be necessary to train multiple models that track each type of animal individually and then integrate the outputs of each model into a single result set.

4.3 Model assessments needed

The process of assessing each prospective model’s accuracy will begin prior to the training phase by splitting the labelled dataset into a training dataset, a validation dataset and a test dataset. Splitting the dataset prior to training means that it is not necessary to source and label additional video data for testing purposes once the model has been trained. The training dataset will be used for the actual training of the model, and the validation dataset will be used to evaluate the accuracy of the model after each round of training. Keeping a separate training and validation dataset in this way helps prevent the model from effectively “memorising” the training dataset, which would result in overfitting of the model and poor inferential performance on other datasets. Once a model with an acceptable accuracy rate has been trained using the training and validation datasets, the test dataset will be used for the final evaluation to determine if the model can produce similarly accurate inferences on a dataset not seen during the training process. This will help ensure that the model is able to generalise well when inferring with other related datasets.

To quantify the accuracy of the model during training and determine the appropriate weight values in each layer of the convolutional neural network, a loss function will be used to calculate a score representing the comparative difference between the predictions and the ground truth. SLEAP includes two loss functions built in by default, a standard mean squared error function which is calculated for each body part using the distance between the predicted position and the labelled position, and an online hard keypoint mining loss function, which is similar to the mean squared error function but applies a higher weight to particular body parts that consistently score poorly, indicating that they are harder to track and therefore require more training to predict accurately. SLEAP also has the capability to use a custom loss function written by the developer, so if it is decided later in the project that the default loss functions are insufficient for our purposes, a more appropriate one can be written and used instead.

The process of identifying the set of model hyperparameters that produces the most accurate model will be performed by defining a common training, validation, and testing data split and then training multiple models using this split to control for split differences between trained models. Due to the long training time required to train a single model as a result of the large input sizes being used in this project, only three key hyperparameters will be tested, the base number of filters, the maximum stride length, and the anchor part. By evaluating trained models with different combinations of values for these hyperparameters, it should be possible to determine the optimal hyperparameter set for this project.

In all machine learning projects, it is important to consider the possibility of algorithmic bias and how it may affect people who may be impacted by the outputs of these models. There is a strong potential for people who have been historically discriminated against to be disadvantaged by these models due to bias being deliberately or inadvertently introduced into the model's training, which can in turn affect the outputs. For instance, a machine learning model designed to assess potential candidates' résumés during a hiring campaign for a traditionally male-dominated industry may have been trained on the résumés of existing staff. This may result in certain assumptions being baked into the model that disadvantage female applicants. However, in the case of this project, the occurrence of algorithmic bias is less likely as humans are not, at least at this stage, measured as part of the input data or directly affected by the outputs. Only videos of weaver ants are used for this project as this is the core focus of the client's research team, however if the subject were to change to the topic of human pose estimation, algorithmic bias could occur if the training data did not consist of a diverse sample of the human population.

5. Evaluation

5.1 Requirements

The client requires that Project APRA produces a model with accurate pose estimation for weaver ants as well as a user-friendly guide in order for the software to be accessible for researchers and non-technical users. The expectation of the model is that it can analyse the posture of at least one ant involved in the formation of a pulling chain.

5.2 Process

In order to validate the model, many test cases will be carried out. The client has provided many videos of weaver ants forming pulling chains and some videos have been cropped to show less ants in order to create a more diverse range of test data. This data will be used to determine the accuracy of the model.

First, training and testing of the model will be performed in the simplest case of a cropped video with a single ant. This approach aims to minimise any complications with limbs overlapping, which are anticipated challenges. The model's performance will be evaluated using predetermined validity criteria, with a threshold set at 90%. Then, the evaluation will progress to more complex scenarios, including more ants walking over chains, multiple ants forming chains as well as multiple chains in one video.

Depending on the outcomes, various next steps will have to be considered. If certain scenarios prove excessively challenging or return validity scores below 85-90%, strategies for enhancing results will be proposed. A potential next step would involve revisiting and altering the tracking methods employed. This could include adjusting parameters or exploring alternative tracking approaches to improve accuracy and reliability, particularly in challenging scenarios like multiple ants forming chains with overlapping movement. Another strategy is to use an alternative video dataset that the client can provide that involve fewer ants. This may allow for easier tracking, however may create new problems depending on the behaviour of the ants that must be tracked.

To measure and compare the accuracy of each trained model, several different metrics will be used to develop a comprehensive idea of each model's accuracy in terms of their ability to correctly identify whether a particular body part is visible in the frame and if so, what its position in the frame is. When determining if a predicted body part position is correct, the raw Euclidean distance between the predicted point and the ground truth label can be used to calculate the average distance for all points, with more accurate models having a lower average distance between the predicted and the actual positions. Averages can also be calculated for individual body part predictions, which can help indicate which parts are more or less difficult to predict.

SLEAP also performs calculations for two additional metrics that can be used to compare model performance, the Percentage of Correct Keypoints (PCK) score, and the Object Keypoint Similarity (OKS) score. The PCK score calculates the percentage of correctly predicted points, with a point being considered correct if the Euclidean distance between the predicted positions and the actual position is below a user-defined threshold value. The OKS score uses the Euclidean distances between all predicted points and their respective ground truth label to calculate the overall similarity of the predicted posture, taking scale and part visibility into account.

For determining the model performance in terms of its ability to identify if a particular body part is visible in a frame, a confusion matrix of the true and false positives and negatives can be used since the visibility is a binary outcome. From this, measures such as the precision and recall can be calculated for the correctness of the model's part visibility predictions. Taken in combination with the metrics above, these measures will be used to assess and compare the accuracy of trained models and decide which model will be used for inference in a Production environment.

Other strategies may involve suggesting modifications to future experiments to align with technological constraints or even using an alternative to SLEAP all together. Additional discussions will also delve into the underlying limitations and constraints, offering valuable insights for the client's future utilisation of SLEAP in future projects.

6. Deployment

Once a model has been developed that the client is satisfied with, the selected model will be handed over to the client in the native SLEAP format. A Jupyter notebook will also be provided that will export the position tracking data in a format suitable for subsequent analysis and illustrate how this process works so that it can be replicated by the client for other research projects. Also, all code and other scripts will be uploaded to GitHub and access will be shared with the client's team.

Additionally, we will provide a comprehensive guide on using SLEAP, covering its installation, training process, and data analysis. Our own trained model and the generated data will be used as a benchmark for reference.

The document will begin with a detailed training guide for installing SLEAP, catering to users of the Windows operating system as this is the operating system used in the client's lab. We will provide a step-by-step walkthrough, addressing common installation challenges encountered during our own experiences. Our goal is to present solutions in a user-friendly manner, suitable for individuals who may not be familiar with coding, ensuring a smooth introduction to the software.

After the installation guide, we will explore training the model, providing insights and solutions to potential challenges encountered during the process. We will explain how to interpret results and highlight the methodology for constructing alternative models similar to the one we provide as a foundational example.

7. References

1. Hotz, N. (2024) What is CRISP DM?, *Data Science Process Alliance*. Available at: https://www.datascience-pm.com/crisp-dm-2/ (Accessed: 28 March 2024).

2. Pereira, T, et al., 2022, SLEAP: A deep learning system for multi animal pose tracking, *New York: Nature Publishing Group*, Available at: https://www.nature.com/articles/s41592-022-01426-1.pdf (Accessed: 26 March 2024).

3. S Ganga Prasath, Souvik Mandal, Fabio Giardina, Jordan Kennedy, Venkatesh N Murthy, L Mahadevan (2022) Dynamics of cooperative excavation in ant and robot collectives, *eLife* 11:e79638. (Accessed: 28 March 2024).

4. Wolf, S. W., Ruttenberg, D. M., Knapp, D. Y., Webb, A. E., Traniello, I. M., McKenzie-Smith, G. C., Leheny, S. A., Shaevitz, J. W., & Kocher, S. D. (2023). NAPS: Integrating pose estimation and tag-based tracking. *Methods in Ecology and Evolution*, 14, 2541–2548. https://doi.org/10.1111/2041-210X.14201. (Accessed: 28 March 2024).

5. Social Leap Estimates Animal poses (SLEAP), *Social LEAP Estimates Animal Poses (SLEAP) - SLEAP (v1.3.3)*. Available at: https://sleap.ai/ (Accessed: 28 March 2024).

6. Google docs: Online document editor | google workspace, *Google*. Available at: https://www.google.com.au/docs/about/ (Accessed: 28 March 2024).

7. Build software better, together, *GitHub*. Available at: https://github.com/about (Accessed: 28 March 2024).

8. Trello brings all your tasks, teammates, and tools together, *Trello*. Available at: https://trello.com/home (Accessed: 28 March 2024).

9. Help - SLEAP (v.1.3.3), *Social LEAP Estimates Animal Poses (SLEAP) - SLEAP (v1.3.3)* Available at: https://sleap.ai/develop/help.html#does-my-data-need-to-be-in-a-particular-format (Accessed: 28 March 2024).

10. loopbio (2018) Video I/O Part 1: An Introduction to Video Compression, *loopbio blog* Available at: http://blog.loopbio.com/video-io-1-introduction.html (Accessed: 28 March 2024).

8. Client Feedback

Meeting date and time: 10-10:30am 26th March 2024
Feedback received:

- References should be moved to the end of the document and document should be populated with additional references throughout the main text.
    - References to SLEAP being used in other research projects should be added.
- Evaluation section needs to be longer and more detailed.
- Captions should be added to the diagrams.
- The definitions section should be removed.
- The handover of the video cropping script should be mentioned in the Deployment section.
- The current skeleton to be used should be explained in more detail (which body and leg segments are going to be tracked).
- References to ‘sponsor’ should be replaced by the term ‘client’.

Team response/action points:

- Additional relevant references have been added to the document.
- Evaluation section has been expanded with further details about how the models will be evaluated.
- Captions explaining the content of figures have been added.
- Definitions section has been removed and abbreviation definitions have been moved to their first mention.
- Handover process for the video cropping script and other development resources has been added to the Deployment section.
- A diagram illustrating the skeleton has been added to the Data Construction and Integration section.
- ‘The term ‘sponsor’ has been replaced with ‘client’.