Project No. ON-00554
Contract No. 4500011380
AWI Project Manager: Carolina Diaz
Contractor Name: neXtgen Agri
Prepared by: Herman Raadsma, Ian Harris, Dacheng Tao, Mehar Khatkar, Junbin Gao, Sarah Thompson, Will Gibson and Mark Ferguson
Publication Date: 1 June 2019

Artificial Intelligence in Wool Production

Contents

Published by Australian Wool Innovation Limited, Level 6, 68 Harrington Street, THE ROCKS, NSW, 2000

Executive Summary

Extraction of information from large-scale digitised data sets through artificial intelligence (AI) is unprecedented both in scale and the rate of change. Novel sources of data capture include digital imaging, GPS location and movement, high resolution biomarkers and bio-sensors, automatic capture of market and environmental data in real-time. The Australian wool industry is ideally placed to evaluate the impact of such novel phenotypes on profitability and advanced farming systems. This project provides a pilot evaluation on the utility of AI, and in particular Deep Learning, in accurately predicting performance outcomes from images, biomarkers and on-animal sensor output.

We developed a semi-automated system that has the capacity to take high resolution images under field/yard conditions and link them to animal electronic identification (EID). The system also allowed the semi-automatic recording of body weight. Using this system, we created an image library of 1,482,041 images from 4072 sheep using 4 camera angles namely front side, top and rear. All sheep were weighed at the time of image capture, and subjectively scored for face cover (1-5), neck wrinkle (1-5), and body wrinkle (1-5) and identified to EID.

Using sub-sets of the images, we applied the digital information for Deep Learning analytical pipelines in particular use of Convolutional Neural Network (CNN) analysis. The models were developed using Keras (https://keras.rstudio.com) and Tensorflow (https://www.tensorflow.org). The data were sub-divided into a training set, an evaluation set and an independent test set to predict how well AI could predict the corresponding phenotypes. Using both side and top camera the predictive algorithms could predict bodyweight with an accuracy of 86% and 87% respectively and with no bias. Combined information from top and side camera resulted in an accuracy of 89%.

For facial recognition AI was trained to detect head shape and body shape for each sheep with an accuracy of 99% provided the sheep were from the same training and test set. Using random subsets of face and body images per sheep, the AI algorithm could match anonymous face and body images with 94% and 98% accuracy to sheep EID, and 99.7% when both face and body information was used. However, when images from the same sheep were tested 5 months later, accuracy was considerably lower (<10%) unless images from both time points were included in the training data set (accuracy increased to 90-98%). This indicated that very large data sets from the same sheep, repeated over time are required in the initial training for facial recognition to detect unique biometrical features for each sheep. Once such initial training data sets are established facial recognition could be applied in novel populations.

For neck and body wrinkle the AI pipelines were able to allocate animals to either a high or low wrinkle class with 73%-90% accuracy pending which camera angle and wrinkle trait was predicted. Using the full scale of wrinkle score (1-5) prediction accuracy was lower at 38%-58%. The AI prediction matched the accuracy of manual scores which was 98%-99% for high and low wrinkle score and 57%-60% for wrinkle score on the expanded 1-5 scale.

For face cover score, initial classifiers delineating between scores 2 and 3 revealed results little better than random. This was largely a function of the distribution of face cover data in the population, where 87% of animals were assigned to a central class and less than 1% of animals were found in the extreme classes. This provided no power for training and validation of the AI algorithms. To test the utility of AI for delineating face cover score, ML classifiers were trained to differentiate between face-cover scores 2 and 4. When multiple areas were cropped from the images, the predictive capacity of the classifier was proven with an accuracy of 87%. With a more balanced data set, where each face cover score is equally represented, it is likely that differentiating between all 5 face cover scores is possible.

A review was conducted of the scope of bio-sensor and bio-marker technologies and their likely utility for the sheep industry to define phenotypes when linked with deep learning AI technologies. Outcomes from global investments in this area are potentially transferrable to the sheep industry and will accelerate the amount of digital data coming on stream with most amenable to AI and Deep Learning pipelines. Within the bio-sensor field on-animal accelerometer and geolocation devices offer the most promise. Within the bio-marker arena, genomics was thought to offer the greatest potential immediate benefits since samples could be collected at an early age and are not affected by physiological state and offer both phenotypic and genetic predictive value for almost all traits from a single sample. Both large scale proteomic (including immunological) and metabolomic investigations offer future promise since they are closely linked to physiological (production/disease) state and amenable to large scale analyses by AI and potentially offer low cost phenotyping for complex traits especially when coupled with on-animal bio-sensors.

A strategy for using data from diverse sources for prediction of on-farm outcomes is presented exemplified by concepts across the production pipeline highlighted from simple trait evaluation/prediction to complex model decision making applications. A strategy plan for ongoing R&D investment in applications of AI technologies for on farm applications is presented with priority areas deemed to have greatest immediate impact and success for on-farm applications detailed. It is concluded that in the emerging area of digital agriculture and precision farming technologies AI technologies will unlock new prospects for the Australian wool industry.

Introduction/hypothesis

Extraction of information from large scale digitised data sets through artificial intelligence (AI) are unprecedented both in scale and the rate of change. Novel sources of data capture includes digital imaging, GPS location and movement, high resolution bio-markers and bio-sensors. In parallel, developments in AI has given rise to ultra-sophisticated platforms to analyse such data sets and make predictive outcomes. In particular Deep Learning, a new Machine Learning technique with state-of-the-art Neural Network architecture, gives rise to efficient exploration of ultra large data sets and predict outcomes not readily seen by conventional human-driven analyses. In combination, this technology is projected to be the most disruptive technological advancement of the current century. This opens many untapped possibilities to be exploited for novel applications and industries.

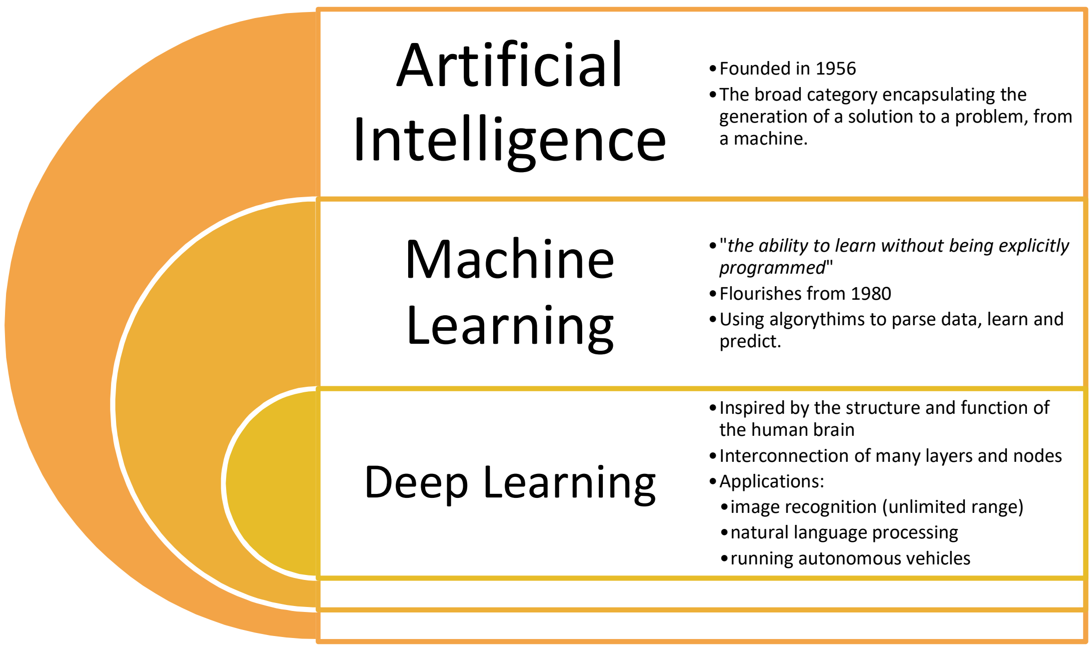
The Australian wool industry is ideally placed to evaluate the impact of such novel phenotypes on profitability and advanced farming systems. Lifetime productivity is a key determinant in sustained profit for the Australian Merino industry. Making efficient low-cost measurements early in an animals' life to predict lifetime productivity or track animal performance in real time, gives breeders and producers a new set of tools in making decisions to capture value for sustained profit. Furthermore, such advanced phenotypes allow better informed selection decisions on which parents to select to breed from. In other words, it gives sheep breeders the best tools to retain the best animals for breeding and production, whilst culling non-profitable sheep at a very young age.

This project piloted the utility of AI and in particular Deep Learning in accurately predicting performance outcomes from images. The project also undertook a global review of new bio-markers and bio-sensors that may have applicability within the sheep industry.

1. Literature Review

Terminology

The data science area is one that is moving quickly. As technologies develop within the artificial intelligence space, new acronyms or descriptors are coined, so in reading this report it's important to have some appreciation of how they relate to each other. The diagram below illustrates some important terms.



Artificial Intelligence (AI)

Artificial Intelligence is intelligence demonstrated by machines. The concept of developing machines to learn and act has been around for centuries. However, the most recent developments have escalated global efforts largely due to unprecedented computing capacity to process and analyse data, enormous volumes of digital data collected at high speed, the need to resolve structures within such data set, availability of open source platforms to apply highly complex statistical and data analytical procedures and algorithms. Applications of outputs from AI are now found in every day human activities and processes. Although most AI applications are used in a predictive capacity with human control on inputs and outputs, there have also been many attempts to define (autonomous) artificial intelligence[1] beyond control of human inputs.

Machine Learning

Artificial Intelligence is encountered today, mostly through a technology field known as Machine Learning. Machine Learning gives computers the ability to learn (or improve their performance) without being explicitly programmed. There are three categories of Machine Learning:

- Unsupervised Learning – The ability to identify groups of observations e.g. cluster 1 vs. cluster 2.
- Supervised Learning – The ability to 'classify' observations e.g. cat's vs dogs.
- Reinforcement Learning – Choosing between actions to maximise a reward.

Deep Learning

Deep Learning is a specific field of machine learning where the emphasis is on learning through successive layers or representations of data. Increasing the number of layers of learning (and complexity) can provide improved accuracy. Deep Learning can be implemented in a very automated way and it can simplify workflow. Until recently

[1] Legg and Hutter 2007

it has been perceived to be very Black Box technology, but over time Convoluted Neural Networks have been dissected and understood.

An example of the application of Deep Learning and its improvement over time from utilising increasing numbers of layers is the Imagenet Large Scale Visual Recognition Challenge (ILSVRC; Figure 1). In 2016, the ILSVRC demonstrated the use of 152 layers of learning to manage the classification of images, surpassing the average capability of humans.

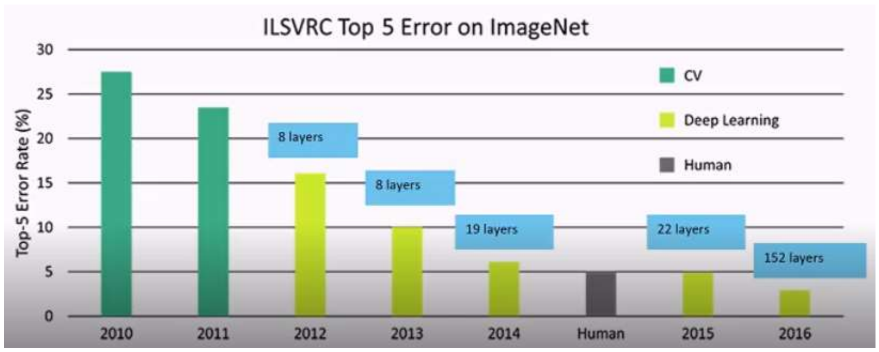


Figure 1.1 https://www.dsiac.org/resources/journals/dsiac/winter-2017-volume-4-number-1/real-time-situ-intelligent-video-analytics

Limitations, Training and Validation

The principal of GIGO (Good information in, Good information out) remains at the core for the AI technology space. Deep Learning appreciates and is well suited to dealing with a lot of data. This is especially fortunate considering the raft of data collection techniques and devices planned for Agri-tech, and the wall of data that will be created.

Knowledge of the data, the situation in which it is collected, how a Deep Learning environment might interpret it and what you expect to see in the results are all requirements. Deep Learning is not a silver bullet. There are numerous examples available demonstrating how Deep Learning has falsely interpreted data. Deep Learning systems can have very high predictive power but can also have low interpretability.

To be effective and accurate, Deep Learning requires a training and validation step where a training dataset is used to develop a Deep Learning model and a separate validation data set is used to validate the model. Further independent testing is required using an independent test set which was not part of the training or validation set. This segregation of data is an important consideration when collecting a Deep Learning dataset. Messy, confounding, or noisy data can be tolerated if modelling, training and validation has been appropriately applied.

Deep Learning technologies are known to overestimate or over-fit results, but techniques and methods have been developed to rectify and reduce this. More data, a simpler model, data augmentation and node-dropout methods can help. The number of experts across this area of artificial intelligence data science is also growing. Applying the technology is relatively straight forward. The modelling, training and validation steps however can take time.

Advantages of Deep Learning

Briefly, some practical advantages of Deep Learning techniques for agriculture are:

- Improved time to solution
- Accuracy
- Low cost in application
- Huge potential for contributing to the understanding of many new types of data for complicated agricultural issues e.g.
    - Pest and weed identification

- Herbage biomass prediction
  - Genomics and phenomics
  - Body weight or condition score prediction
  - Health traits
  - Body composition analysis
- Using images, video, text, audio and many other data types within a Deep Learning application where there is spatial or temporal structure
- Reusability of Deep Learning infrastructure across a variety of tasks or analyses
- Integration with diverse data sources including those collected through conventional sources, supply chains or collected for alternative uses.

New recording data

Performance recording of livestock for the purposes of selecting and developing animals that are genetically more efficient and more productive than previously, is a path of continual improvement. Whether using technologies and techniques that have developed in other industries, or creating new ideas and applications from within, livestock species have changed constantly. The application of Deep Learning in agriculture will be no different. It is a field that has been demonstrated to allow industries and sectors to develop and grow because of the way in which new forms of data can be collected, analysed and then used to inform. Biological based industries such as sheep and lamb production are excited about the application of Deep Learning as new analysis techniques emerge along with new data collection methods.

Traditional types of recording information such as weights, measures and scores are very common in many species, but they are relatively analogue and manual in nature. The proposition of capturing more digital information, or capturing many more of the traditional weights, measures and scores in a digital and automated format, has captured the imagination of farmers and scientists alike. The end of collecting the weights, measures and scores has not yet been reached, but we do need more cost effective, accurate, and unbiased ways of capturing them.

Devices and their integration with the internet have the potential to flood measurement systems and genetic evaluations with multitudes of new volumes and types of data. On the face of it, this would be a daunting challenge without the prospect of new analysis techniques such as Deep Learning. Cameras recording specific events of animals in their natural state; videos capturing footage as humans force behaviour such as movement; audio recordings able to detect heightened levels of stress as animals are preyed upon or separated from each other; sensors either on the body or in the body monitoring spatial, temporal, hormonal or cyclical change, are common examples of new forms of data heading towards animal scientists to discover new aspects of behaviour and performance that previously were simply too difficult to appreciate.

In addition, this new data will not require the presence of a human technician to scribe the information. Instead it has the potential to be uploaded to cloud-based platforms constantly or when in the vicinity of nodes and receivers strategically placed to minimise the interference with natural activity and behaviour. The effort globally to create devices, pipelines and integrated services for Agritech to take advantage of the Internet of Things (IoT) and the raft of data to come from it, is significant and new inventions and systems are appearing almost daily.

There is a growing amount and complexity of data being generated by fully automated, high-throughput data recording or phenotyping platforms, including digital images, sensor and sound data, unmanned systems, and information obtained from real-time non-invasive computer vision.

Deep Learning, Genetics and on-farm performance monitoring

Deep Learning networks have the capacity to not only receive and analyse a lot of data, but the data need not be as structured or organised as thoroughly as previously collected. Deep Learning systems have a tolerance for missing data, and they can learn about what to do in such a case. As described above, training and validation Deep Learning models that subscribe to unsupervised methods can automatically circumvent such cases, allow for them and continue to improve.

Depending on the situation, this may be beneficial. Perhaps one of the more frustrating aspect to recording animals is the haphazard nature of their activity and the potential to miss events and occurrences. The real world constantly challenges the animal scientist with situations and applications with data capturing scenarios that are difficult and having a technology that can accept less than perfect datasets would be very welcome.

In years to come as devices, sensors, connectivity and analysis systems develop and mature, big 'agricultural' data will have a welcome home for uncovering new biological knowledge, discoveries and predictions. The recognition of patterns and regularities in the world around us lies at the heart of scientific and technological progress. It's how we advance and how we innovate. It's also an area where deep learning excels.
The potential for transforming the way we collect traditional phenotypes and then analyse them, is only limited by the imaginations of future animal scientists. A challenge will be keeping abreast of inventions and technologies coming from other industries and sectors and working out how to apply them in sheep production systems.

2. Project Objectives

The overall aim of this project is to evaluate the use of advanced phenotypes and artificial intelligence technologies for the prediction of lifetime performance at young ages, management of performance changes in real time, and to provide advanced, highly predictive phenotypes as inputs for ongoing selection decisions by commercial and seed stock sectors. Using low-cost but highly predictive technologies, advanced phenotypes allow sheep breeders to make commercial decisions to cull unprofitable sheep at an early age, provide interactive and instant management decisions to prevent unfavourable production and health outcomes, and make highly informed breeding decisions to select those animals best suited to generate future lamb drops.

The immediate objective of the project is to provide a proof of concept that novel phenotyping technologies based on image analysis, bio-marker and bio-sensor technologies combined with deep learning AI technologies will unlock new prospects for the Australian wool industry.

The specific deliverables of this project, which are to be completed to the satisfaction of AWI Key Personnel, are as follows:

1. A semi-automated system that has the capacity to take high resolution images and link them to animal EID as suitable for deep learning pipelines.
2. An image library of sheep linked to their measured performance.
3. Evaluation of the ability for deep learning to extract meaningful information from digital images. This will be completed on the following; identity of individual animals through facial recognition, neck wrinkle, body wrinkle, face cover and bodyweight.
4. A review of the scope of bio-sensor and bio-marker technologies and their likely utility for the sheep industry to define phenotypes when linked with deep learning AI technologies, and completion of a report documenting the same.
5. A strategy for analytical approaches to integrate data from all sources – on farm production and management data combined with predicted outputs from image capture, bio-sensor and bio-marker data into an integrated phenotype prediction to track long-term outcomes as inputs for ongoing selection as well as phenotype changes in real time for adaptive management.
6. A strategy plan for ongoing R&D investment in applications of AI technologies for on farm applications.

Success in Achieving Objectives

This project has met all its stated objectives through the deliverables identified below. Detailed results and outcomes are contained in the report further on. A brief caption of the deliverables is included in this section.

- A semi-automated system that has the capacity to take high resolution images and link them to animal EID as suitable for deep learning pipelines: *a semi-automatic image capture system was developed for sheep and used in yard/field conditions and linked to automatic EID recording and performance (body weight) recording. The interim milestone report contains a full description of the system.*
- An image library of sheep linked to their measured performance: *an image library of 4072 sheep sampled from 8 flocks was constructed using 4 cameras simultaneously capturing on average 400 images per sheep. The library contains over 1,482,041 images.*
- Evaluation of the ability for deep learning to extract meaningful information from digital images. This will be completed on the following; identity of individual animals through facial recognition, neck wrinkle, body wrinkle, face cover and bodyweight. *This was achieved using a convoluted neural network analysis of all images from 3 cameras (front, side, top) on 4025 sheep scored for neck wrinkle, body wrinkle and face cover, electronically identified and weighed at time of image capture. The success and accuracy against each trait is presented in detail below.*
- A review of the scope of bio-sensor and bio-marker technologies and their likely utility for the sheep industry to define phenotypes when linked with deep learning AI technologies, and completion of a report documenting the same. *Report containing review and assessment of potential utility of most commonly used bio-markers and bio-sensors completed for use in sheep.*
- A strategy for analytical approaches to integrate data from all sources – on farm production and management data combined with predicted outputs from image capture, bio-sensor and bio-marker data into an integrated phenotype prediction to track long-term outcomes as inputs for ongoing selection as well as phenotype changes in real time for adaptive management. *A strategy for using data from diverse sources for prediction of on-farm outcomes is presented with concepts highlighted from simple trait evaluation/prediction to complex model decision making applications.*
- A strategy plan for ongoing R&D investment in applications of AI technologies for on farm applications. *A key strategy document for R&D investment and priority areas deemed to have greatest immediate impact and success for on-farm applications has been detailed.*

3. Methodology

A semi-automated system that has the capacity to take high resolution images and link them to animal EID as suitable for deep learning pipelines

We had initially planned to design and build a crate specifically for this project. However, we found a commercially available crate, the Breed Elite manual crate, that could be modified for our requirements. The crate had four cameras attached and necessary camera attachments added as required. The four cameras are placed to take an image from each of the front, back, top and side of the animal. We also added four purpose-built lights to the crate. The lights are built from PVC tube, lined with aluminium foil, fitted with a diffuser at the front and then a normal bulb holder at the back. LED bulbs are added. Through the initial testing stages some changes to the crate had to be made to ensure camera positions were in the correct place and images were suitable. This included altering the front door set up and latching system. The crate is now efficient at handling the sheep safely and capturing the required images. The cameras used on the crate were Logitech C920 HD Pro webcams. These cameras were sealed from moisture and dust using silicon tape and attached. A micro-computer with sufficient USB ports was purchased to handle the cameras and image capture.

Figure 3.1 Sheep crate in use.

Custom software was developed to collect images from the webcams and store the jpegs on the computer's hard drive along with the sheep's EID and weight. The software was a desktop application with a graphical user interface (GUI) developed in Java. The Java OpenCV library was used to interface with the webcams.
Bluetooth allowed the EID to be directly inserted into a field in the GUI. The weight was entered manually. Image files were stored with the farm name, sheep's EID and weight, camera ID and timestamp all embedded in the file name so that a separate data source was not required i.e. all the required information was either in the image file or in its name.

Each photo is labelled with enough information to ensure it is always unique.
A sample image name is:
"F_Wallaloo -E_982 123707956915-DT_2018-10-31-11-48-10-N_260-C_3-W_41.6.jpg"
A definition of these components is
Farm (F): Wallaloo
EID (E): 982 123707956915
Date and time (DT): 2018-10-31-11-48-10 (Format YYYY-MM-dd-HH-mm-ss)
Sequential number of photos for that particular sheep (N): 260
Camera number (C): 3
Weight (W; kg): 41.6

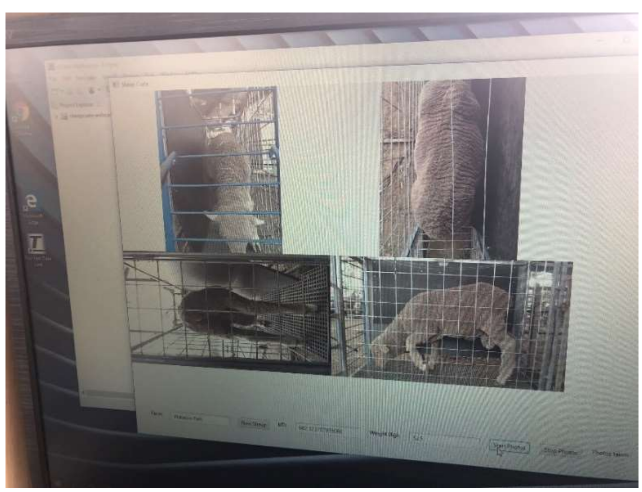


Figure 3.2 Sheep Crate app running.

The software displays an image from each camera, capturing and storing approximately 3 images per second from each of the 4 cameras, resulting in a total of 12 images per second from all the cameras. The images were stored in jpeg format with a resolution of 1,920 by 1,080 pixels.
The software is designed with a stop and start button for taking photos. Once the sheep are in the crate the start button is pressed. The software has a counter and once that counter reaches 100 photos (from each camera) the operator presses the stop photos button and the sheep are changed over.

All images are saved to the hard drive of the computer. At the end of a day's image capture, all images are downloaded to an external hard drive and replicated. They are then deleted from the computers hard drive to make enough room for the next day's images. The system can collect images of approximately 500 sheep per day.

Description of the data and QC

During two rounds of phenotypic recording, a total of 1,482,041 images were collected from 4072 sheep across eight mobs. These images were taken from four different cameras i.e. one camera each taking images from front, top, side and back of sheep as described under the image capture system above. The images are structured in subfolders and made available on solid state HDD, representing over 500GB of data.

Developing system of processing the images

The nature of AI is based on analysis and processing of Big Data and as such requires efficient analytical pipelines. The data collected so far on this project also fits in this category. A total of 500Gb of data is represented by these images and were uploaded to the University of Sydney HPC (High Performance Computing) Linux server.

Scripts were written to allow different attributes from the meta-data imbedded in the file names to be extracted automatically viz. farm ID, animal EID, date and time of recording, image index, camera (0, 1, 2 and 3) body weight and folder name of the file. This allowed a fully indexed header file to be used for further data processing and automatic image extraction and data sub-setting. In particular, the QC processing is a preliminary requirement for AI training data. Descriptive summary statistics written for each data set were used to check for duplications, mean and variance distribution and presence of outlier data. The data were organised into twelve folders as shown in Table 4.1. These were grouped further into eight groups based on the mob of origin as shown in Table 4.2. All matched phenotypes related to the images were similarly processed.

Development of Deep Learning models for prediction of phenotypes from sheep images

The images were randomly divided into three datasets viz., a training set, a validation set and a test set. The training set was used for developing prediction models using convolutional neural networks (CNNs), whereas validation and test sets were used for testing the performance of the models. The models were developed using Keras (https://keras.rstudio.com) and Tensorflow (https://www.tensorflow.org). TensorFlow is an open source software library for numerical computation originally developed by the Google Brain Team. The images were resized for CNN analysis. Various models with different numbers of hidden layers and nodes were trained and tested.

All sheep were visually scored on a scale of 1-5 for neck and body wrinkle based on the standard scoring system with increasing scores depicting increased wrinkle development (Figure 3.1). Similarly face-cover was visually assessed using a score from 1-5 with increasing score depicting increased cover or conversely decreased openness (Figure 3.2).

Figure 3.1 The schematic scale for neck, body and breech wrinkle.

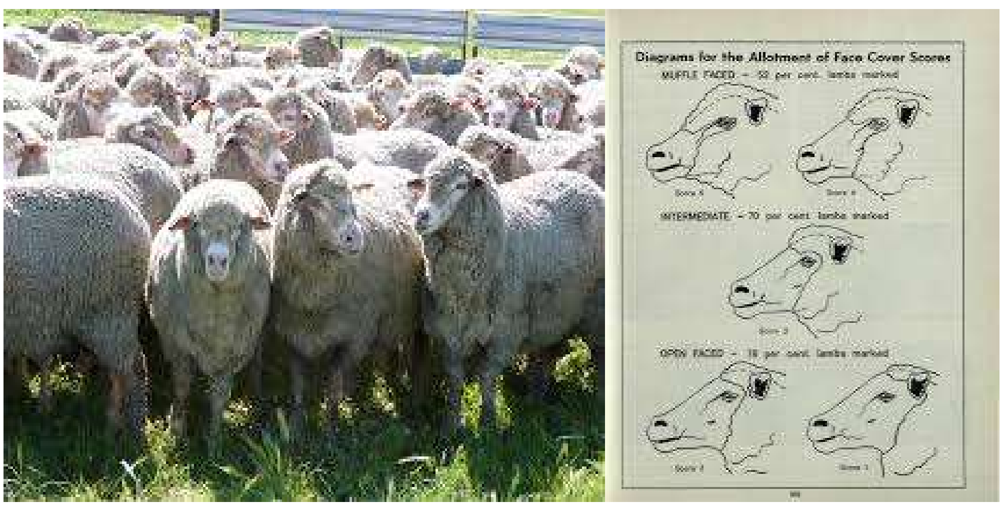


Figure 3.2 Sheep depicting variation in wool face cover and the standard scoring system for face cover.

Method for Facial recognition and identification

As illustrated in Figure 3.3, the proposed system includes five modules, namely, (1) ***CNN-based Detector***, (2) ***Data Clean & Pre-processing,*** (3) ***CNN Feature Extractor***, (4) ***Feature Fusion***, and (5) ***Classifier.*** The CNN-based detector is designed to detect sheep body and sheep face images for each frame of the input video. Based on the detection results, image frames with low detection scores are treated as noise data and removed from the dataset. The remaining the images are pre-processed into two fixed size images, which are facial crop and body crop. Both crops are then fed to the CNN feature extractor to extract facial features and body features simultaneously. These two feature vectors are then fused and sent to the classifier to identify the sheep ID of the corresponding frame. Each component is described in more detail in the following subsections.

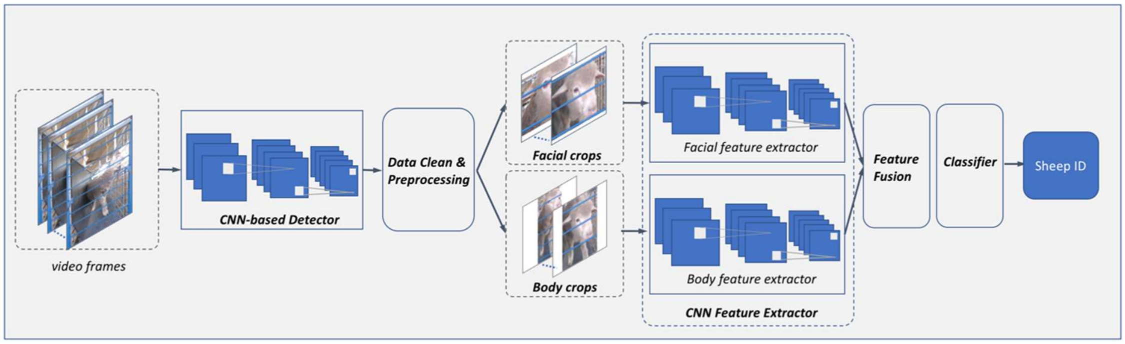


Figure 3.3 Simple pipeline for sheep identification

1. CNN-Based Detector

YOLOv3 is a single-stage and lightweight object detection model. In order to make predictions in real time, we employ YOLOv3 model to detect sheep body and face from the images. Although the off-the-shelf YOLOv3 has a detector for sheep, it does not perform well on the real-world images taken from the farm. Also, the original YOLOv3 detector cannot detect sheep faces. Thus, we annotate a small dataset with bounding box for both sheep body and face, and then fine-tune the YOLOv3 model on our own dataset to achieve our aim of detecting sheep body and faces from images.

2. Data Cleaning & Pre-processing

Since the face of a sheep is one of the main biometrics for sheep identification, we assume that sheep faces can be seen in the input images to the recognition system. Thus, after running the detector on input images, we remove all image frames with low confidential score for face detection and crop the sheep body and the face from the rest of the images according to detection results. We then resize the image crops to make the input size compatible with the neural network.

3. Feature Extractor

To extract useful information from the input images, we use a two-stream CNN feature extractor. One stream takes the face crop as input to extract features only related to the facial region, the other stream is fed with body crop to find the holistic features related to body size and shape. Also note, that each branch is a separate Resnet50-based network initialized with pre-trained weights.

4. Feature Fusion

In this component, the feature sets originating from the facial and body branch are consolidated into a single feature set, which is expected to be more robust than the individual one. Here, we use the standard feature fusion method, feature concatenation.

5. Classifier

In our experiments, we tried two types of classifiers for sheep identification, Softmax and Cosine. Suppose we have an input image x_i, y_i is the corresponding label, $f(x_i)$ is the feature vector from the feature fusion component, *C* is the number of sheep, *W* and *b* are the trainable weights and bias in the network.

Softmax Classifier. A softmax classifier is learned by minimizing the standard softmax loss function:

$$\mathcal{L}_{softmax} = -\frac{1}{N}\sum_{1}^{N} log \frac{e^{W_{y_i}^T f(x_i)+b_{y_i}}}{\sum_{j=1}^{C} e^{W_j^T f(x_i)+b_j}}$$

Cosine Classifier. Slightly different to softmax classifier, a cosine classifier can be taught by adding constraints that the feature and weight are both unit vector.

$$\mathcal{L}_{cosine} = -\frac{1}{N}\sum_{1}^{N} log \frac{e^{W_{y_i}^T f(x_i)}}{\sum_{j=1}^{C} e^{W_j^T f(x_i)}}$$

$$\textit{subject to } \|f(x_i)\| = 1, \|w_j\| = 1$$

4. Results

Summary description of sheep images & data

A total of 1,482,041 images have been stored online and on separate HDD.
Table 4.1 shows the source of images and the number of sheep from which the data was captured.

Table 4.1. Folder-wise counts of sheep and images.

Folder	N Sheep	N Images	Mob ID
images_batch1/Wallaloo_Park/Day_1	234	97131	1_Wallaloo_Small_Rams
images_batch1/Wallaloo_Park/Day_2/last_of_Day1_Mob	68	29054	1_Wallaloo_Small_Rams
images_batch1/Wallaloo_Park/Day_2/Main_Rams	201	85143	2_Wallaloo_Main_Rams
images_batch1/Wallaloo_Park/Day_2/Small_Ewe_Lambs	170	72385	3_Wallaloo_Small_Ewes
images_batch1/Wallaloo_Park/Day_3b/Main_Rams	448	194612	2_Wallaloo_Main_Rams
images_batch1/Wallaloo_Park/Day_4	496	210867	4_Wallaloo_Main_Ewes
images_batch1/Wallaloo_Park/Day_5	382	164833	4_Wallaloo_Main_Ewes
images_batch2/Photo_Crate/Curlew/Day_1_Wethers	558	161773	5_Curlew_wethers
images_batch2/Photo_Crate/Curlew/Day_2_Ewes	498	150337	6_Curlew_ewes
images_batch2/Photo_Crate/Kurra_Wirra/Day_1	496	157474	7_Kurra_Wirra_ewes
images_batch2/Photo_Crate/Kurra_Wirra/Day_2	428	113266	7_Kurra_Wirra_ewes
images_batch2/Photo_Crate/Yama	95	45166	8_Yama_lambs

Table 4.2 shows the range and mean bodyweights for each mob at time of image collection and were used as inputs for Deep Learning of body weight prediction.

Table 4.2. Mob-wise body weight of animals.

	Mob ID	N	Mean	Min	Max	SD
1	1_Wallaloo_Small_Rams	297	46.4	22.6	85.5	7.77
2	2_Wallaloo_Main_Rams	647	52.4	27.6	71.0	6.41
3	3_Wallaloo_Small_Ewes	169	35.1	22.8	44.2	4.36
4	4_Wallaloo_Main_Ewes	858	43.6	25.4	61.0	5.24
5	5_Curlew_wethers	552	42.3	25.0	60.0	5.44
6	6_Curlew_ewes	485	45.0	28.0	62.0	5.39
7	7_Kurra_Wirra_ewes	906	40.0	23.6	58.8	4.71
8	8_Yama_lambs	92	40.1	28.0	53.0	5.22

Table 4.3 shows the distribution of subjectively scored traits in all sheep for which image data was available. It includes the classes represented and the number of sheep counted across the full data set.

Table 4.3. Summary of face-cover, neck and body wrinkle.

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	face_cover [integer]	Mean (sd) : 3 (0.4) min < med < max: 1 < 3 < 6 IQR (CV) : 0 (0.1)	1: 5 (0.1%) 2: 301 (7.4%) 3: 3516 (86.7%) 4: 210 (5.2%) 5: 24 (0.6%) 6: 1 (0.0%)		4057 (99.63%)	15 (0.37%)
2	neck_wrinkle [integer]	Mean (sd) : 2.7 (0.9) min < med < max: 1 < 3 < 33 IQR (CV) : 1 (0.3)	1: 150 (3.7%) 2: 1586 (39.1%) 3: 1646 (40.6%) 4: 619 (15.3%) 5: 55 (1.4%) 33: 1 (0.0%)		4057 (99.63%)	15 (0.37%)
3	body_wrinkle [integer]	Mean (sd) : 2.9 (1) min < med < max: 1 < 3 < 5 IQR (CV) : 2 (0.3)	1: 336 (8.3%) 2: 922 (22.7%) 3: 1638 (40.4%) 4: 1013 (25.0%) 5: 145 (3.6%)		4054 (99.56%)	18 (0.44%)

Figure 4.1 shows the distribution of bodyweights within each mob as bell distributions with mean, one SD landmarks and the full range of observed weights for each mob.

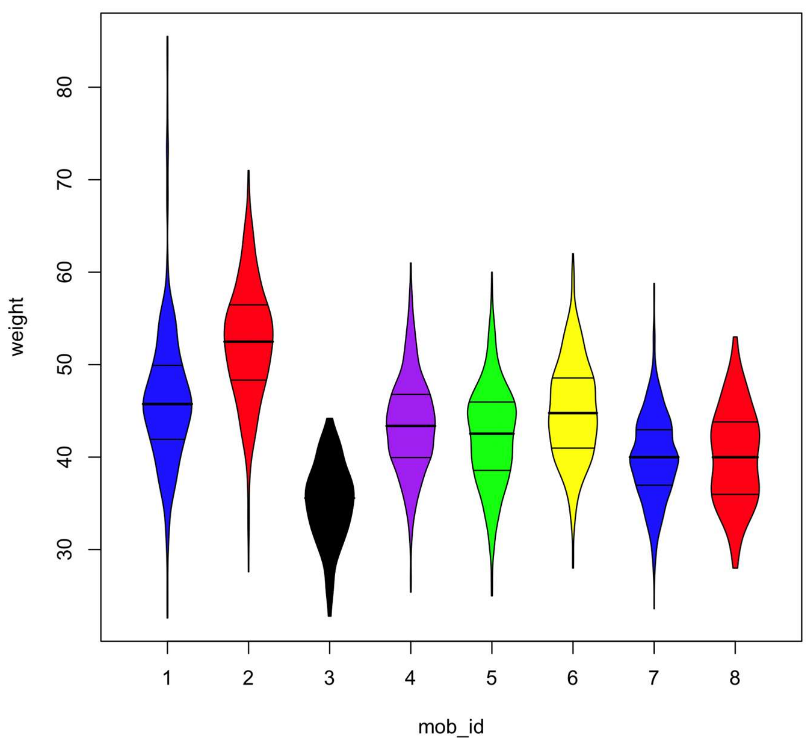


Figure 4.1 The distribution of body weight among the eight mobs used for the deep learning data set to predict body weight from visual images.

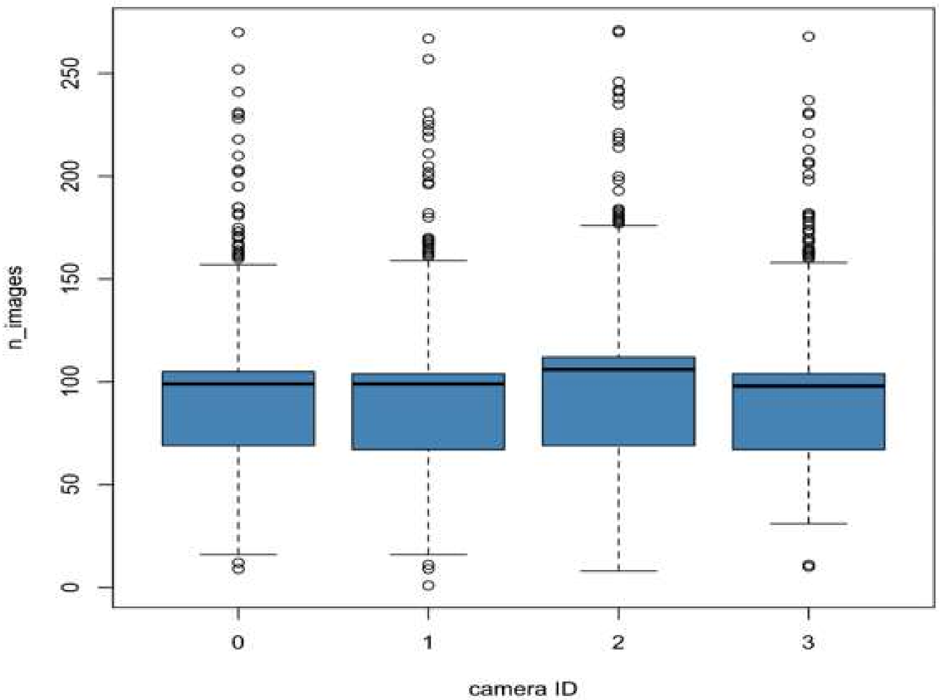


Figure 4.2. Camera-wise counts and distribution of number of images taken for each individual sheep.

Table 4.4. Camera-wise number of images.

Camera	Mean	Min	Max	Median
0	91.5	9	270	99
1	90.4	1	267	99
2	95.8	8	271	106
3	90.2	10	268	98

Data editing was conducted to set the minimum number of images to 10 per sheep per camera and further editing to delete all images with no matching phenotype records.

Furthermore, sheep from mob 3 representing a cohort of small ewe lambs which had not been shorn were removed from further analyses. All other sheep were approximately 2-3 weeks off shears and weighed straight off pasture.

Results for Prediction of Body weight

Description of CNN model: A convolutional neural network (CNN), also known as convnets, was used to predict the weight from the images. The CNN model trained here consisted of a stack of four layer_conv_2d layers, four layer_max_pooling_2d layers, two dense fully connected layers and one dropout layer. The layer_conv_2d is a 2D convolution layer (e.g. spatial convolution over images) which creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs. The first layer_conv_2d takes in images as input tensors (height X width X channels = 200X 400X3). The width and height dimensions shrank as the CNN goes deeper. For prediction of weight the network ends with a single unit with linear activation. The dropout layer was applied by randomly dropping out (setting to zero) 50 % of output features of the layer during the training. In total there were 3,797,985 trainable parameters in this model. The network was trained by minimising the mean squared errors (MSE) of prediction in the training examples. The model was trained for 200 epochs where one iteration over all the training data is called an epoch. However, the losses were tracked in the validation set for early stopping and selection of best model.

The models were trained using 10 images from each sheep (the training set). In the test set, for each sheep 10 predicted values of weight were obtained using 10 different images. The 10 images for a sheep were selected randomly from all the images available for that sheep from the camera being analysed. Following estimates are based on the mean of 10 weights per sheep. The results from 10 individual images are provided in the Appendix.

For both the side and the top camera, the models in the training set fitted the data with a high degree of accuracy (0.94 and 0.95 respectively; Table 4.5 and 4.7) and was confirmed in the validation data (0.86 and 0.87 respectively Table 4.5 and 4.7). The models fitted the data with negligible bias (all estimates were close to 1.00 which is bias free Table 4.5 and 4.7 for side and top camera.) For both the side and top camera the predicted weights matched the test (observed) weights with a high degree of accuracy (0.86 and 0.87) and negligible bias (Table 4.5 and 4.7).

To make sure that the final CNN model was trained on the features of the images directly related to the outcome variable, body weight in this case, additional models with the same architecture were trained on the randomly shuffled trait values for comparison. These null models are not expected to show any association or predictive power. In deed when the weight data were permuted (i.e. body weight values was randomly shuffled over the animals/images), and the same neural network was trained in similar fashion it showed that the accuracy of predicted weights was not significantly different from 0 (Pearson correlation coefficient 0.001 and 0.006, respectively for side and top camera Table 4.6 and 4.8, Figure 4.5 and 4.7). These null models could not differentiate any variation in the body weight from the images and predicted close to the mean weight for all the images shown as horizontal spread of predicted values in Figure 4.5 and 4.7. This suggests that the trained CNN models on the real data (Table 4.5 and Table 4.7) were able to extract and use relevant features from the images associated with the body weight measurements. In other words, the neural network was trained to a high degree of accuracy based on the information presented to it.

Using images from both side and top cameras combined made a small but detectable improvement in the accuracy of fitting the data (0.96) and in the predicted weights (0.89) in the test set but the improvements should be considered marginal at the 2nd decimal place. The marginal improvement of using information from 2 cameras was also shown in MAPE -the mean absolute % error (as a deviation from predicted and observed) which was 6.45% and 6.14% for side and top camera respectively (Tables 4.5 and 4.7) and 5.97% combined (Table 4.9).

One factor that may have reduced accuracy in the predicted test set when compared to the fit in the training data was a small % of bodyweights may have been recorded with small random errors (Appendix A). Furthermore, the bodyweight of yarded sheep (non-fasted) may have changed slightly throughout the course of the day as the sheep emptying out, which may also affect the accuracy of bodyweights in the training data (not tested).

Body weight prediction from Side Camera

Table 4.5 Performance of the CNN model 1 for prediction of weight using images from the side camera.

Data Set	*n*	*MAE*	*MAPE*	*MSE*	*COR*	*Bias*	*Trait Mean*	*Trait SD*	*Predicted Mean*
Train	2292	1.975	4.712	6.427	0.941	1.003	44.088	6.912	44.457
Val	581	2.702	6.249	12.574	0.858	0.998	44.565	6.851	44.818
Test	**912**	**2.783**	**6.449**	**12.824**	**0.858**	**0.994**	**44.522**	**6.878**	**44.631**

Where MAE is mean absolute error; MAPE is mean absolute percent error; MSE is mean squared error; COR is accuracy of prediction as correlation coefficient between predicted and actual trait value; Bias is regression coefficient of regression of predicted value on trait value.

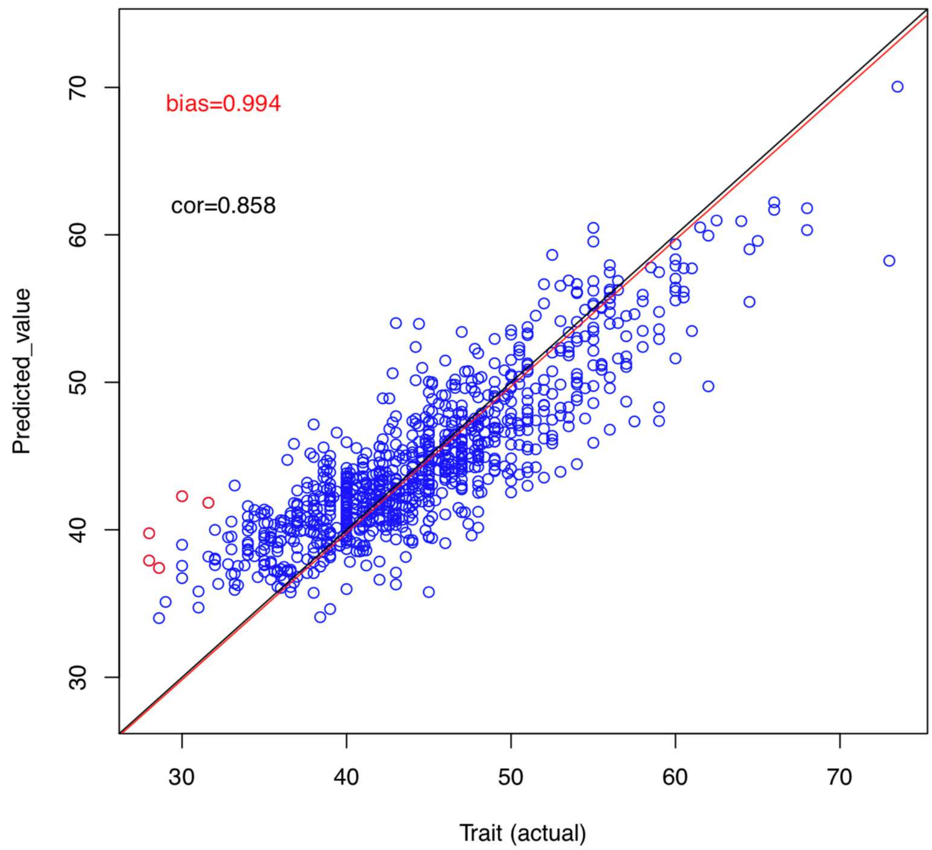


Figure 4.4 Performance of the CNN model 1 in test set for prediction of weight using images from the side camera.

Table 4.6. Performance of the CNN model 1 for prediction of weight using images from the side camera.
The weights were reshuffled randomly for this analysis.

Data Set	*n*	*MAE*	*MAPE*	*MSE*	*COR*	*Bias*	*Trait Mean*	*Trait SD*	*Predicted Mean*
Train	2292	5.41	12.511	47.415	0.163	0.980	44.34	6.965	44.496
Val	581	5.181	12.041	46.32	0.074	0.984	44.191	6.821	44.497
Test	**912**	**5.394**	**12.63**	**46.476**	**0.001**	**0.986**	**44.126**	**6.775**	**44.545**

where MAE is mean absolute error; MAPE is mean absolute percent error; MSE is mean squared error; COR is accuracy of prediction as correlation coefficient between predicted and actual trait value; Bias is regression coefficient of regression of predicted value on trait value.

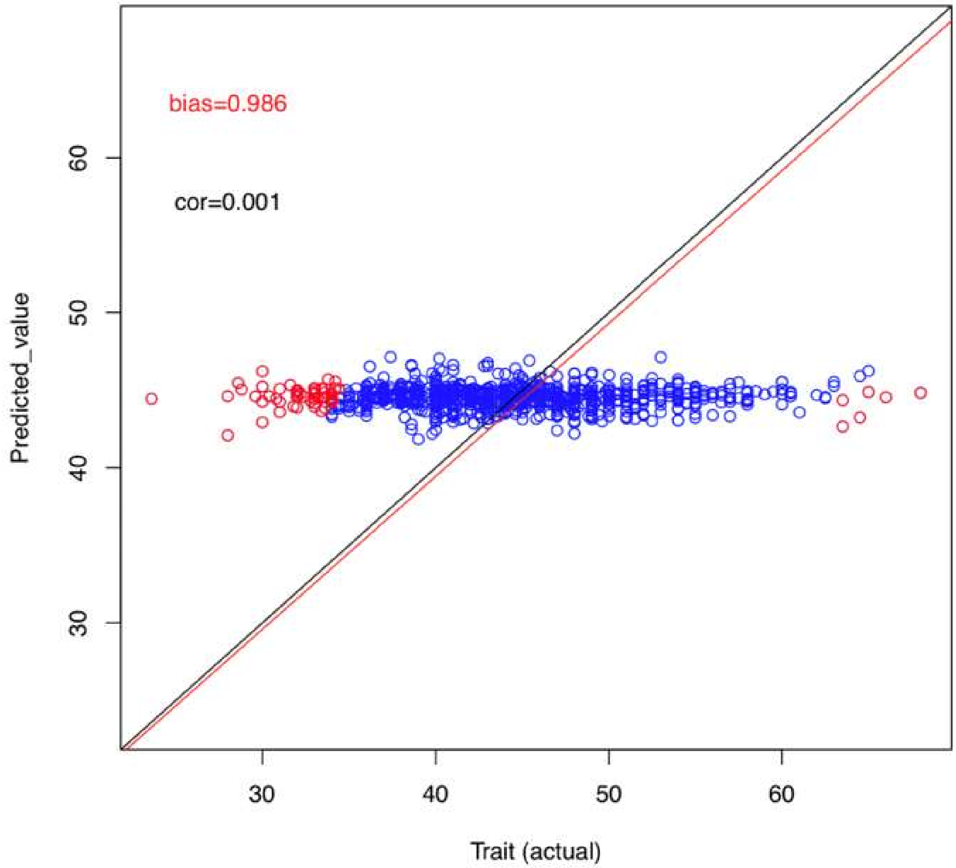


Figure 4.5 Performance of the CNN model 1 in test set for prediction of weight using images from the side camera. The weights were reshuffled randomly for this analysis.

Body weight prediction from Top Camera

Table 4.7 Performance of the CNN model 1 for prediction of body weight using images from the top camera.

Data Set	*n*	*MAE*	*MAPE*	*MSE*	*COR*	*Bias*	*Trait Mean*	*Trait SD*	*Predicted Mean*
Train	2292	1.754	4.149	4.975	0.954	0.999	44.088	6.912	44.267
Val	581	2.486	5.697	11.349	0.872	0.992	44.565	6.851	44.486
Test	**912**	**2.662**	**6.144**	**11.685**	**0.871**	**0.991**	**44.522**	**6.878**	**44.439**

where MAE is mean absolute error; MAPE is mean absolute percent error; MSE is mean squared error; COR is accuracy of prediction as correlation coefficient between predicted and actual trait value; Bias is regression coefficient of regression of predicted value on trait value.

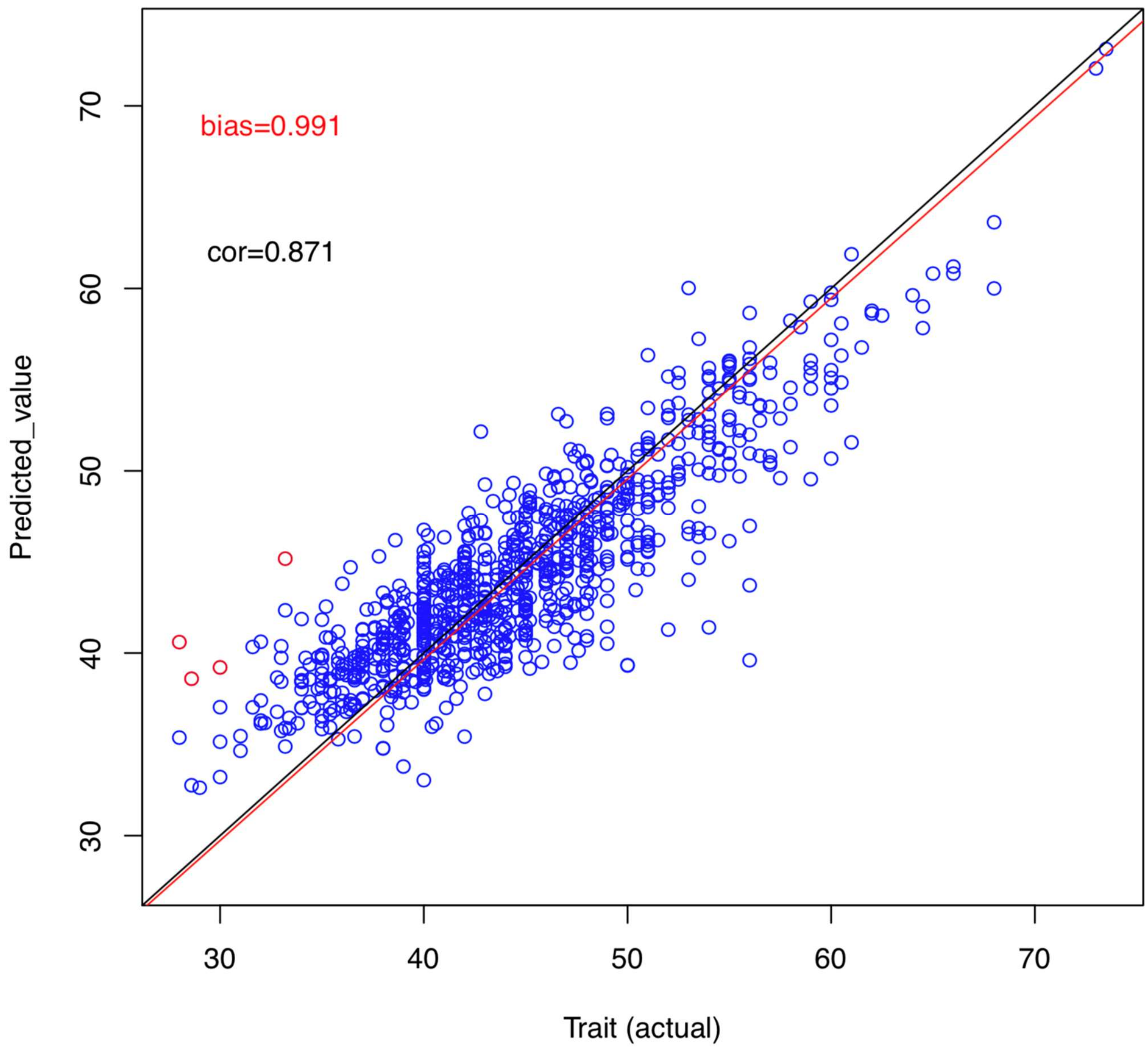


Figure 4.6 Performance of the CNN model 1 in test set for prediction of weight using images from the top camera.

Table 4.8 Performance of the CNN model 1 for prediction of weight using images from the top camera.
The weights were reshuffled randomly for this analysis.

Data Set	*n*	*MAE*	*MAPE*	*MSE*	*COR*	*Bias*	*Trait Mean*	*Trait SD*	*Predicted Mean*
Train	2292	5.352	12.182	47.25	0.237	0.964	44.34	6.965	43.792
Val	581	5.18	11.847	46.878	0.006	0.968	44.191	6.821	43.794
Test	**912**	**5.327**	**12.265**	**46.199**	**0.006**	**0.970**	**44.126**	**6.775**	**43.795**

where MAE is mean absolute error; MAPE is mean absolute percent error; MSE is mean squared error; COR is accuracy of prediction as correlation coefficient between predicted and actual trait value; Bias is regression coefficient of regression of predicted value on trait value.

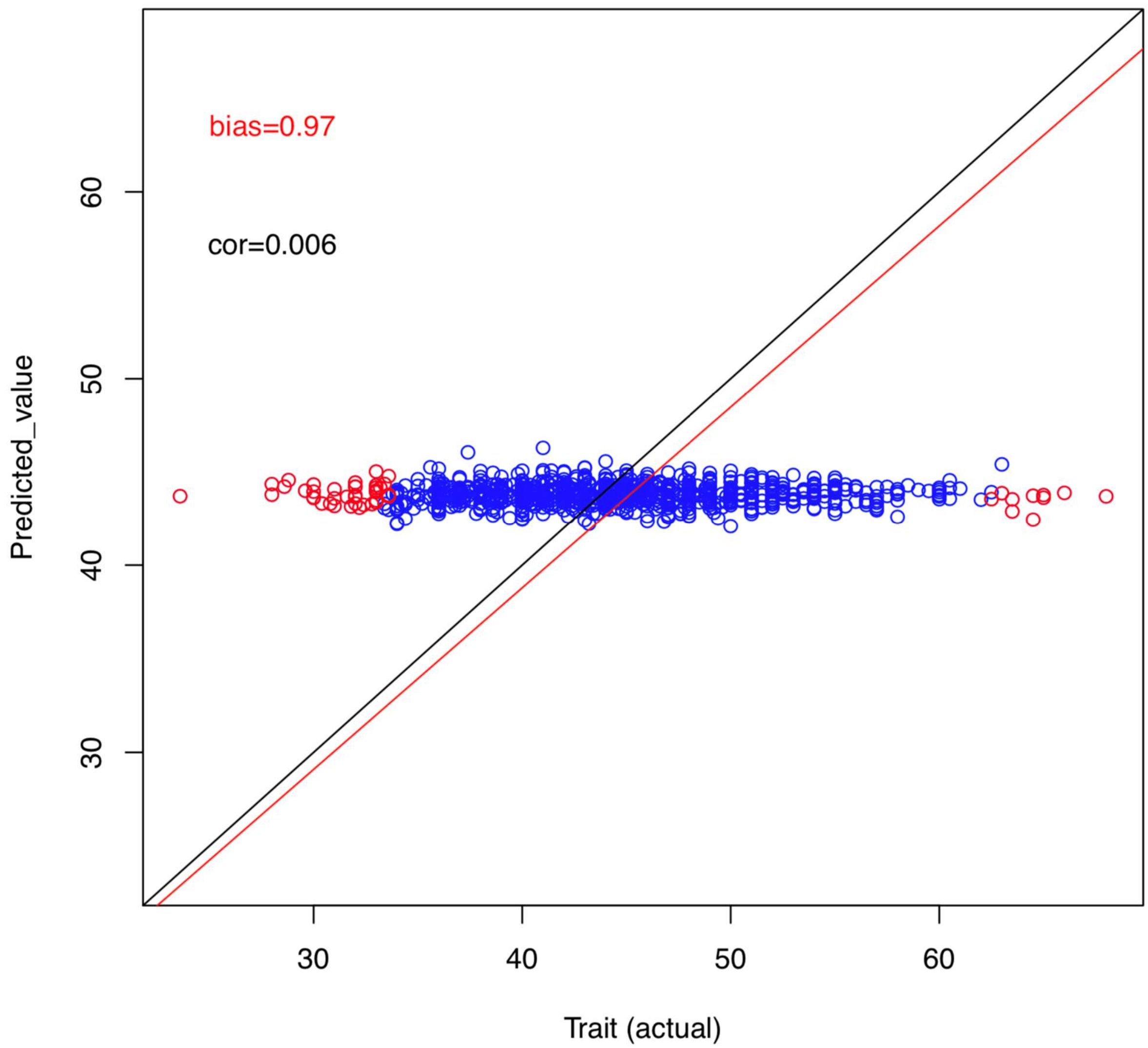


Figure 4.7 Performance of the CNN model 1 in test set for prediction of weight using images from the top camera. The weights were reshuffled randomly for this analysis.

Body weight prediction by using averaging prediction of side and top camera

Table 4.9 Performance of the CNN model 1 for prediction of weight using images. The mean of predicted weights from side and top camera were used in this analysis i.e. mean of 20 images.

Data Set	*n*	*MAE*	*MAPE*	*MSE*	*COR*	*Bias*	*Trait Mean*	*Trait SD*	*Predicted Mean*
Train	2292	1.723	4.114	4.832	0.960	1.001	44.088	6.912	44.362
Val	581	2.427	5.601	10.724	0.882	0.995	44.565	6.851	44.652
Test	**912**	**2.578**	**5.971**	**10.852**	**0.886**	**0.993**	**44.522**	**6.878**	**44.535**

where MAE is mean absolute error; MAPE is mean absolute percent error; MSE is mean squared error; COR is accuracy of prediction as correlation coefficient between predicted and actual trait value; Bias is regression coefficient of regression of predicted value on trait value.

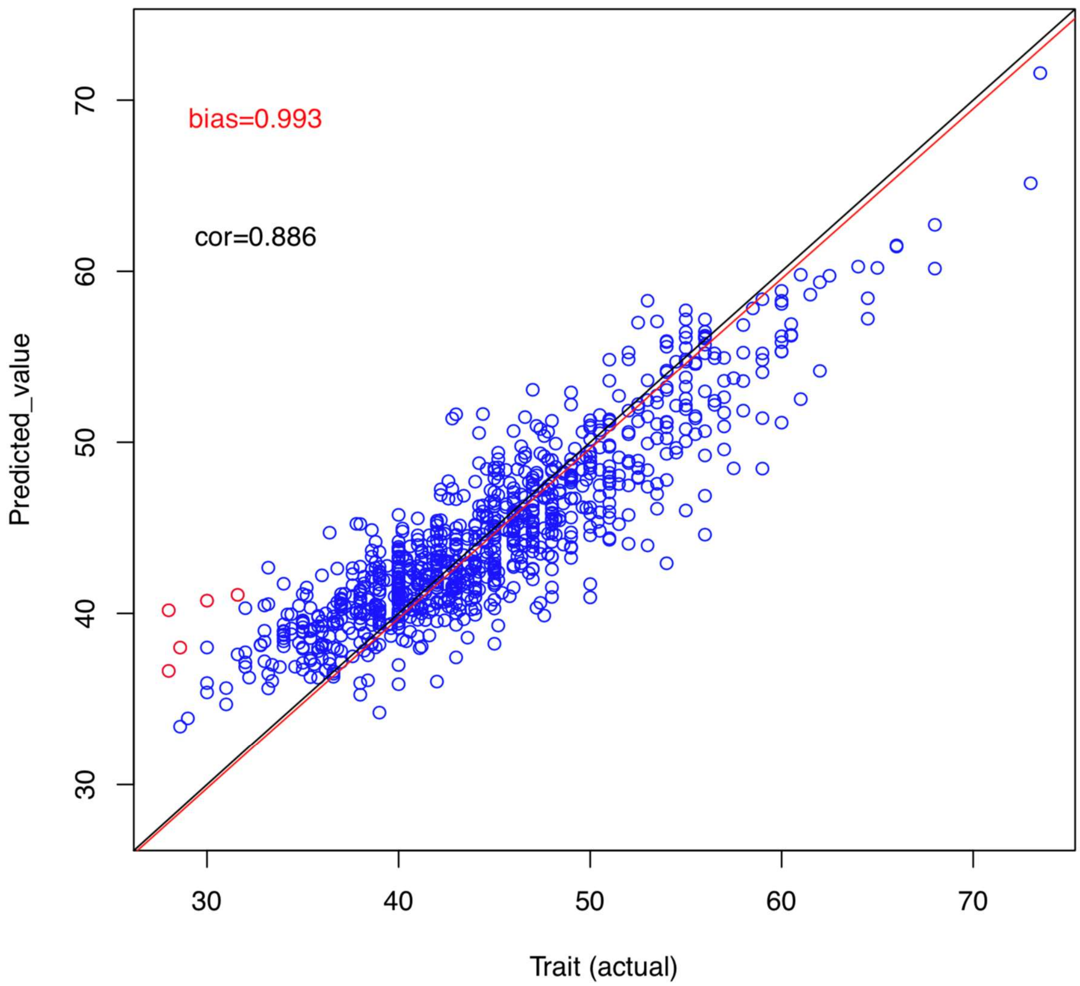


Figure 4.8 Performance of the CNN model 1 in test set for prediction of weight using images. The mean of predicted weights from side and top camera were used in this analysis i.e. mean of 20 images.

Results for facial recognition

Sheep Body and Face Detection

To train the detector, we labelled 1362 images, in which 1117 images were randomly selected from Wallaloo Park, whilst 245 images were taken from a combination of farms. We first trained the detector on the 600 images of Wallaloo Park and used 517 images from the same farm and 245 images from the other farms for testing respectively, results are shown in Table 4.10. Although the detector performs well on test images from Wallaloo Park (accuracy 0.987-0.990) the performance drops considerably when testing images from other farms (0.794-0.895). Thus, we retrained the detector on a mixed dataset containing images from multiple farms. The results are displayed in Table 4.11 with a detection accuracy of 0.979 to 0.990.

Table 4.10 Detection accuracy of the model trained on images from Wallaloo Park and tested against images from the same farm and other farms.

	Images from mixed farms
Face	0.979
Full body	0.995
Face and body	0.987

Table 4.11 Detection accuracy of model trained on images from multiple farms combined.

	Images from the Same Farm	Images from other farms
Face	0.987	0.794
Full body	0.990	0.895
Face and Body	0.988	0.845

The results showed that identification of specific components (head or body) from sheep images could be achieved with 98% accuracy from images provided by the data capture platform described above. The capture of the targets shown in the photos below of head and body segments were identified and isolated by yellow and red boxes respectively. These targets were used as inputs for the identification testing described below.

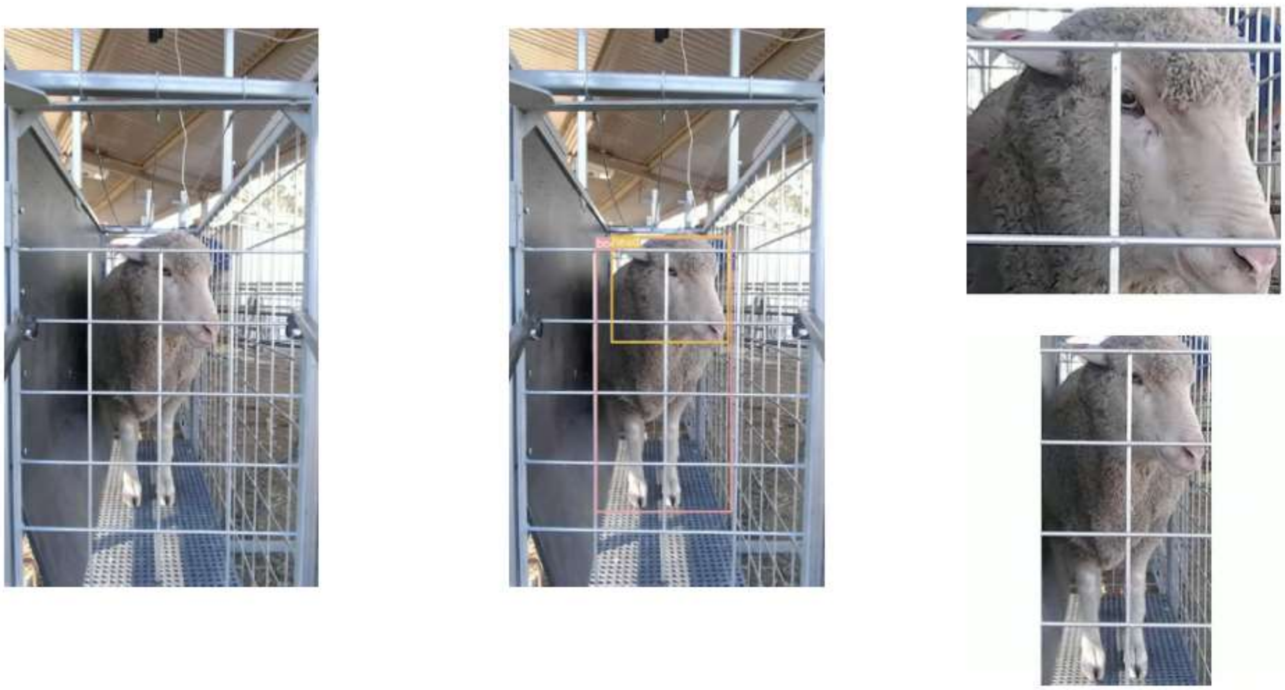


Figure 4.9. Example of head and body capture from the original image.

Sheep Identification

There were approximately 110 images per sheep available for sheep identification. Half of these images were used for training, the other half for testing. Only images taken from the front camera (camera_0), were used in this evaluation. We used different numbers of sheep from the same farm for training the model. The results are provided in Table 4.12.

Table 4.12 Accuracy of identification model trained on images from one farm

Number of Sheep	20	500	1000	1900
Head (Top 1 Accuracy %)	**94.07**	**95.49**	**96.54**	**97.21**
Body (Top 1 Accuracy %)	**97.35**	**97.70**	**98.05**	**98.38**
Head + Body (Top 1 Accuracy %)	**98.65**	**99.30**	**99.55**	**99.66**

The proposed method achieves impressive accuracy on sheep identification (Table 4.12). Across the range 94.1-99.7% of images could be matched to their corresponding identity. However, the images used for testing and those used for training are collected at the same time. Therefore, the differences between images belonging to the same sheep are relatively small. In order to test whether the model has better generalized capabilities and the ability to handle environmental change (eg. wool growth, test conditions), we trained an identification model on the images. These images were captured during the same time period, then evaluated the model on another dataset of images collected for the same animal six months later. In this case, there was a large variation between training images and testing images of the same sheep. It is not surprising that the trained model does not perform well in this setting (accuracy of capturing images assigned to the same sheep was 6%; Table 4.13). This suggests that training the model on data with a small variance, will lead to overfitting issues (ie the model is too specific to the dataset and lacks broader universal application/utility). To tackle this problem, a simple solution was to create a new training dataset with a large variation between images. By combining data from the two different time periods the network can focus on learning the true biometric information for each individual sheep. Therefore, we retrained a model on a mixed dataset containing images collected from the two time periods. Using the mixed dataset, images could be assigned to their correct identity across both time periods with 90-98% accuracy (Table 4.13).

Table 4.13. Accuracy of identification model trained on the mixed dataset, the result is obtained by evaluating the model on the 800 sheep images, which were collected 6 months later.

	Model Trained only on batch1 data (1000 sheep) and tested on same sheep (batch2-900 sheep)	Model trained on mixed data from batch 1 and 2 (1900 sheep)
Head	0.05	89.98
Body	0.08	93.98
Head + Body	0.06	98.47

This work aimed at providing a proof of concept that individual sheep identification can be addressed by using computer vision pipelines based on deep neural networks. For identification, we showed that building a good quality training dataset is essential for a convolutional network to learn unique patterns and structures of individual sheep from images.

Future work to improve the current system includes:

(1) Enhancing the training dataset. The acquisition of sheep images for each subject should consider different conditions, such as pose variation, distance variation, background variation, illumination variation and appearance variation.

(2) Designing a better network architecture for feature extractor. We can consider incorporating feature from early convolutional layers with features from later layers, since early layers encode local information that might be useful for differentiating sheep.
(3) Sheep facial landmarks prediction and pose estimation. Facial landmarks and pose estimation have demonstrated their effectiveness in many computer vision tasks, such as face recognition, and action recognition. Thus, it is a promising way to improve the sheep identification model by learning these two auxiliary tasks.

Results on neck and body wrinkle score

For the 4045 sheep which were scored for neck and body wrinkle, it was apparent that most were distributed in the middle class and relatively few in the extreme classes. Convolutional Neural Net training is sensitive to data inputs from extreme contrasts in order to define classifying features. The higher number in the middle class contributed higher weightage in the loss function. Furthermore from repeat scores by manual assessors it was apparent that assignment to extreme classes had some noise in that animals may drift by one class category. For this reason the wrinkle traits were analyses as a categorical trait on the original scale of 1-5 classes, and as a binary trait where animals from the two extreme low (class 1 and 2) and two extreme high classes (class 4 and 5) were combined in a single low and high wrinkle class respectively. This was termed wrinkle as a binary (0-1) trait by omitting the middle class. In other words we asked and tested whether CNN could classify animals as being of high or low wrinkle respectively. This doubled the analyses required but showed distinct advantages of letting CNN predictions classify animals with extreme phenotypes. To compare if CNN analyses were performing better than would be expected at random, we re-assigned images animals randomly to a score (permuted) and would expect that CNN could not arrive at accurate predictions in the training, validation and most importantly the test set i.e. the accuracies were expected to be zero. Finally, accuracy was set as proportion of animals correctly assigned to the manual score assigned by assessors. Images from both top and side camera were used for these analyses.

The CNN models trained for prediction of wrinkle scores consisted of a stack of five layer_conv_2d, five layer_max_pooling_2d layers, two dense fully connected layers and four dropout layers. The first layer_conv_2d takes in images as input tensors (height X width X channels = 200X 400X3). The width and height dimensions shrank as CNN goes deeper. The same CNN architecture was used for wrinkle as categorical and binary traits. For prediction of the score categories the network ends with units equal to the number of categories to be predicted with "softmax" activation function which means it will return a vector of probability scores for all the categories (summing to 1). In total there were 933,506 trainable parameters in this model. The network was trained by minimising the “categorical crossentropy" for predicted classes in the training examples, where in binary classification “binary_crossentropy” loss was minimised. The categorical_crossentropy measures the distance between probability distribution of output by CNN and the true distribution of the given class labels/scores. The model was trained for 200 epochs. However, applying a large number of epochs can result in overfitting in the training set, hence, the losses were tracked in the validation set for early stopping and selection of the best model. This strategy along with applying dropout helped to avoid overfitting and to generalise the model for predicting the classes in the unseen data. Similar to weight prediction, 10 images for each sheep were used in the training, validation and test sets. The results presented here are based on the median prediction of classes from 10 images.

Table 4.14a-d show the predicted accuracy for neck and body wrinkle using the top and side cameras. For body and neck wrinkle, high levels of accuracy were observed when wrinkle was classified as a binary trait – test accuracies were in the range of 0.732-0.899 (Table 4.14a-d). When body and neck wrinkle were analysed as categorical traits however, accuracy of prediction dropped significantly from to 0.368-0.583 (Table 4.14a-d).

The utility of prediction (Kappa value) indicates the probability of a correct prediction occurring by chance. The closer the kappa value is to 1, the least likely the correct prediction occurred due to chance. The kappa value for neck and body wrinkle, using images from the top camera and the binary system, was moderate to high, suggesting the correct wrinkle prediction did not occur due to chance (0.472-0.798; Table 4.14a and 4.14c). When wrinkle was classified as categorical, the kappa values were substantially lower, especially body wrinkle score detected by side camera, which was not significantly different from zero (0.019), suggesting the correct prediction probably occurred by chance (Table 4.14b).

Table 4.14a-d The accuracy of prediction for body and neck wrinkle fitted as a binary trait (0-1) and as a categorical trait (1-5) under a CNN model using images from both top and side camera.

Table 4.14a

	Body Wrinkle Low vs high (0-1)				Body Wrinkle score 1-5			
Top camera	N	Accuracy	Kappa	P value	N	Accuracy	Kappa	P value
Train	1353	0.991	0.982	0	2292	0.682	0.517	3.65E-145
Validation	333	0.910	0.820	9.80E-54	581	0.454	0.184	0.00011481
Test	**585**	**0.899**	**0.798**	**6.19E-94**	**912**	**0.487**	**0.240**	**1.47E-12**

Table 4.14b

	Body Wrinkle Low vs high (0-1)				Body Wrinkle score 1-5			
Side camera	N	Accuracy	Kappa	P value	N	Accuracy	Kappa	P value
Train	1353	0.900	0.800	4.73E-198	2292	0.459	0.113	1.61E-05
Validation	333	0.775	0.550	5.99E-22	581	0.373	0.017	0.61644748
Test	**585**	**0.732**	**0.463**	**1.10E-29**	**912**	**0.368**	**0.019**	**0.62000751**

Table 4.14 c

	Neck Wrinkle Low vs high (0-1)				Neck Wrinkle score 1-5			
Top camera	N	Accuracy	Kappa	P value	N	Accuracy	Kappa	P value
Train	1324	0.902	0.759	1.13E-70	2292	0.596	0.322	4.25E-69
Validation	326	0.782	0.481	5.41E-05	581	0.554	0.287	7.17E-17
Test	**574**	**0.808**	**0.472**	**5.48E-05**	**912**	**0.581**	**0.304**	**4.83E-21**

Table 4.14d

	Neck Wrinkle Low vs high (0-1)				Neck Wrinkle score 1-5			
Side camera	N	Accuracy	Kappa	P value	N	Accuracy	Kappa	P value
Train	1324	0.804	0.511	2.74E-18	2292	0.598	0.325	2.18E-70
Validation	326	0.758	0.420	0.00212949	581	0.554	0.286	7.17E-17
Test	**574**	**0.815**	**0.498**	**9.46E-06**	**912**	**0.583**	**0.308**	**1.36E-21**

Accuracy = proportion correctly classified in original score dataset
Kappa = (Total Accuracy-random Acc)/(1-randomAcc)

The accuracy of the prediction of wrinkle scores by the CNN model is further validated when the best fit model for body wrinkle prediction derived from images taken by the top camera was compared to using the same images but randomly allocated to sheep ID and their scores (ie random control model). The fitted model had an accuracy of 0.899 with a kappa prediction value of 0.789. Whereas the random control had an accuracy of 0.53 and a kappa value of 0.000, showing the correct results occurred due to chance alone. This showed that the wrinkle prediction was indeed derived by the trained CNN model on features associated with the wrinkles contained in the images (Table 4.15).

Table 4.15 The accuracy of prediction for body wrinkle fitted as a binary trait under a CNN model compared to the images fitted randomly (permuted).

	Score type	Camera _View	Data Set	n	Accuracy	Kappa	Accuracy PValue
Body wrinkle fitted	Binary	top	train	1353	0.991	0.982	0
Body wrinkle fitted	Binary	top	valid	333	0.910	0.820	9.80E-54
Body wrinkle fitted	**Binary**	**top**	**test**	**585**	**0.899**	**0.798**	**6.19E-94**
Body wrinkle (random)	Binary	top	train	1371	0.500	0.0	0.51077501
Body wrinkle (random)	Binary	top	valid	358	0.525	0.0	0.52144945
Body wrinkle (random)	**Binary**	**top**	**test**	**542**	**0.530**	**0.0**	**0.5174952**

The capacity for CNN to predict wrinkle scores and the improvement of using wrinkle as a binary (0-1) trait vs a categorical trait (1-5) could also in part be explained by the robustness of the visual assessment by manual scores. When body and neck wrinkle were assessed on repeated scores of 200 sheep the average accuracy was 0.566 and 0.595 for body and neck wrinkle respectively (Table 5.16), with a low predictive value for kappa (0.407 and 0.355, Table 4.16). When body and neck wrinkle scores were collapsed as a binary trait in low and high classes, accuracy of manual assessors was substantially higher (0.992 and 0.811 for body and neck wrinkle respectively, Table 4.16) corresponding with higher kappa values 0.978 and 0.890 (table 4.16. This suggests that manual assessors were able to accurately distinguish between high wrinkle and low wrinkle sheep and this improved visual benchmark binary phenotype used training Deep Learning model resulted in higher accuracies for the CNN derived predictions. This confirms that highly accurate phenotypes are required as input parameters for Deep Learning models. Good information in= good information out.

Table 4.16 The accuracy of subjective wrinkle and face cover scores by repeat assessment on 190 sheep.

Trait scored	n	Accuracy	Kappa	AccuracyPValue
body_wrinkle 1-5	189	0.566	0.407	1.91E-06
body_wrinkle_binary	118	0.992	0.978	1.04E-14
neck_wrinkle 1-5	190	0.595	0.355	0.017395212
neck_wrinkle_binary	111	0.982	0.890	0.004893489
face_cover 1-5	190	0.811	0.206	0.617430752

Results for Prediction of Face cover

Face-cover Scoring Machine Learning Model

The goal was to create a machine learning model that can process a photo of a sheep and predict its face-cover score. To create the ML model involves:

- Collecting a dataset of photos of sheep with clear views of their head
- Having a human expert classify the sheep from their photos with a face-cover score: 1 – 5
- Group the photos with the same face-cover score into directories
- Train a Convolutional Neural Network (CNN) with the photos in their classification directories
- Hold back the photos from a few sheep to be used as a test set to calculate the accuracy of the CNN classifier

Human Scoring

Approximately 400 photos were taken of each sheep. It wasn't practical for an expert to look through all the photos, so a desktop GUI application was created to display 12 images from a single sheep at once. The GUI also included fields to input various parameters like face-cover score, neck wrinkle and body wrinkle scores. A csv linking the sheep's EID and the scores was the output of the application. This csv was then used as the input to a script that grouped the photos with the same scores into directories.

Data Selection

Table 4.17 The distribution of face-cover scores.

Face Cover Score	Number of Sheep
1	5
2	302
3	3532
4	212
5	24

The data used was collected during late 2018 and early 2019. Face-cover scores 1 and 5 had too few sheep to be to be included in the model. Additionally, some initial work had indicated that delineating between 2, 3 and 4 with a relatively small data set would not be effective. In order to test the ability of ML models to classify face-cover scores, a model was created to delineate between just 2 and 4. The rationale being that trying the easier classification task would have more chance of providing a clue as to the validity of the concept. Examples of a face score 2 (Figure 4.10) and 4 (Figure 4.11) are provided.

Figure 4.10 Example of a face cover score 2.

Figure 4.11 Example of face-cover score 4.

Cropped Head Shots

In the side photos the area of sheep's heads uses approximately 3% of the pixels. To remove superfluous information the pictures were cropped around the sheep's head. The original images with dimensions 1,920 by 1,080 where cropped down to 380 by 370 pixels. The sheep were often very active in the crate resulting in lots of photos where their head was angled in a way that their face-cover was obscured. Consequently, the process of manually cropping the images also became a process of selecting images where the sheep's head was close to side on. The selecting and cropping of the images was performed with a custom-built desktop application which allowed over 45,000 photos to be processed in less than 10 hours. An example of a photo that was not used is provided in figure 4.12.

Figure 4.12 Example of a photo that was not used because the sheep's head was not side on.

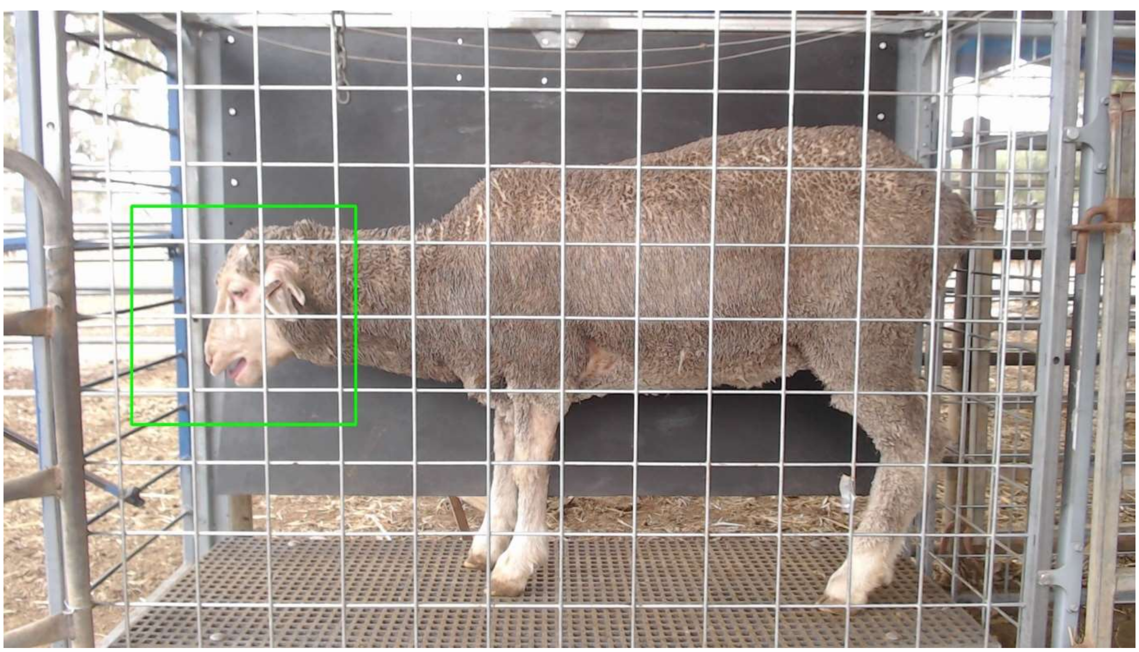


Figure 4.13 Example of a photo that was selected for training the ML model and the cropped area.

Table 4.18 The number of raw images and cropped images for each face-cover score.

Face-cover Score	Number of raw images	Number of cropped images
2	26,650	5,449
4	19,644	3,311

The cropped photos were selected from approximately 200 sheep for each of the two scores. Photos from ten sheep from both scores were separated out to be used to test the ML model. The remaining photos were reduced to 3,081 photos for each score so that the training sets for each score were balanced.

Transfer Learning with Inception V3 and ImageNet

Modern image recognition models have millions of parameters. Training from scratch requires a lot of labelled training data and can consume hundreds of Graphical Processer Unit (GPU) hours. Transfer learning is a technique that shortcuts much of this process by taking most of a model that has already been trained on a related task and re-using it in a new model. Although it is not as good as training the full model, it is effective for many applications, works with moderate amounts of data and can run on a laptop.

The model for the sheep face-cover score classifier used the Inception V3 architecture trained on the ImageNet dataset with the final layer having been trained on the cropped photos. The Inception V3 architecture is a deep (48 layers) convolutional neural network that has been improved with the addition of inception layers which are multiple concatenated convolutions. ImageNet is a large image data set of over 14 million labelled images. An Inception V3 model with all but its final layer trained on ImageNet is available from Google. That left the final layer to be trained with the cropped photos. This process took between several hours to several days depending on the number of photos used for training. In all cases 4,000 training steps were used.

Cropped Headshot Results

Photos from ten sheep from both scores (2 and 4) were used to test the ML learning model created from the cropped head shots. Considering a random result would have an accuracy of 0.5, the result of 0.73 was reasonably poor. The model ended up being heavily biased towards a face-cover score of 2 i.e. almost all the actual 2's were detected correctly but almost half the 4's were labelled as 2's.

Confusion Matrix

		Predicted	
		2	4
Actual	2	156	4
	4	101	127

Precision, recall and F_1 (the harmonic mean of precision and recall) are often better measures than accuracy. They are defined as:

$$accuracy = \frac{True\ Positives}{Total}$$

$$precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

Precision, recall and F_1

		Precision	Recall	F_1
Score	2	0.607	0.975	0.748
	4	0.970	0.557	0.708

Cropped Pictures Multiplied

An advantage of cropping the images is that the cropped area can be jittered slightly, creating a far bigger data set. Nine cropped areas where selected from each photo. The original manually selected area was chosen as was two photos above, below and from both sides. This resulted in nine cropped photos from each original photo. The spacing between adjacent cropped areas was twelve pixels. This process resulted in approximately 29,800 photos for each face-cover score.

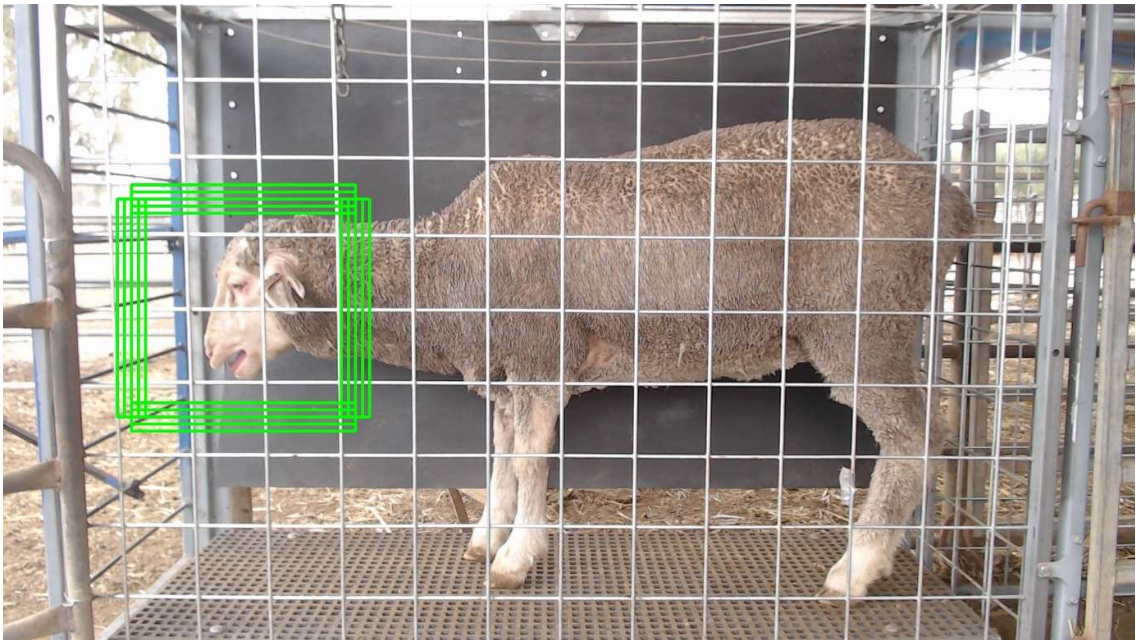


Figure 4.14 Positioning of the nine cropped photos selected from an original photo in a cross configuration.

Confusion Matrix

		Predicted 2	Predicted 4
Actual	2	142	17
	4	35	201

		Precision	Recall	F_1
Score	2	0.802	0.893	0.845
	4	0.922	0.852	0.885

The results were far better, boosting the accuracy from 0.729 to 0.868. As multiplying the number of images by nine had greatly improved the predictive accuracy of the binary classifier, the number was increased again by another factor of nine. 81 cropped images were selected in a square grid. The spacing between cropped areas was eight pixels. The number of photo's in the resulting training set for each face-cover score was approximately 268,000.

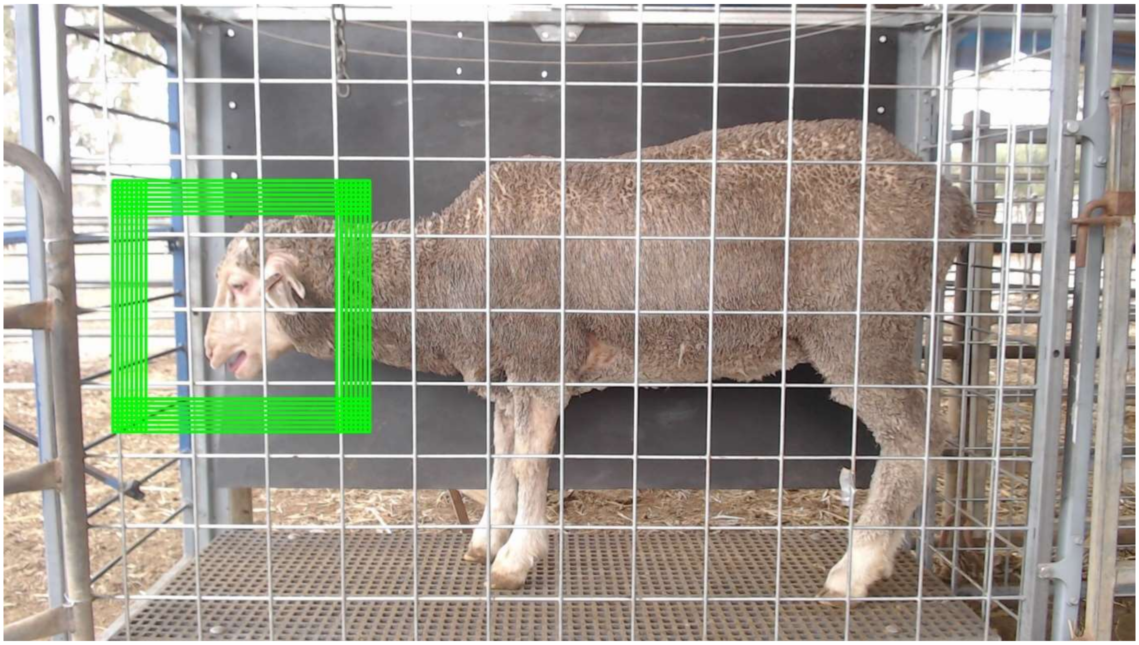


Figure 4.15 Positioning of the 81 cropped photos selected from an original photo in a grid configuration.

Confusion Matrix

		Predicted 2	Predicted 4
Actual	2	143	16
	4	44	192

		Precision	Recall	F_1
Score	2	0.765	0.899	0.827
	4	0.923	0.814	0.865

Unfortunately, the further increase in cropped images from a factor of nine to 81 did not improve the performance of the model. In fact, the performance dropped slightly from 0.868 to 0.848.

Conclusions

Machine learning can be used to create a model that can predict the face-cover score of sheep from photos. The real question is prediction accuracy. The majority (87%) of the sheep photographed had a face-cover score of 3. Training ML models is best done with balanced data sets i.e. each classification having roughly the same number of photos. Some preliminary tests were done creating classifiers delineating between scores 2 and 3. The results were little better than random. To make the task easier ML classifiers were trained to differentiate between face-cover scores 2 and 4.

Using the full images of the photos with scores 2 and 4 also produced random results. Cropping the photos created a classifier that was better than random with an accuracy of 0.729. Cropping multiple areas from the photos dramatically increased the accuracy to 0.868. Creating classifiers to delineate between scores 2 and 4 was a good place to start to gain some insight into the techniques required to increase the accuracy i.e. cropping and multiplying. Using these techniques, it is likely that differentiating between 2, 3 and 4 is possible with the current data set although the accuracy will drop. The real solution is to gather a much larger dataset balanced across all 5 face-cover score values.

Review Bio-sensor and bio-markers

Bio-sensor

The term biosensors encompasses devices that have the potential to quantify physiological, immunological and behavioural responses of multiple animal species. Novel biosensing methodologies offer highly specialised monitoring devices for the specific measurement of individual and multiple parameters covering an animals' physiology as well as monitoring of an animals' environment. These devices are not only highly specific and sensitive to the parameters being analysed, but they are also reliable and easy to use, and can accelerate the monitoring process. Novel biosensors in livestock management provide significant benefits and applications in disease detection and isolation, health monitoring and detection of reproductive cycles, as well as monitoring physiological wellbeing of the animal via analysis of the animals' environment. With the development of integrated systems and the Internet of Things, continuous monitoring devices are expected to become affordable. The data generated from integrated livestock monitoring is anticipated to assist farmers and the agricultural industry to improve animal productivity in the future. The data is expected to reduce the impact of the livestock industry on the environment, while at the same time driving the new wave towards the improvements of viable farming techniques. This review focusses on the emerging technological advancements in monitoring of livestock health for detailed, precise information on productivity, as well as physiology and well-being. Biosensors will contribute to the 4th revolution in agriculture by incorporating innovative technologies into cost-effective diagnostic methods that can mitigate the potentially catastrophic effects of infectious outbreaks in farmed animals (Suresh Neethirajan1, Sheng-Tung Huang2, Satish K.Tuteja1, David Kelto Recent Advancement in Biosensors Technology for Animal and Livestock Health Management BioRXIV ttp://dx.doi.org/10.1101/128504doi: bioRxiv preprint first posted online April 2017)

Bio-markers

Bio-markers are indicators of biological processes and physiological states that can reveal a variety of health and production associated traits. Although global efforts in bio-marker research has primarily focussed on human medical outcomes as early diagnostic or prognostic disease markers, they now find more widespread application in livestock industries. Biomarkers may be broadly classified on basis of their biological characteristics but in many cases, responses among them are overlapping and interrelated, (ie proteomic responses may be strongly influenced by genetic factors and thus linked to genomic and transcriptomic responses or environmental factors and physiological state and thus linked to metabolomic responses). The following broad classification of bio-markers can cover most applications in livestock:

- Genomic- DNA based marker profiles
- Transcriptomic- RNA based marker profiles
- Proteomic- protein based marker profiles
- Metabolomic- small molecule metabolite profiles

Sampling may be from bio-fluids (blood, semen, saliva, urine, milk), tissue or cell-based origins, and potentially gaseous outputs. Diagnostic laboratory procedures are usually semi-automated and may be high target specific (ie single molecule/mutation specific) or multi-targets with many thousands of targets processed simultaneously. The abundance of target outputs makes these molecular phenotypes an ideal source of data input for Deep Learning since immediate structures in the data or association to economic phenotypes are not usually obvious. Combined statistical and Deep learning approaches are useful here to develop predictive models which can be tested for robustness, accuracy, specificity and sensitivity and in many cases have potential for multiple phenotypes from a single sample thus greatly increasing cost efficiency. In all cases bio-markers should have the following characteristics:

- Be accurate, sensitive and specific for a biological (disease/production) trait
- Be robust and unaffected by unrelated conditions
- Be easy to measure and sample from accessible body fluids and tissue
- Be easy and low cost in diagnosis
- Potentially pluri-specific for a wide range of traits (ie DNA as a source for genetic markers)
- Potentially have prognostic (pre-event) predictive capability
- Potentially be able to be used for phenotypic prediction and genetic evaluation

There is now also global effort to integrate bio-sensor and bio-marker analysis and reporting to the point where *in vivo* diagnostic implants of miniaturized sensors can report on biological states and molecules in real time. At present, to integrate such platforms for sheep is lagging behind those developed in human or large animals (dairy, beef cattle). Nevertheless, rapid transition across biological species is highly likely given the similarity in broad underlying biological processes.

Despite the technological advances to undertake large scale discovery of genomic, proteomic and metabolomic investigations, relatively few have been conducted in sheep. The upfront cost of the research could be seen as high but, equally important is the potential payoff if large scale predictions could be made at low cost against a broad range of traits. The biomarker profiles themselves are not informative unless matched against data bases of high quality phenotypes and potentially breeding values of individuals which are well characterised for not only core production traits, but also the complex traits for which we aim to seek predictive outcomes –such as disease resistance, reproductive potential, carcass quality and feed conversion efficiency. The general requirements to conduct such research is presented below. The capacity for Deep Learning to analyse such data sets is relatively novel but in principle entirely consistent with the analytical which are currently available already. At present the only bio-marker platform which would have prospective potential for widespread use is a genomic (DNA) based approach but proteomic and metabolomic approaches are warranted.

Examples of Biosensor and Biomarker technology

Researching the development and application of new biosensor and biomarker technologies specifically in sheep systems has identified some examples. However, it is clear there are many other efforts in species other than sheep. Innovation in biomarker and biosensor technology in humans is abundant. Dairy cattle, followed by beef, is also receiving attention as is the more intensive livestock such as poultry and pigs. Ovine applications are certainly in research and development stages whereas commercial application is more advanced and common elsewhere.

Below are examples of technologies being applied to sheep production systems or close to being applied:

What is it?	Detecting pain
What phenotype is it measuring?	Pain, welfare, disease, mastitis
Biomarker/Biosensor	Biosensors, imagery
Species	Sheep
The problem	While a measurement scale (SPFES) to access pain in sheep has been developed, its use is time consuming and sometimes biased by humans.
How does it work?	Uses imagery to detect whether a sheep is in pain. There are five main things which happen to sheep's faces: their eyes narrow, their cheeks tighten, their ears fold forwards, their lips pull down and back, and their nostrils change from a U shape to a V shape. The use of an AI system with the SPFES can rank these characteristics on a scale to measure the severity of pain.
Measure real time or sent back to lab?	Relies on the analysis of an image, so not real time yet.
$$$	N/A.
Being used/still in trials?	Trials
Reference	• AI system to assess pain levels in sheep https://www.cam.ac.uk/research/news/researchers-design-ai-system-to-assess-pain-levels-in-sheep • Estimating Sheep Pain Level Using Facial Action Unit Detection https://www.cl.cam.ac.uk/~pr10/publications/fg17.pdf

What is it?	Predicting Lameness in Sheep
What phenotype is it measuring?	Lameness
Biomarker/Biosensor	Biosensor
Species	Sheep
The problem	Lameness is a clinical symptom associated with sheep diseases around the world, having adverse effects on weight gain, fertility, and lamb birth weight, and increasing the risk of

	secondary diseases. Current methods to identify lame animals rely on labour intensive visual inspection.
How does it work?	A collar, leg, and/or ear attached tri-axial accelerometer to discriminate between sound and lame gait movement in sheep.
Measure real time or sent back to lab?	Currently relies on the analysis of Biosensor, so not real time.
$$$	N/A.
Being used/still in trials?	Trials
Reference	• Predicting Lameness in Sheep Activity Using Tri-Axial Acceleration Signals https://www.researchgate.net/publication/322410714_Predicting_Lameness_in_Sheep_Activity_Using_Tri-Axial_Acceleration_Signals

What is it?	Tracking sheep
What phenotype is it measuring?	Pedigree, mothering ability, behaviour, windchill, feeding, predation
Biomarker/Biosensor	Biosensor, Solar powered smart ear tag
Species	Sheep
The problem	Provides another different technical method of tracking
How does it work?	Trilateration
Measure real time or sent back to lab?	Potentially real time
$$$	N/A.
Being used/still in trials?	Trials
Reference	• Digibale http://www.woolindustries.org/2._EN_-_Marcus_Majass_-IWTO-presentation.pdf

What is it?	Classification of sheep urination events using accelerometers to aid improved measurements of livestock contributions to nitrous oxide emissions
What phenotype is it measuring?	Urination behaviour
Biomarker/Biosensor	Biosensor
Species	Sheep
The problem	Livestock emissions account for 74% of nitrous oxide contributions to greenhouse gases in the UK. However, it remains uncertain how much is directly attributable to localised sheep urination events, which could generate nitrous oxide emission 'hot spots'
How does it work?	Accelerometers
Measure real time or sent back to lab?	Requires analysis of data
$$$	N/A.
Being used/still in trials?	Trials
Reference	• Classification of sheep urination events using accelerometers to aid improved measurements of livestock contributions to nitrous oxide emissions https://www.sciencedirect.com/science/article/pii/S0168169917313017

What is it?	Genomics
What phenotype is it measuring?	Contributes to Multiple traits
Biomarker/Biosensor	Biomarkers
Species	Multiple
The problem	Trying to enhance genetic evaluation systems through the inclusion of genome information

How does it work?	Genotyping, Imputation, SSGBLUP
Measure real time or sent back to lab?	Laboratory process, real time mini sequencers have been commercialised especially for accessing environmental DNA. But the data still needs to be incorporated into an evaluation.
$$$	Varies depending on genotype density, service provider and genotyping technology
Being used/still in trials?	Commercially available
Reference	• Predicting phenotypes from genotypes using Deep Learning https://www.biorxiv.org/content/biorxiv/early/2017/12/31/241414.full.pdf • Low density panels https://www.researchgate.net/publication/320600909_Using_a_very_low-density_SNP_panel_for_genomic_selection_in_a_breeding_program_for_sheep • Imputation https://gsejournal.biomedcentral.com/articles/10.1186/s12711-016-0244-7

What is it?	Mastitis Detection using neural networks
What phenotype is it measuring?	Mastitis
Biomarker/Biosensor	Biosensor in automatic milking systems
Species	Dairy Cattle
The problem	Accurately detecting the stage of progression of mastitis in a milking quarter
How does it work?	Neural network analysis of milking data
Measure real time or sent back to lab?	Not real time
$$$	N/A.
Being used/still in trials?	Trial
Reference	• Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks https://www.cambridge.org/core/journals/journal-of-dairy-research/article/detection-of-mastitis-and-its-stage-of-progression-by-automatic-milking-systems-using-artificial-neural-networks/892676CAA3E691509A2BFE0AB228063A

What is it?	Microbiome
What phenotype is it measuring?	Potentially feed efficiency, animal health, performance, and productivity (e.g. milk lactate and milk yield)
Biomarker/Biosensor	Bio marker
Species	Currently dairy cattle
The problem	Limitations in current bioinformatics-based approaches to identifying patterns of gene covariation in the microbiome to predict animal phenotypes
How does it work?	Novel data mining and machine learning approaches are critical for future investigations on the microbiome to improve animal production and phenotype prediction in animal agriculture.
Measure real time or sent back to lab?	A laboratory process that utilises next-generation sequencing methods.
$$$	N/A.
Being used/still in trials?	Research
Reference	• Precision Animal Agriculture: https://academic.oup.com/jas/article/96/4/1540/4828311

What is it?	Use of GPS tracking collars and accelerometers for rangeland livestock production research
What phenotype is it measuring?	Grazing behaviour in extensive situations; genetic selection for grazing distribution
Biomarker/Biosensor	Biosensor (GPS)
Species	Cattle
The problem	The use of feed supplement placement in areas far from water and on steep slopes measured with GPS tracking and corresponding impacts on distribution patterns
How does it work?	GPS
Measure real time or sent back to lab?	Potentially real time
$$$	N/A.
Being used/still in trials?	Commercial application
Reference	• Use of GPS tracking collars and accelerometers for rangeland livestock production research https://academic.oup.com/tas/article/2/1/81/4824982

What is it?	Image analysis
What phenotype is it measuring?	Body weight
Biomarker/Biosensor	Biosensor
Species	Cattle
The problem	Livestock body weight is critical for nutritional and breeding management because it is a direct indicator of animal growth, health status, and readiness for market. Therefore, accurate body weight estimation is essential to research and genetic evaluation. Potentially also mitigates health and safety issues of livestock and traditional crush equipment.
How does it work?	Uses machine vision technology and image analysis to predict.
Measure real time or sent back to lab?	Real time or delayed
$$$	
Being used/still in trials?	More research being progressed but has been productised
Reference	• Precision Animal Agriculture: https://academic.oup.com/jas/article/96/4/1540/4828311 • Agroninja: http://agroninja.com/#/beefie

What is it?	Behaviour
What phenotype is it measuring?	Grazing, sleeping, rumination, flight (perhaps if being attacked), suckling, lambing, well-being, illness.
Biomarker/Biosensor	Bio sensor
Species	Sheep, cattle, pigs, etc
The problem	There are phenotypes which livestock industries have identified for some time as being very valuable to collect, but the act of collecting them can disrupt the phenotype itself. For example, knowing the grazing patterns of animals in extensive operations and associating that with weight gain.
How does it work?	Many of the research projects or applications that attempt to monitor animal behaviour and movement, rely on the use of tri-axial accelerometers. These are sensors that capture movement in 3 dimensions. • Smart Sensors: https://www.youtube.com/watch?v=JJFyPHpEfpo • Sheep Wellbeing: https://thoughtexperiment.co.nz/tag/deep-learning/ • Behaviour: https://www.dairycareaction.org/uploads/2/4/2/6/24266896/4.3_radeski.pdf • https://www.sciencedirect.com/science/article/abs/pii/S1871141317301543

	• Actiwatch Mini Biosensor https://www.camntech.com/products/actiwatch-mini/actiwatch-mini-overview • Ear tag deployed accelerometer successfully infers sheep behaviour https://zenodo.org/record/995731#.XEs02VwzY2w • Applications of machine learning in animal behaviour studies https://www.sciencedirect.com/science/article/pii/S0003347216303360 • Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour https://royalsocietypublishing.org/doi/full/10.1098/rsos.171442 • Accelerometer articles http://www.citeulike.org/user/Tony54/tag/accelerometer • Automatic Detection of Suckling Events in Lamb through Accelerometer Data Classification http://sendronet.com/downloads/suckling.pdf
Measure real time or sent back to lab?	Data is captured on the sensor and depending on the rate of capture and the battery life, is analysed later.
$$$	Collars for cattle are about $200, but the cost of sheep applications are unknown
Being used/still in trials?	Commercial applications exist

What is it?	Counting applications
What phenotype is it measuring?	Livestock: Mob information per paddock or across the farm; stock reconciliations for tax auditing; containment or lack of containment; stock rustling. Microbes or pathogens: count of parasite eggs in a faecal sample (WEC); count of spores in a sward sample; count of somatic cells in a milk sample, etc.
Biomarker/Biosensor	Biomarkers in the form of cells, eggs or spores, etc.
Species	Multiple
The problem	The process of counting things like spores and eggs is currently manual and laborious. It also relies on an expert and often laboratory facilities.
How does it work?	Use of images (from satellites or drones or microscopes) and classification technology to automate what is currently a manual task. "*CNN models have the capacity to automatically learn the distinctive features of different object classes from a large number of annotated images*"[2]. • Faecal Egg Counting https://www.nzherald.co.nz/the-country/news/article.cfm?c_id=16&objectid=12155320 • Counting Whales https://www.biorxiv.org/content/biorxiv/early/2018/10/16/443671.full.pdf • Counting livestock: https://diydrones.com/forum/topics/is-there-a-software-to-count-cattle
Measure real time or sent back to lab?	Depends on how quickly the images can be sent back to a server (in the cloud) for counting, or whether the counting algorithm can be onboarded.
$$$	N/A.
Being used/still in trials?	Trials and research

What is it?	Identification; facial recognition
What phenotype is it measuring?	It could substitute the manual reading of an ID or ear tag number, which are integral in genetic evaluation in corresponding with any number of phenotypes. It may not replace physical tags in registered or performance recorded livestock, but it could be an accurate enough management tool in commercial flocks.
Biomarker/Biosensor	Biomarker
Species	Cattle, Sheep, Goats especially those that are patterned.

[2] https://www.biorxiv.org/content/biorxiv/early/2018/10/16/443671.full.pdf

The problem	If identifiers must be sought, read and scribed by humans, then there is inherently error. Performance recording can easily carry 5% of error when tying events with individual animals.
How does it work?	As an event occurs (live weight, fleece weight, EMA, etc) a camera is synchronised to capture an image or a video so that the recording can be associated with an individual. The technology space is called Computer Vision, the science and technology of machines that 'see'. • Holstein Friesian Cattle http://openaccess.thecvf.com/content_ICCV_2017_workshops/papers/w41/Andrew Visual_Localisation_and_ICCV_2017_paper.pdf
Measure real time or sent back to lab?	Still camera or video architecture would influence the return rate of an individual's identifier. The task or event being performed may not rely on immediate feedback, if an image is captured, and an event record is associated to the image file name. e.g. Commercial farmers taking photos of animals in the field, to be detained months later when in the yards.
$$$	N/A.
Being used/still in trials?	Trials and research

5. General Discussion

This project has clearly demonstrated that artificial intelligence approaches do have considerable potential in the wool industry. This project was designed to test machine leaning methods in a wool industry context and determine the potential utility of these methods. This project had to create the entire pipeline of information from data capture through to information analysis. The project was carried out in an iterative way to allow for trial and error. The first challenge with this project was the collection of accurate phenotypic information that could be fed into the Machine Learning environment. One of the limitations of the project was to try and build an image set that was representative of the Merino population. While the range in weights achieved was large, the subjective traits were mainly limited to 1 or 2 dominant scores. The project team went to a range of different flocks with different genetic backgrounds in an attempt to increase the range in subjective scores. Despite this effort, the data set that we ended up with was very much skewed towards the median scores.

This project has clearly demonstrated that with the correct training data set, machine learning models will be very powerful in predicting a range of informative traits from image-based inputs of sheep. It has also shown the importance of collecting highly representative data sets as training sets. In addition, it has shown the importance of accuracy in collecting these data sets. The trials completed that tested the repeatability of human scoring were concerning and showed that this was a significant source of error. If we are to assume that this repeatability is typical among stud breeders and service providers that routinely do these scores, this presents a significant opportunity for AI to outperform human equivalents. By carefully building a database of agreed scores (and associated images), that was balanced across all scores, we are confident we could develop a model that has higher repeatability than human equivalents. Furthermore such assessments could be done at far greater speeds than possible by humans-potentially on the run through a race-since image capture and analysis would be almost instantaneous.

This project has shown that sheep can be successfully identified from an image similar to humans and other species. Interestingly, the inclusion of additional features (whole body from the front angle), added additional accuracy to identification prediction. Within this data we only had sheep at two separate time points. The prediction capacity between time points was very low if images from both times were not included in the training set. However, including images from both times in the training revealed highly accurate identification prediction. This project did not investigate how many different time points (or how many representative photos from each time point) that would be required to develop a system that could routinely identify an animal regardless of time of the year or age. Once the training models accurately identifies the bio-metric features associated from ID it would use those in novel populations for tracking individual animals in real time. Once again substantially large training data sets with appropriate editing and sub training to extract the informative features would be required.

This work was completed under a range of sheep yard conditions (portable yards through to permanent) making it a useful prototype process for the sheep industry. The crate conditions under which the images were collected are probably better than what would be expected for a commercial application of an image and machine learning tool. That is, if you need to put the animals in a crate to take liveweight by cameras, you may as well just weigh them. This project did not investigate how to apply the findings here to a more 'real-life' application where animals were assessed either in the yards or in the paddock. However a object detection algorithms improve, it is now possible to capture images from moving objects and provide sufficient information for Deep learning pipelines, and once the data sets are sufficiently large and robust to be representative of population at large, it is entirely feasible to collect data in commercial settings from moving objects and make assessments instantaneously allowing for drafting and classification "on-the run" . In addition movement may provide additional information on sheep health and welfare attributes.

6. Impact of Wool Industry – Now & in 5 years' time

This pilot project has clearly demonstrated the potential of deep learning approaches to a range of tasks in the wool industry. None of the models developed within this pilot project are ready for instant deployment in the industry. Therefore, the immediate industry impact is small. However, the potential in the technology, that has been demonstrated by the favourable results here, is enormous over a 5-year time frame. The technology has the potential to reduce labour requirements and improve management precision across a range of aspects of the wool industry. There are two main areas where AI can contribute, firstly, in automating, improving the accuracy or increasing the frequency of traditional sheep phenotypes and secondly, in combining a complex information to aid in decision making at the systems level.

Traditional sheep phenotypes

A potential list of phenotypes which may have augmented low cost predictions based on Deep learning are summarised in Table 6.1. The traits fit in with those currently routinely measured or of interest on-farm. In particular, access for low cost performance phenotyping in commercial farms could benefit management decisions and flow back to seed stock sector for increased genetic improvement.

Table 6.1 Potential list of phenotypes that could be 'measured' using Deep Learning applications.

Phenotype	Australia	Deep Learning Potential
Weight	Birth, weaning, 200 days, yearling, 18 months, adult weight	Estimating or predicting the weight of an animal from imagery or video, particularly for commercials farmers. Using these predictions to then better manage feed and pregnant ewe management.
Carcass/Meat	Eye muscle depth and fat depth. Intra-muscular fat, eating quality, carcass yield, carcass weight	New in-plant measurement data captured with sensor technology throughout the slaughtering chain or biomarkers determining the market suitability of carcasses after rapid genetic or chemical analysis.
Wool	Fleece weight, Fibre Diameter, CV of fibre diameter, Staple Strength, Staple Length, Curvature,	Very high definition imagery supporting the determination of quality aspects of wool and its manufacturing path. Potential for *insitu* scanning or in shed in real time.

	Colour, Style, Vegetable matter	
Health	Worm egg count, flystrike, condition score, Dag, footrot,	Biosensor and biomarker devices capturing new behaviour, chemical, immunological, or pathogenic data; Satisfying the consumer demand for welfare and/or intervention free aspects of livestock production. Low cost laboratory diagnostics- on farm diagnostics. Breech strike susceptibility based on image data pre-mulesing.
Reproduction/ Survival	Fertility, fecundity, lamb survival	Unbiased behavioural data combined with very local microclimate information and animal health or endocrine biosensor monitoring. Prediction of accurate gestation length and lambing dates for improved genetic evaluation from pregnancy scan data. Ewe scan to predict cryptic variation in maternal ability and fertility.

Potential scenarios for AI to be used in complex decision-making processes.

AI could also be used in more complex scenarios to support smart-decision making processes and combine complex data from multiple sources. As AI will find applications in mainstream commerce and agricultural applications, it will become the norm rather than the exception. Below are some potential scenario's and their requirements for development and evaluation.

a. Prediction lifetime performance

Merino ewes are usually kept on farm for 4-7 years after being selected as replacements. In modern Merino enterprises, there are normally a lot more ewe lambs weaned than are required to replace the oldest age group. Therefore, a selection process is completed to determine which ewes are retained for breeding. This decision is usually made when sheep are between 6-12 months of age. However, there is very little information available to make this decision and it is largely made on phenotypic appearance of the animal (ie a manual visual classification). We know from the Merino Lifetime Productivity trial that the lifetime value derived from ewes can be hundreds of dollars different between individuals. However there is very little opportunity for a commercial wool producer to determine which ewes are the most profitable to retain with any degree of confidence.

There is significant potential to collect information at an early age which has predictive power for lifetime performance. While, each characteristic may have limited accuracy as a standalone indicator, combined they may prove to be useful as a potential selection criterion. Potential indicators could include DNA based information (individual production potential as well as pedigree and litter /rearing information), early fleece and body weight data measured objectively. If combined with pedigree and behaviour information that are outputs of accelerometer and geolocation data further flock dynamics maybe unravelled. Once selection has occurred, ongoing evaluation of predicted performance against actual performance can be completed as production data is collected annually. This information would enable the predictive capacity of the AI model to improve over time. The prediction system will be ongoing decision making, updating decision as soon as new information becomes available.

Furthermore, there maybe scope to use visual data captured at an early age to be used in training data sets matched against lifetime performance. It is suggested to focus on these features (data) which have been available and measurable. There are some ways to use genetic information for selecting for breading to improve productivity. Once the diverse sources of information have been determined, it is a matter to choose appropriate AI/Machine Learning approach for decision making such as clustering or classification strategies or model/algorithms.

To build such a model would require:

- Detailed priority ranking and evaluation of possible early life indicator traits- will need pilot evaluation
- Very substantial data set of early life indicators /assessment n> 20,000 lambs
- Matched lifetime productivity /performance data
- Sound details of fixed effects- farm, geographical data, seasonal data, sheep type etc.

b. Prediction of optimal resource use on farm

Farmers are faced with many complex decisions which often have to be made in real time and often with a 12-month time horizon. One of the more complex ones is how to use the feed base on farm and match it against production requirements of different stock classes. The main variable the farmer has control over is to change stocking rate and preferentially allocate feed availability and feed quality to different stock classes-ie growing sheep, pregnant and lactating ewes, dry sheep and sheep destined for slaughter. Additionally, decisions need to be made to harvest and store feed for future use or to bring in additional feed-both at substantial cost.

The potential to match on farm feed availability to individual mob requirements is gaining momentum where image analysis (information such as multispectral images and/or LiDAR data of the farm land to decide the quality of on-farm biomass) allows for prediction of feed availability/quality that can be made relatively quickly and easily and matched to stock requirements. Improvements using body size and condition of sheep in mobs is also more accessible potentially through visual image capture and analysis. Climate data for on farm use may also be added to make predictions of short-term changes in the feed base. Individual productivity and mob-based productivity information could be collected annually as drivers for on farm profitability.

Using digital data collected on farm could potentially be used in training data sets to optimise resource allocation and utilisation against potential profitability of each cohort of sheep to combine in optimising on farm profit. The data inputs would be substantial and once again use of AI data processing could allow for on-farm decision support networks to be generated in real time. On farm decisions could be changed and adapted to seasonal and market changes as new information comes to hand during the production cycle.

To build such a model would require:

- Detailed knowledge of methods to capture feed availability and quality easily and at low cost in real time
- Access to remote sensing data or on farm drone image data.
- Detailed knowledge on stock nutritional requirements for each stock class and changes throughout the production cycle
- Matched data capture-digitally – for feed availability and stock requirements
- Very substantial on-farm data for training sets across a wide range of geographical locations and farm types

c. Prediction optimal selling times for wool

It is common for wool sheep to be shorn annually and for wool sold by auction for spot price at or about the time of sale. There are some forward selling options but this only accounts for a small portion of the annual clip. Variability in price is determined through multifactorial variables, the main ones which include wool quality/type, demand and availability in the market. Farmers have some capacity to store wool harvested on farm and sell when prices may be optimal for each wool type.

Historical records of wool sales records are substantial and updated records come into the market on weekly/daily basis. Use of market information may potentially be modelled through AI to make short term predictions in the market allowing farmers to decide optimal times to sell and potentially harvest wool on farm.

To build such a model would require:

- Historical sales/price records for individual wool types and market volume
- Price prediction models based on AI training data and validation throughout the data base (ie exclude subsets of data for independent validation which are not part of the training data)
- Modelling/economic expertise
- Can be done as a desktop project

7. Conclusions & Recommendations

This project has piloted the use of Deep Learning in a wool industry context. The project results clearly demonstrate that there is an enormous potential for the use of AI augmented measurement and identification in the Australian wool industry. While this project has not delivered any 'farm-ready' outcomes it has paved the way for future work to refine and expand the work completed here.

We recommend that a follow up project should be considered that tackles one of the complex combinations of information sources to improve decision making within the wool industry. We have provided three potential scenarios but there are others that could be considered. Furthermore improvements on accuracy and cost of measurement of single traits already measured by industry as identified in Table 6.1 are all amenable to be developed for AI based prediction, allowing for more effective use of industry information. It is recommended that a number of such traits are developed further fur use in AI pipelines to a point where they can be used in industry applications.

8. Bibliography

9. List of abbreviations and/or glossary

10. Appendices

a. Appendix 1 – AWI Communication Report Template (see attached)

Artificial Intelligence in Wool Production
Name of research body
neXtgen Agri Ltd.
Name(s) of any other project co-funding bodies and funding split
N/A
Name(s) of any organisations involved (and specify how they are involved)
University of Sydney – Development of machine learning models for facial and trait recognition
Project start date
15 June 2018
Project end date
29 March 2019
Other key dates (eg key milestones report(s), events , product launch)
Interim progress report on all objectives and appropriate sub activities commenced and on track (28 September 2018)
Main objectives of the project (approx. 150 words)
The overall objective of this project is to provide sheep breeders with tools to use advanced phenotypes and artificial intelligence technologies for prediction of lifetime performance at young ages and management of performance changes in real time. The objective is also to provide advanced, highly predictive phenotypes as inputs for ongoing selection decisions by commercial and seed stock sectors. The immediate objective is to provide a proof of concept that novel phenotyping technologies based on image analysis, bio-marker and bio-sensor technologies combined with deep learning AI technologies will unlock a new horizon for the Australian sheep industry.
Project description (approx. 250 words)
This project has the long-term aim to evaluate the use of advanced phenotypes and artificial intelligence (AI) technologies for the prediction of lifetime performance at young ages, management of performance changes in real time, and provide advanced highly predictive phenotypes as inputs for ongoing selection decisions. These longer-term and highly sought-after aims will require significant investment in phenotype capture and the development of associated AI algorithms. Before this investment can be considered there is a need to determine the probability of success of this work. This project will provide a proof of concept that semi-automated image capture combined with machine learning techniques can be used to determine identification (facial recognition), wrinkle scores, head cover and liveweight in sheep. The project will also investigate other novel phenotyping technologies that can be adopted to commercial sheep farming systems as a conduit for AI technologies. This project will combine these findings to present a strategy plan for ongoing R&D investment in applications of AI technologies for on farm purposes.
Project (and key milestones) outcomes and outputs (approx.. 250 words)
The specific deliverables of this project are: 1. A semi-automated system that has the capacity to take high resolution images and link them to animal EID as suitable for deep learning pipelines 2. An image library of sheep linked to their measured performance 3. Demonstrated capacity for deep learning to extract meaningful information from digital images. This will be completed on 2 traits which are biologically robust and 3 that are more challenging. 4. A review of the scope of bio-sensor and bio-marker technologies and their likely utility for the sheep industry to define phenotypes when linked with deep learning AI technologies 5. A strategy for analytical approaches to integrate data from all sources – on farm production & management data combined with predicted outputs from image capture, bio-sensor and bio-marker data into an integrated phenotype prediction to track long-term outcomes as inputs for ongoing selection as well as phenotype changes in real time for adaptive management. A strategy plan for ongoing R&D investment in applications of AI technologies for on farm applications
Benefits for woolgrowers and wool industry (approx. 150 words)
This project aims to provide a proof of concept and investment strategy for further consideration. The benefits for woolgrowers and the wool industry from this project will flow in subsequent projects to this one. These benefits include the potential to remotely and automatically weigh and identify animals without extensive infrastructure. It

will also lay the foundation for completely new ways to assess traits in sheep without additional time and effort from managers. The concepts initiated in this project will transform the decision-making capacity of the Australian sheep industry in both the tactical management of sheep within a production season as well as the strategic breeding decisions. Once completely developed the concepts initiated here will augment decisions being made by sheep managers on a daily basis.
Is the project related to other AWI-funded or other past/present research
This is a new area of investment for AWI, there is no other research of this kind in sheep that we are aware of.
Potential/real next steps in the research/project
If this proof of concept is successful, the next steps are to develop systems using these techniques that can be deployed in sheep yards or paddocks to assist sheep producers measure and manage their animals. The investment strategy developed as part of this project will clarify the steps that AWI could consider.
Names(s)/roles(s)/contact details of the potential spokesperson/people
Mark Ferguson, project leader, mark@nextgenagri.com, +64 21 496 656
Names(s)/roles(s)/contact details of the key personnel in the project that can be contacted for information for communication purposes (if different from above)
Current images/video assets and potential opportunities
Image capture on farm will provide image and video opportunities. There is also an opportunity to capture images of the process of marking up photos for the face recognition work. No current image or video assets.