Ben-Gurion University of the Negev
Faculty of Engineering Sciences
Department of Industrial Engineering and Management

A machine vision system and models to predict individual feed intake of dairy cows: adapting to different feeds in a cowshed

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE M.Sc. DEGREE

By: May Saar

Supervised by: Prof. Yael Edan and Prof. Ilan Halachmi

September 2021

BEN-GURION UNIVERSITY OF THE NEGEV
THE FACULTY OF ENGINEERING SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT

A machine vision system and models to predict individual feed intake of dairy cows: adapting to different feeds in a cowshed

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE M.Sc. DEGREE

By: May Saar

Supervised by: Prof. Yael Edan and Prof. Ilan Halachmi

Author:……………………………………… Date: 14.9.2021

Supervisor:………………………………….. Date: 14.9.2021

Supervisor:………………………………….. Date: 14.9.2021

Chairman of Graduate Studies Committee………………… Date: 19/9/2021

September 2021

Acknowledgements

This study was supported by grants from the Israeli Chief Scientist of Agriculture, projects 459-451415 (‘‘Kendel”), the Israel Dairy Board, fund number 459-4490 and the TechCare grant agreement number 459-6715Y51. This research was partially supported by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering at Ben-Gurion University of the Negev.

Special thanks to Ran Bezen for his support through this study.

I would like to thank Almog Hitelman for her support, encouragement and for sharing with me her vast knowledge throughout my studies and research.

I gratefully thank Assaf Godo for designing the real-time system and Joseph Lepar for building and installing it.

We thank all members of the Precision Livestock Farming (PLF) Lab and the staff of the Volcani research farm for their assistance in the study.

I wish to express my gratitude and appreciation to my academic advisors, Prof. Yael Edan and Prof. Ilan Halachmi, who provided me guidance, support, and insights over the course of the research.

Abstract

A real-time machine vision system including models that are able to adapt to multiple types of feed was developed to predict individual feed intake of dairy cows. Using a Red-Green-Blue-Depth (RGBD) camera, images of feed piles of two different feed types (lactating cows' feed and heifers' feed) were acquired in a research dairy farm, for a range of feed weights under varied configurations and illuminations. Several models were developed to predict individual feed intake: two Transfer Learning (TL) models based on Convolutional Neural Networks (CNN), one CNN model trained on both feed types, and one Multilayer Perceptron and Convolutional Neural Network (MLP-CNN) model trained on both feed types, along with categorical data. We also implemented a statistical method to compare these four models using a Linear Mixed Model (LMM) and a Generalized Linear Mixed Model (GLMM), revealing that all models performed significantly different. The TL models performed best. The best models trained on both feeds with TL methods, achieved Mean Absolute Errors (MAE) of 0.12 and 0.13 kg per meal with RMSE of 0.18 and 0.17 kg per meal for the two different feeds, when tested on varied data collected manually in a cowshed. Testing the model with actual cows' meals data automatically collected by the system in the cowshed, resulted in an MAE of 0.14 kg per meal and RMSE of 0.19 kg per meal. These results suggest the potential of measuring individual feed intake of dairy cows in a cowshed using RGBD cameras and Deep Learning (DL) models that can be applied and tuned to different types of feed.

Keywords: Individual feed intake, precision livestock farming (PLF), deep learning, RGBD camera, transfer learning.

Publications

Journal papers submitted

Saar, M., Edan, Y., Godo, A., Lepar, J., Parmet, Y., Halachmi, I. A machine vision system to predict individual cow feed intake of different feeds in a cowshed. Accepted November 2021. (Appendix A).

Conference abstract and presentations

Saar, M. A machine vision system to predict individual cow feed intake of different feeds in a cowshed. Best research award in Cattle Science Conference, Jerusalem, Israel, 2022.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Description of the problem

Individual cow feed intake is a significant factor for dairy management; more than 60% of farm expenses are devoted to feed (Bloch et al., 2019; Buza et al., 2014; Halachmi et al., 2016). This major economic impact of feed intake in dairy production has motivated genetic studies on moderately heritable traits of feed intake and nutrient utilization efficiency (Korver, 1988; Vandehaar, 1998; Pryce et al., 2014). Despite this, genetic selection based on individual feed efficiency has not been widely applied, mainly due to the high cost and practical limitations of individual feed intake measurements (Berry et al., 2014; Seymour et al., 2019). Feed conversion effectiveness can be determined using information about a cow's feed intake, and milk production and composition (National Research Council, 2001 and 2007; Volden, 2011). Hence, monitoring feed intake can improve farm management decisions (Shalloo et al., 2004), which is potentially beneficial for farm productivity (Buza et al., 2014; Halachmi et al., 2016; Herd et al., 2003).

Different feed intake measurement systems have been developed, including electronic scales in the feeding stalls to measure the feed consumed by each cow. These weighing systems have been used by several researchers (Bach et al.,2004; Halachmi et al., 1998; Waghorn et al., 2012;). The systems are available primarily for research institutions, rather than for commercial cowsheds, because of their high cost, additional infrastructure, high maintenance, and frequent cleaning requirements, all of which make them impractical for most commercial farms (Stajnko et al., 2010; Wang et al., 2006).

In order to evaluate the mass of the feed, an imaging algorithm can be utilized. Feed mass evaluations based on cameras were performed either by using Structured Light Illumination (SLI) methods (Shelley, 2013), by implementing Light Detection and Ranging (LIDAR) sensing methods (Shelley et al., 2016), or by using 3D Time-of-Flight camera when protected from the sun (due to infrared light contained in sunlight) (Borchersen et al., 2018; Lassen et al., 2018). These methods are impractical on a commercial farm mainly due to their sensitivity to sunlight. Bloch et al. (2019) attempted to overcome the sunlight issue using a photogrammetry method resulting in estimated errors of 0.483 kg for heaps up to 7 kg under laboratory conditions, and 1.32 kg for heaps up to 40 kg in a cowshed. The method requires multiple high-quality Red-Green-Blue (RGB) cameras per feed pile measurement along with colored markers which make it impractical for a cowshed on a commercial farm.

Machine vision (MV) and deep learning methods have made technological advances in recent years (Szegedy et al., 2016). Deep learning and specifically convolutional neural networks (CNN) are a discipline in the machine learning field and can be used for complicated MV tasks such as regression, classification, and detection (Bezen et al., 2020). CNNs are based on non-linear training providing the ability to learn millions of parameters. Thus, they require large amount of diverse data (Ros et al., 2016). Few studies were conducted in the field of feed intake measurements using neural networks (Bezen et al., 2020; Chen et al., 2020; Shen et al., 2021). In a recent study (Bezen et al., 2020), an MV system using a Red-Green-Blue-Depth (RGBD) camera was designed, and a CNN was compared with and without RGB function (RGBD vs. depth, Bezen et al., 2020). The MAE and RMSE obtained per meal were 0.127 and 0.184 kg respectively (Bezen et al., 2020). However, the performance of the model was measured using images of heaps manually spread and not images of actual cows' meals.

1.2 Objectives

The aim of this study was to develop a new MV system for monitoring individual feed intake in an outdoor cowshed: the MV system included a new method and models including a calibration station providing the ability to adapt to multiple feeds in commercial conditions. Several learning models were compared; the best model was validated on actual cows' meals.

2 Literature review

This section reviews relevant literature on precision livestock farming (section 2.1), the importance of feed intake measurements (section 2.2), individual feed intake measurement systems (section 2.3), data collection in a cowshed (section 2.4), computer vision (section 2.5) and deep learning (section 2.6).

2.1 Precision livestock farming

As agriculture production grows, the production units increase, making it difficult for farmer to closely monitor their farm. As a result, animals move from being known and recognized by their farmer to be undifferentiated units (Werkheiser, 2018). Precision livestock farming (PLF) can be described as livestock production management using the technology and principles of process engineering (Wathes et al., 2008). The central purpose of PLF is to develop sensors that measure the most crucial process parameters of the different livestock production systems more accurately, automatically and continuously (Wathes et al., 2008). With that being said, it is important to claim that the goal of these tools is not to replace the farmers but to support them in animal management (Berckmans, 2004). PLF refers to the use of various technologies for managing individual animals in order to improve farm performance and management strategies (Shelley, 2013).

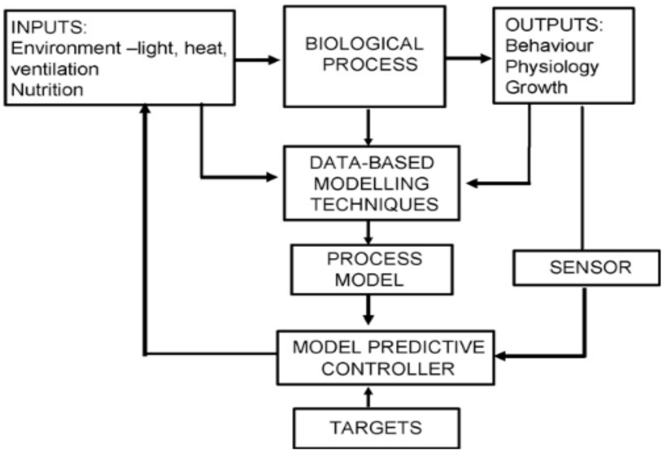


Figure 1: *Schematic overview of the basis of PLF for biological processes (Aerts et al. 2003).*

The increased use of information technology (IT) products in support of livestock management is what originated PLF (Mertens et al., 2011). The main benefit from maintaining PLF systems is ensuring every process in a livestock activity is controlled and optimized (Banhazi and Black, 2009). In order to monitor and control livestock production processes, the PLF approach uses modern monitoring and control theory. Three conditions must be satisfied in order to achieve favorable monitoring and control of such processes (Berckmans, 2004): The first condition is that the animal's variables (e.g., weight, activity, behavior, feed intake, physiological variables etc.) must be measured and analyzed continuously. The second condition is that a reliable prediction on how the animal variables will modify or how the animal will respond to environmental changes, must be available at all times (Berckmans, 2004). The third and last condition is to integrate this prediction together with the on-line measurements in an analyzing algorithm in order to monitor and manage the animals automatically and to achieve on-line monitoring of animal health, welfare, or take control actions (e.g. climate control and feeding strategies) (Berckmans, 2004; Werkheiser, 2018).

Precision dairy farming technologies are relatively new to the dairy farming industry (Shelley, 2013). In the last few years progress has been made towards adoption of precision dairy farming into everyday operations including technologies such as computerized milk yield recordings, biometric identification and health monitoring systems (Shelley, 2013). This thesis is part of further adoption of precision dairy farming technologies providing increased dairy industry production and efficiency.

2.2 Feed intake measurement importance

Individual cow productivity is getting much attention in modern dairy farming (González et al., 2008). Individual's feed intake and feed efficiency have a great economic value for farm management as the feed holds about 40%-60% of the total cost of milk production.

As a result, even the smallest improvement in feed efficiency will have a major economic effect for the farmer (Buza et al., 2014; Connor et al., 2019; Halachmi et al., 2016; Holtenius et al., 2018; Lassen et al., 2018; Shelley, 2013; Waghorn et al., 2012). Combining information of feed intake and milk production and composition can help estimating the individual's cow feed efficiency (Volden, 2011).

In order to keep up with today's expected milk production, dairy farmers must rise feed efficiency by selecting effective cows and managing cows for increased productivity (VandeHaar et al., 2016). The selection requirements are measurements of feed efficiency in order to identify efficient and inefficient cows that may be retained or removed from the herd, respectively (Waghorn et al., 2012).

Farm management decision-making can improve using proper monitoring of feed intake, as the productivity of dairy cows and the profitability of the farm can be determined (Shalloo et al., 2004). This has the potential to increase the farm's productivity (Buza et al., 2014; Halachmi et al., 2016; Herd et al., 2003; Vandehaar, 1998) and assist in farm budgeting, milk yield calculating and feed nutrition rationing (Richter, 2018). Over-Feeding of cows causes a loss of profitability to the dairy farm as money spent on extra feed cannot be returned as instead of being utilized by the cow for milk production, the feed goes to waste (Shelley, 2013.).

Studies that took place in the dairy farm of Beit Dagan, Israel, revealed high variance and diversity between cows up to 30-40% in feed intake for producing an equal amount of milk (Richter, 2018). The economic value of an individual cow to any stage of lactation can be estimated using dry matter intake (DMI) and milk yield (Halachmi et al., 2004). DMI is also essential for nutritional reasons, when feed is allocated individually through self-feeders that are computer controlled. In robot-milking dairies self-feeders are a crucial part of the dairy system (Halachmi et al., 2004).

Additionally, it is important to consider the environmental aspects (Boadi et al., 2004; Knapp et al., 2014): when animal efficiency is improved, methane gas (CH_4) production per unit of milk is reduced. The amount of feed energy related with animal maintenance is around 50% in dairy cows and the remaining 50% is used for production (Boadi et al., 2004). As productivity increases, methane gas emissions rises but methane gas emissions per unit of milk decreases (Boadi et al., 2004; Knapp et al., 2014). Moreover, we can state that if we use fewer cows in order to produce the same amount of milk annually (i.e. increasing feed efficiency), the total methane (CH_4) emission will be decreased (Knapp et al., 2014; Montes et al., 2013).

2.3 Individual feed intake measurement systems

Individual feed intake measurement systems research dealt with system design, development or validation (Bezen et al., 2019) (Table 1). The systems were developed with various sensors and algorithms (Bach et al., 2004; Bezen et al., 2020; Bloch et al., 2019; Bloch et al., 2021; Chapinal et al., 2007; Chizzotti et al., 2015; Halachmi et al., 1998; Shelley, 2013; Shelley et al., 2016;). Besides individual feed intake measurement systems, statistical models for daily DMI prediction were developed. These models are based on data collected using mechanical weighing systems (Gabler, 2002; Halachmi et al., 2016; Halachmi et al., 2004; Holtenius et al., 2018; Jensen et al., 2015; NRC, 2001; Volden, 2011).

Different sensors were used for developing individual feed intake measurement systems:

Electronic scales

This sensor is the most straightforward method to measure feed intake in a cowshed. However, electronic scales are used almost solely in researchers and not in commercial cowsheds (Bezen, 2019)(Halachmi et al., 2019) (Table 1). A scale is located at each feeding station and measures the mass of feed intake of each meal from each bin (Figure 2).

Table 1: Advantages and disadvantages of electronic scales used for measuring feed intake in a cowshed.

Advantages	Disadvantages
The accuracy and reliability of the weight sensor (Bach et al., 2004).	High price and frequent cleaning required (Stajnko et al., 2010; Wang et al., 2006).
Feed type and environmental conditions do not affect the measure ability of the sensor (Bezen, 2019).	The mechanical parts often require a great deal of maintenance (Bezen, 2019).
	Suitable infrastructure is required (Bezen, 2019).

Figure 2: An electronic weight system for individual feed intake in dairy farm, Volcani research institute ((Halachmi et al., 1998)).

Structured light illumination (SLI)

The SLI method includes a camera and a light projector (Figure 3). The surface information are generated from the deformation of the projected pattern (Bezen, 2019; Geng, 2011) with the Z axis as depth.

This method can evaluate the volume and surface of the objects in the scene (Bezen, 2019) but it requires controlled lighting conditions and indoors conditions in order to work properly (Geng, 2011). Previous research (Shelley, 2013) which utilized this system to determine the volume and weight of feed in a bin before and after feeding dairy cows, was tested on 272 feed heaps in a laboratory. According to the results, there was a high difference between predicted values and actual values. There were only 72%

of results within four pounds of the difference between the weight measured on the scale and the image weight calculated.

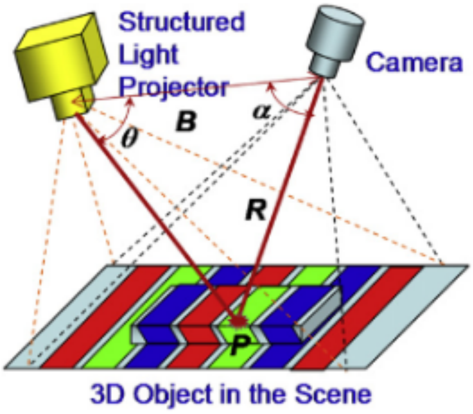


Figure 3: Illustration of structured light (Geng, 2011).

Calibrated stereo cameras

Photogrammetry (Mikhail et al. 2001) is a method of producing a 3D model of an object surface by taking a number of pictures from different angles using Red, Green, Blue (RGB) cameras located at different locations. This is possible because RGB pictures are less sensitive to sunlight. The images are processed using specialized software. Triangulation is used to locate the coordinates of features (tie points) that are common to a number of pictures in the space photographed. An animal feed heap point cloud is thereby generated, from which the feed volume is estimated. The feed volume, along with the feed density, is then used to estimate the feed mass (Bloch et al., 2019) (Figure 4).

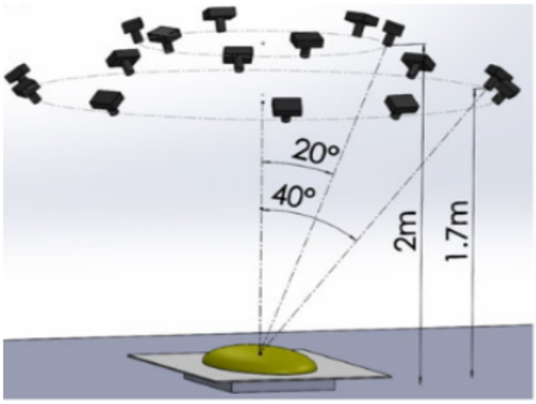


Figure 4: *Camera orientation around the cow feed heap (Bloch et al., 2019).*

The method was tested (Bloch et al., 2019) in a laboratory and in cowshed conditions, with 125 and 60 feed heaps respectively. The estimated error for calculating the mass under laboratory conditions was

0.483 kg for feed heaps up to 7 kg. The SD for the cowshed experiment was 0.44 kg, resulting in a total error of 1.32 kg for heaps up to 40 kg in the cowshed. At least eight images of the feed pile are required to calculate the volume, which is a big disadvantage of this system. This system is impractical due to its need of at least eight cameras per pile of feed and due to the fact that the coloured markers used for the point cloud processing would not be operative in a cowshed on a working farm as dirt may change their colours and tractors may move them from their spot.

RGBD cameras and IR sensors

RGBD cameras provide a couple of images in four channels: Red, Green, Blue and Depth (Cyganek and Siebert, 2011) which provide depth information per-pixel in the RGB image using infrared or near-infrared projector (Figure 5) (Bezen, 2019).

Few RGBD feed intake calculation algorithms and methods have been developed through the years in different conditions: indoor (Shelley et al., 2016), outdoor (Bezen, 2019; Borchersen, Hansen, Borggaard, 2018; Lassen et al., 2018), and open cowshed (Bezen, 2019; Lassen et al., 2018).

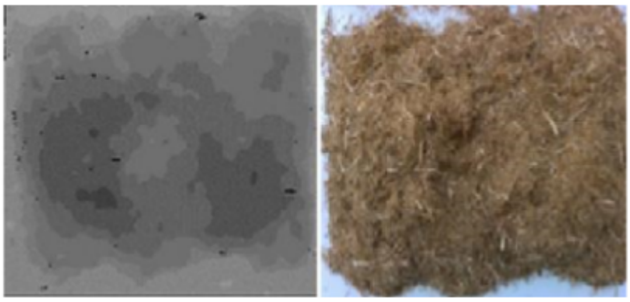


Figure 5: *Images of a feed pile. Left: depth image; right: RGB image (Bezen, 2019).*

The depth sensor based on IR or near-IR projector provides a serious advantage for object surface assessment. Nevertheless, the IR sensor is sensitive to sunlight and influenced by it (Bezen, 2019).

In a previous study (Shelley et al., 2016), A minimum error of 0.2 kg for a single heap's weight was achieved using IR sensors in laboratory conditions.

In order to overcome the sunlight effect on the IR sensor a most recent study for measuring feed intake used a RGBD camera (Bezen, 2019) implementing a deep learning algorithm on images acquired in an open cowshed. The system directly measured feed intake of a single meal which was in the range of 0-8 kg of TMR (Total Mix Ration) and resulted in a mean absolute error of 0.127 kg per single meal. Bezen (2019) applied a deep convolutional neural network (CNN) model to predict feed intake. In order to extract the features from raw images, convolutional layers were applied. The features included four dimensional tensors based on depth and colour images (the tensor contained the results of image subtraction of feed piles images before and after the meal).

2.4 Data collection in cowsheds

Precision dairy farming technologies for animal status monitoring and management continues to raise (Gargiulo et al., 2018). Different sensors and sensing techniques for more precise and accurate information from animals and animal variables have been developed (Borchers and Bewley, 2015; Werkheiser, 2018; Richter, 2018).

One of the sensors systems worth mentioning is the automatic milking system that includes milk yield (MY) and milk components measurements (Richter, 2018) which along with feed consumption measurements can indicate a cow's feed efficiency (Volden, 2011). Positive relationship between feed intake and milk yield and between feed intake and milk components has been noted (Brown et al., 1977).

2.5 Computer vision

Computer vision (CV) is an interdisciplinary scientific field with strong connections to mathematical and computer sciences (Andrew, 2001). CV is concerned with multiple tasks aiming to enable computer systems to automatically see, identify, understand and represent visual world, imitating the same way as human vision does (Feng et al., 2019; Spencer et al., 2019).

CV methods started to be used during the 1960s. The main efforts in applying the methods were invested in extracting shapes information about objects using edges and basic shapes (LG, 1963). With the development of diverse representations of image patterns, CV methods began to deal with more complex perception problems (Spencer et al., 2019).

CV focuses on precisely interpreting scenes in pictures and giving a meaning to what is happening in them. Machine learning in general and deep neural network algorithms in particular are favorable tools to increase the capabilities of CV towards the goal of making the computer draw information from images that a human could have concluded (Guo et al., 2016).

3D computer vision

3D computer vision refers to the consolidation of the color channels of an image (RGB) and the depth dimension of the object presented in the image (RGBD) in order to better understand the environment

(Cyganek and Siebert, 2011). RGBD cameras are sensing systems that are able to capture RGB images together with depth information per-pixel (Henry et al., 2012).

RGBD data has been extensively used for variety of applications in both industry and academic fields (Brook et al., 2012; Johnson, 2018; Martin and Thrun, 2002; Orteu, 2009; Preising and Hisa, 1995; Sels et al., 2019; Seo et al., 2019; Zhang et al., 2018) and in particular for PLF tasks such as measurements of dairy cows' feed heaps (Bezen et al., 2020; Bloch et al., 2019; Shelley et al., 2016; Shelley, 2013).

RGBD sensors use two leading approaches: Time of Flight (ToF) and Structured Light (SL). Recently, RGBD sensors are increasingly used for 3D modeling due to their low costs (Darwish et al., 2017).

Intel RealSense depth camera D435

The camera is suitable for indoor and outdoor environment and includes a dedicated low power vision processor for real time depth sensing (Intel, 2018). Besides the vision processor, the camera includes a RGB sensor, an infrared (IR) projector and both left and right IR cameras (Intel, 2018). The IR projector projects a static IR pattern on the scene in order to boost the texture of low-texture scenes (figure 6). The vision processor gets the captured scene from both right and left imagers and calculates the depth values for each pixel in the image. The depth frame is generated by the depth pixels values (Carfagni et al., 2019). RealSense technology is supported by an open source software developments kit (SDK) which ease the support for software developers (Carfagni et al., 2019).

Figure 6: The static IR dot pattern projected on a wall (Carfagni et al., 2019).

2.6 Deep learning

Deep learning is a subfield of machine learning which aims to learn high level features in data using hierarchical architectures (Guo et al., 2016). This approach has been extensively applied in traditional artificial intelligence domains and in computer vision specifically. There are three main reasons for the popularity of deep learning nowadays: first of all, the growth of computational capabilities of computer

Graphics Processing Units (GPUs), secondly, the relatively low cost of computing hardware, and last- the great advances in machine learning algorithms (Guo et al., 2016; Tsai et al., 2018).

Convolutional Neural Networks (CNN)

CNN is the most common approach used in variety of computer vision tasks and applications. CNNs are built from three main neural layers which play different roles: convolutional layers, pooling layers and fully connected layers (Guo et al., 2016; Zhang et al., 2018), and trained in a robust way which was found highly effective (Guo et al., 2016).

Training a CNN includes two stages: 1) Forward stage which aims to represent the input image with weights and bias in each layer of the network and then output a prediction which is used for calculating the loss function using the ground truth labels. 2) Backward stage which is based on the loss cost and computes the gradients of each parameter of the network using the chain rule. All of the parameters are updated based on the gradients' calculations (Guo et al., 2016).

Convolutional layers

Convolutional layers include multiple feature maps that derive various features from tensors inserted to them, each feature map is responsible to recognize a specific feature from an image. Feature maps are generated by multiple kernels convolving the whole image and the intermediate feature maps (Figure 6). Recognizing features from an image is achieved by using a filter for every area of the image (Guo et al., 2016; Zhang et al., 2018).

The convolution operation has three main advantages: (1) Reduced number of parameters by weight sharing mechanism in the same feature map; (2) Correlations among nearby pixels is being learned; (3) The location of the object isn't important (Guo et al., 2016).

Standard convolution layer has a computational cost of $h_i * w_i * d_i * d_j * k * k$ where $h_i * w_i * d_i$ is the input tensor (where h_i is for height, w_i for width, d_i for depth), k is the kernel size and $h_i * w_i * d_j$ is the output tensor (where h_j is for height, w_j for width, d_j for depth) (Sandler et al., 2018).

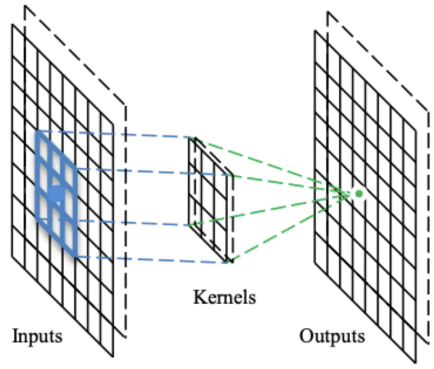


Figure 7: The operation of the convolutional layer (Guo et al., 2016).

Depthwise convolutional layers

Depthwise convolutional layers perform independently over every channel of the input (Kaiser, Gomez, Chollet, 2017).

Depthwise convolution layers empirically work almost as well as standard convolution layers and their computational cost is only: $h_i * w_i * d_i * (d_j + k^2)$ (where h_i is for height, w_i for width, d_i and d_j for depths, k is the kernel size) (Sandler et al. , 2018).

Pooling layers

Pooling layers can be used to reduce the network parameters by reducing the dimensions of feature maps in the network. Those layers summarize small areas in the image into a single output. Pooling layers are defined by their aggregation functions, usually maximum or average (Guo et al., 2016; Zhang et al., 2018).

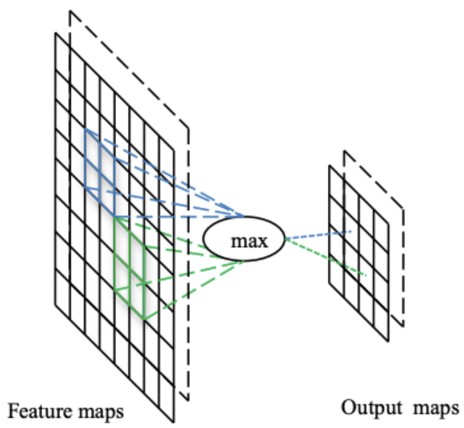


Figure 8:** The operation of the max pooling layer (Guo et al., 2016).**

Fully-connected layers

Fully-connected layers are placed after the last pooling layer in the network converting the 2D features maps into a 1D feature vector with a predefined length. Those layers contain about 90% of the parameters in a CNN (Guo et al., 2016).

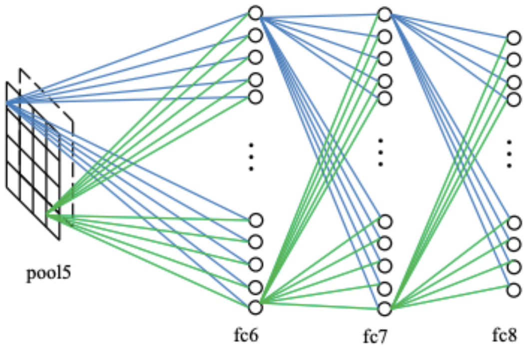


Figure 9: The operation of the fully-connected layer (Guo et al., 2016).

Dropout and DropConnect

In order to improve the generalization ability of a network, during each training case the algorithm randomly leaves out some of the feature detectors. When training with dropout, a randomly selected set of activations are set to zero in each layer. DropConnect is derived from dropout and randomly drops weights and not activations (Guo et al., 2016; Wan et al., 2013).

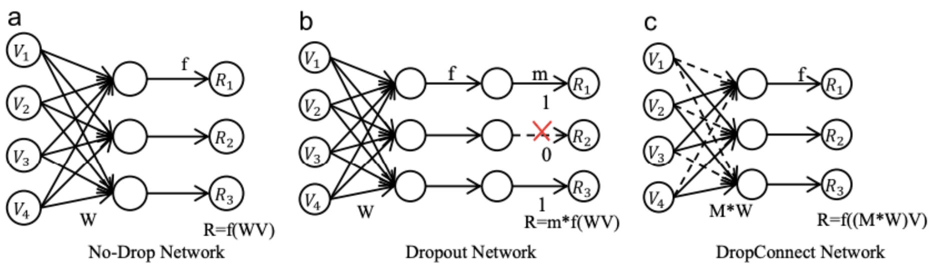


Figure 10: A comparison of (a) No-Drop Network, (b) Dropout Network and (c) DropConnect network (Guo et al., 2016).

Batch Normalization

Batch normalization allows us to be less careful about initialization and use higher learning rates by normalizing the layers input. The normalization is done by scaling and adjusting the activations and shifting the inputs mean and variance to zero and one respectively (Ioffe and Szegedy, 2015).

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Figure 11: Batch normalization transform, applied to activation x over a mini-batch (Ioffe and Szegedy, 2015).

Activation functions

Activation functions aim to transform the activation level of a neuron into an output signal (Sibi et al., 2013). The choice of activation functions in CNN has a significant effect on the task performance (Ramachandran et al., 2017). Thanks to non-linear activation functions, stronger learning of networks can be achieved. Activation functions are executed after the input vector is multiplied with the weight vector (equation 1).

$$x^{[l]} = f(W^{[l]} \cdot x^{[l-1]}) \qquad (1)$$

Where f is an activation function applied to each of its elements.

Different nonlinear activation functions can be used:

Sigmoid: $y = \sigma(x) = \frac{1}{1+e^{-x}}$ (Han and Moraga, 1995) (2)

ReLu: $y = max(0, x)$ (Nair and Hinton, 2010) (3)

Swish: $y = x \cdot \sigma(x) = x \cdot \frac{1}{1+e^{-x}}$ (Ramachandran et al., 2017) (4)

Skip connections

Deeper neural networks are more difficult to train, and they have vanishing gradient problems, which causes trouble reaching convergence. An optional solution for this problem is to skip connections which connect the output of one layer to the input of a prior one (Figure 10). Residual function uses the difference between a mapping applied to the input and the original input instead of learning the direct mapping applied to the input which makes it easier to optimize the residual function (He et al., 2016).

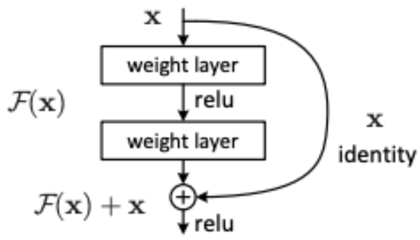


Figure 12: *Residual learning (He et al., 2016).*

EfficientNets models

EfficientNets models were designed based on the idea that carefully balancing network depth, width and resolution can lead to better performance rather than scaling only one the channels (Tan and Le, 2019). EfficientNets' architecture main building block is mobile inverted bottleneck (MBConv). Moreover, it uses skip connections for training deeper networks. Each block is composed of convolutional, depthwise convolutional and batch normalization layers (Sandler et al., 2018; Tan and Le, 2019).

EfficientNets achieved better accuracy and efficiency than previous convolutional networks. EfficientNet-B7 achieved state-of-the-art accuracy on ImageNet while being 8.4x smaller and 6.1x faster on inference than the best existing convolutional networks (Tan and Le, 2019).

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Figure 13: EfficientNet-B0 baseline network - each row represents a stage i with Li layers, input resolution (Hi (height),Wi (width)) and output channels Ci (Tan and Le, 2019).

Training CNN with transfer learning

Transfer learning (TL) is one of the most popular methods in computer vision research (Huh et al., 2016). TL is used when there is enough training data for one learning task, but limited training data for a different task that is similar to the first task but not identical (Wang and Schneider, 2014). By using TL, we try to store the knowledge learned in the form of feature extractor, when solving the first (source) task in the source domain and then apply it to new problems of interest. The knowledge learned in the source domain is applied as a feature extractor in the first layers of the network, those features are global and can be used for different tasks, particularly in tasks where the training data is limited (Yosinski et al., 2014). Training a CNN from scratch on a small training set is likely to lead to overfitting which is the main reason for using TL. Overfitting may be caused due to the large number of parameters to be learned in the CNN which is larger than the number of input images (LeCun et al., 2015). A process noted as fine-tuning is the process of retraining a network with new data. The weights of the network are updated according to the new task.

Multiple inputs and Mixed data

When handling different types of data (i.e., images, numerical values, video, text or categorical value) it is not available to input all the data to the same single network (Yuan et al., 2020). In their research, Yuan et al. (2020) proposed a general hybrid deep neural network architecture to separately handle each input with a suitable network in order to perform feature learning. At the end of the feature learning part, the learned features are concatenated into an ensemble feature which contains information from diverse inputs, and is fed to subsequent neural network to perform target learning (Yuan et al., 2020). As inputs may vary, the neural network used for each data type might be multiple layer perceptron (MLP), CNNs, recurrent neural networks etc. (Yuan et al., 2020).

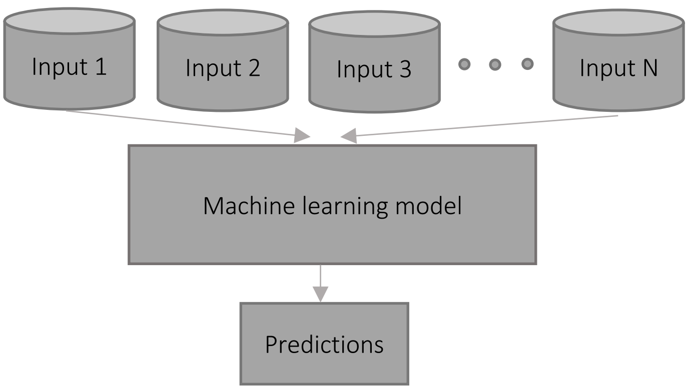


Figure 14: Machine-learning model with multiple inputs.

3 Methods

3.1 Data acquisition and system design

There were two phases of data collection: manual collection for training, testing and statistically comparing several deep learning models, and automatic collection for validation of the best DL model obtained. Both manual and automatic data collections were conducted at a research cowshed at the Volcani Agricultural Research Organization (Israel); A group of 60 Holstein cows participated in the automatic data collection process.

Two feeds were used in the manual collection phase and a single feed type (eaten by lactating cows) was used in the automatic collection phase:

Feed eaten by lactating cows (feed type A) included the following components: wheat silage (37.7%), ground corn grain (18.1%), wheat hay (9.4%), lactose waste (7.8%), gluten feed (7.5%), corn distilled dry grain (4.7%), rapeseed meal (3.7%), wheat grain (2.5%), soybean meal (2.5%), barley grain (0.8%) and vitamins and micro-elements.

Feed eaten by heifers (feed type B) included the following components: straw (33.2%), gluten feed (12.2%), ground corn grain (11.5%), wheat hay (11.1%), wheat grain (8.8%), cotton seed (8.8%), sunflower meal (7%), vitamins and trace minerals (3.9%), lactose waste (3.5%).

3.1.1 Manual data acquisition for training, testing, and comparing models

In order to obtain a varied dataset for training, data of two feed types were collected in September 2020, over a period of 10 days in 10 different sessions (5 sessions for each feed type). The weights of the piles were in the range of 0-45 kg for feed type A, and 0-22 kg for feed type B. To ensure a diverse dataset, images of each pile of the same weight were acquired multiple times, in multiple pile arrangements, illuminations, and at different time periods during the day (as described in Table 2).

The acquisition process was as follows: an off-the-shelf Intel RealSense depth camera (D435, Intel, USA) was installed on an aluminium rod 130 cm above the feed lane. The camera was connected to a computer (equipped with an Intel Core i7-7500U processor) (Figure 15). A Python script (van Rossum, 1995) was developed to operate the camera. An electronic scale was used to weigh the feed piles using a 1 000 kg loadcell with 0.023% precision (SQB, Keli CEE, Poland). In addition, a 17w led bulb was used when collecting data in the dark in order to minimizes the noise in the images and to uniformly scatter the light above the pile. The lighting was empirically derived for this purpose. Each pile was manually weighed and then manually spread on the ground before acquiring an image.

Table 2: Manual data collected at a research cowshed for both feeds

Session	Feed diet	Condition	Number of images
1	A	Daylight – morning hours	185
2		Daylight – afternoon hours	185
3		Daylight – afternoon hours	170
4		Dark using 17w led bulb – night hours	159
5		Direct sun - morning hours	66
6	B	Daylight – morning hours	180
7		Daylight – afternoon hours	170
8		Daylight – morning hours	186
9		Dark using 17w led bulb – night hours	170
10		Direct sun - morning hours	66
			1,537

Figure 15: *Manual data acquisition setup.*

The specific weights of each feed diet were empirically derived as the average of 10 different measurements as follows: two buckets of 3 and 10 liters were filled up with feed and weighed. The weight per liter was calculated by dividing the weight in each bucket by its volume. The average kilogram per liter of feed was calculated for each feed diet, resulting in 0.302 and 0.095 kg per liter for feed diets A and B respectively.

3.1.2 Automatic data acquisition

An automatic system for measuring feed intake was designed, built, and installed for about 4 weeks in March 2021. First, in order to fine tune the trained model with data collected under different conditions (different surface and diverse illumination conditions caused by the different season), 300 images of feed piles of feed type A were manually acquired and labelled. Thereafter, the system operated automatically collecting data of feed type A.

The system included two feeding stations (Figure 16). Each station was equipped with a camera, and both cameras were connected to the same computer. In addition, a 17w LED bulb was positioned next to each camera and illuminated each feed pile during dark hours. Thus, the feed piles were lit uniformly, and shadows were minimized. A station included a weighing palette alongside the camera was used as a calibration station. The weighing palette was attached to an electronic scale built of four loadcells. The signal from the load cells was amplified by a load cell amplifier (HX711, SparkFun Electronics, USA) and read by an Arduino microcontroller (Mega, Arduino, Italy), which streamed the weight readings to the computer.

Figure 16: *Illustration of the real-time system consisting of two feeding stations.*

The weighing palette measured 1x1.5 meters, corresponding to the size of a single feeding area at the dairy farm in which the study was conducted. In both stations, when a cow entered the feeding station, motion was detected with a motion detection algorithm that uses Gaussian blurring and binary thresholding (Bezen et al., 2020). Data from the RGBD camera and the weight of the feed pile were acquired before and after a cow ate from the feed piles. A total of 2 000 entrances to the calibration station were detected, resulting in 2 000 actual meals eaten by cows.

Figure 17: *A calibration station consisting of a camera facing a weighing palette.*

3.2 Datasets for training and fine-tuning

The datasets included tensors (i.e., multidimensional arrays) representing single meals. Each tensor was assembled by subtracting a lower pile weight image from a higher one (for RGB and depth channels, i.e., four channels for each tensor). Moreover, for each tensor, two categorical variables representing the type of feed and the period of time during the day in which the image was taken (morning/afternoon/night) were created.

The tensor creation process included the following steps: (a) Assembling meals in the range of 0-6 kg fresh weight per a meal. (b). Data augmentation: before subtraction of two images, horizontally and vertically flipping augmentations were performed on the minuend image, such that from each subtraction of two images, two different meals emerged. Augmentation was randomly performed on some of the images (about 30%) to increase the dataset while avoiding creation of a too homogenous dataset. (c) Concatenating the subtracted RGB and depth images to create 4-channel tensors. (d) Resizing the tensors (160, 120, 4). (e) Coding the categorical variables as follows: the type of feed variable was binary coded (1 for feed A and 0 for feed B) and the time period variable was one-hot encoded. The one-hot encoding technique is one of the most common ways to transform categorical features into numerical data which is a suitable format used as input for neural networks (Seger, 2018).

Approximately 30 000 RGBD tensors were created from the manually collected data for each feed type. From these tensors, three datasets were created for the training phase: (1) Tensors of feed type A, (2) Tensors of feed type B, (3) Tensors of both feed types (50% tensors from each type). All datasets were distributed approximately uniformly between the different weights in the range of 0-6 kg. Additional 7 000 RGBD tensors were created for model fine-tuning, using the 300 images manually collected in March 2021.

3.3 Analysis

The sensitivity of the models to the train and test sets, obtained from the manually collected data, was evaluated using 5-fold cross-validation. The overall performance of each model (mean absolute error (MAE) and RMSE) was computed by averaging the outcomes of all the five folds. To examine if one of the models had significantly better performance compared to the rest, a linear mixed model (LMM) and a generalized linear mixed model (GLMM) were used (Laird and Ware, 1982): $y - \hat{y} = model + (1|sample\ id)$ and $(y - \hat{y})^2 = model + (1|sample\ id)$ respectively. The LMM examined each model's bias, and the GLMM determined which model resulted in the lowest squared residual. All analyses were performed using the R statistical package at the 0.05 significance level.

4 Learning models

4.1 Overview

Several models were developed trained, tested, and compared. Models (1) and (2) use the CNN architecture detailed in section 4.2. The third model uses the MLP-CNN architecture detailed in section 4.3 (Appendix B).

(1) Combined model: training using tensors of both feed types. The model was trained using a data set of 40 000 tensors where 50% of the data was from each of the feed types. This was done without indicating to the model which feed type was captured in each tensor

(2) Transfer learning: (a) A CNN model was trained using 30 000 tensors of feed type A, and finetuned to adjust this model to predict weights of feed type B, using a dataset of 22 670 tensors of feed type B. (b) A CNN model was trained using 30 000 tensors of feed type B, and finetuned to adjust this model to predict weights of feed type A, using a dataset of 24 000 tensors of feed type A. The sizes of the datasets for fine-tuning were randomly selected.

(3) Multilayer Perceptron and Convolutional Neural Network (MLP-CNN) model using multiple inputs of mixed data: This model was trained using the same dataset as model (1), with two additional categorical variables in the model's input, representing the feed diet in each tensor and the period time of the day the image was taken at.

4.2 CNN

A CNN architecture inspired by the EfficientNet B0 baseline model (Tan and Le, 2019) was developed during this study (Figure 18, Appendix G). The architecture was composed of six inverted residual blocks. Each block included a batch normalization layer, a convolutional layer, and a depth-wise convolutional layer which greatly reduces the number of parameters learned by the model. Each block was followed by another batch normalization layer, an activation layer, and a max pooling layer. The blocks had different numbers of filters and different inputs dimensions. In addition, learning rate annealing technique was used such that the starting learning rate was relatively high, and it was gradually lowered during training (Table 3). Furthermore, to avoid overfitting on the training set, an early stopping method was used to stop the training process when the model's performance stopped improving. in addition, a dropout layer was used to help avoiding overfitting. Finally, the loss function was mean squared error (MSE), and the optimizer was root mean square propagation (RMSprop).

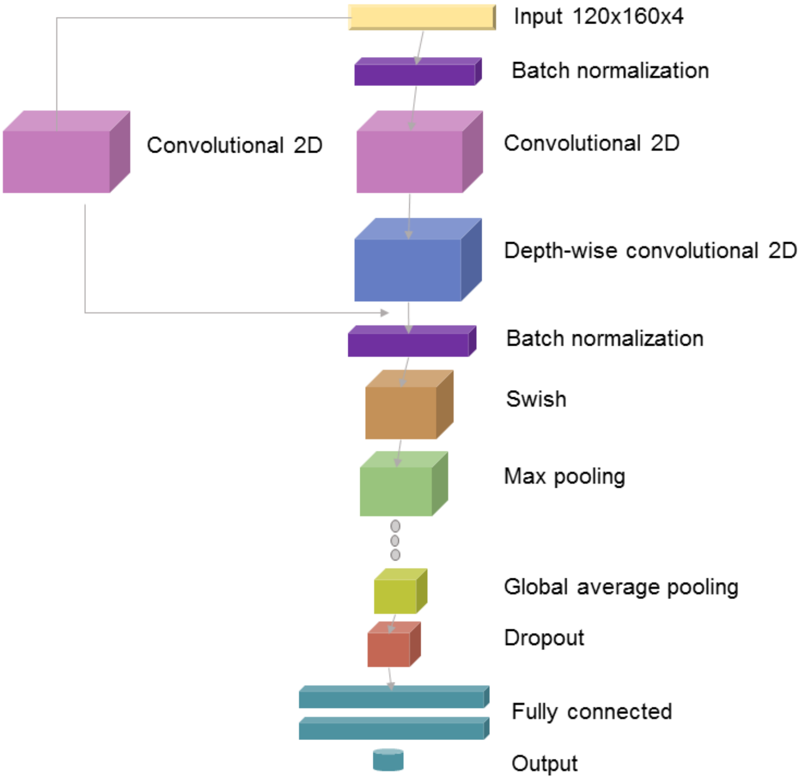


Figure 18: *CNN developed: the network developed was inspired by EfficientNet B0 baseline model.*

Table 3: CNN models' hyperparameters values.

Hyperparameter	Value
Learning Rate (maximum)	0.001
Learning Rate (minimum)	$6.25\ 10^{-5}$
Batch size	16
Dropout rate	0.25
Regularizers	L2 0.01

4.3 MLP-CNN

A MLP-CNN architecture was developed during this study (Figure 19, Appendix G) for mixed-data and multiple inputs (categorical and image data). The MLP network was used to handle the categorical data (i.e., type of feed and time period) and the CNN was used to extract features from the tensors. The MLP network was composed of multiple Fully Connected (FC) layers. The CNN was similar to the one used in models (1) and (2) besides the final output layer which was removed.

Finally, the outputs of both networks were concatenated and inserted as the input of multiple FC layers, in order to make the final prediction of the weight of each meal.The CNN model was trained, evaluated and tested on 40,000 RGBD tensors in which each one of them included 4 channels of data: RGB and depth metrics, each with 120x160 pixels. The MLP model was trained, evaluated and tested on two categorical variables which represent the feed diet captured in each image and the period time of the day the images were taken at.

The outputs of the models were continuous values (feed intake per meal in kg). The dataset of each model was split into training and test sets. Twenty percent of the data were randomly selected and used to test the model, and the remaining 80% of the data were used for training. All the models were trained on NVIDIA GeForce GTX 1080ti GPU, Intel Core™ i7-8700, 64-bit six-core 3.2GHz CPU, 32 GB memory running on Microsoft Windows 10 system.

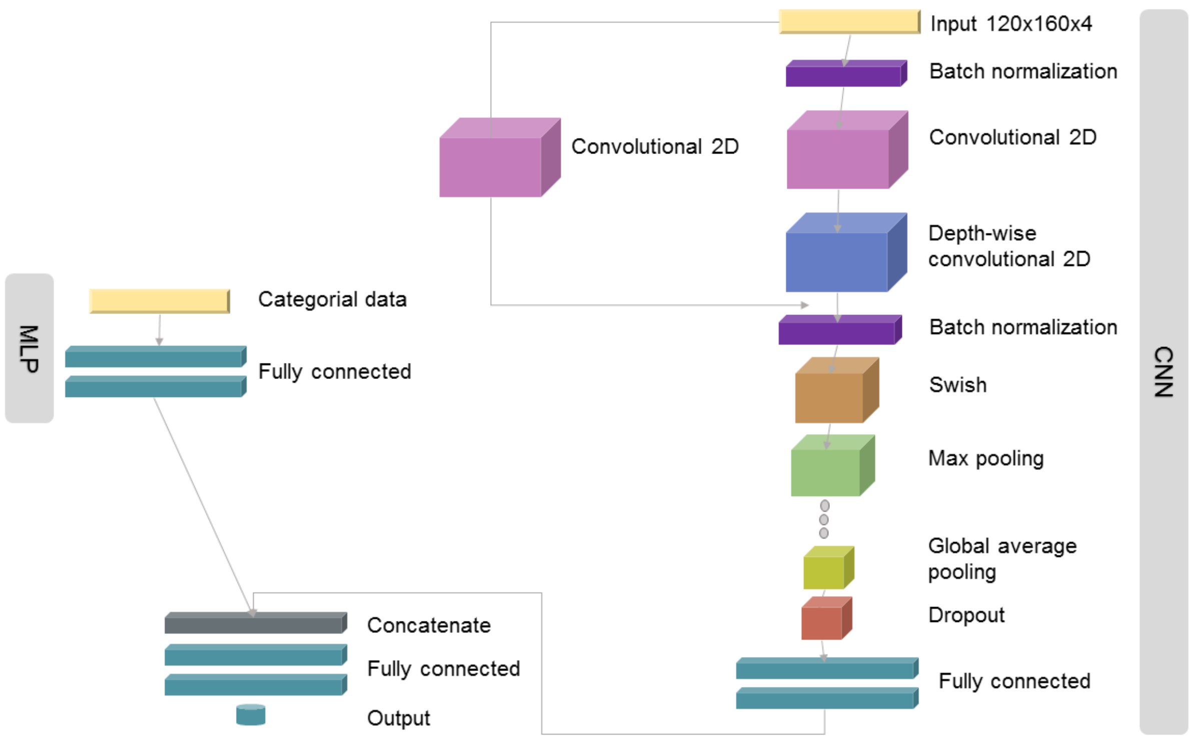


Figure 19: *MLP-CNN developed: combining CNN and MLP together where the MLP addition allows us to insert categorical data to the network.*

5 Results and discussion

5.1 Results

5.1.1 Manual data acquisition

The minimum average error rate using 5-fold cross-validation was 0.12 kg MAE and 0.18 kg RMSE per meal for feed type A, and 0.13 kg MAE and 0.17 kg RMSE per meal for feed type B, for an average meal weight of 2.92 kg (model 2a, Table 4). (Appendix D, Appendix E).

Table 4: Average performance (kg per meal) of the different models on five different datasets created from the manually collected data;

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \ ; \ \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Learning Model	MAE	RMSE	SD
Model 1 - Combined	0.18	0.26	0.26
Model 2a - TL			
Feed diet A	0.12	0.18	0.18
Fine-tune to feed diet B	0.13	0.17	0.17
Model 2b - TL			
Feed diet B	0.1	0.14	0.13
Fine-tune to feed diet A	0.16	0.22	0.21
Model 3 - MLP-CNN	0.17	0.25	0.25

MAE = mean absolute error; TL = transfer learning; MLP-CNN = Multilayer Perceptron and Convolutional Neural Network.
Model 1 – tensors of both feeds; Model 2a,2b – TL from one feed to the other; Model 3 - tensors of both feeds in addition to two categorical variables.

5.1.2 Statistical analysis

Post hoc analysis revealed significant differences between the learning models (Table 5). Model 2a resulted in the lowest bias when predicting the error, and it underestimated the weights of the predicted meals based on the LMM analysis (Table 6). Both models 2a and 2b resulted in the smallest residual in predicting the squared error based on the GLMM analysis (Table 6) (Appendix F).

Table 5: Pairwise comparisons between the models using Tukey method.

Learning Models compared	*P*
Model 1 – Model 2a	<0.0001 ***
Model 1 – Model 2b	<0.0001 ***
Model 1 – Model 3	<0.0001 ***
Model 2a – Model 2b	<0.0001 ***
Model 2a – Model 3	<0.0001 ***
Model 2b – Model 3	<0.0001 ***

Table 6: Estimated marginal means of the linear mixed model and generalized linear mixed model indicate the model with the lowest deviation when predicting the error, and the models with the smallest residuals when predicting the squared errors, respectively.

Model	Estimated marginal means	
	LMM	GLMM
Model 1	0.02	0.04
Model 2a	**-0.004**	*0.02*
Model 2b	-0.02	*0.02*
Model 3	0.006	0.04

* The lowest bias when predicting error is in **bold**, and the models with the smallest residuals when predicting squared errors are in *italics*.

5.1.3 Automatic data acquisition

Using model 2a with the automatic collected data resulted in an MAE of 0.14 kg per meal, and an RMSE of 0.19 kg per meal for an average meal weight of 2.75 kg. Correlation between the actual values and the model predictions can be seen (Figure 20).

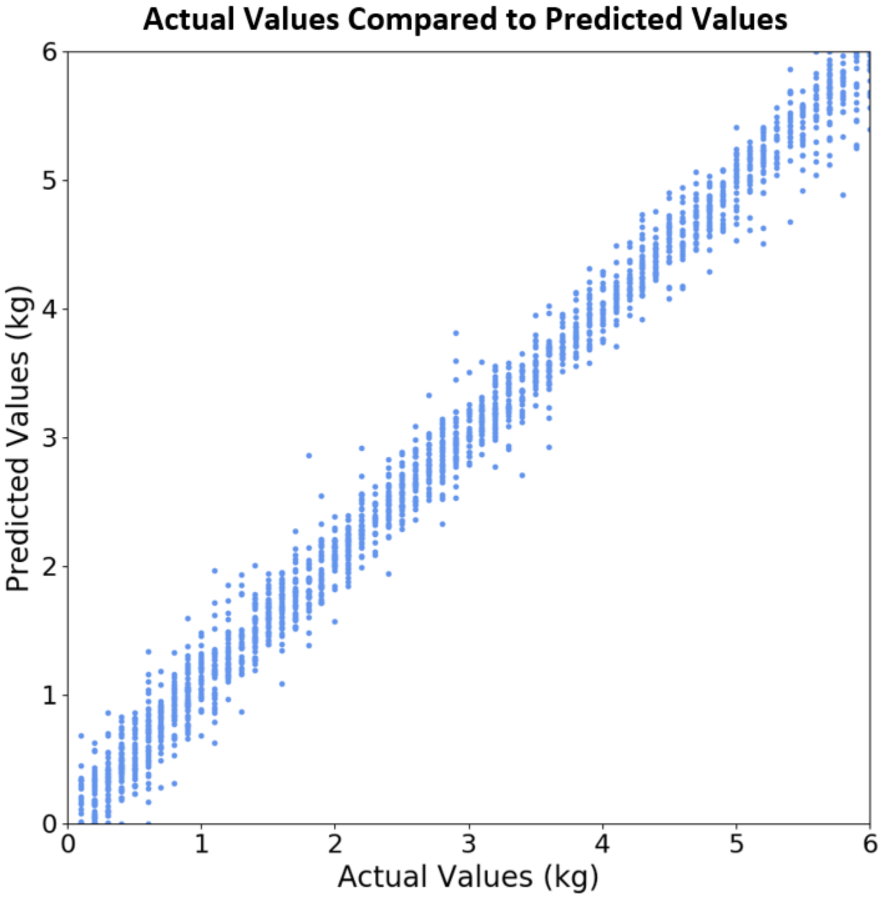


Figure 20: *Predicted values vs. actual values. X and Y axes are the weight in kg.*

5.2 Discussion

The minimum error (MAE of 0.14 and RMSE of 0.19 kg per meal) reported in this study, when tested on data collected automatically by the MV automatic system in an outdoor cowshed, was lower than that achieved in earlier studies (Shelly et al., 2016; Bloch et al., 2019). The minimum error was close to the error reported in a previous study (Bezen et al., 2020). However, Bezen's et al. (2020) system was tested on feed piles manually brought from the farm. In Bezen et al. (2020) there were no actual cows below the camera, only simulated cows' meals that were manually picked from the feed piles. In the current study an automatic system was developed to acquire images of actual cows' meals in the dairy farm; during both day and night; under shade, direct sunlight, and lamplight. In their study, Bloch et al. (2021) reported 120g accuracy and concluded that it is sufficient for cow ranking. The current study obtained a bit higher error (MAE of 0.14 and RMSE of 0.19 kg per meal), but this error was achieved with an MV system and not a mechanical weighing system as used by Bloch et al. (2021).

Training a model with data of feed type A, followed by fine tuning the model with data of feed type B, was superior to the rest of the models. In further research or commercial application, we would advise the same method in applying and utilizing feed intake models for different feeds or different conditions (i.e., different animals, locations). Results suggest that our model's error was higher (MAE=0.23, RMSE=0.27 kg) for data collected in the afternoon comparing to the rest of the day (MAE=0.14, RMSE=0.19 kg for data collected in the morning-noon hours and MAE=0.12, RMSE=0.16 kg for data collected in the evening-night hours). Further research may consider other illumination aspects.

One of the advantages of the MV system is that it requires a single calibration station which enables to dynamically adjust the model to changes in the feed mix. Tuning the model to changes in the feed mix requires less data and less training time than training a new model from scratch. Future study may examine the response time between changing the feed mix and updating the model. Another advantage of this system is that a single computer can run multiple cameras simultaneously, collecting data from multiple feeding stations to estimate feed intake of various of cows.

However, a few caveats may be noted. First, the feed in a commercial cowshed is distributed as a single unseparated long pile along the feeding lane where there is no separation between the feeding stations and therefore, the system is unable to cope with cows that eat from each other's station at the same time. Another disadvantage is the need of a single camera per feeding station. Additional research is necessary to address this issue, to decrease the total system costs (299$ per camera) by utilizing a single camera in multiple feeding stations. A programmatic separation between the stations could handle this issue. Another drawback of the developed MV system is that it does not include a cow detection ability. However, this feature can be added and coupled with the feed intake system to determine how much feed each individual cow consumes.

6 Conclusions and future work

Measuring cows' individual feed intake was conducted in an outdoor cowshed, at various times during the day, using newly developed machine vision and deep learning models. The selected model achieved an MAE of 0.14 kg per meal, and an RMSE of 0.19 kg per meal, for actual cows' meals automatically collected by the automatic system installed at a research dairy farm. These results suggest the potential of measuring individual feed intake of dairy cows in a cowshed using RGBD cameras and deep learning models that can be applied and tuned to different types of feed.

Both models trained and tested using TL methods were revealed to be superior to the other models, according to the LMM and GLMM analysis (Table 5). Additionally, the TL models were more stable (i.e., had a smaller variance) when evaluating their sensitivity using 5-fold cross-validation (Table 3). The MLP-CNN model achieved a lower error than the combined model (Table 3). This demonstrates the advisability of using categorical variables representing the type of feed and the image acquisition time as additional inputs to the model.

Future studies should focus on adding an automatic calibration system and validating the system with a larger number of cameras in a commercial farm. Future work should focus on development of individual cow detection, the use of one camera for few feeding stations and development of a software to handle the division of the feed lane into separate feeding stations. Moreover, the model performance can improve using more inputs to the network such as eating time and exact hour of eating. The MV system and models may be adapted to estimate feed intake of other animals such as beef cattle, small ruminants, and pigs.

7 References

Aerts, J.M., Wathes C.M., Berckmans, D. (2003). Dynamic Data-based Modelling of Heat Production and Growth of BroilerChickens: Development of an Integrated Management System. Biosyst. Eng., 84 (2003), pp. 257-266

Andrew, A. M. (2001). Multiple View Geometry in Computer Vision. In *Kybernetes* (Vol. 30). https://doi.org/10.1016/S0143-8166(01)00145-2

Bach, A., Iglesias, C., & Busto, I. (2004). Technical Note: A Computerized System for Monitoring Feeding Behavior and Individual Feed Intake of Dairy Cattle. *Journal of Dairy Science*, *87*(12), 4207–4209. https://doi.org/10.3168/jds.s0022-0302(04)73565-1

Banhazi, T. M., & Black, J. L. (2009). Precision Livestock Farming: A Suite of Electronic Systems to Ensure the Application of Best Practice Management on Livestock Farms. *Australian Journal of Multi-Disciplinary Engineering*, *7*(1), 1–14. https://doi.org/10.1080/14488388.2009.11464794

Bäuerle, A., & Ropinski, T., (2019). Net2Vis: Transforming Deep Convolutional Networks into Publication-Ready Visualizations. arXiv preprint arXiv:1902.04394

Berckmans, D. (2004). Automatic on-line monitoring of animals by precision livestock farming. *Animal Production in Europe: The Way Forward in a Changing World*.

Berry, D. P., Coffey, M. P., Pryce, J. E., Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J. J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K., Veerkamp, R. F., 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. Journal of Dairy Science 97 (2014), 3894-3905.

Bewley, J. M., & Russell, R. A. (2010). *Reasons for Slow Adoption Rates of Precision Dairy Farming Technologies: Evidence from a Producer Survey*.

Andrew, A. M. (2001). Multiple View Geometry in Computer Vision. In *Kybernetes* (Vol. 30). https://doi.org/10.1016/S0143-8166(01)00145-2

Bach, A., Iglesias, C., & Busto, I. (2004). Technical Note: A Computerized System for Monitoring Feeding Behavior and Individual Feed Intake of Dairy Cattle. *Journal of Dairy Science*, *87*(12), 4207–4209. https://doi.org/10.3168/jds.s0022-0302(04)73565-1

Berckmans, D. (2004). Automatic on-line monitoring of animals by precision livestock farming. *Animal Production in Europe: The Way Forward in a Changing World*.

Bezen, R., Edan, Y., & Halachmi, I. (2020). Computer vision system for measuring individual cow intake using RGB-D camera and deep learning algorithms. Computers and Electronics in Agriculture, 172 (2020), Article 105345. https://doi.org/10.1016/j.compag.2020.105345

Bloch, V., Levit, H., & Halachmi, I. (2019). Assessing the potential of photogrammetry to monitor feed intake of dairy cows. *Journal of Dairy Research*, *86*(1), 34–39. https://doi.org/10.1017/S0022029918000882

Boadi, D., Benchaar, C., Chiquette, J., Masse, D., (2004) Mitigation strategies to reduce enteric methane emissions from dairy cows: update review. *Canadian Journal of Animal Science* **84**, 319–335.

Borchers, M. R., & Bewley, J. M. (2015). An assessment of producer precision dairy farming technology use, prepurchase considerations, and usefulness. *Journal of Dairy Science*, *98*(6), 4198–4205. https://doi.org/10.3168/jds.2014-8963

Borchersen, S., N.W. Hansen, and Borggaard. C. (2018). *System for determining feed consumption of at least one animal* (Vol. 103).

Brook, A., Cristofani, E., Vandewal, M., Matheis, C., & Jonuscheit, J. (2012). 3-D radar image processing methodology for non-destructive testing of aeronautics composite materials and structures. 2012 IEEE Radar Conference, 806–811. IEEE.

Brown, C. a., Chandler, P.T. & Holter, J.B. (1977). Development of Predictive Equations for Milk Yield and Dry Matter Intake in Lactating Cows. Journal of Dairy Science, 60(11), pp.1739–1754. http://linkinghub.elsevier.com/retrieve/pii/S0022030277840988.

Buza, M. H., Holden, L. A., White, R. A., & Ishler, V. A. (2014). Evaluating the effect of ration composition on income over feed cost and milk yield. *Journal of Dairy Science*, *97*(5), 3073–3080. https://doi.org/10.3168/jds.2013-7622

Carfagni, M., Furferi, R., Governi, L., Santarelli, C., Servi, M., Uccheddu, F., & Volpe, Y. (2019). Metrological and critical characterization of the intel D415 stereo depth camera. Sensors (Switzerland), 19(3). https://doi.org/10.3390/s19030489

Chapinal, N., Veira, D. M., Weary, D. M., & Von Keyserlingk, M. A. G. (2007). Technical note: Validation of a system for monitoring individual feeding and drinking behavior and intake in group-housed cattle. *Journal of Dairy Science*, *90*(12), 5732–5736. https://doi.org/10.3168/jds.2007-0331.

Chen, L., Yang, X., Sun, C., Wang, Y., Xu, D., Zhou, C., 2020. Feed intake prediction model for group fish using the MEA-BP neural network in intensive aquaculture. Information Processing in Agriculture 7, 261-271.

Chizzotti, M., Valente. E.E.L., Machado , F.S. , Pereira , L.G.R. (2015).Technical note: Validation of a system for monitoring individual feeding behavior and individual feed intake in dairy cattle. Journal of Dairy Science, 98 , pp. 3438-3442.

Connor, E. E., Hutchison, J. L., Van Tassell, C. P., & Cole, J. B. (2019). Defining the optimal period length and stage of growth or lactation to estimate residual feed intake in dairy cows. Journal of Dairy Science, 102(7), 6131–6143. https://doi.org/10.3168/jds.2018-15407

Cyganek, B., Siebert, J.P. (2011). An Introduction to 3D Computer Vision Techniques and Algorithms. *John Wiley & Sons*.

Darwish, W., Tang, S., Li, W., & Chen, W. (2017). A new calibration method for commercial RGB-D sensors. Sensors (Switzerland), 17(6), 1–15. https://doi.org/10.3390/s17061204

Feng, X., Jiang, Y., Yang, X., Du, M., & Li, X. (2019). Computer vision algorithms and hardware implementations : A survey. *Integration, the VLSI Journal*, *69*(July), 309–320. https://doi.org/10.1016/j.vlsi.2019.07.005

Gabler, M.T., and A. J. H. (2002). *2002. 2001, 1989 NRC requirements for dairy heifers compared*.

Gargiulo, J. I., Eastwood, C. R., Garcia, S. C., & Lyons, N. A. (2018). Dairy farmers with larger herd sizes adopt more precision dairy technologies. *Journal of Dairy Science*, *101*(6), 5466–5473. https://doi.org/10.3168/jds.2017-13324

Geng, J. (2011). Structured-light 3D surface imaging: a tutorial. *Advances in Optics and Photonics*, *3*(2), 128. https://doi.org/10.1364/aop.3.000128

González, L. A., Tolkamp, B. J., Coffey, M. P., Ferret, A., & Kyriazakis, I. (2008). Changes in Feeding Behavior as Possible Indicators for the Automatic Monitoring of Health Disorders in Dairy Cows. *Journal of Dairy Science*, *91*(3), 1017–1028. https://doi.org/10.3168/jds.2007-0530

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Halachmi, I., Ben Meir, Y., Miron, J., & Maltz, E. (2016). Feeding behavior improves prediction of dairy cow voluntary feed intake but cannot serve as the sole indicator. *Animal*, *10*(9), 1501–1506. https://doi.org/10.1017/S1751731115001809

Halachmi, I., Edan, Y., Maltz, E., Peiper, U. M., Moallem, U., & Brukental, I. (1998). A real-time control system for individual dairy cow food intake. In *Computers and Electronics in Agriculture* (Vol. 20).

Halachmi, I., Edan, Y., Moallem, U., & Maltz, E. (2004). Predicting Feed Intake of the Individual Dairy Cow. *Journal of Dairy Science*, *87*(7), 2254–2267. https://doi.org/10.3168/jds.S0022-0302(04)70046-6

Han, J., & Moraga, C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. Natural to Artificial Neural Computation. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 195–201.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Henry, P., Krainin, M., Herbst, E., Ren, X., & Fox, D. (2012). *RGB-D mapping : Using Kinect-style depth cameras for dense 3D modeling of*. https://doi.org/10.1177/0278364911434148

Herd, R., Archer, J., & Arthur, P. (2003). Reducing the cost of beef production through genetic improvement in residual feed intake: Opportunity and challenges to application. *Journal of Animal Science*, *81*(13_suppl_1), E9–E17. https://doi.org/10.2527/2003.8113_suppl_1E9x

Holtenius, K., Andrée O'hara, L., & Karlsson, J. (2018). *The influence of milk yield, body weight and parity on feed intake by dairy cows*.

Huh, M., Agrawal, P., & Efros, A. A. (2016). What makes ImageNet good for transfer learning? ArXiv Preprint ArXiv:1608.08614.

Intel. (2018). RealSense D400 Series (D435). Retrieved from https://www.intel.com/content/www/us/en/support/articles/000026827/emerging-technologies/intel-realsense-technology.html

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ArXiv Preprint ArXiv:1502.03167*.

Jensen, L. M., Nielsen, N. I., Nadeau, E., Markussen, B., & Nørgaard, P. (2015). Evaluation of five models predicting feed intake by dairy cows fed total mixed rations. *Livestock Science*, *176*, 91–103. https://doi.org/10.1016/j.livsci.2015.03.026

Johnson, J. W. (2018). Adapting Mask-RCNN for Automatic Nucleus Segmentation. 2, 608–618. https://doi.org/10.1007/978-3-030-17798-0

Kaiser, L., Aidan, N.G., Francois, C. (2017). Depthwise Separable Convolutions for Neural Machine Translation. arXiv preprint arXiv:1706.03059.

Knapp, J.R., Laur, G.L., Vadas, P.A., Weiss, W.P., Tricarico, J.M. (2014). Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. J Dairy Sci. Elsevier;97: 3231–3261.

Laird, N. M., Ware, J. H., 1982. Random-Effects Models for Longitudinal Data. Biometrics 38, 963—974.

Lassen, J., Rind Thomasen, J., Hansen, R. H., Gunnar Brink Nielsen, G., Olsen, E., René, P., … Borchersen, S. (2018). *Individual measure of feed intake on in-house commercial dairy cattle using 3D camera system*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436.

LG, R. (1963). *Machine perception of three-dimensional solids.pdf*.

Martin, C., & Thrun, S. (2002). Real-time acquisition of compact volumetric 3D maps with mobile robots. Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292), 1, 311–316. IEEE.

Mertens, K., Decuypere, E., De Baerdemaeker, J., De Ketelaere, B. (2011). Statistical control charts as a support tool for the management of livestock production. The Journal of Agricultural Science, 149(03), pp.369–384.

Mikhail, E.M., J.S. Bethel, and J. C. M. (2001). Introduction to modern photogrammetry. John Wiley & Sons, New York.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on Machine Learning (ICML-10), 807–814.

NRC. (2001). *Nutrient Requirements of Dairy Cattle : Seventh Revised Edition*.

Orteu, J.-J. (2009). 3-D computer vision in experimental mechanics. Optics and Lasers in Engineering, 47(3–4), 282–291.

Preising, B., & Hisa, T. C. (1995). Robot performance measurement and calibration using a 3D computer vision system. Robotica, 13(4), 327–337.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). *SWISH: A SELF-GATED ACTIVATION FUNCTION*. Retrieved from http://arxiv.org/abs/1710.05941

Ricther, I. Modeling Individual Feed Intake of a Dairy Cow Based on Feeding Behavior. Ben-Gurion university of the negev, faculty of Engineering sciences department of Industrial Engineering and Management.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A. M., 2016. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 3234–3243.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018a). *MobileNetV2: Inverted Residuals and Linear Bottlenecks*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018b). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Seger, C., 2018. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. Master’s thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

Sels, S., Verspeek, S., Ribbens, B., Bogaerts, B., Vanlanduit, S., Penne, R., & Steenackers, G. (2019). A CAD matching method for 3D thermography of complex objects. Infrared Physics and Technology, 99, 152–157. https://doi.org/10.1016/j.infrared.2019.04.014

Seo, J., Sa, J., Choi, Y., Chung, Y., Park, D., & Kim, H. (2019). A YOLO-based Separation of Touching-Pigs for Smart Pig Farm Applications. International Conference on Advanced Communication Technology, ICACT, 2019-Febru, 395–401. https://doi.org/10.23919/ICACT.2019.8701968

Seymour, D. J., Cánovas, A., Baes, C. F., Chud, T. C. S., Osborne, V. R., Cant, J.P., Brito, L.F., Gredler-Grandl, B., Finocchiaro, R., Veerkamp, R. F., de Haas, Y., Miglior, F., 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. Journal of Dairy Science 102, 7655-7663.

Shalloo, L., Dillon, P., Rath, M., & Wallace, M. (2004). Description and Validation of the Moorepark Dairy System Model. *Journal of Dairy Science*, *87*(6), 1945–1959. https://doi.org/10.3168/jds.S0022-0302(04)73353-6

Shelley, A.N., Lau, D. L., Stone, A. E., & Bewley, J. M. (2016). Short communication: Measuring feed volume and weight by machine vision. *Journal of Dairy Science*, *99*(1), 386–391. https://doi.org/10.3168/jds.2014-8964

Shelley, A.N. (2013). *UKnowledge MONITORING DAIRY COW FEED INTAKE USING MACHINE VISION*. Retrieved from https://uknowledge.uky.edu/ece_etds/24

Shen, W., Li, G., Wei, X., Fu, Q., Zhang, Y., Qu, T., Chen, C., Wang, R., 2021. Assessment of dairy cow feed intake based on BP neural network with polynomial decay learning rate. Information Processing in Agriculture, doi.org/10.1016/j.inpa.2021.04.008, Published online by Science Direct 29 April 2021.

Sibi, P., Allwyn Jones, S., & Siddarth, P. (2013). Analysis of different activation functions using back propagation neural networks. *Journal of Theoretical and Applied Information Technology*, *47*(3), 1344–1348.

Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, *5*(2), 199–222. https://doi.org/10.1016/j.eng.2018.11.030

Stajnko, D., Vindiš, P., Janžekovič, M., & Brus, M. (2010). *Non Invasive Estimating of Cattle Live Weight Using Thermal Imaging*. Retrieved from www.intechopen.com

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826.

Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Retrieved from http://arxiv.org/abs/1905.11946

Tsai, H., Ambrogio, S., Narayanan, P., Shelby, R. M., & Burr, G. W. (2018). Recent progress in analog memory-based accelerators for deep learning. *Journal of Physics D: Applied Physics*, Vol. 51. https://doi.org/10.1088/1361-6463/aac8a5

Vandehaar, M. J. (1998). Efficiency of Nutrient Use and Relationship to Profitability on Dairy Farms. *Journal of Dairy Science*, *81*(1), 272–282. https://doi.org/10.3168/jds.s0022-0302(98)75576-6

VandeHaar, M. J., Armentano, L. E., Weigel, K., Spurlock, D. M., Tempelman, R. J., & Veerkamp, R. (2016). Harnessing the genetics of the modern dairy cow to continue improvements in feed efficiency. *Journal of Dairy Science*, *99*(6), 4941–4954. https://doi.org/10.3168/jds.2015-10352

van Rossum, G., 1995. Python tutorial, Technical Report CS-R9526. CWI, Amsterdam, the Netherlands.

Volden, H. (2011). *NorFor- The Nordic feed evaluation system*.

Waghorn, G. C., Macdonald, K. A., Williams, Y., Davis, S. R., & Spelman, R. J. (2012). Measuring residual feed intake in dairy heifers fed an alfalfa (Medicago sativa) cube diet. *Journal of Dairy Science*, *95*(3), 1462–1471. https://doi.org/10.3168/jds.2011-4670

Wan, L., Zeiler, M., Zhang, S., Lecun, Y., & Fergus, R. (2013). *Regularization of Neural Networks using DropConnect*.

Wang, X., & Schneider, J. (2014). Flexible Transfer Learning under Support and Model Shift. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27 (pp. 1898–1906). Curran Associates, Inc.

Wang, Z. ;, Nkrumah, J. D., Li, ;, & Basarab, J. A. ; (2006). Test duration for growth, feed intake, and feed efficiency in beef cattle using the GrowSafe System1. In *Journal of Animal Science* (Vol. 84).

Werkheiser, I. (2018). Precision Livestock Farming and Farmers' Duties to Livestock. *Journal of Agricultural and Environmental Ethics*, *31*(2), 181–195. https://doi.org/10.1007/s10806-018-9720-0

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27(pp. 3320–3328). Curran Associates, Inc.

Yuan, Z., Jiang, Y., Li, J., & Huang, H. (2020). Hybrid-DNNs: Hybrid deep neural networks for mixed inputs. arXiv:2005.08419. http://arxiv.org/abs/2005.08419

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, Vol. 42, pp. 146–157. https://doi.org/10.1016/j.inffus.2017.10.006

8 Appendices

8.1 Appendix A – Paper submitted

A machine vision system to predict individual cow feed intake of different feeds in a cowshed

Submitted to: Animal

Authors: May Saar, Yael Edan, Assaf Godo, Josheph Lepar, Yisrael Parmet, Ilan Halachmi

A machine vision system to predict individual cow feed intake of different feeds in a cowshed

M. Saar [a,b], Y. Edan [a], A. Godo [b], J. Lepar [b], Y. Parmet [a], I. Halachmi [a,b]

[a] *Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, 1 Ben Gurion Avenue, P.O.B. 653, Be'er Sheva 8410501, Israel*

[b] *Precision Livestock Farming (PLF) Lab., Institute of Agricultural Engineering, Agricultural Research Organization (A.R.O.) – The Volcani Center, 68 Hamaccabim Road, P.O.B 15159 Rishon Lezion 7505101, Israel*

Abstract

Data on individual feed intake of dairy cows, an important variable for farm management, is currently unavailable in commercial dairies. A real-time machine vision system including models that are able to adapt to multiple types of feed was developed to predict individual feed intake of dairy cows. Using a Red-Green-Blue-Depth (**RGBD**) camera, images of feed piles of two different feed types (lactating cows' feed and heifers' feed) were acquired in a research dairy farm, for a range of feed weights under varied configurations and illuminations. Several models were developed to predict individual feed intake: two Transfer Learning (**TL**) models based on Convolutional Neural Networks (**CNN**), one CNN model trained on both feed types, and one Multilayer Perceptron and Convolutional Neural Network model trained on both feed types, along with categorical data. We also implemented a statistical method to compare these four models using a Linear Mixed Model and a Generalized Linear Mixed Model, showing that all models are significantly different. The TL models performed best and were trained on both feeds with TL methods. These models achieved Mean Absolute Errors (**MAE**) of 0.12 and 0.13 kg per meal with RMSE of 0.18 and 0.17 kg per meal for the two different feeds, when tested on varied data collected manually in a cowshed. Testing the model with actual cows' meals data automatically collected by the system in the cowshed, resulted in an MAE of 0.14 kg per meal and RMSE of 0.19 kg per meal. These results suggest the potential of measuring individual feed intake of dairy cows in a cowshed using RGBD cameras and Deep Learning models that can be applied and tuned to different types of feed.

Keywords: individual feed intake, precision livestock farming, deep learning, transfer learning, Red-Green-Blue-Depth camera.

Implications

Data on individual feed intake of dairy cows, an important variable for farm management, is currently unavailable in commercial dairies. We developed a system utilizing a Red-Green-Blue-Depth camera for measuring an individual cow's feed intake in a cowshed, and a Convolutional Neural Network with transfer learning methods, enabling adaptation to different feeds. In an experiment conducted in a research cowshed, the system error was found to be sufficiently low. This system can be utilized to identify individual eating behavior, and efficient and inefficient cows. Adapting to different feeds is an important feature for dairy farms.

Introduction

Individual cow feed intake is a significant factor for dairy management; more than 60% of farm expenses are devoted to feed (Bloch et al., 2019; Buza et al., 2014; Halachmi et al., 2016). This major economic impact of feed intake in dairy production has motivated genetic studies on moderately heritable traits of feed intake and nutrient utilization efficiency (Korver, 1988; Vandehaar, 1998; Berry et al., 2014). Despite this, genetic selection based on individual feed efficiency has not been widely applied, mainly due to the high cost and practical limitations of individual feed intake measurements (Berry et al., 2014; Seymour et al., 2019). Feed conversion effectiveness can be determined using information about a cow's feed intake, and milk production and composition (National Research Council, 2001 and 2007; Volden, 2011). Hence, monitoring feed intake can improve farm management decisions (Shalloo et al., 2004), which is potentially beneficial for farm productivity (Buza et al., 2014; Halachmi et al., 2016; Herd et al., 2003).

Different feed intake measurement systems have been developed, including electronic scales in the feeding stalls to measure the feed consumed by each cow. These weighing systems have been used by several researchers (Bach et al.,2004; Halachmi et al., 1998; Chapinal et al., 2007). Both self-designed weighing systems and commercial systems are available primarily for research institutions, rather than for commercial cowsheds, due to their high cost, additional infrastructure, high maintenance, and frequent cleaning requirements, all of which make them impractical for most commercial farms (Stajnko et al., 2010; Wang et al., 2006).

In order to evaluate the mass of the feed, an image processing algorithm can be utilized. Feed mass evaluations based on cameras were performed by using structured light illumination methods (Shelley, 2013), by implementing light detection and Ranging sensing methods (Shelley et al., 2016), and by using 3D Time-of-Flight camera when protected from the sun (due to infrared light contained in sunlight) (Borchersen et al., 2018;

Lassen et al., 2018). Those methods are impractical on a commercial farm mainly due to their sensitivity to sunlight. Bloch et al. (2019) attempted to overcome the sunlight issue using a photogrammetry method resulting in estimated errors of 0.483 kg for heaps up to 7 kg under laboratory conditions, and 1.32 kg for heaps up to 40 kg in a cowshed. However, this method requires multiple high-quality Red-Green-Blue (**RGB**) cameras per feed pile measurement along with colored markers; hence it is impractical for a cowshed on a commercial farm.

Machine vision (**MV**) and deep learning methods have made technological advances in recent years (Szegedy et al., 2016). Deep learning and specifically convolutional neural networks (**CNN**) are a discipline in the machine learning field and can be used for complicated MV tasks such as classification, detection, and recognition (Bezen et al., 2020). CNNs are based on non-linear, end-to-end training which requires the learning of many parameters. Thus, they require large amount of diverse data (Ros et al., 2016). Few studies have been conducted in the field of feed intake measurements using neural networks (Bezen et al., 2020; Chen et al., 2020; Shen et al., 2021). In a recent study (Bezen et al., 2020), an MV system using a Red-Green-Blue-Depth (**RGBD**) camera was designed, and a CNN was compared with and without RGB function. The MAE and RMSE obtained per meal were 0.127 and 0.184 kg respectively. However, the performance of the model was measured using images of heaps spread manually and not images of actual cows' meals. The aim of this study was to develop a new MV system for monitoring individual feed intake in an outdoor cowshed. Several new learning models were developed, trained, and compared; the best model was validated on actual cows' meals in a real environment.

Material and methods

Data collection

There were two phases of data collection: manual collection for training, testing and statistically comparing several deep learning models, and automatic collection for validation of the best deep learning model obtained. Both manual and automatic data collections were conducted at a research cowshed at the Volcani Agricultural Research Organization (Israel). A group of 60 Holstein cows participated in the automatic data collection process.

Two feeds were used in the manual collection phase and a single feed type (eaten by lactating cows) was used in the automatic collection phase:

Feed eaten by lactating cows (feed type A) included the following components: wheat silage (37.7%), ground corn grain (18.1%), wheat hay (9.4%), lactose waste (7.8%), gluten feed (7.5%), corn distilled dry grain (4.7%), rapeseed meal (3.7%), wheat grain (2.5%), soybean meal (2.5%), barley grain (0.8%) and vitamins and micro-elements.

Feed eaten by heifers (feed type B) included the following components: straw (33.2%), gluten feed (12.2%), ground corn grain (11.5%), wheat hay (11.1%), wheat grain (8.8%), cotton seed (8.8%), sunflower meal (7%), vitamins and trace minerals (3.9%), lactose waste (3.5%).

Manual data collection for training, testing, and comparing models

In order to obtain a varied dataset for training, data of two feed types were collected in September 2020, over a period of 10 days in 10 different sessions (5 sessions for each feed type). The weights of the piles were in the range of 0-45 kg for feed type A, and 0-22 kg for feed type B. To ensure a diverse dataset, images of each pile of the same weight were acquired multiple times, in multiple pile arrangements, illuminations, and at different time periods during the day (as described in Table 1).

The acquisition process was as follows: an off-the-shelf Intel RealSense depth camera (D435, Intel, USA) was installed on an aluminium rod 130 cm above the feed lane. The camera was connected to a computer (equipped with an Intel Core i7-7500U processor). A Python script (van Rossum, 1995) was developed to operate the camera. An electronic scale was used to weigh the feed piles using a 1 000 kg loadcell with 0.023% precision (SQB, Keli CEE, Poland). Each pile was manually weighed and then manually spread on the ground before acquiring an image.

Automatic data collection for validation

An automatic system for measuring feed intake was designed, built, and installed for about 4 weeks in March 2021 in the research cowshed. First, in order to fine-tune the trained model with data collected under different conditions (different surface and diverse illumination conditions caused by the different seasons and times along the day), 300 images of feed piles of feed type A were manually acquired and labelled. Thereafter, the system operated automatically collecting data of feed type A.

The system included two feeding stations (Fig. 1). Each station was equipped with a camera, and both cameras were connected to the same computer. In addition, a 17w LED bulb was positioned next to each camera and illuminated each feed pile during dark hours. Thus, the feed piles were lit uniformly, and shadows were minimized. A station including a weighing palette alongside the camera was used as a calibration station. The weighing palette was attached to an electronic scale built of four loadcells. The signal from the load cells was amplified by a load cell amplifier (HX711, SparkFun Electronics, USA) and read by an Arduino microcontroller (Mega, Arduino, Italy), which streamed the weight readings to the computer. The weighing palette measured 1x1.5 meters, corresponding to the size of a single feeding area at the dairy farm in which the study was conducted. In both stations, when a cow entered the feeding station, motion was detected with a motion detection algorithm that uses Gaussian blurring and binary thresholding (Bezen et al., 2020). Data from the RGBD camera and the weight of the feed pile were acquired before and after a cow ate from the feed piles. A total of 2 000 entrances to the calibration station were detected, resulting in 2 000 actual meals eaten by cows.

Table 1

Manual data collected at a research cowshed for both feeds.

Session	Feed type	Condition	Number of images
1	A	Daylight – morning hours	185
2	A	Daylight – afternoon hours	185
3	A	Daylight – afternoon hours	170
4	A	Dark using 17w led bulb – night hours	159
5	A	Direct sun - morning hours	66
6	B	Daylight – morning hours	180
7	B	Daylight – afternoon hours	170
8	B	Daylight – morning hours	186
9	B	Dark using 17w led bulb – night hours	170
10	B	Direct sun - morning hours	66
			1 537

Abbreviations: A = lactating cows' feed; B = heifers' feed.

Fig. 1. Illustration of the automatic system installed in the cowshed consisting of two feeding stations, cameras, computer, weighing palette, LED lights.

Datasets for training and fine-tuning

The datasets included tensors (i.e., multidimensional arrays) representing single meals. Each tensor was assembled by subtracting a lower pile weight image from a higher one (for RGB and depth channels, i.e., four channels for each tensor). Moreover, for each tensor, two categorical variables representing the type of feed and the period of time during the day in which the image was taken (morning/afternoon/night) were created.

The tensor creation process included the following steps: (a) Assembling meals in the range of 0-6 kg fresh weight per meal (i.e., 'as fed', not DM). (b). Data augmentation: before subtraction of two images, horizontally and vertically flipping augmentations were performed on the original image, such that from each subtraction of two images, two different meals emerged. Augmentation was randomly performed on some of the images (about 30%) to increase the dataset while avoiding creation of a too homogenous dataset. (c) Concatenating the subtracted RGB and depth images to create 4-channels tensors. (d) Resizing the tensors (160, 120, 4). (e) Coding the categorical variables as follows: the type of feed variable was binary coded (1 for feed A and 0 for feed B) and the time period variable was one-hot encoded. The one-hot encoding technique is one of the most common ways to transform categorical features into numerical data which is a suitable format used as input for neural networks (Seger, 2018).

Approximately 30 000 RGBD tensors were created from the manually collected data for each feed type. From these tensors, three datasets were created for the training phase: (1) Tensors of feed type A, (2) Tensors of feed type B, (3) Tensors of both feed types (50% tensors from each type). All datasets were distributed approximately uniformly between the different weights in the range of 0-6 kg. An additional 7 000 RGBD tensors were created for model fine-tuning, using the 300 images manually collected in March 2021.

Developing learning models to adapt to different feed types

The following models were developed, trained, tested, and compared:

(1) Combined model: trained using tensors of both feed types.

The model was trained using a data set of 40 000 tensors where 50% of the data was from each of the feed types. This was done without indicating to the model which feed type was captured in each tensor.

(2) Transfer Learning (**TL**): (a) A CNN model was trained using 30 000 tensors of feed type A, and fine-tuned to adjust this model to predict weights of feed type B, using a dataset of 22 670 tensors of feed type B. (b) A CNN model was trained using 30 000 tensors of feed type B, and fine-tuned to adjust this model to predict weights of feed type A, using a dataset of 24 000 tensors of feed type A.

(3) Multilayer Perceptron and Convolutional Neural Network (**MLP-CNN**) model using multiple inputs of mixed data: this model was trained using the same dataset as model (1), with two additional categorical variables in the model's input, representing the type of feed in each tensor and the time period at which the image was taken.

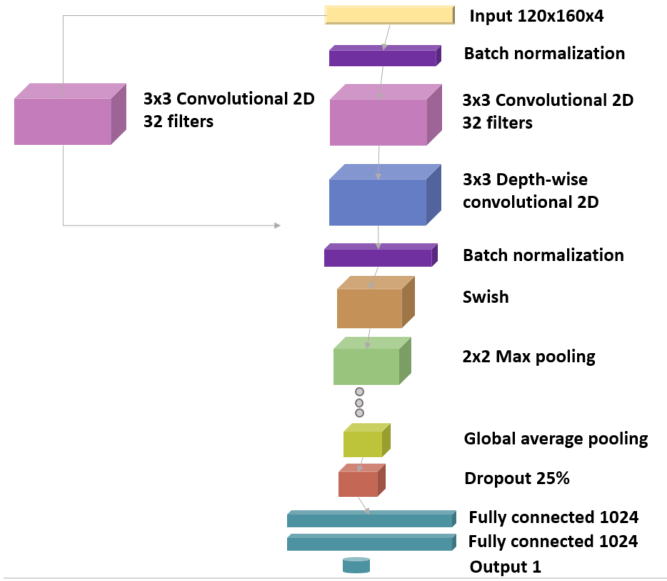
Models (1) and (2a, 2b) used an architecture which was developed during this study (Fig. 2a, Table 2) and inspired by the EfficientNet B0 baseline model (Tan and Le, 2019). The architecture was composed of six inverted residual blocks. Each block included a batch normalization layer, a convolutional layer, and a depth-wise convolutional layer. To avoid overfitting on the training set, an early stopping method was used to stop the training process when the model's performance stopped improving. Finally, the loss function was mean squared error, and the optimizer was root mean square propagation.

Model (3) used an MLP-CNN architecture for mixed data (categorical and images data) which was developed during this study (Fig. 2b). The MLP network was used to handle the categorical data (i.e., type of feed and time period) and the CNN was used to extract features from the tensors. The MLP network was composed of multiple Fully Connected (**FC**) layers. The CNN was similar to the one used in models (1) and (2) besides the final output layer which was removed. Finally, the outputs of both networks were concatenated and inserted as the input of multiple FC layers, in order to make the final prediction of the weight of each meal.

The outputs of the models were continuous values (feed intake per meal in kg). The dataset of each model was split into training and test sets. Twenty percent of the data were randomly selected and used to test the model, and the remaining 80%

of the data were used for training. All the models were trained on NVIDIA GeForce GTX 1080ti GPU, Intel Core™ i7-8700, 64-bit six-core 3.2GHz CPU, 32 GB memory running on Microsoft Windows 10 system.

(a)



(b)

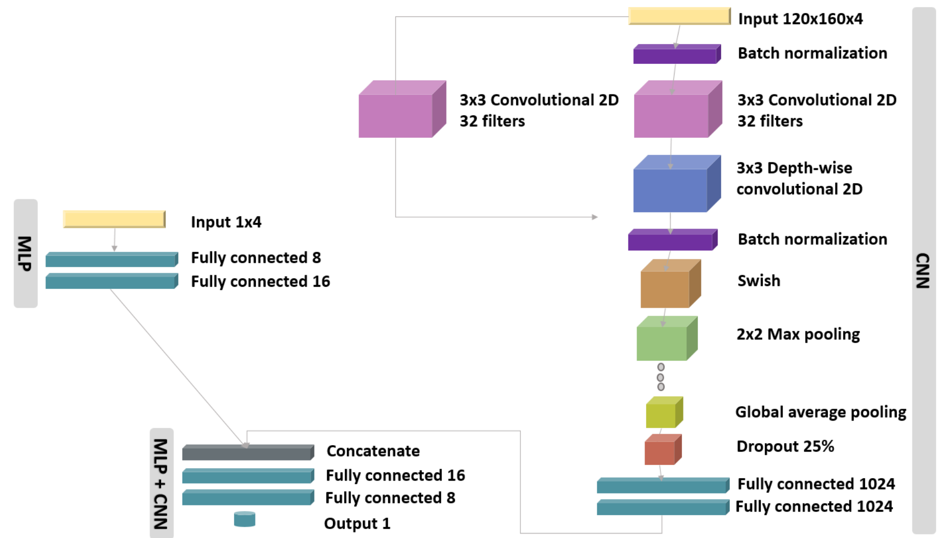


Fig. 2. Development of Convolutional Neural Network (CNN) (a): the network developed was inspired by EfficientNet B0 baseline model. Development of Multilayer Perceptron and Convolutional Neural Network (MLP-CNN) (b): combining CNN and MLP together where the MLP addition allows us to insert categorical data to the network.

Table 2

CNN models' hyperparameters values.

Hyperparameter	**Value**
Learning Rate (maximum)	0.001
Learning Rate (minimum)	6.25 10^{-5}
Batch size	16
Dropout rate	0.25
Regularizer	0.01

Abbreviation: CNN = Convolutional Neural Network.

Analysis

The sensitivity of the models to the training and test sets, obtained from the manually collected data, was evaluated using 5-fold cross-validation. The tensors of each fold were randomly selected after the tensors' creation process. The overall performance of each model (mean absolute error (**MAE**) and RMSE) was computed by averaging the outcomes of all the five folds. To examine if one of the models had significantly better performance compared to the rest, a linear mixed model (**LMM**) and a generalized linear mixed model (**GLMM**) were used (Laird and Ware, 1982): $y - \hat{y} = model + (1|sample\ id)$ and $(y - \hat{y})^2 = model + (1|sample\ id)$ respectively. The LMM examined each model's bias, and the GLMM determined which model resulted in the lowest squared residual. Moreover, *post hoc* pairwise comparisons were conducted using Tukey method to determine whether the models were significantly different. All analyses were performed using the R statistical package at the 0.05 significance level.

Results

Manual data collection

The minimum average error rate using 5-fold cross-validation was 0.12 kg MAE and 0.18 kg RMSE per meal for feed type A, and 0.13 kg MAE and 0.17 kg RMSE per meal for feed type B, for an average meal weight of 2.92 kg (model 2a, Table 3). *Post hoc* analysis revealed significant differences between the learning models (P<0.0001***). Model 2a resulted in the lowest bias when predicting the error, and it underestimated the weights of the predicted meals based on the LMM analysis (Table 4). Both models 2a and 2b resulted in the smallest residual in predicting the squared error based on the GLMM analysis (Table 4).

Automatic data collection

Using model 2a with the automatically collected data resulted in an MAE of 0.14 kg per meal, and an RMSE of 0.19 kg per meal for an average meal weight of 2.75 kg. Correlation between the actual values and the model predictions is presented in Fig. 3.

Table 3

Average performance (kg per meal) of the models utilizing five different datasets created from the manually collected data.

Learning Model	MAE	RMSE	SD
Model 1 - Combined	0.18	0.26	0.26
Model 2a - TL			
Feed type A	0.12	0.18	0.18
Fine-tune to feed type B	0.13	0.17	0.17
Model 2b - TL			
Feed type B	0.1	0.14	0.13
Fine-tune to feed type A	0.16	0.22	0.21
Model 3 - MLP-CNN	0.17	0.25	0.25

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \; ; \; \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Abbreviations: MAE = mean absolute error; TL = transfer learning; MLP-CNN = Multilayer Perceptron and Convolutional Neural Network; A = lactating cows' feed intake; B = heifers' feed intake.

Model 1 – tensors of both feeds; Models 2a,2b – TL from one feed to the other; Model 3 - tensors of both feeds in addition to two categorical variables.

Table 4

LMM and GLMM analysis.

Model (Table 3)	Estimated marginal means	
	LMM	GLMM
Model 1	0.02	0.04
Model 2a	-0.004	0.02
Model 2b	-0.02	0.02
Model 3	0.006	0.04

Abbreviations: LMM = linear mixed model; GLMM = generalized linear mixed model.

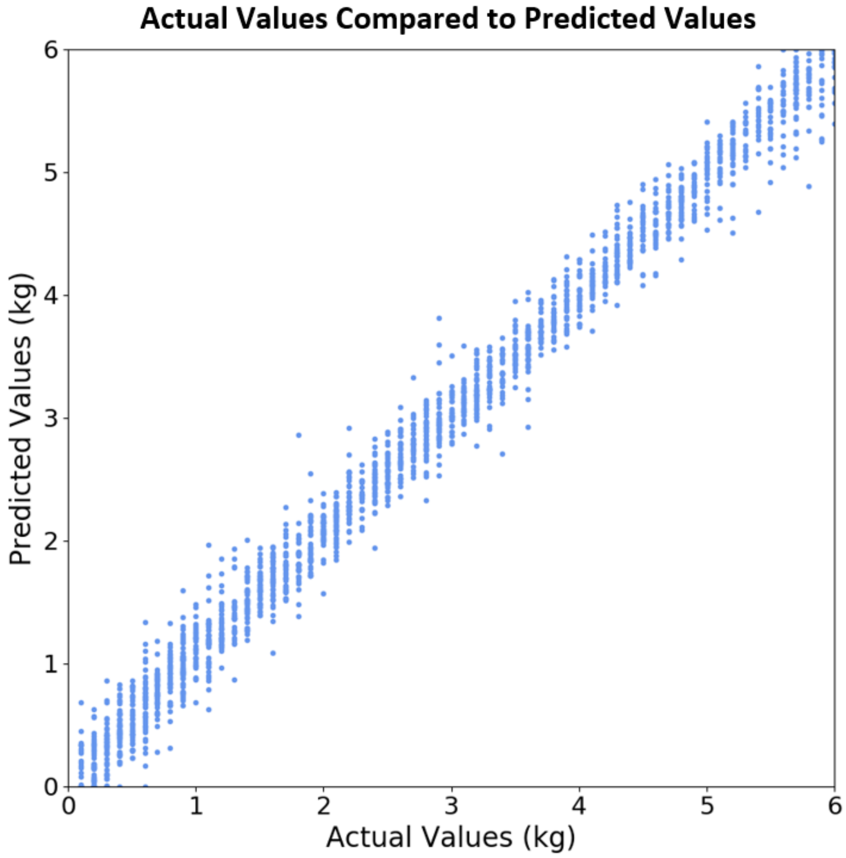


Fig. 3. Lactating cows' feed intake predicted values vs. actual values. X and Y axes are the weight in kg.

Discussion

The minimum error (MAE of 0.14 and RMSE of 0.19 kg per meal) reported in this study, when tested on data collected automatically by the MV system in an outdoor cowshed, was lower than that achieved in earlier studies (Shelly et al., 2016; Bloch et al., 2019). The minimum error was close to the error reported in a previous study (Bezen et al., 2020). However, the system of Bezen et al. (2020) was tested on feed piles that were created manually without using actual piles created by cows. In the current study the images for validation were automatically acquired from actual cows' meals in a dairy farm with a variety of lighting conditions: during both day and night; under shade, direct sunlight, and lamplight. Bloch et al. (2021) reported 120g accuracy and concluded that it was sufficient for cow ranking under commercial conditions. The current study obtained a higher error (MAE of 0.14 and RMSE of 0.19 kg per meal), but this error was achieved with an MV system and not a mechanical weighing system as used by Bloch et al. (2021).

Training a model with data of feed type A, followed by fine-tuning the model with data of feed type B, was superior to the rest of the models

(Table 3, Table 4). In further research or commercial application, we would advise the same method in applying and utilizing feed intake models for different feeds or different conditions (i.e., different illuminations, housing, species). Results suggest that our model's error was higher (MAE=0.23, RMSE=0.27 kg) for data collected in the afternoon comparing to the rest of the day (MAE=0.14, RMSE=0.19 kg for data collected in the morning-noon hours and MAE=0.12, RMSE=0.16 kg for data collected in the evening-night hours). Further research may consider other illumination aspects.

One of the advantages of the developed MV system is that it requires a single calibration station which enables dynamic adjustment of the model to changes in the feed mix. Tuning the model to changes in the feed mix requires less data and less training time than training a new model from scratch. Future study may examine the response time required to update the model once the feed mix is changed. Another advantage of this system is that a single computer can run multiple cameras simultaneously, collecting data from multiple feeding stations to estimate feed intake of various of cows.

However, a few caveats may be noted. First, the feed in a commercial cowshed is distributed as a single unseparated long pile along the feeding lane where there is no separation between the feeding stations and therefore, the system is unable to cope with cows that eat from each other's station at the same time. Another disadvantage is the need of a single camera per feeding station. Additional research is necessary to address this issue, to decrease the total system costs (299$ per camera) by utilizing a single camera in multiple feeding stations. A programmatic separation between the stations could handle this issue. Another drawback of the developed MV system is that it does not include automatic cow detection. However, this feature can be added and coupled with the feed intake system to determine how much feed each individual cow consumes.

Conclusions

Measuring cows' individual feed intake was conducted in an outdoor cowshed for multiple types of feed, at various times during the day, using newly developed machine vision and deep learning models. The selected model achieved an MAE of 0.14 kg per meal, and an RMSE of 0.19 kg per meal, for actual cows' meals automatically collected by the system installed at a research dairy farm. These results suggest the potential of measuring individual feed intake of dairy cows in a cowshed using RGBD cameras and deep learning models that can be applied and tuned to different types of feed.

Both models trained and tested using TL methods were revealed to be superior to the other models, according to the LMM and GLMM analysis (Table 4). Additionally, the TL models were more stable (i.e., had a smaller variance) when evaluating their sensitivity using 5-fold cross-validation (Table 3). The MLP-CNN model achieved a lower error than the combined model (Table 3). This demonstrates the importance of including categorical variables representing the type of feed and the image acquisition time as additional inputs to the model. Future studies should focus on adding an automatic calibration system. The system should be validated with a larger number of cameras in a commercial farm and include individual cow detection, the use of one camera for few feeding stations and development of software to handle the division of the feed lane into separate feeding stations.

Ethics approval

All the procedures in this study were carried out in accordance with the accepted ethical and welfare standards of the Israel Ethics Committee (approval number IL-801/18).

Data and model availability statement

None of the data were deposited in an official repository. The data that support the study findings are confidential.

Author ORCIDs

May Saar: https://orcid.org/0000-0002-2105-4476
Yael Edan: https://orcid.org/0000-0002-7430-8468
Ilan Halachmi: https://orcid.org/0000-0002-2303-1016

Author contributions

May Saar: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration.
Yael Edan: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.
Assaf Godo: Mechanical design, Mechanical engineering, Writing - reviewing.
Joseph Lepar: System building, Writing – reviewing.
Yisrael Parmet: Formal analysis, Writing – reviewing & editing.
Ilan Halachmi: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of interest

None.

Acknowledgements

Special thanks to Ran Bezen for his support throughout this study. We gratefully thank Assaf

Godo, Joseph Lepar, Harel Levit and all other members of the Precision Livestock Farming (PLF) Laboratory and the staff of the Volcani Research Farm for their assistance in the study. The authors thank the editor Prof. Angelika Haeussermann and the anonymous reviewers of Animal for their crucial valuable comments that improved the paper readability and content.

Financial support statement

This study was supported by grants from the Israeli Chief Scientist of Agriculture, projects 459-451415 (‘‘Kendel”), the Israel Dairy Board, fund number 459-4490 and the TechCare grant agreement number 459-6715Y51. This research was partially supported by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering at Ben-Gurion University of the Negev.

References

Bach, A., Iglesias, C., Busto, I., 2004. Technical Note: A computerized system for monitoring feeding behavior and individual feed intake of dairy cattle. Journal of Dairy Science 87, 4207–4209.

Berry, D. P., Coffey, M. P., Pryce, J. E., Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J. J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K., Veerkamp, R. F., 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. Journal of Dairy Science 97, 3894-3905.

Bezen, R., Edan, Y., Halachmi, I., 2020. Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. Computers and Electronics in Agriculture 172, 105345.

Bloch, V., Levit, H., Halachmi, I., 2019. Assessing the potential of photogrammetry to monitor feed intake of dairy cows. Journal of Dairy Research 86, 34–39.

Bloch, V., Levit, H., Halachmi, I., 2021. Design a system for measuring individual cow feed intake in commercial dairies. Animal 15, doi: 10.1016/j.animal.2021.100277.

Borchersen, S., Hansen, N.W., Borggaard, C., 2018. System for determining feed consumption of at least one animal. U.S. Patent No. 9,861,081. Patent and Trademark, Office Washington, DC, USA.

Buza, M. H., Holden, L. A., White, R. A., Ishler, V. A., 2014. Evaluating the effect of ration composition on income over feed cost and milk yield. Journal of Dairy Science 97, 3073–3080.

Chapinal, N., Veira, D.M., Weary, D.M., von Keyserlingk, M.A.G., 2007. Technical Note: Validation of a System for Monitoring Individual Feeding and Drinking Behavior and Intake in Group-Housed Cattle. Journal of Dairy Science 90, 5732– 5736.

Chen, L., Yang, X., Sun, C., Wang, Y., Xu, D., Zhou, C., 2020. Feed intake prediction model for group fish using the MEA-BP neural network in intensive aquaculture. Information Processing in Agriculture 7, 261-271.

Halachmi, I., Ben Meir, Y., Miron, J., Maltz, E., 2016. Feeding behavior improves prediction of dairy cow voluntary feed intake but cannot serve as the sole indicator. Animal 10, 1501–1506.

Halachmi, I., Edan, Y., Maltz, E., Peiper, U. M., Moallem, U., Brukental, I., 1998. A real-time control system for individual dairy cow food intake. Computers and Electronics in Agriculture 20, 131-144.

Herd, R.M., Archer, J.A., Arthur, P.F., 2003. Reducing the cost of beef production through genetic improvement in residual feed intake: Opportunity and challenges to application. Journal of Animal Science 81, 9–17.

Korver, S., 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: A review. Livestock Production Science 20, 1–13.

Laird, N. M., Ware, J. H., 1982. Random-Effects Models for Longitudinal Data. Biometrics 38, 963—974.

Lassen, J., Thomasen, J. R., Hansen, R. H., Nielsen, G. G. B., Olsen, E., Stentebjerg, P. R. B., Hansen, N.W., Borchersen, S., 2018. Individual measure of feed intake on in-house commercial dairy cattle using 3D camera system. Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, 11-16 February 2018, Auckland, New Zealand, pp. 635-640.

National Research Council, 2001. Nutrient Requirements of Dairy Cattle : Seventh Revised Edition. The National Academies Press, Washington, D.C., USA.

National Research Council, 2007. Nutrient Requirements of Small Ruminants: Sheep, Goats, Cervids, and New World Camelids. The National Academies Press, Washington, D.C., USA.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A. M., 2016. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. Proceedings of the 29th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 27-30 June 2016, Las Vegas, NV, USA, pp. 3234–3243.

Seger, C., 2018. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

Seymour, D. J., Cánovas, A., Baes, C. F., Chud, T. C. S., Osborne, V. R., Cant, J.P., Brito, L.F., Gredler-Grandl, B., Finocchiaro, R., Veerkamp, R. F., de Haas, Y., Miglior, F., 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. Journal of Dairy Science 102, 7655-7663.

Shalloo, L., Dillon, P., Rath, M., Wallace, M., 2004. Description and Validation of the Moorepark Dairy System Model. Journal of Dairy Science 87, 1945–1959.

Shelley, A.N., Lau, D. L., Stone, A. E., Bewley, J. M., 2016. Short communication: Measuring feed volume and weight by machine vision. Journal of Dairy Science 99, 386–391.

Shelley, A.N., 2013. Monitoring Dairy Cow Feed Intake Using Machine Vision. Master's thesis, University of Kentucky, Lexington, Kentucky, USA.

Shen, W., Li, G., Wei, X., Fu, Q., Zhang, Y., Qu, T., Chen, C., Wang, R., 2021. Assessment of dairy cow feed intake based on BP neural network with polynomial decay learning rate. Information Processing in Agriculture, doi: 10.1016/j.inpa.2021.04.008, Published online by Science Direct 29 April 2021.

Stajnko, D., Vindiš, P., Marjan, J., Maksimiljan, B., 2010. Non-invasive estimating of cattle live weight using thermal imaging. In: Joo, Meng (Ed.), New Trends in Technologies: Control, Management, Computational Intelligence and Network System. IntechOpen, Rijeka, Croatia, pp. 243–256.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, 27-30 June 2016, Las Vegas, NV, USA, pp. 2818–2826.

Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, 9-15 June 2019, Long Beach, CA, USA, pp. 6105-6114.

Vandehaar, M. J., 1998. Efficiency of Nutrient Use and Relationship to Profitability on Dairy Farms. Journal of Dairy Science 81, 272–282.

van Rossum, G., 1995. Python tutorial, Technical Report CS-R9526. CWI, Amsterdam, the Netherlands.

Volden, H., 2011. NorFor- The Nordic feed evaluation system, 130. Wageningen Academic Publishers, Wageningen, Netherlands.

Wang, Z., Nkrumah, J.D., Li, C., Basarab, J.A., Goonewardene, L.A., Okine, E.K., Crews, D.H., Moore, S.S., 2006. Test duration for growth, feed intake, and feed efficiency in beef cattle using the GrowSafe System. Journal of Animal Science 84, 2289–2298.

8.2 Appendix B – Code

- CNN models:

Final_model_builder.
py

- CNN train and test:

Final_estimator_reg.p
y

- MLP-CNN models:

Final_model_builder_
MLP_CNN.py

- MLP-CNN train and test:

Final_estimator_reg_
MLP_CNN.py

- General functions:

custom.py Final_general_functio
ns.py Final_general_functio
ns.py

- Data preparation:

2types_image_prep_s
ub_npz.py

- Data collection (manually):

Final_weightLable_G
UI.py

- Data collection (Real-time system):

Final_weightLabel_M
otionDetection.py Final_collectData_ca
mera.py

8.3 Appendix C – Data

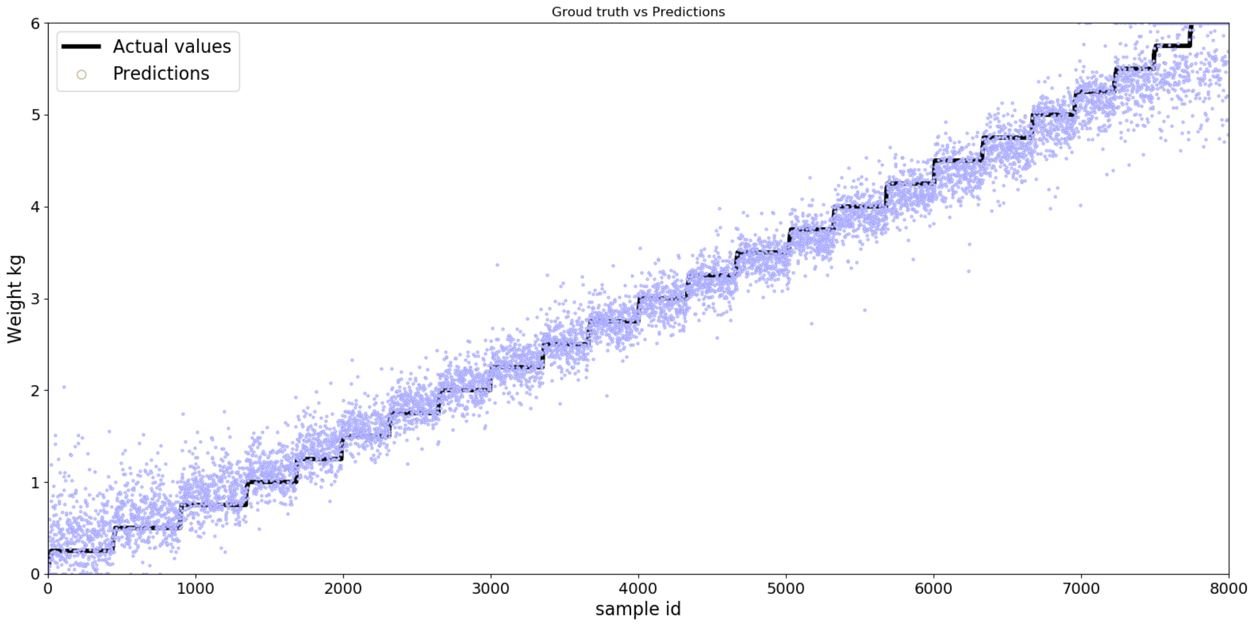
http://res-sharepoint/plf_lab/cattle/_layouts/15/start.aspx#/Shared%20Documents/Forms/AllItems.aspx?RootFolder=%2Fplf%5Flab%2Fcattle%2FShared%20Documents%2FMay%20Saar%20%2D%20Thesis%2FData&FolderCTID=0x012000324B591A09BBB5418F7EF05286F6B85E&View=%7B54FC2795%2DE55A%2D4613%2DB1F7%2D0D93E145203D%7D

If needed, contact Ilan Halachmi for access: halachmi@volcani.agri.gov.il.

8.4 Appendix D – Graphs of ground truth compared to each model's predictions on the test sets (results of fold 1 are shown as an example).

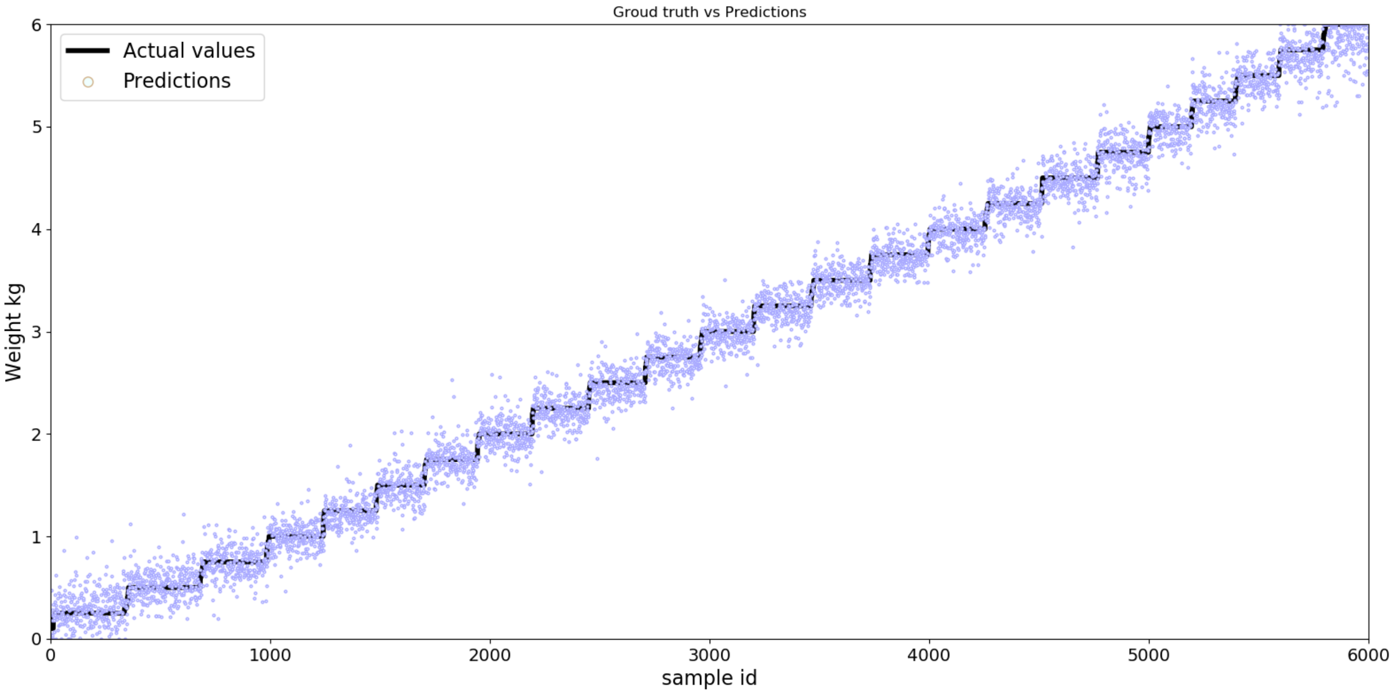
Weight predictions of data collected manually in the cowshed during preliminary collection. Correlation between the actual values and the model predictions can be seen. The observations are arranged in ascending order on the X-axis, according to the actual value; the Y-axis is the weight in kg.

Model 1 – Combined:

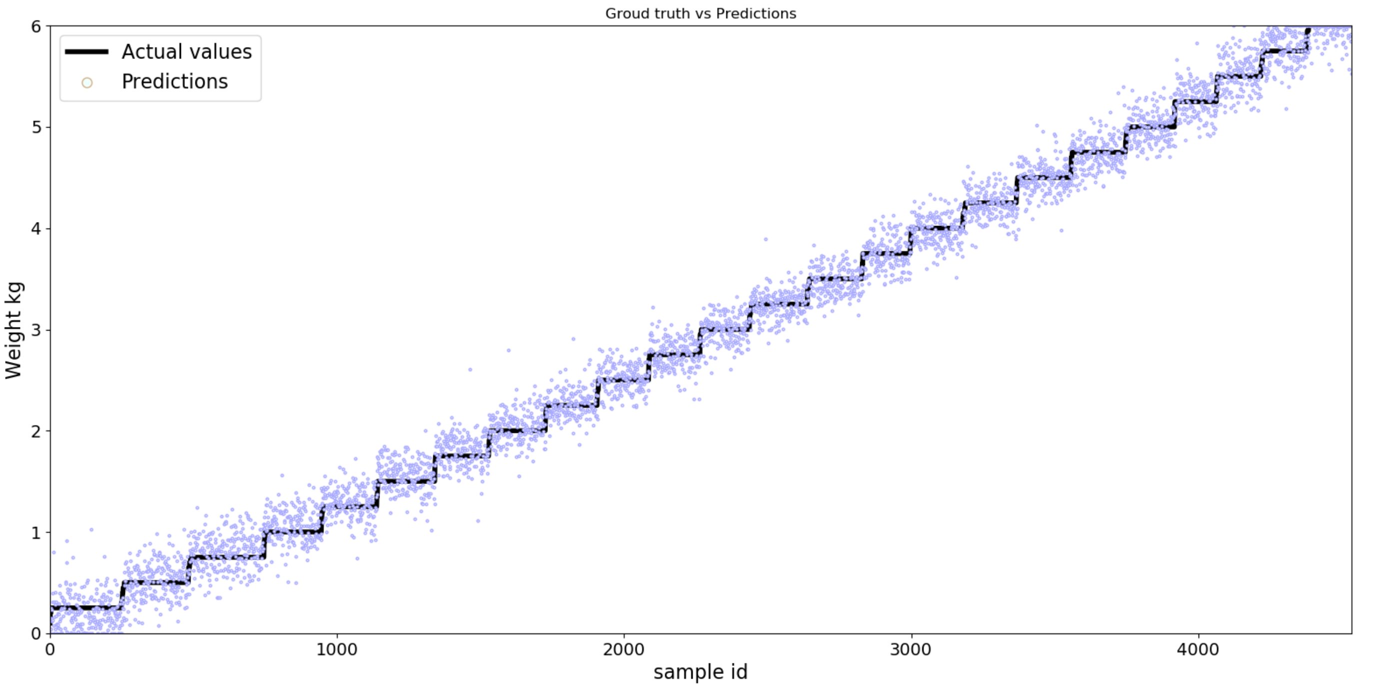


Model 2a – TL: Feed diet A; Fine tune to feed diet B

A:

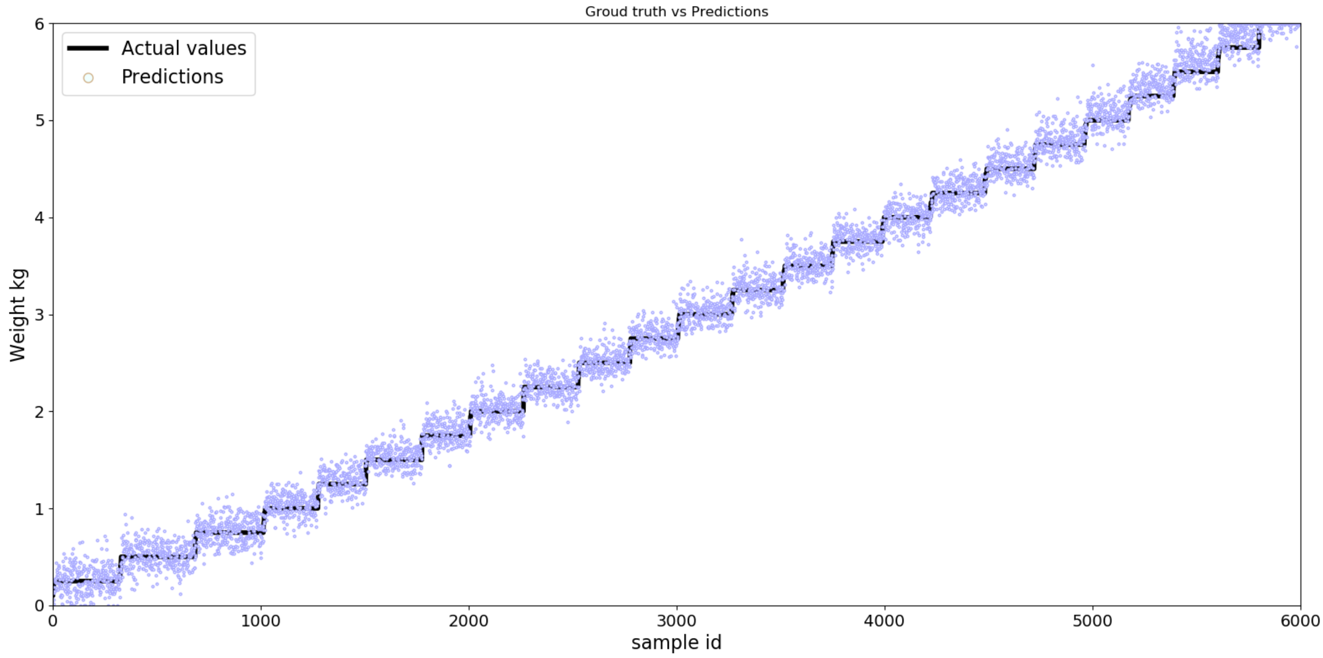


B:

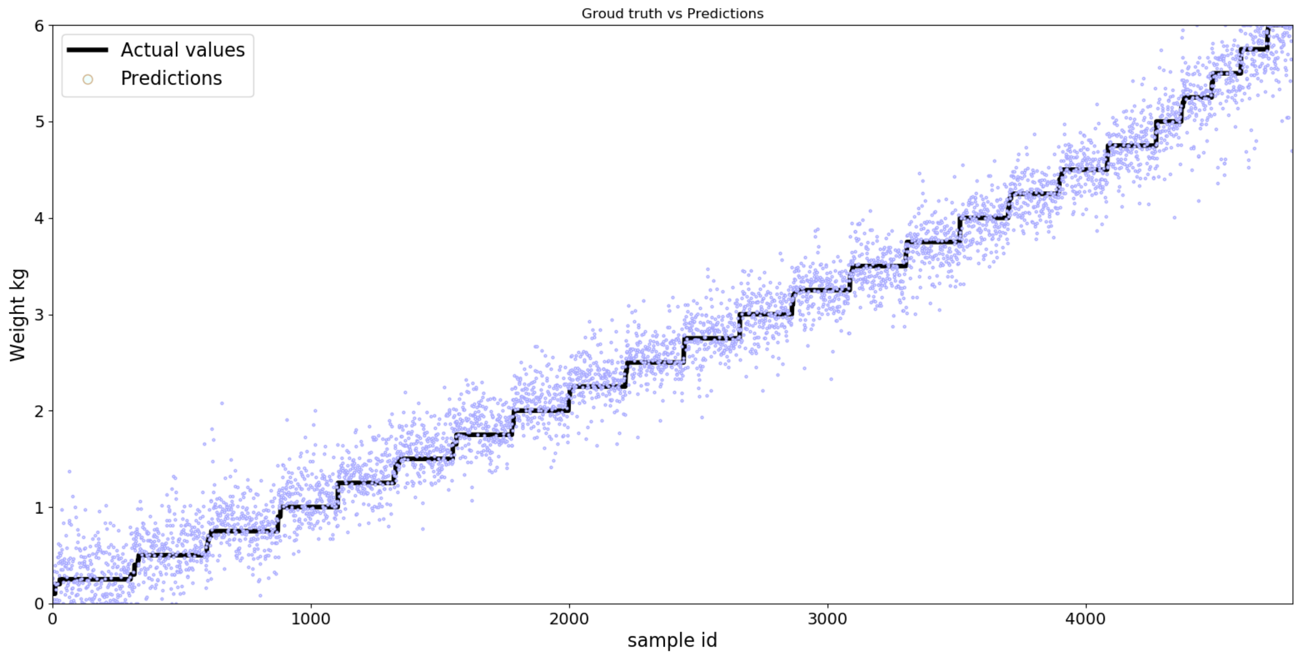


Model 2b – TL: Feed diet B; Fine tune to feed diet A

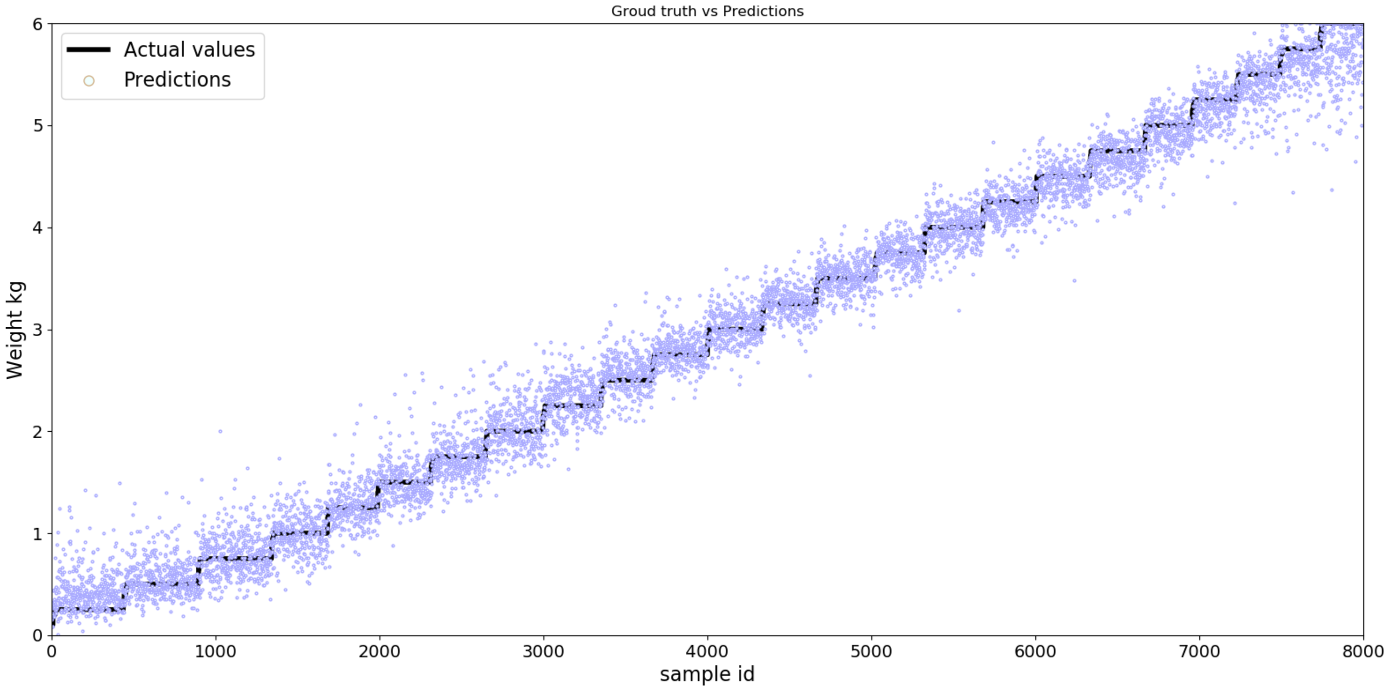
B:



A:



Model 3 – MLP-CNN:



8.5 Appendix E – 5-fold cross validation of each model tested on preliminary data.

	Model 1 – Combined: both feed diets								
	Fold 1	*Fold 2*	*Fold 3*	*Fold 4*	*Fold 5*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
MAE	0.16	0.18	0.2	0.18	0.22	0.188	0.0204	0.16	0.22
RMSE	0.23	0.25	0.28	0.25	0.31	0.264	0.028	0.23	0.31
Maximum Error	1.7	1.5	1.6	2.1	2.3	1.84	0.307	1.5	2.3
Std	0.23	0.25	0.28	0.25	0.3	0.262	0.0248	0.23	0.3

	Model 2a – TL: feed diet A								
	Fold 1	***Fold 2***	***Fold 3***	***Fold 4***	***Fold 5***	***Mean***	***Standard deviation***	***Minimum***	***Maximum***
MAE	0.1	0.11	0.13	0.12	0.17	0.126	0.0241	0.1	0.17
RMSE	0.14	0.18	0.19	0.17	0.24	0.184	0.0326	0.14	0.24
Maximum Error	0.8	1.2	1.16	0.94	1.5	1.12	0.2396	0.8	1.5
Std	0.14	0.18	0.19	0.16	0.23	0.18	0.0303	0.14	0.23

	Model 2a – TL: feed diet B								
	Fold 1	*Fold 2*	*Fold 3*	*Fold 4*	*Fold 5*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
MAE	0.12	0.14	0.13	0.14	0.14	0.134	0.008	0.12	0.14
RMSE	0.15	0.18	0.17	0.18	0.18	0.172	0.0116	0.15	0.18
Maximum Error	0.8	0.75	0.75	0.84	0.8	0.788	0.0343	0.75	0.84
Std	0.15	0.18	0.17	0.18	0.18	0.172	0.0116	0.15	0.18

	Model 2b – TL: feed diet B								
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Standard deviation	Minimum	Maximum
MAE	0.09	0.11	0.13	0.1	0.1	0.106	0.01356	0.09	0.13
RMSE	0.12	0.15	0.17	0.13	0.13	0.14	0.0179	0.12	0.17
Maximum Error	0.5	0.9	0.7	0.5	0.7	0.66	0.1496	0.5	0.9
Std	0.11	0.14	0.17	0.13	0.13	0.136	0.0196	0.11	0.17

	Model 2b – TL: feed diet A								
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Standard deviation	Minimum	Maximum
MAE	0.17	0.17	0.15	0.17	0.16	0.164	0.008	0.15	0.17
RMSE	0.22	0.23	0.21	0.24	0.23	0.226	0.0102	0.21	0.24
Maximum Error	1.3	1.5	1.3	1.6	1.3	1.4	0.1265	1.3	1.6
Std	0.21	0.22	0.2	0.23	0.22	0.216	0.0102	0.2	0.23

	Model 3 – MLP-CNN: both diet types								
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Standard deviation	Minimum	Maximum
MAE	0.14	0.18	0.19	0.18	0.2	0.178	0.0204	0.14	0.2
RMSE	0.2	0.25	0.28	0.26	0.29	0.256	0.03137	0.2	0.29
Maximum Error	1.6	1.8	1.73	1.9	2.3	1.866	0.238	1.6	2.3
Std	0.2	0.25	0.28	0.27	0.29	0.258	0.0319	0.2	0.29

8.6 Appendix F – Statistical tests

Code:

final_StatisticalTests.
R

Results – feed intake weight estimation (all models, all conditions, both fed diets):

final_Cholvot_Eglot_C
ombined_CNNMLP.cs

LMM: $y - \hat{y} = \boldsymbol{model} + (\mathbf{1}|\boldsymbol{sample\ id})$

Anova:

```
> Anova(lmm)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: data$diff
           Chisq Df Pr(>Chisq)
Algorithm 1457.7  3  < 2.2e-16 ***
```

Estimated marginal means:

```
> summary(emmeans_lmm)
 Algorithm     emmean       SE  df asymp.LCL asymp.UCL
 CholvotEglot -0.00442 0.000954 Inf  -0.00629  -0.00255
 Combined      0.02628 0.001079 Inf   0.02416   0.02839
 EglotCholvot -0.02087 0.000938 Inf  -0.02271  -0.01903
 MLPCNN        0.00645 0.001079 Inf   0.00433   0.00856

Degrees-of-freedom method: asymptotic
Confidence level used: 0.95
```

Post-hoc contrasts pairwise comparisons:

```
> summary(lmm_tuckey)
 contrast                    estimate      SE  df z.ratio p.value
 CholvotEglot - Combined      -0.0307 0.00124 Inf -24.846 <.0001
 CholvotEglot - EglotCholvot   0.0165 0.00117 Inf  14.049 <.0001
 CholvotEglot - MLPCNN        -0.0109 0.00124 Inf  -8.792 <.0001
 Combined - EglotCholvot       0.0472 0.00126 Inf  37.352 <.0001
 Combined - MLPCNN             0.0198 0.00129 Inf  15.341 <.0001
 EglotCholvot - MLPCNN        -0.0273 0.00126 Inf -21.641 <.0001

Degrees-of-freedom method: asymptotic
P value adjustment: tukey method for comparing a family of 4 estimates
```

GLMM: $(\boldsymbol{y} - \widehat{\boldsymbol{y}})^2 = \boldsymbol{model} + (\mathbf{1}|\boldsymbol{sample\ id})$

Anova:

```
> Anova(glmm)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: SQR_diff
           Chisq Df Pr(>Chisq)
Algorithm 8693.7  3  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Estimated marginal means:

```
> summary(emmeans_glmm)
 Algorithm    response       SE  df asymp.LCL asymp.UCL
 CholvotEglot   0.0246 0.000180 Inf    0.0242    0.0249
 Combined       0.0442 0.000367 Inf    0.0435    0.0450
 EglotCholvot   0.0218 0.000163 Inf    0.0215    0.0221
 MLPCNN         0.0416 0.000349 Inf    0.0409    0.0422

Confidence level used: 0.95
Intervals are back-transformed from the log scale
```

Post-hoc contrasts pairwise comparisons:

```
> summary(glmm_tuckey)
 contrast                    ratio      SE  df z.ratio p.value
 CholvotEglot / Combined     0.555 0.00515 Inf -63.388 <.0001
 CholvotEglot / EglotCholvot 1.127 0.00979 Inf  13.800 <.0001
 CholvotEglot / MLPCNN       0.591 0.00552 Inf -56.316 <.0001
 Combined / EglotCholvot     2.030 0.01900 Inf  75.591 <.0001
 Combined / MLPCNN           1.065 0.01034 Inf   6.464 <.0001
 EglotCholvot / MLPCNN       0.525 0.00493 Inf -68.674 <.0001

P value adjustment: tukey method for comparing a family of 4 estimates
Tests are performed on the log scale
```

8.7 Appendix G – Models hyper-parameters

Model 1 - Combined

Optimizer: RMSProp ; Loss Function: MSE ; Maximum Learning rate: 0.001 ; Minimum Learning rate: 0.0000625 ; Number of epochs: 100 ; Batch size: 16 ; Dropout rate: 0.25 ; Number of Tensors: 40k ; Input size: (160,120)

Model 2 - TL

<u>Base model</u>

Optimizer: RMSProp ; Loss Function: MSE ; Maximum Learning rate: 0.001 ; Minimum Learning rate: 0.0000625 ; Number of epochs: 50 ; Batch size: 16 ; Dropout rate: 0.25 ; Number of Tensors: 30k ; Input size: (160,120)

<u>Fine-tuning</u>

Optimizer: RMSProp ; Loss Function: MSE ; Learning rate: 0.00005 ; Number of epochs: 30 ; Batch size: 16 ; Dropout rate: 0.25 ; Number of Tensors: 23k ; Input size: (160,120)

Model 3 - MLP-CNN

Optimizer: RMSProp ; Loss Function: MSE ; Maximum Learning rate: 0.001 ; Minimum Learning rate: 0.0000625 ; Number of epochs: 190 ; Batch size: 16 ; Dropout rate: 0.25 ; Number of Tensors: 40k ; Input size: (160,120)

תקציר

מערכת ראייה ממוחשבת הכוללת מודלים המסוגלים להתמודד עם מספר בלילי מזון שונים פותחה על מנת לחזות צריכת מזון פרטנית של פרה ברפת. באמצעות מצלמה הכוללת ערוצי צבע ועומק, צולמו תמונות של ערימות מזון של שני סוגי בלילים שונים (בליל עגלות ובליל חולבות). התמונות צולמו ברפת חלב מחקרית באופן ידני ואוטומטי תחת סידורים שונים ותאורות שונות במשך כ-4 שבועות וכללו ערימות במגוון משקלים.

מודלים שונים פותחו במטרה לחזות את צריכת המזון הפרטנית של פרה ברפת: שני מודלים מבוססי העברת למידה המבוססים על רשת נוירונים קונבולוציונית, מודל אחד מסוג רשת נוירונים קונבולוציונית אשר אומן על שני בלילי המזון בו-זמנית, ומודל אחד מסוג קולטן רב-שכבתי ורשת נוירונים קונבולוציונית שאומן על שני בלילי המזון בו-זמנית בצירוף נתונים קטגוריים.

בנוסף, פיתחנו שיטה סטטיסטית להשוואת 4 המודלים הללו באמצעות מודל לינארי מעורב ומודל לינארי מוכלל מעורב אשר הראתה כי כל המודלים שונים באופן מובהק זה מזה. המודלים אשר אומנו באמצעות העברת למידה הניבו את הביצועים הטובים ביותר.

המודלים הטובים ביותר (בעלי השגיאה הנמוכה ביותר) אשר אומנו באמצעות שיטות העברת למידה, השיגו שגיאה ממוצעת מוחלטת של 0.12 ו- 0.13 קילוגרם לארוחה עם שורש שגיאה ריבועית ממוצעת של 0.18 ו- 0.17 קילוגרם לארוחה עבור שני בלילי המזון כאשר נבחנו על מאגר נתונים מגוון שנאסף באופן ידני ברפת. בחינת המודל על נתונים שנאספו באופן אוטומטי על ידי המערכת ברפת, הניבה שגיאה ממוצעת מוחלטת של 0.14 קילוגרם לארוחה ושורש שגיאה ריבועית ממוצעת של 0.19 קילוגרם לארוחה. תוצאות אלו מצביעות על כך שהמודלים שאומנו באמצעות שיטות העברת הלמידה במסגרת המחקר עשויים להיות מעשיים להערכת צריכת מזון פרטנית של פרה ברפת עבור בלילי מזון שונים.

מילות מפתח: צריכת מזון פרטנית, חקלאות מדייקת בחיות משק, למידה עמוקה, מצלמת עומק וצבע, העברת למידה.

בנגב גוריון-בן אוניברסיטת
ההנדסה למדעי הפקולטה
וניהול תעשייה להנדסת המחלקה

מערכת ראייה ממוחשבת ומודלים למדידת צריכת מזון פרטנית ברפת עבור סוגי מזון שונים

ההנדסה במדעי מגיסטר תואר לקבלת מהדרישות חלק מהווה זה חיבור

מאי סער:מאת
הלחמי אילן אידן ופרופ' יעל" פרופ בהנחיית:

המחבר:............................ תאריך: 14.9.2021 חתימת

המנחה:.............................. תאריך: 14.9.2021 אישור

המנחה:.............................. תאריך: 14.9.2021 אישור

גליה רביד

מחלקתי:.......................... תאריך:.......................... 19/9/2021 תואר ועדת ר"יו אישור

תשפ"ב ,אלול
2021,ספטמבר

בנגב גוריון-בן אוניברסיטת
ההנדסה למדעי הפקולטה
וניהול תעשייה להנדסת המחלקה

מערכת ראייה ממוחשבת ומודלים למדידת צריכת מזון פרטנית ברפת
עבור סוגי מזון שונים

ההנדסה במדעי מגיסטר תואר לקבלת מהדרישות חלק מהווה זה חיבור

מאי סער:מאת
הלחמי אילן אידן ופרופ' יעל" פרופ בהנחיית:

תשפ"ב ,אלול
ספטמבר,2021