Journal of Agricultural Engineering

https://www.agroengineering.org/

Sheep pose estimation via image analysis and body measurements derived from key points

Cafer Tayyar Bati

Department of Animal Science, Faculty of Agriculture, Van Yuzuncu Yil University, Van, Turkey

Corresponding author: Cafer Tayyar Bati, Department of Animal Science, Faculty of Agriculture, Van Yuzuncu Yil University, Van, Turkey. *E-mail:* cafertayyarbati@gmail.com

Please cite this article as doi: 10.4081/jae.2025.1719

Submitted: 4 February 2025
Accepted: 5 August 2025

Sheep pose estimation *via* image analysis and body measurements derived from key points

Cafer Tayyar Bati

Department of Animal Science, Faculty of Agriculture, Van Yuzuncu Yil University, Van, Turkey

Corresponding author: Cafer Tayyar Bati, Department of Animal Science, Faculty of Agriculture, Van Yuzuncu Yil University, Van, Turkey. E-mail: cafertayyarbati@gmail.com

Conflict of interest: the author declares no competing interests.

Abstract

Pose estimation and body measurement in livestock play an important role in various agricultural applications such as health monitoring, breeding and management. In this study, we propose a novel approach for body measurement and pose estimation of sheep using object detection and supervised machine learning algorithms. From a dataset of sheep, 40 side-view videos for each class (grazing, standing, and sitting), each lasting approximately 8 seconds, are used for model training. Firstly, pose classification was performed on the sheep images using the YOLOv8 pose object detection framework. Simultaneously, keypoint training was performed on the images to detect key points for anatomical landmarks. Then, using these keypoints, various body measurements of the sheep in the images were measured and a comprehensive dataset was created. Six different supervised machine learning algorithms were trained on this dataset to further improve pose estimation. Furthermore, the models were tested on frontal images to evaluate their performance against different image angles and dataset feature. The experimental results show that the supervised machine learning algorithms trained on the body measurement data perform better for both side and frontal images (mAP; K nearest neighbour algorithm 1.00 for side images, Support vector machines 0.94 for frontal images) outperform state-of-the-art networks such as YOLOv8 (mAP; 0.94, 0.89 for side and frontal images respectively), EfficientNet (mAP; 0.93, 0.91), RetinaNet (mAP; 0.91, 0.88) and Faster R-CNN (mAP; 0.89, 0.87) based on image data only. This approach can play an important role in improving the accuracy and efficiency of livestock management systems by supporting practical applications such as animal welfare monitoring, herd health assessment, and precision agriculture through more accurate position estimation and body measurements.

Keywords: Body measurement, keypoint detection, livestock management, object keypoint similarity, supervised machine learning

Introduction

By 2050, with the expected increase in the world population, the global demand for meat and animal products is expected to increase significantly. This emphasises the need for more efficient production methods in the livestock industry (Forslund *et al*., 2023). Therefore, there is a need to focus on the development of new and innovative solutions to utilise resources more efficiently (Hamadani and Ganai, 2023). Sheep farming has an important position among agricultural activities and animal welfare monitoring and evaluation is an important requirement in this industry. Although contact-based animal behaviour detection methods for sheep welfare monitoring are capable of high accuracy and fast data processing, they are not practical for sheep farming due to the increased cost associated with equipping each sheep with sensors. Furthermore, such sensors are not suitable for long-term group animal behaviour monitoring and require regular manual battery replacement (Xu *et al*., 2023). Machine learning techniques such as deep learning are gaining attention for their ability to automatically predict sheep posture and behavior. These methods offer an alternative to contact-based sensor technologies that can negatively impact animal welfare (Bati and Ser, 2023a). Recent studies have increasingly utilized machine learning approaches to monitor animal behavior and predict physical characteristics. These methods have shown promising results in tasks such as estimating body composition from images, predicting body condition from 3D scans, and classifying general sheep behavior under various environmental conditions (Jin *et al*., 2022; González-Baldizón *et al*., 2022; Stephansen *et al*., 2023; Shalaldeh *et al*., 2023). The common purpose of researchers and practitioners in studies such as this one in recent years is to be able to convert their attitudes towards the welfare of farm animals into practical improvements with remote monitoring capability. While these studies demonstrate the applicability of image-based analysis in animal science, they generally address posture recognition and body measurement as separate tasks. In this context, the current study presents a novel two-stage approach that combines supervised learning with deep learning-based keypoint detection to derive body measurements, thereby providing a hybrid and interpretable system for accurately classifying sheep poses.

Accurately assessing sheep movements enables farmers to monitor performance non-invasively. As the behaviour of farm animals has the potential to reflect their response to environmental and health conditions, enabling detailed observation of this behaviour is of significant value in animal science. For example, in hot weather, sheep may increase lying time (Pollard *et al*., 2004) and decrease walking speed (Horie *et al*., 2023), which may reduce feed intake and increase the probability of health problems (Schütz *et al*., 2024). Therefore, monitoring and analysing the behaviour of sheep allows their health and welfare to be determined and can help to take proper measures (Jin *et al*., 2022; Hu *et al*., 2023). The process of monitoring a sheep's behaviour can also involve the classification and segmentation of different activities. Monitored behaviours for sheep usually include grazing, lying, standing, walking, ruminating, running, etc. (Jin *et al*., 2022). For the detection of these behaviours, a set of key points and body measurements between these points are usually used. These key points represent different joints of the sheep and these key points and body measurements can be used for pose estimation. Pose parameters can be used as a tool for more detailed analyses and allow many sheep behaviour studies to be automated. For example, it is important to distinguish between grazing and resting behaviour of sheep and to analyse grazing behaviour. Because the identification and classification of grazing behaviour in free-ranging ruminants will help to improve the efficiency of animal production (Alvarenga *et al*., 2016). However, in the sheep industry, estimation of body measurements is very important as it provides a rough indication of the muscle and fat cover of the live animal. These measurements can influence the decision to sell or retain the animal and can help to estimate carcass yield and quality. Body measurements can also be used to calculate many growth traits that can help determine the productivity of animals and identify superior animals for breeding (Kenyon *et al*., 2014).

While body measurements are so important in animal husbandry, taking these measurements manually, monitoring and analyzing animal behaviour using traditional methods can be intrusive, time-consuming and subjective, especially on large-scale farms (Meckbach *et al*., 2021; Stephansen *et al*., 2023; Gao *et al*., 2023). Therefore, the use of measurements obtained by image analysis (instead of traditional methods) to classify sheep posture and predict their pose has the potential to overcome the intrusive disadvantages of traditional measurement methods. Pose estimation on video sequences using machine learning methods is a relatively new research area. Body measurements taken via keypoints with image analysis offer a powerful way to easily and quickly observe animal movements in a natural environment, without the biases that human observation can introduce.

Deep learning-based pose estimation, despite its widespread use in recent studies, most methods focus solely on visual features extracted from images without incorporating structural body measurements that could improve classification accuracy. Integrated systems that seamlessly combine pose classification with keypoint-based body measurements are still lacking. To address this gap, this study proposes a hybrid approach that leverages both image-based detection and quantitative biometric features to enhance the robustness and accuracy of sheep pose predictions. In line with this motivation, this study presents a two-stage computer vision-based framework specifically designed for the identification and classification of sheep postures. First, a YOLOv8 pose-based network is trained to classify sheep behaviors and detect key points of the sheep's body. Second, a predictive, supervised machine learning model is designed to recognize animal behaviors based on the detected key points and automatically calculated body measurements. Furthermore, the proposed methodology is compared with recently used object detection models and comparative experiments are conducted. The methods proposed in the manuscript present an innovative approach to automate sheep posture classification and body measurements using computer vision techniques.

Experimental setup

In this section, the experimental datasets and each stage of the experimental setup are described in detail. The workflow of the study is given in Figure 1.

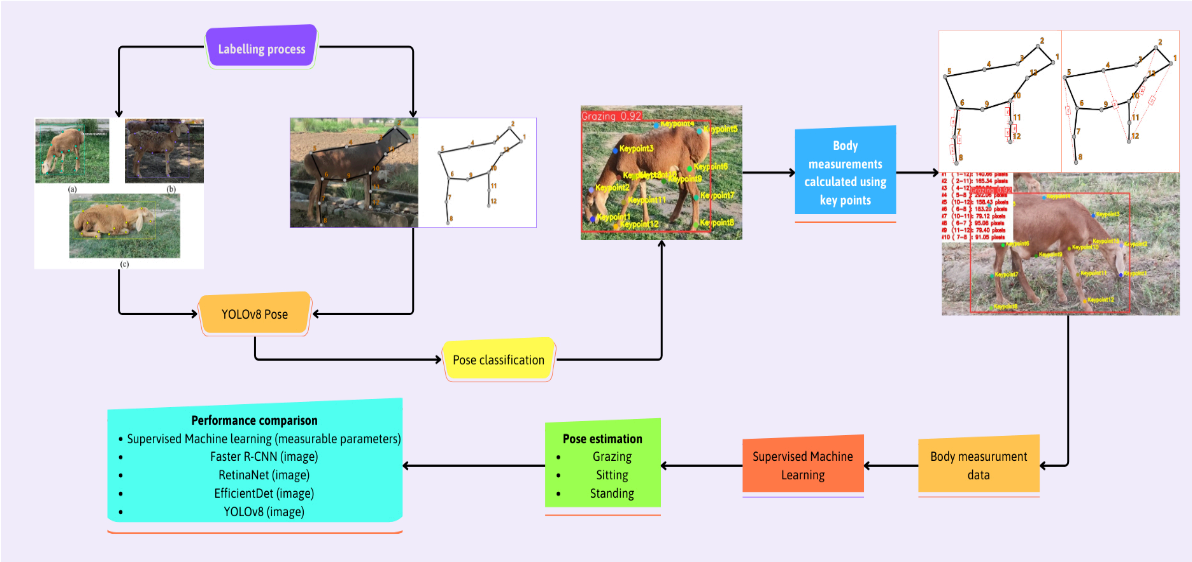


Figure 1. Workflow of the study

Video data

In this study, a dataset consisting of 417 sheep videos covering five activity classes, including grazing, running, sitting, standing, and walking, was used (Khan and Kelly 2023a; 2023b). Each behaviour was recorded from front and side perspectives. Further information about the dataset can be found in Kelly *et al*. (2024) (see Figure 2 for example frames).

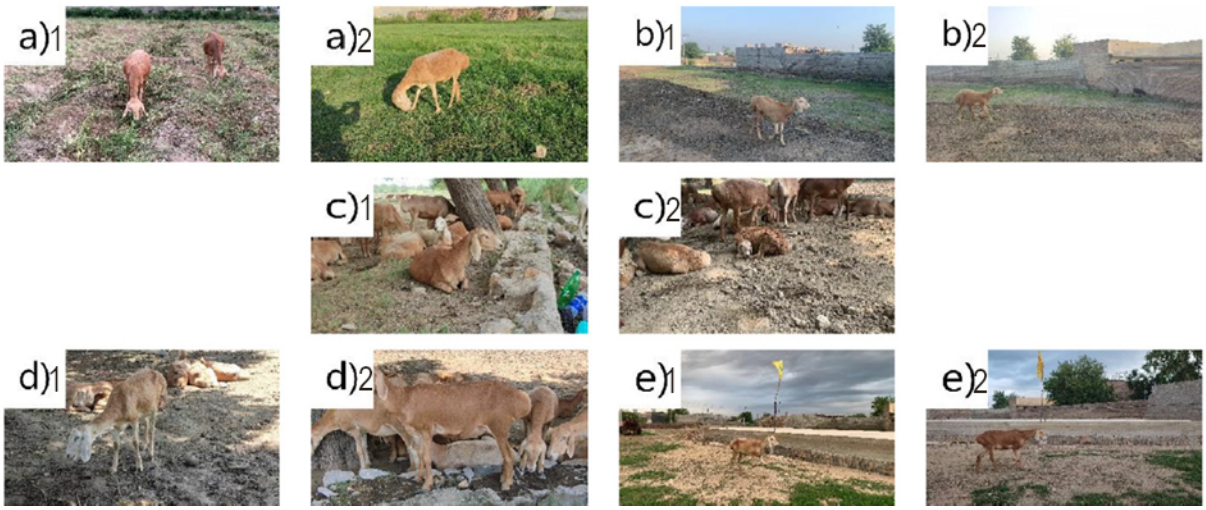
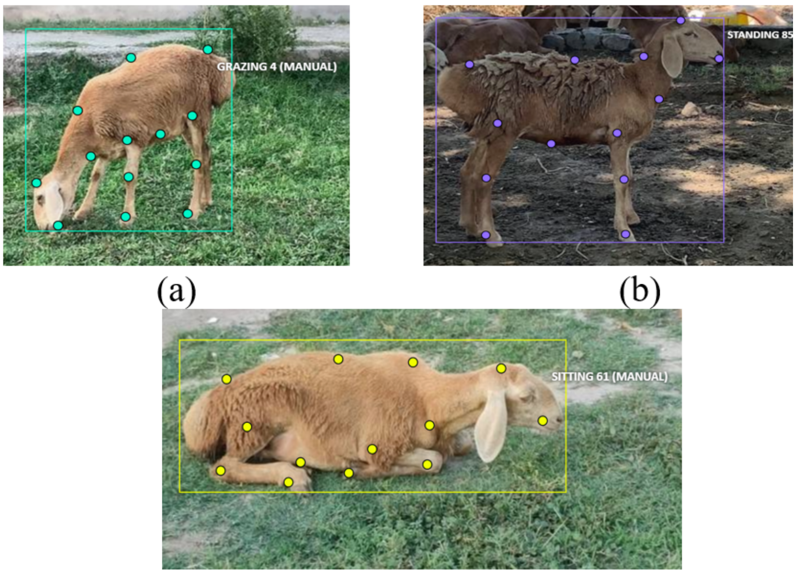


Figure 2. Sample images from the dataset. a) grazing, b) running, c) sitting, d) standing, e) walking, 1) frontal, 2) side Kelly *et al*. (2024).

Image processing and labelling

In the experimental setup, a total of 120 videos were used by randomly selecting 40 side-recording videos from each of the grazing, standing and sitting behaviour classes in the data set. Afterwards, 5 frames per second were captured from the videos, each of which lasted 8 seconds on average. Thus, an image training dataset containing 5015 frames in total was created. Then, a test dataset consisting of frontal images was prepared in order to test the models on datasets with different features. The test set contains 1146 images (382 per class). Each high-resolution image (3840x2160) occupied approximately 3 MB of memory. Considering the computer processing power during the labelling and analyses, each image was rescaled (800x450) and reduced to approximately 180 KB. Each of the images to be used for training and testing purpose was labelled as a class using the Computer Vision Annotation Tool (CVAT) (https://cvat.ai/) (Figure 3) and 13 different key points of each sheep were labelled (Figure 4). A total of 13 key points were selected based on distinct anatomical reference points that are typically visible in different poses of sheep and are necessary for both pose classification and the extraction of body measurements. These key points include the head, neck, shoulders, elbows, knees, heels, and tail base. The number and placement of these points were determined based on previous pose estimation studies and verified through visual inspection to ensure consistent visibility and sufficient anatomical coverage in various poses.



(a) (b) (c)

Figure 3. Sample images from the labelling process of the dataset used in the study. a) grazing, b) standing, c) sitting

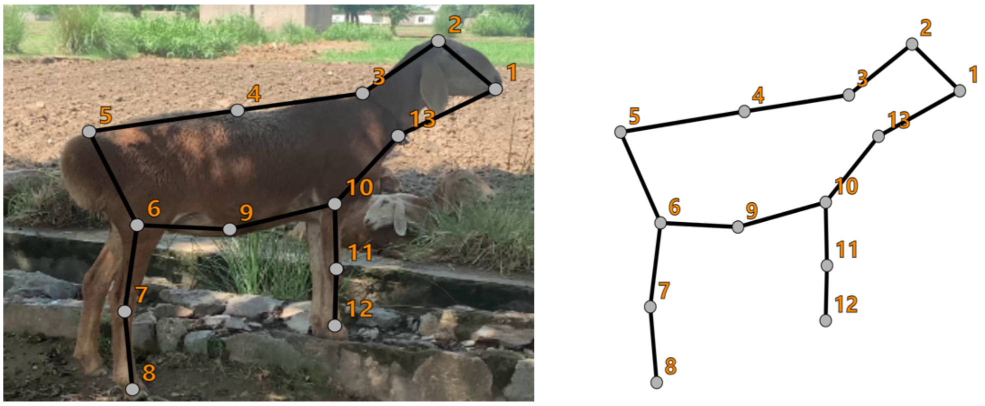


Figure 4. Labelling of key points

Body measurement from key points

For the supervised machine learning analyses of the study, Euclidean distance was used to calculate body measurements over the images (Eq. 1). Some additional coding was made to the "predict" module of the YOLOv8 model for the detection of key points on the images and then automatic measurement of the determined distances. Accordingly, the measurements calculated for each image were recorded in an excel file respectively. Finally, this excel file was merged for all images and the body measurements data set was saved in .csv format.

$$\boldsymbol{d(Keypoint_i, Keypoint_j)} = \sqrt{(\boldsymbol{x_j} - \boldsymbol{x_i})^2 + (\boldsymbol{y_j} - \boldsymbol{y_i})^2} \qquad \text{(Eq. 1)}$$

Figure 5 shows the calculated body measurements of sheep. Accordingly, the measurement between Keypoint 1 and Keypoint 12 is Measurement 1 (M1), the measurement between Keypoint 2 and Keypoint 11 is M2, the measurement between Keypoint 4 and Keypoint 12 is M3, the measurement between Keypoint 5 and Keypoint 8 is M4, the measurement between Keypoint 10 and Keypoint 12 is M5, the measurement between Keypoint 6 and Keypoint 8 was assigned to variable M6, between Keypoint 10 and Keypoint 11 to variable M7, between Keypoint 6 and Keypoint 7 to variable M8, between Keypoint 11 and Keypoint 12 to variable

M9 and between Keypoint 7 and Keypoint 8 to variable M10. Class values were assigned to these measurements obtained for all images and the data set in Table 1 was created.

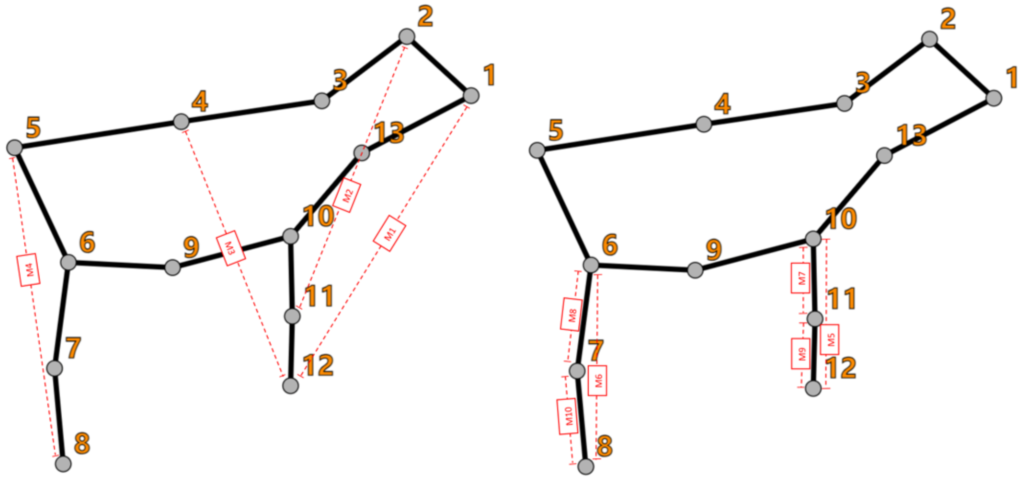


Figure 5. Body measurements calculated using key points

Table 1. Body measurements data.

Frame	M1	M2	M3	…	M9	M10	Class
1	64.008	95.708	207.060	…	44.283	49.254	Grazing
2	69.029	101.203	206.620	…	43.829	49.254	Grazing
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
2446	133.405	148.337	112.432	…	49.092	38.910	Sitting
2447	134.848	147.574	110.345	…	49.163	39.812	Sitting
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
5014	260.946	225.996	228.543	…	59.682	71.589	Standing
5015	275.409	237.059	224.154	…	55.444	68.411	Standing

Model configurations

YOLOv8 for pose classification

YOLOV8 pose model was used in the image analysis part of the study. The dataset is divided into 70% for training, 20% for validation and 10% for testing. Accordingly, 28 of the 40 videos in each class were used for training, 8 for validation and 4 for testing. As a result, a total of 3561 side images were used for training, 954 validation data points for validation, and 500 data points for testing. The models have a batch size of 16, learning rate of 0.01, momentum of 0.937, epoch of 100, image size of 640, and optimisation method of SGD. The training at this stage took approximately 3.5 hours. The hyperparameters used (e.g., batch size, learning rate, momentum) were initially selected based on the default YOLOv8 settings reported in the relevant literature and then fine-tuned using pre-validation experiments. For example, learning rates of 0.001, 0.005, and 0.01 were tested, and 0.01 provided the best balance between convergence speed and accuracy. During the training stage, default data augmentation techniques integrated into the YOLOv8 training pipeline were applied. These include random horizontal flipping, HSV color jittering, image scaling, and translation. No additional manual augmentations such as Gaussian noise or customized brightness/contrast changes were applied.

Supervised machine learning algorithms for pose estimation based on body measurements

Six different supervised machine learning algorithms were used in this part of the study. These are Support Vector Machines (SVM), Naïve Bayes (NB), Classification and Regression Tree (CART), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN) and Random Forest (RF) algorithms. In artificial neural networks, a hidden layer of size (128,128) was used, relu was used as the activation function and sgd was used as the optimisation method. In addition, gini index was used in CART and RF algorithms and rbf was used as the kernel in SVM. In the RF algorithm, the number of estimators was taken as 100, the minimum sample split was taken as 2 and the minimum number of sample leaf was taken as 1. In the KNN algorithm, the number of neighbours was taken as 5 and 'minkowski' was used as the metric. In the NB algorithm, the 'Gaussian Naive Bayes' model and the variance smoothing value of 1e-9, which is the default value, were used. At this stage of the study, 10% of the data set was reserved for testing. In addition, 10-fold cross-validation was used in model analyses. The hidden layer size in the artificial neural network (128, 128) was experimentally determined by comparing several configurations (e.g., 64×64, 128×128, and 256×256). The selected size provides a balance between model complexity and performance.

Performance evaluation metrics and system utilization

The assessment methodology utilized to evaluate the effectiveness of the models encompasses Accuracy, Precision, Recall and mAP (mean average precision). These performance metrics are derived from the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values extracted from the confusion matrix, as outlined in Eq. 2 through 5 (Bati and Ser, 2023b).

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{tp+tn}}{\boldsymbol{tp+tn+fp+fn}} \qquad \text{(Eq. 2)}$$

$$\boldsymbol{Precision} = \frac{\boldsymbol{tp}}{\boldsymbol{tp+fp}} \qquad \text{(Eq. 3)}$$

$$\mathrm{Recall} = \frac{\mathrm{tp}}{\mathrm{tp+fn}} \qquad \text{(Eq. 4)}$$

$$\mathrm{mAP} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{AP_i} \quad , \qquad \text{where} \quad \text{AP: average precision} \qquad \text{(Eq. 5)}$$

Object Keypoint Similarity (OKS) is widely regarded as the standard metric for keypoint evaluation, as noted by Maji *et al*. (2022). It is measured on a scale from 0 to 1, where higher scores correspond to a closer match between the predicted keypoints and the ground truth. In the context of pose estimation models, a prediction is classified as "positive" or accurate if the predicted keypoints lie within a predefined distance threshold from the actual ground truth keypoints (Jocher *et al*., 2023). The Keypoint Similarity (KS) for each keypoint is determined by applying the Euclidean distance between the predicted and ground truth keypoints to an unnormalized Gaussian distribution with a standard deviation of σ_i.

Formally, keypoint similarity for the i-th keypoint type can be expressed as in Eq. 6 (Dutta and Dawn, 2023).

$$KS_i = \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) \quad \text{(Eq. 6)}$$

Where,

d_i denotes the Euclidean distance between the ground truth and the predicted i-th keypoint, k is a constant associated with the i -th keypoint,

s represents the scale of the ground truth object, such that s^2 corresponds to the segmented area of the object.

OKS is ultimately calculated as the arithmetic mean of all annotated keypoints in a given instance.

The mathematical representation for OKS is provided in Eq. 7 (Dutta and Dawn, 2023).

$$OKS = \frac{\sum_i KS_i.\delta(v_i>0)}{\sum_i \delta(v_i>0)} \quad \text{(Eq. 7)}$$

Where,

KS_i denotes the Keypoint Similarity for the i-th type keypoint,

v_i represents the ground truth visibility flag for the i-th keypoint,

$\delta(v_i > 0)$ is the Dirac-delta function, which is 1 if the i-th keypoint is annotated and 0 otherwise. Unannotated parts (where $v_i = 0$) do not affect the OKS calculation.

The determination of true positives, false positives, and false negatives is based on an OKS threshold, analogous to the Intersection over Union (IoU) metric. A detection is considered a true positive if the OKS score between the ground truth and the prediction exceeds the OKS threshold; otherwise, it is counted as a false positive. All unmatched ground truths are considered false negatives (Ronchi and Perona, 2017; Dutta and Dawn, 2023). Similar to object detection, precision and recall values are calculated, followed by the computation of the final metric, mean average precision (mAP) (Dutta and Dawn, 2023).

The open-source Python 3.8.3 (Van Rossum and Drake, 2009) package program and Pytorch 1.11.0 (Paszke *et al*., 2019), a high-performance deep learning library for YOLO v8 pose, were used in deep learning analyses. The hardware features of the computer used in the study are as follows; processor: Intel Core i7-9750H, graphics: NVIDIA GeForce RTX 2070, 8 GB GDDR6 Dedicated VRAM, and memory: 32 GB DDR4 2666 MHz.

Results

Evaluation of pose classification accuracy based on image analysis

In the analyses for pose classification, the YOLOv8 pose model was trained for 100 epochs. The results obtained after 3.5 h of training are given in Figure 6 a,b) show the prediction-recall curves for class and keypoint prediction on the validation data, respectively. The model achieved a mean average precision (mAP) of 0.99 for pose classification and 0.96 for keypoint localization. When the accuracy measure is calculated for all classes, 881 of the total 954 validation data points were correctly classified, and a classification accuracy of 93% was achieved (Figure 6 c-d). Figure 6e shows the accuracy, mAP and loss curves for the training and validation data recorded during the training process.

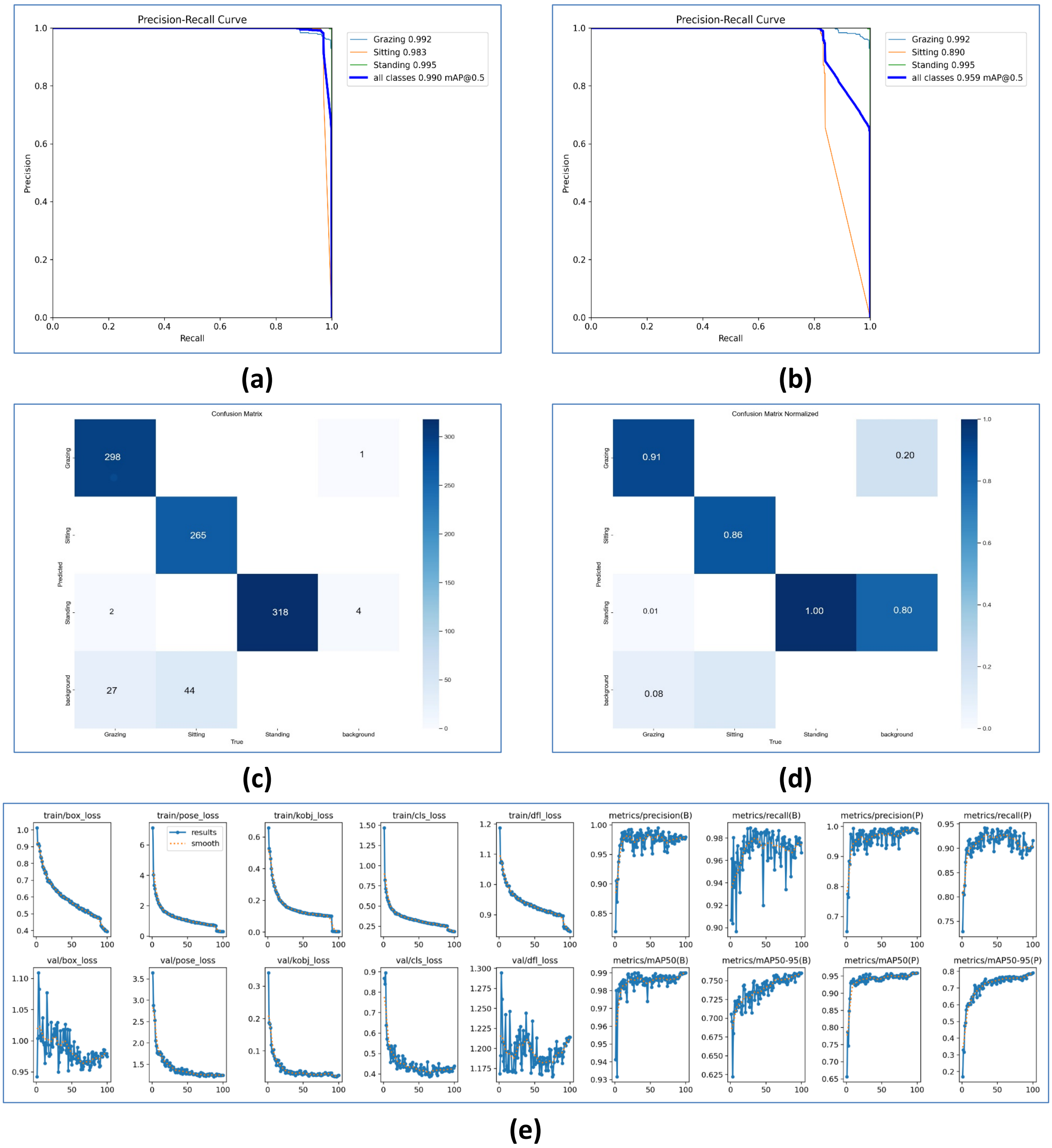
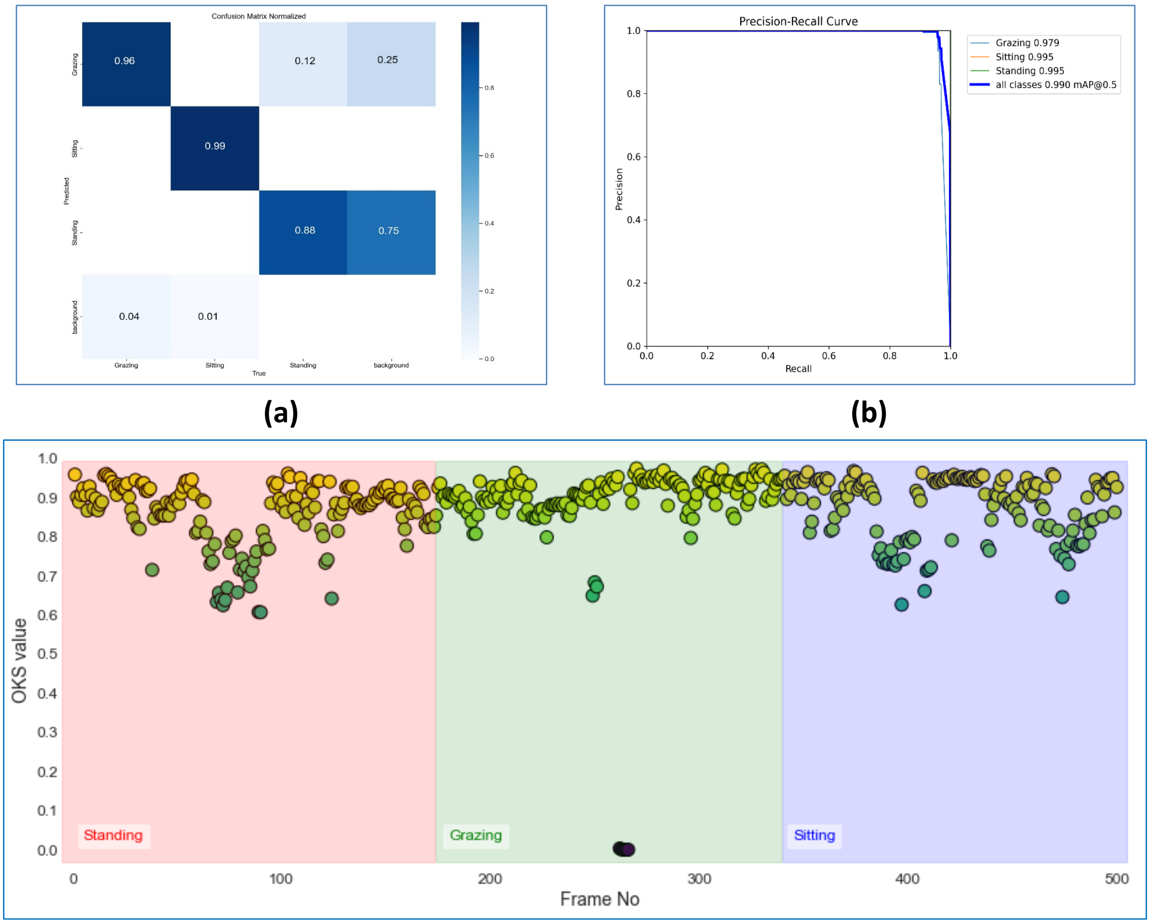


Figure 6. Results obtained during the training process for pose classification

Model testing

Test on side images: As shown in Figure 7a, the model achieved 94% overall accuracy on the side test images and demonstrated high performance across all classes. The precision-recall (PR) and OKS visualizations (Figures 7 b,c) show that most predictions achieved high key point similarity and that 99% (total 495) exceeded the 0.50 OKS threshold. Figure 8 presents the ground truth along with the various predictions made by the model for both keypoints and class labels, as well as the corresponding OKS values.

In Figure 8, some OKS values are obtained depending on the keypoint placements of the model. The predicted frame for the sitting class, when compared with the ground truth, demonstrates highly accurate keypoint placement, as evidenced by an OKS value of 0.98, further supporting the precision of the prediction. In some cases, minor keypoint placement errors occurred (2^{nd} and 3^{rd} keypoints), but some keypoint placements even exceeded the ground truth (e.g., the 12^{th} keypoint in the standing class).



(a) **(b)**

(c)

Figure 7. Confusion matrix (a) and precision-recall curve (b) obtained from side view test images for the pose classification task, along with the graph of OKS values (c).

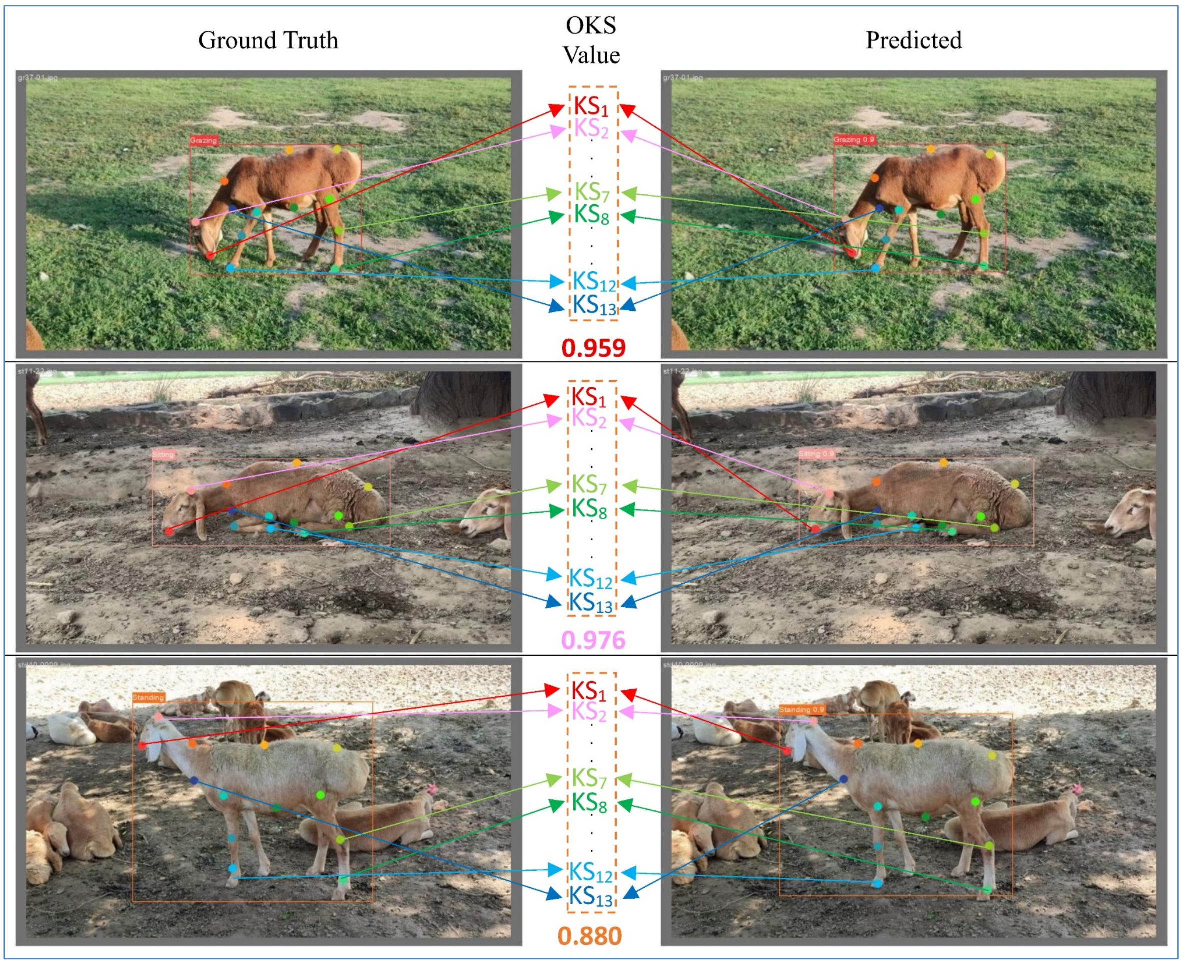


Figure 8. Comparison of ground truth and predicted frames with some OKS values

Test on frontal images

In order to test the model on a different dataset, a dataset was created using the frontal images (Figure 2 - a1, c1, d1) of the sheep in the original dataset different from the training dataset. A total of 1146 images, 382 images of each class, were used as test data.

Figure 9a presents the confusion matrix showing the performance of the YOLOv8 model on the frontal test images. The model achieved an overall accuracy of 89% on frontal images, with performance differences observed across classes. Figure 9b presents the precision-recall (PR) curve for the key points, while the OKS values are presented in Figure 9c. Figure 9c illustrates that 92% of the OKS values (1049 values in total) are higher than 0.50; these values are indicated by a transition from green to yellow.

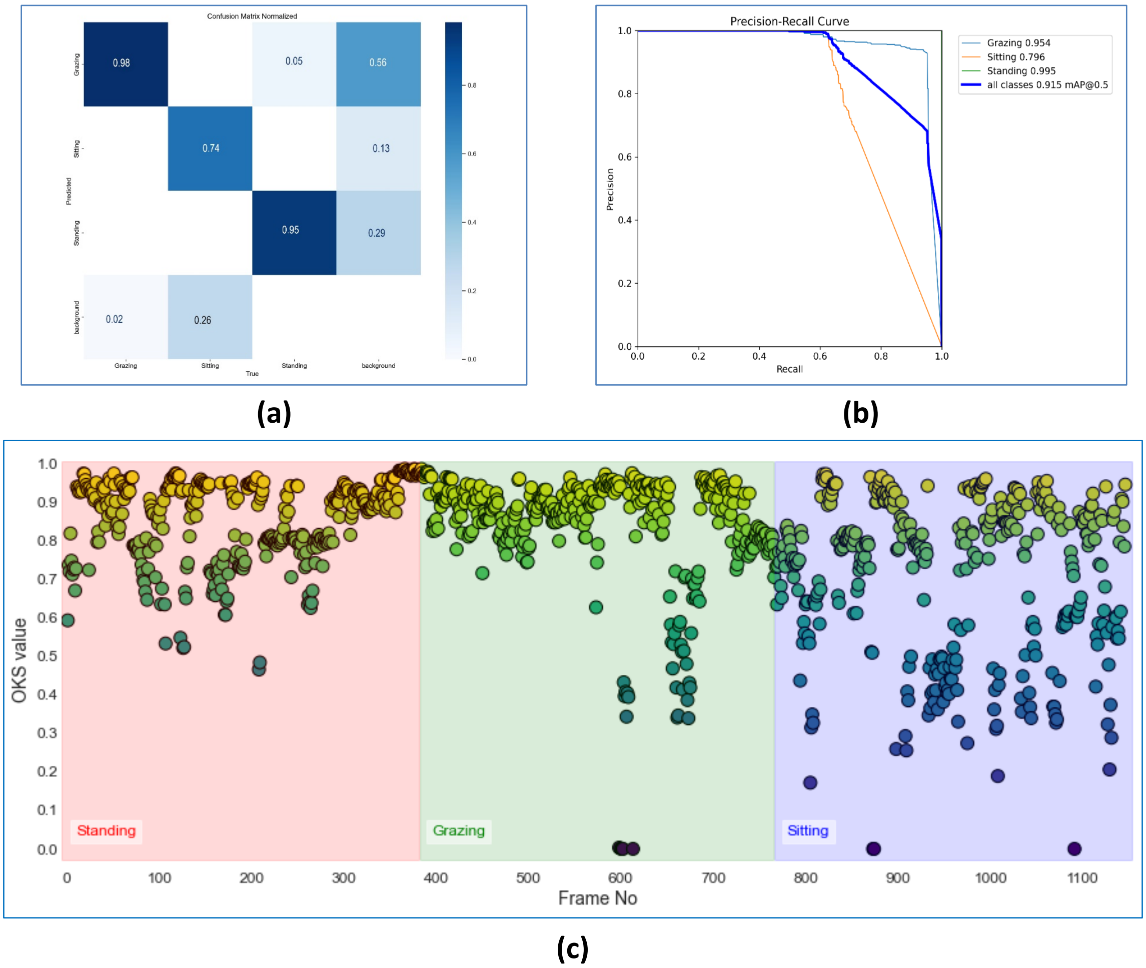


Figure 9. Confusion matrix (a) and precision-recall curve (b) obtained from frontal view test images for the pose classification task, along with the graph of OKS values (c).

Supervised machine learning pose estimation accuracy based on body measurements

Figure 10 shows samples of images for the calculation of body measurements over images using keypoints and presents the box plot of the body measurements dataset. According to the graph, the mean values of M1, M2, M3 and M4 variables are higher, while the mean values of M7, M8, M9 and M10 variables are naturally lower.

(a)

(b)

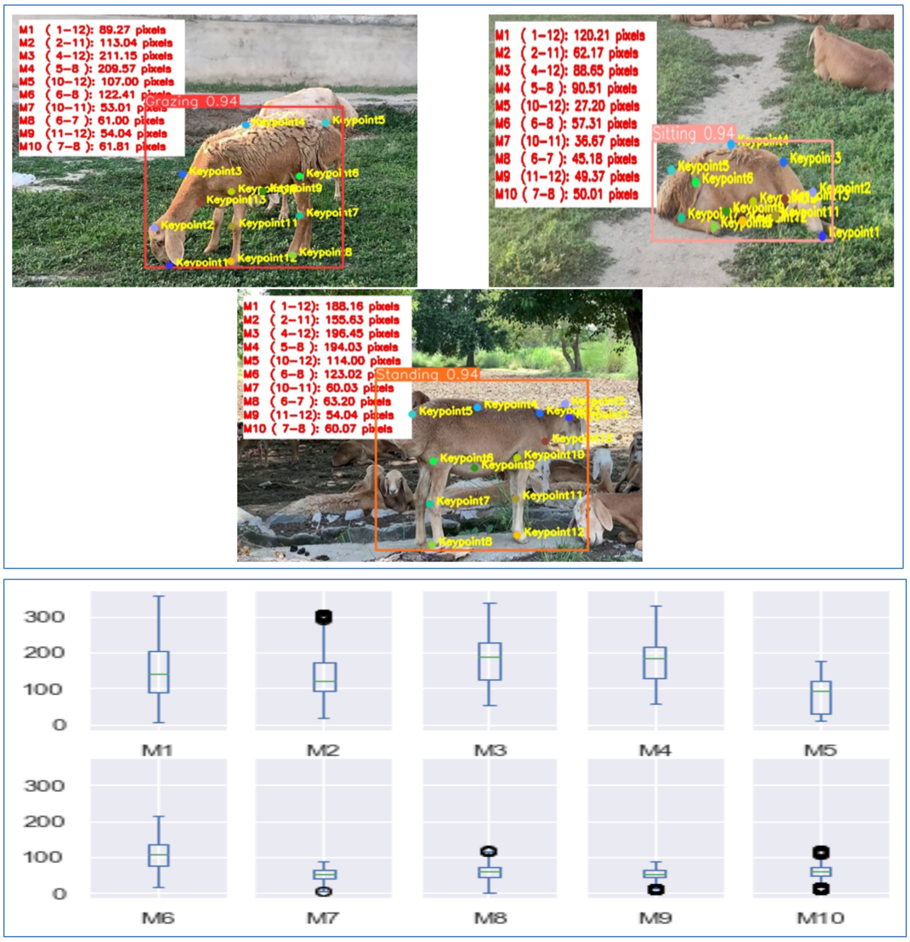
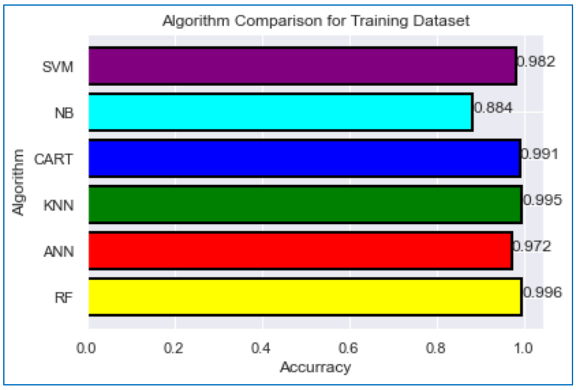


Figure 10. Sample images of body measurements automatically and instantly calculated using keypoints (a), and the box plot of the body measurements dataset (b).

Figure 11 shows the pose estimation performances on the body measurements data set of 5015 sheep. According to the training results, the two most successful models were RF and KNN. It is seen that the models other than NB have over 97% pose estimation success.



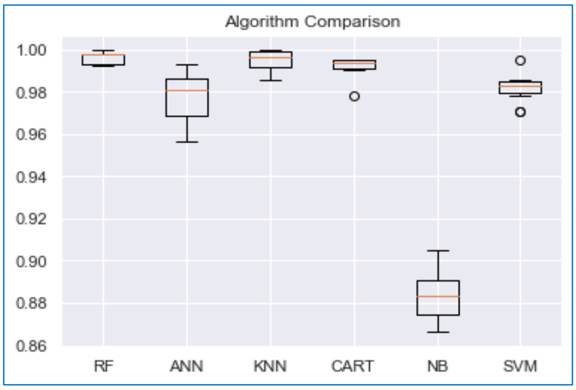


Figure 11. Training performance of supervised machine learning algorithms

Model testing

Test on side images: Figure 12 and Table 2 show the test pose estimation results of the supervised machine learning algorithms on side images. As shown in Figure 12, KNN achieved the highest accuracy among the tested models, and most models performed above 97%. Table 2 shows the confusion matrix and performance metric results of all models. As can be seen in Table 2, it is understood that the K Nearest Neighbour model correctly classifies all test sheep with 100% accuracy.

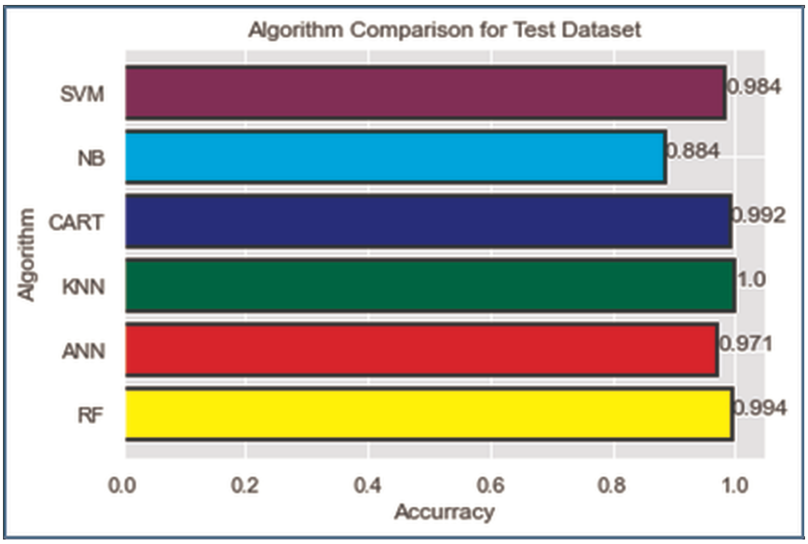
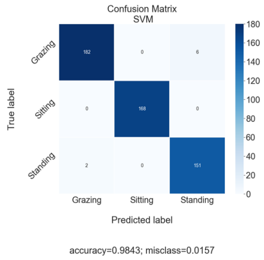
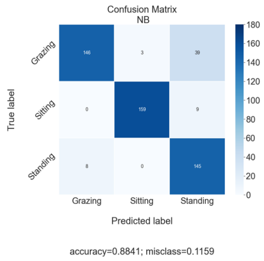
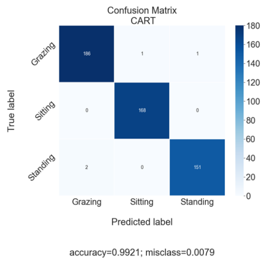
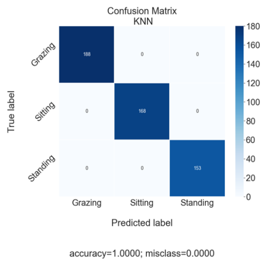
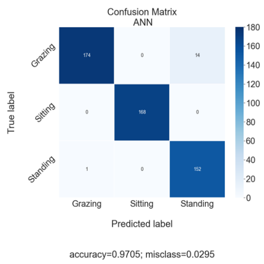
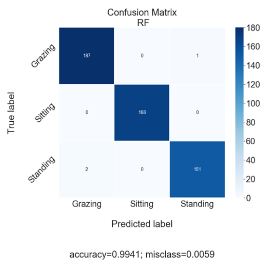


Figure 12. Test performance of supervised machine learning algorithms on side images.

Table 2. Performance metrics of supervised machine learning algorithms with confusion matrix on side images.

Algorithms			Precision	Recall	F1-score
SVM	 	Grazing	0.99	0.97	0.98
		Sitting	1.00	1.00	1.00
		Standing	0.96	0.99	0.97
NB	 	Grazing	0.95	0.78	0.85
		Sitting	0.98	0.95	0.96
		Standing	0.75	0.95	0.84
CART	 	Grazing	0.99	0.99	0.99
		Sitting	0.99	1.00	1.00
		Standing	0.99	0.99	0.99
KNN	 	Grazing	1.00	1.00	1.00
		Sitting	1.00	1.00	1.00
		Standing	1.00	1.00	1.00
ANN	 	Grazing	0.99	0.93	0.96
		Sitting	1.00	1.00	1.00
		Standing	0.92	0.99	0.95
RF	 	Grazing	0.99	0.99	0.99
		Sitting	1.00	1.00	1.00
		Standing	0.99	0.99	0.99

Test on frontal images: In this section, the supervised machine learning algorithms are tested with a dataset of body measurements taken from frontal images with 1146 samples. The test performances of the models on this dataset are presented in Table 3. SVM achieved the highest performance (94%) on frontal images based on body measurements, outperforming models trained only on image data (Tables 3 and 4).

Table 3. Performance metrics of supervised machine learning algorithms on frontal images

Algorithms (accuracy)		**Precision**	**Recall**	**F1-score**	**Algorithms (accuracy)**		**Precision**	**Recall**	**F1-score**
SVM (0.94)	Grazing	0.82	1.00	0.90	KNN (0.93)	Grazing	0.85	0.90	0.88
	Sitting	1.00	0.84	0.92		Sitting	0.93	0.89	0.91
	Standing	1.00	1.00	1.00		Standing	1.00	1.00	1.00
NB (0.90)	Grazing	0.86	0.77	0.81	ANN (0.93)	Grazing	0.87	0.87	0.87
	Sitting	0.91	0.91	0.91		Sitting	1.00	0.91	0.95
	Standing	0.93	1.00	0.96		Standing	0.91	1.00	0.95
CART (0.92)	Grazing	0.84	0.87	0.86	RF (0.93)	Grazing	0.87	0.87	0.99
	Sitting	0.91	0.89	0.90		Sitting	0.91	0.91	0.91
	Standing	1.00	1.00	1.00		Standing	1.00	1.00	1.00

Comparison with state of the art models

Table 4 shows the performance comparison of the proposed methods with different state of the art algorithms. To ensure a fair and consistent comparison, all models were trained and tested on the same dataset. The input image size (640×640), batch size (16), number of epochs (100), and basic data augmentation techniques were standardized across all models. Since optimizers are the default SGD for many models, SGD was used for all models. Evaluation was performed using the same metrics under the same test conditions.

Table 4. Performance comparison of different algorithms

Algorithms	Side Images			Frontal Images		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Image-Based Models						
Faster R-CNN (Ren *et al*., 2015)	0.89	0.90	0.89	0.87	0.88	0.88
RetinaNet (Lin *et al*., 2017)	0.91	0.91	0.90	0.88	0.88	0.87
EfficientDet (Tan *et al*., 2020)	0.93	0.93	0.94	0.91	0.91	0.90
YOLOv8 (Jocher *et al*., 2023)	0.94	0.94	0.93	0.89	0.88	0.89
Measurable Parameter-Based Models						
YOLOv8 pose + KNN	**1.00**	**1.00**	**1.00**	0.93	0.93	0.93
YOLOv8 pose + SVM	0.98	0.98	0.99	**0.94**	**0.94**	**0.95**

Table 4 shows the performance comparison of Faster R-CNN (Ren *et al*., 2015), RetinaNet (Lin *et al*., 2017), EfficientDet (Tan *et al*., 2020) and YOLOv8 (Jocher *et al*., 2023) models with the models proposed in the paper. According to the table, all models have shown successful performances in sheep pose classification with a performance of 87% and above. The KNN algorithm trained with the body measurements dataset obtained with the YOLOv8pose model trained with side images was the most successful model and outperformed the other models by providing 100% pose accuracy on the side images. All models were tested on frontal images with weights trained on side images without training on frontal images. The results are shown in the frontal images section of Table 4. According to the results, it is understood that the models show a decrease in performance against a different data set. In the test we performed on the frontal images dataset, it was found that the models trained with measurable parameters performed better than the other state-of-the-art models. The SVM trained with measurable parameters outperformed the other models with an accuracy of 94% in sheep pose estimation.

Discussion

In this study, an image-based measurement system, developed to overcome the difficulties of traditional manual measurement methods, offers instant and effortless measurement to improve animal welfare and working conditions. The proposed methods performs pose estimation of sheep with automatically taken body measurements and introduces an important innovation to the existing literature in this respect. With this approach, pose estimation of sheep with non-contact measurements and estimation of additional parameters such as body weight are possible. This is a significant advantage with the potential to optimize animal welfare and farm management.

In case recorded body weight data of sheep are available, models for body weight estimation can be developed and the performance of these models can be evaluated by using body measurements obtained with these data. In this context, the proposed methods are not only suitable for non-contact pose estimation, but also for body weight estimation, and are suitable for further development with further studies. In the sheep industry, body measurements are an important tool to assess the physical condition of animals. These measurements provide information on the relationship between the animal's muscle mass and fat cover, giving clues to the overall health and productivity status of the animal. This information is taken into account especially when making important decisions such as the sale or breeding of animals. It also plays an important role in estimating carcass yield and quality. Live weight is an important parameter in livestock management and is used to determine factors such as feeding, drug doses, mating timing. In addition, the study of the relationship between body measurements and live weight is important for the determination of genetic potential and breed standards and the development of breeding programmes (Castillo *et al*., 2023). Sowande and Sobola (2008) used body measurements of West African dwarf sheep for body weight estimation in their study. In the study, body measurements taken by hand were estimated by allometric and linear regression methods. In this study, the coefficient of determination of the allometric regression model was reported to be in the range of 0.87-0.99. He *et al*. (2023) proposed a sheep body weight estimation approach based on convolutional neural networks using RGB-D images. Using a dataset of 726 RGB-D images of sheep, they showed that a lightweight convolutional neural network model trained on RGB-D images can obtain an acceptable weight estimation result. In this study, as in our study, a non-contact approach is proposed.

Deng *et al*. (2021) performed sheep pose estimation in their study. In this study, as in the present study, they classified the standing, grazing and lying behaviours of sheep. They obtained a mAP value of 0.93 in pose classification over 1500 images with YOLOv3. Hu *et al*. (2023), in their YOLOv5-based grazing sheep behaviour recognition study, classified standing, grazing and lying behaviours similar to our study and obtained a mAP value of 0.92. In this study, unlike the present study, both still and handheld camera images were used. At the same time, images from different times of the day were also included in the study. It is known that different features of the dataset improve the training performance of the models. In this study, although the original dataset contains both side and frontal images, only side profile images are used in the training stage, while frontal images are reserved for testing the model. The main reason for this approach is to enable the proposed models to be tested directly on the frontal images dataset without training with frontal images and to achieve the objectives of the study by comparing these results with existing state-of-the-art methods. Since the frontal images are different from the features of the training dataset, the model performances decreased slightly in the test on this dataset (Table 4). This is expected for the networks used. Because these networks are usually tested on images similar to the dataset they are trained on. Therefore, although the test performance of the models is high, they may not perform adequately on images with different features (Bati and Ser, 2024). Some researchers have stated that models should be trained with more diverse and different images to overcome this problem (Shorten and Khoshgoftaar 2019; Yang *et al*. 2022). It is thought that as the diversity in the training data of the model increases, the test performance will also improve. However, when the characteristics of the training and test data are different from each other, even increasing the size of the dataset may not be sufficient to achieve high accuracy and recall rates (Cheng *et al*. 2022). Although the models based on measurable parameters proposed in this study have some performance degradation on frontal images, they still perform acceptably and are quite successful compared to other state-of-the-art methods. This shows that the models have the generalisation capability and can be used in different data sets.

The study's novelty lies in using contactless body measurements (e.g., M1, M2, M3) derived from keypoints for pose estimation, combined with supervised machine learning models. The correct placement of key points plays a major role in the success of pose estimation. When the pose PR curves of the test data are examined (Figure 7), it is seen that the mAP value is 0.99 in side images and 0.92 in frontal images in parallel with the OKS values. The correct placement of key points improves the quality of the dataset by increasing the accuracy of body measurements (M1, M2, ...). In such networks, as the quality of the data set increases, the results obtained are more successful. The positive relationship between keypoint placements and body measurements is similarly observed between these two factors and the pose estimates of supervised machine learning algorithms. While 99% correct keypoint placement was achieved in the side images, the KNN algorithm made pose estimation with 100% accuracy (Figure 7 and Table 2). In the frontal images, 92% correct keypoint placement resulted in 94% accuracy of the SVM algorithm in pose estimation (Figure 9 and Table 3). These results show how the correct placement of keypoints improves the quality of the dataset by increasing the accuracy of the body measurements. Thus, it can be seen that the better the quality of the data set, the better the results obtained in such networks.

In the test results on the frontal images, especially when the OKS values of the 'sitting' class are analysed (Figure 9c), it is seen that the model exhibits a lower precision (Figure 9b). This poor performance is related to the fact that many key points of the 'sitting' class are not visible in the frontal images, which leads to low OKS values. Due to the nature of the sitting position, the key points, especially in the leg and foot areas, are not visible, which may have contributed to this decrease in the model's performance. In order to increase the sensitivity to keypoint localisation error, it may be useful to adjust the sigma values used in the OKS calculation (Eq. 6). Too small sigma values mean that the model focuses on localising keypoints more precisely. However, this also requires the model to adhere to stricter constraints, which can complicate the training process (Jocher *et al*., 2023).

The main similarity between the findings of the studies that perform pose classification over images and the present study is that although different object detection models are used, similar mAP values are generally obtained (YOLOv8 mAP 0. 94, Ren *et al*. (2015) Faster R-CNN mAP 0.89, Lin *et al*. (2017) RetinaNet mAP 0.91, Tan *et al*. (2020) EfficientDet mAP 0.93, Hu *et al*. (2023) YOLOv5 mAP 0.92 and Deng *et al*. (2021) YOLOv3 mAP 0.92 for side images). This makes it necessary to combine these object detection models with plug-ins or other models that can improve model performance.

In the present study, the method, which is based on body measurements obtained from images with the use of key points, is not limited to measurements on images. At the same time, videos (Figure 10a) or real-time tracking can be used to quickly detect changes in the position of the sheep due to changes in measurements, so that unusual behaviour can be detected immediately. Furthermore, different measurements can be used to detect various specific behaviours. For example, the measurement value between points 1-9 (Figure 4 and Figure 5) can give an idea about a sheep's self-bending behaviour. Otherwise, changes in the behaviour

of a pregnant ewe as parturition approaches (Waters *et al*., 2021) can be assessed based on measurements between different key points.
The challenges and some limitations encountered in the study can be summarised as follows; the camera was not stable during the creation of the dataset used, which caused shaking, which caused difficulties especially in keypoint labelling. This may have led to some labelling errors in ground truth labelling (Figure 8). Furthermore, the change in performance of the models on datasets containing frontal images shows that the models are sensitive to different camera angles. In particular, it is clear that they perform better in side views where key points can be observed more clearly. In addition, since the distance of each sheep to the camera was not constant (some close to the camera and some far away), there may have been some fluctuations between sheep in the body measurements made on the images. However, this did not affect the ratio of body measurements for each sheep. In real-world scenarios, various practical challenges may arise, such as poor lighting, obstruction caused by swarming behavior, inconsistent camera angles, and motion blur caused by handheld cameras. These issues can reduce keypoint detection accuracy and, consequently, measurement accuracy. To mitigate these issues, future studies could benefit from additional training with augmented and diverse datasets, integration of depth or stereo vision systems, and the use of lightweight models on edge computing platforms. Additionally, incorporating small amounts of locally collected data for fine-tuning could help adapt the model to specific farm conditions.

This study evaluated in detail the relationship between pose estimation and body measurements, as well as the performance of different supervised machine learning algorithms. The findings provide new ways for automatic identification and tracking of animal behaviour with image-based measurements and increase the potential of precision agriculture applications in livestock management. In future work, it is planned to further develop the current 2D keypoint-based approach to obtain 3D pose estimation and body measurements using stereo cameras or depth sensors (such as RGB-D cameras). These technologies will enable real-time tracking of animals in 3D space and more precise measurements. In this way, both pose estimation and body measurements will become more accurate and practical. Furthermore, by working on models that can distinguish behaviours such as walking and running, we aim to provide metric information such as running distance and walking distance. By incorporating additional functions such as weight estimation based on body measurements into our system, we plan to increase the use of this model in precision agriculture practices. These advances will expand and optimise the use of contactless measurement and tracking systems in the livestock industry.

Conclusions

This study presents a two-stage hybrid approach that integrates supervised learning on measurable anatomical parameters with keypoint-based image analysis to classify sheep poses. The proposed method outperforms traditional and deep learning-only models, achieving high classification accuracy across multiple postures and view angles. By incorporating body measurements into the learning process, the model enhances interpretability and generalization capabilities, offering a promising direction for automatic livestock monitoring. Beyond sheep pose classification, the developed system has practical potential in precision agriculture. Its contactless, continuous observation capability makes it suitable for early detection of musculoskeletal or behavioral abnormalities, body weight estimation, and overall health monitoring, thereby contributing to welfare-focused livestock management. From an economic perspective, the system can reduce labor costs by automating daily monitoring tasks and improve overall farm efficiency. The non-invasive nature of the solution, which operates without physical contact, ensures stress-free monitoring while protecting animal privacy, as data collection occurs in controlled farm environments.

References

Alvarenga, F.A.P., Borges, I., Palkovič, L., Rodina, J., Oddy, V.H., Dobos, R.C. 2016. Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. Appl. Anim. Behav. Sci. 181:91-99.

Bati, C.T., Ser, G. 2023a. SHEEPFEARNET: Sheep fear test behaviors classification approach from video data based on optical flow and convolutional neural networks. Comput. Electron. Agr. 204:107540.

Bati, C.T., Ser, G. 2023b. Effects of data augmentation methods on YOLO v5s: application of deep learning with Pytorch for individual cattle identification. YYU J. Agr. Sci. 33:363-376.

Bati, C.T. Ser, G. 2024. Improved sheep identification and tracking algorithm based on YOLOv5 + SORT methods. SIViP 18:6683–6694.

Castillo, P.E., Macedo, R.J., Arredondo, V., Zepeda, J.L., Valencia-Posadas, M., Haubi, C.U. 2023. Morphological description and live weight prediction from body measurements of Socorro Island Merino lambs. Animal 13:1978.

Cheng, M., Yuan, H., Wang, Q., Cai, Z., Liu, Y., Zhang, Y. 2022. Application of deep learning in sheep behaviors recognition and influence analysis of training data characteristics on the recognition effect. Comput. Electron. Agric. 198:107010.

Deng, X., Yan, X., Hou, Y., Wu, H., Feng, C., Chen, L., et al. 2021. Detection of behaviour and posture of sheep based on YOLOv3. Inmateh. Agr. Eng. 64:457-466.

Dutta, T., Dawn, K. 2023. Object keypoint similarity in keypoint detection. Accessed on: 31 July 2024. Available from: https://learnopencv.com/object-keypoint-similarity/#disqus_thread

Forslund, A., Tibi, A., Schmitt, B., Marajo-Petitzon, E., Debaeke, P., Durand, JL., et al. 2023. Can healthy diets be achieved worldwide in 2050 without farmland expansion? Glob. Food. Secur. 39:100711.

Gao, G., Wang, C., Wang, J., Lv, Y., Li, Q. Ma, Y., et al. 2023. CNN-Bi-LSTM: A Complex environment-oriented cattle behavior classification network based on the fusion of CNN and Bi-LSTM. Sensors (Basel) 23: 714.

González-Baldizón, Y., Pérez-Patricio, M., Camas-Anzueto, J.L., Rodríguez-Elías, O.M., Escobar-Gómez, E.N., Vazquez-Delgado, H.D., et al. 2022. Lamb behaviors analysis using a predictive CNN model and a single camera. App. Sci. 12:4712.

Hamadani, A., Ganai, N.A. 2023. Artificial intelligence algorithm comparison and ranking for weight prediction in sheep. Sci. Rep. 13: 3242.

He, C., Qiao, Y., Mao, R., Li, M., Wang, M. 2023. Enhanced LiteHRNet based sheep weight estimation using RGB-D images. Comput. Electron. Agr. 206:107667.

Horie, R., Miyasaka, T., Yoshihara, Y. 2023. Grazing behavior of Mongolian sheep under different climatic conditions. J. Arid. Envion. 209:104890.

Hu, T., Yan, R., Jiang, C., Chand, N.V., Bai, T., Guo, L., Qi, J. 2023. Grazing sheep behaviour recognition based on improved YOLOv5. Sensors (Basel) 23:4752.

Jin, Z., Guo, L., Shu, H., Qi, J., Li, Y., Xu, B., et al. 2022. Behavior classification and analysis of grazing sheep on pasture with different sward surface heights using machine learning. Animals (Basel) 12:1744.

Jocher, G., Chaurasia, A., Qiu, J. 2023. Ultralytics YOLO (Version 8.0.0) [Computer software]. vailable from: https://github.com/ultralytics/ultralytics

Kelly, N.A., Khan, B.M., Ayub, M.Y., Hussain, A.J., Dajani, K., Hou, Y., Khan, W. 2024. Video dataset of sheep activity for animal behavioral analysis via deep learning. Dat. Brief. 52:110027.

Kenyon, P.R., Maloney, S.K., Blache, D. 2014. Review of sheep body condition score in relation to production characteristics. N. Z. J. Agr. Res. 57:38-64.

Khan, B., Kelly, N. 2023a. Video dataset of sheep activity (grazing, running, sitting). Mendeley Data V1. Available from: https://data.mendeley.com/datasets/h5ppwx6fn4/1

Khan, B., Kelly, N. 2023b. Video dataset of sheep activity (standing and walking). Mendeley Data V1. Available from: https://data.mendeley.com/datasets/w65pvb84dg/1

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P. 2017. Focal loss for dense object detection. IEEE T Pattern Anal 42:318-327.

Maji, D., Nagori, S., Mathew, M., Poddar, D. 2022. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition, New Orleans. pp. 2637-2646.

Meckbach, C., Tiesmeyer, V., Traulsen, I. A. 2021. promising approach towards precise animal weight monitoring using convolutional neural networks. Comput. Electron. Agr. 183:106056.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. 2019. PyTorch: an imperative style, high-performance deep learning library. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

Pollard, J., Cox, N., Hogan, N., Huddart, F., Webster, J., Chaya, W., et al. 2004. Behavioural and physiological responses of sheep to shade. MAF Policy Project FMA 123.

Ren, S., He, K., Girshick, R., Sun, J. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. Available from: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf

Ronchi, R.M., Perona, P. 2017. Benchmarking and error diagnosis in multi-instance pose estimation. Proc. IEEE Int. Conf. on Computer Vision, Venice. pp. 369-378.

Schütz, K.E., Saunders, L.R., Huddart, F.J., Watson, T., Latimer, B., Cox, N.R. 2024. Effects of shade on the behaviour and physiology of sheep in a temperate climate. Appl. Anim. Behav. Sci. 272:106185.

Shalaldeh, A., Page, S., Anthony, P., Charters, S., Safa, M., Logan, C. 2023. Body composition estimation in breeding ewes using live weight and body parameters utilizing image analysis. Animals (Basel) 13:2391.

Shorten, C., Khoshgoftaar, T.M. 2019. A survey on image data augmentation for deep learning. J. Big Data. 6:1-48.

Sowande, O.S., Sobola, O.S. 2008. Body measurements of West African dwarf sheep as parameters for estimation of live weight. Trop. Anim. Health Pro. 40:433-439.

Stephansen, R.B., Manzanilla-Pech, C.I., Gebreyesus, G., Sahana, G., Lassen, J. 2023. Prediction of body condition in Jersey dairy cattle from 3D-images using machine learning techniques. J. Anim. Sci. 101:skad376.

Tan, M., Pang, R., Le, Q.V. 2020. Efficientdet: Scalable and efficient object detection. Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, Seattle. pp, 10781-10790.

Van Rossum G., Drake, F.L. 2009. Python 3 Reference Manual. Scotts Valley, CreateSpace.

Waters, B.E., McDonagh, J., Tzimiropoulos, G., Slinger, K.R.., Huggett, Z.J., Bell, M.J. 2021. Changes in sheep behavior before lambing. Agriculture 11:715.

Xu, Y., Nie, J., Cen, H., Wen, B., Liu, S., Li, J., et al. 2023. Spatio-temporal-based identification of aggressive behavior in group sheep. Animals (Basel) 13:2636.

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F. 2022. Image data augmentation for deep learning: A survey. arXiv:220408610.