

---

This is the **accepted version** of the journal article:

Larsen, Mona L. V.; Wang, Meiqing; Willems, Sam; [et al.]. «Automatic detection of locomotor play in young pigs : A proof of concept». Biosystems Engineering, Vol. 229 (May 2023), p. 154-166. DOI <https://doi.org/10.1016/j.biosystemseng.2023.03.006>

---

This version is available at <https://ddd.uab.cat/record/292686>

under the terms of the  license

# Automatic detection of locomotor play in Young pigs: A proof of concept

Mona L.V. Larsen <sup>a,b,\*</sup>, Meiqing Wang <sup>a</sup>, Sam Willems <sup>a</sup>, Dong Liu <sup>a</sup>, Tomas Norton <sup>a,\*\*</sup>

<sup>a</sup> Department of Biosystems, KU Leuven, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium

<sup>b</sup> Department of Animal and Veterinary Sciences, Aarhus University, Blichers All\_e 20, 8830 Tjele, Denmark

Play behaviour is considered an indicator of animal welfare in young pigs. However, as play behaviour events are short-lasting and occur sporadically, continuous monitoring is necessary. This study presents a first attempt at automatic detection of locomotor play behaviour in young pigs from video by classifying locomotor play from other solitary behaviours including standing, walking, and running. Two methods were developed, compared, and sequentially combined: (1) a less computational heavy method utilising the Gaussian Mixture Model for quantification of movement combined with the calculation of contour features and standard machine learning classifiers (*FEATURES*); (2) a computational heavy method utilising a deep learning classifier taking both spatial and temporal features into account (*DEEP*).

The *DEEP* classifier outperformed the *FEATURES* classifier and obtained values of internal validation recall, precision, and specificity of 94%, 88% and 96%, respectively. When combining the two classification methods, almost similar performance was retained, whilst 44% of the other behaviours were correctly classified without the need for deep learning methods. The combination thereby decreased the computational power needed to run the algorithm. Thus, locomotor play can be automatically detected in young pigs and the combination of a less computational heavy method with a deep learning method can reduce the computational requirements for the classification and detection of complex behaviours. Future work should focus on the segmentation of single pigs during high-speed activity in order to enable the play detection algorithm to work in real-life settings.

## Introduction

Animal welfare refers to the quality of an animal's life as it is experienced by the animal itself. Therefore, only the animal's own perception of its condition is important. Current animal welfare quality schemes used across EU countries still rely mostly on resource-based indicators (Stygar et al., 2022). However, it is clear that only animal-based indicators, including behaviour and physiology, can truly measure how the animal perceives its conditions. The use of technology to monitor the animals throughout the production period has potential to assist in the production of continuous and objective data of animal-based indicators, a process which is not possible to do in practice with manual audio-visual observation. This data-driven approach can ultimately ensure better welfare for production animals.

Play behaviour has the potential to be indicative of both the lack of negative experiences and also the presence of positive experiences. Play is often expressed in the absence of welfare and health threats and it can drop out of the animal's behavioural repertoire in more challenging conditions (Held & Špinka, 2011). Furthermore, play seems to be a self-rewarding behaviour, meaning that the reward to the animal is in the expression of the behaviour itself (Held & Špinka, 2011). In young pigs, the performance of play behaviour has, for example, been shown to decrease immediately after weaning (Franchi, Larsen, Winters, Jensen, & Pedersen, 2022) which is one of the most stressful experiences in the life of a growing pig. Also, when air quality was found to reduce, with ammonia levels rising from 5 to 20 ppm, the performance of play behaviour in young pigs was found to decrease (O'Connor et al., 2010). Multiple studies have also shown that increased space and the provision of certain enrichment can increase the performance of play (e.g., Nakamura, Tanaka, Nishida, & Uetake, 2011;

Ocepek, Newberry, & Andersen, 2020; Oostindjer, van den Brand, Kemp, & Bolhuis, 2011). The performance of play in young pigs has also been positively related to weight gain, feeding and drinking behaviour, indicating that better nourished pigs perform more play (Franchi et al., 2023).

The Welfare Quality® protocol for pigs already encourages the use of more animal-based indicators, including complex behaviours, in welfare assessments (Welfare Quality, 2009). However, these complex behaviours are currently assessed by scanning the pen with 2 min intervals for a 10 min period and noting how many pigs perform a specific behaviour at each scan; a method not suitable for short-lasting and sporadically occurring behaviours such as aggression and play behaviour. However, this instantaneous sampling method is often used when observing play behaviour in research studies. Moreover, due to the very high workload of manual observation, play behaviour is often only observed for a few days, with the observation lasting just over a few hours during each of the days. Since pigs show changes in their activity based on age and diurnal patterns (Jensen, Toft, & Kristensen, 2017; Villagrà et al., 2007) ideally their behaviour should be monitored continuously during the day and across the production period. This emphasises the importance of developing automatic methods to observe play behaviour in pigs. Only with automatic continuous methods can we get a true insight into the effects of play behaviour in pigs, its development with time and across the day, and its negative and positive relationships to animal welfare.

Pigs perform four types of play including social play, locomotor play, rotational play, and object play. The three latter ones are solitary behaviours, although solitary play behaviour can be contagious (Held & Špinka, 2011). Recent studies have attempted to automatically detect the social behaviours in pigs, but without distinguishing social play from other social interactions (Gan et al., 2021; Rodriguez-Baena et al., 2020). A recent study developed an algorithm for the automatic detection of object engagement in pigs (Chen, Zhu, Steibel, et al., 2020) but to the authors knowledge, no studies have so far attempted to automatically detect locomotor play in young pigs, although this is probably the type of play performed the most by pigs (Newberry, Wood-Gush, & Hall, 1988).

The current study presents a first attempt on developing an algorithm for the automatic detection of locomotor play in young pigs with the following aims: (1) to test the performance of a low computational footprint method utilising the Gaussian Mixture Model (GMM) for quantification of movement combined with standard machine learning classifiers; (2) to test the performance of a computational heavy method utilising a deep learning classifier (CNN + LSTM) on the raw segmented video; (3) to compare the two methods; and (4) to combine the two classifications to decrease the computational power needed, by excluding non-play events that are classified correctly by the first method from being input to the deep learning classifier.

## Materials and methods

The animal experimental procedures and care of animals under study were carried out in accordance with the Ministry of Food, Agriculture and Fisheries, The Danish Veterinary and Food Administration under act 474 of 15. May 2014 and executive order 2028 of 14. December 2020, and under consideration of the Arrive Guidelines (Du Sert et al., 2020).

### Animals and housing

The animals used in the current study were part of a larger study conducted from February to November 2020 at the experimental pig herd, Aarhus University, Foulum, Denmark, investigating alternative weaning management to decrease weaning stress in pigs (Winters et al., 2023). The current study used data from six pens across two rounds of pigs (round 1: March 2020; round 2: August 2020). The pens varied in weaning age, average weaning weight, group size as well as the genetic hybrid of the sow giving birth to the pigs of a particular pen (see Table 1). Video data were available on day 1 and 2 after weaning (day0), with each litter of pigs being weaned in their home pen with the sow being removed. The days that were included for each pen can be seen in Table 1.

Each home pen was a farrowing pen designed for loosehoused sows (FT30, Jyden A/S, Denmark), measuring 3.0 × 2.2 m with 41% slatted and 59% solid concrete flooring. Each pen was equipped with a covered piglet creep area (0.9 m<sup>2</sup>; not heated after weaning), a low 800 × 280 mm polyconcrete feed trough (Jyden A/S, Denmark) providing access to a pelleted standardised weaner diet (Prime Midi Piller, DLG, Denmark, metabolisable energy: 14.8 MJ kg<sup>-1</sup>, 19.3% crude protein), with no medical zinc oxide added to it, delivered once daily in the morning, and a 310 × 170 × 110 mm pig water trough (Aqua-Level system with hinged trough, Jyden A/S, Denmark) positioned next to the feed trough on the slatted floor providing ad libitum access to water. Every morning after weaning, all pens were provided with approximately 130 g of chopped wheat straw on the solid floor and approximately 400 g of sawdust in the creep area. The room temperature was 24 °C and the artificial light was off from 22:00 h to 07:00 h. No pigs were treated with antibiotics until day 3 after weaning.

### Play behaviour data

For each pen, a 2D camera (HIKVISION, Hangzhou, China, model DS-2CD2145FWD-I, 2.8 mm lens) was installed 2.3 m above the floor providing a top-down view of the entire pen, except for under the creep area (see Fig. 1). Video was recorded continuously throughout the day using the Blue Iris 5 software (<https://blueirissoftware.com/>) with a resolution of 1280 × 720 pixels and 15 frames s<sup>-1</sup> and stored on a HP Z1 Entry Tower G5 computer with i7-8700 processor, 16 GB ram and an extra internal hard drive of 4 TB. The recorded videos were backed-up on 2 TB external hard drives and stored for later analysis.

Video data was labelled for locomotor play on dayp1 and dayp2 relative to weaning (day 0 at 14:00 h) from 14:00 to 22:00 h as these were the hours with artificial light and without disturbance from other experimental procedures or general care taking of the pigs. See Table 2 for the definition of locomotor play used in the current study. Locomotor play was labelled using the Noldus Observer XT 15 (Noldus Information Technology, The Netherlands) software package. This included behavioural sampling and continuous recording (Bateson & Martin, 2021) of the individual pig using exact start and stop times. Labelling of the other behaviours included in Table 2 are described in section 2.3.3. After a descriptive analysis of the labelled play behaviour data, it was chosen to focus on the evening hours from 18:00e22:00 h as 80% of locomotor play occurred during those hours. Furthermore, in two of the included pens, the pigs performed almost no locomotor play on dayp1 and thus, this day was not included for

those specific pens (see Table 1). A total of 40 h of video data was labelled for developing and validating the algorithm.

### **Play detection algorithm**

All coding were performed in the Python 3.9.7 environment. CNN and LSTM were implemented with Keras 2.6 using TensorFlow 2.6 as the backend. All calculations were performed on a computer with an AMD Ryzen 9 3900X 12-Core CPU 3.80 GHz with 32 GB RAM memory running a Microsoft Windows 11 64-bit, \_64-based processor operating system. The graphic card was NVIDIA RTX 2080-Ti with 11 GB of physical memory. See Supplementary Video 1 for a visualisation of the developed play detection algorithm. Supplementary video related to this article can be found at <https://doi.org/10.1016/j.biosystemseng.2023.03.006>

### **Quantification of movement**

Prior to video processing, 40 h-long videos were split into halfhour video sequences (N = 80). To quantify behavioural intensity, a GMM algorithm (Zivkovic & Van Der Heijden, 2006) was applied to each half-hour video, utilising the `cv2.createBackgroundSubtractorMOG2()` function of the OpenCV library including shadow detection (shadow is considered as noise due to artificial lighting condition rather than pig movement). The GMM algorithm classifies static pixels as background and changing pixels as foreground, thereby assigning each background, shadow, and foreground pixel an intensity value of 0, 127 and 255, respectively. In the current setting, only the pigs are moving and thus, the GMM algorithm results in frames with only the moving part of pigs having a pixel intensity of 255. To remove shadows, all pixel intensity values of 127 were changed to zero values, making the shadows part of the background. The first 500 frames of each video sequence were used to initialise the GMM algorithm and thus, not used in the further analysis. Afterwards, the background detection is updated, depending on the set learning rate (LR): the lower the LR, the less often is the background updated, meaning that a lower LR will result in a more detailed movement detection. In the current study, a LR = 0.001 was used to produce 'all activity' video sequences, while a LR = 0.02 was used to produce 'high activity' video sequences. To smooth the edges of each pixel, each frame of each video sequence was eroded with a 3 × 3 kernel. 'Low activity' video sequences were then produced by subtracting the 'high activity' from the 'all activity' video sequence. The raw, 'low activity' and 'high activity' halfhour video sequences were cropped into 1-sec video sequences. For each 1-sec, the respective 'low activity' (pixel intensity = 85) and 'high activity' (pixel intensity = 170) video sequences were summarised into 15-frames (1 s) movement maps, with dark grey pixels representing low activity (pixel intensity = 85), light grey pixels representing high activity (pixel intensity = 170), and white pixels representing an overlap of low and high activity within the 15 frames (pixel intensity = 255). The concept of using the GMM to extract high and low activity in order to derive features of animal behaviour follows the work of Willems et al. (2022). See Fig. 1 for an example of a movement map. The creation of the movement maps by means of GMM on half hour videos required 0.35 s of computational time per 1-sec video (15 frames). Both the 1-sec raw video sequences and the connected movement maps were used in the further analysis.

**Table 1 – Characteristics of pens included in the study (TN = Topigs Norsvin), which days relative to weaning (day0) were included for each pen, and whether the pen was used as train or test data.**

Round	Pen	Hybrid	Weaning age (d)	Avg. weaning body mass (kg)	Group size	Day+1	Day+2	Data
1	1	Danbred	25	6.64	12	x	x	Train
1	2	TN-70	28	9.30	10	x	x	Train
1	3	TN-70	27	7.13	15	x	x	Train
2	1	Danbred	25	6.04	10	x	x	Train
2	2	TN-70	25	7.96	14	x	x	Train
2	3	TN-70	24	5.89	15	x	x	Test



**Fig. 1 e** Example of a movement map including a pig performing locomotor play. Left: first raw frame (out of 15) used to construct the movement map. Right: 15 frames (1 s) movement map with the green solid lines representing the contours detected and the red rectangles representing the minimum area rectangle for each contour.

**Table 2 – Description of the behaviours labelled from the video recording. Locomotor play behaviour was labelled continuously from the full video, while the other behaviours were labelled from 1 s contour videos (see section 2.3.3.).**

Behaviour	Description	Adapted from
Standing	Movement of the body and no more than three legs within the 1-s video.	
Walking	Slow forward movement with one leg at a time, moving all four legs once within the 1-s video.	
Running	Fast forward movement without hops or gallop-like movement, moving the legs more than once within the 1-s video (running) or short-lasting exaggerated movements.	
Locomotor play	At least 2 forward hops or gallop-like energetic forward movement (the two forelimbs move in phase, followed by the two hind limbs). Often associated with vigorous ear flapping, moving across a large area of the pen, and occasionally bouncing into other pigs.	Donaldson, Newberry, Špinková, and Cloutier (2002) Brown, Klaffenböck, Nevison, and Lawrence (2015) Martin, Ison, and Baxter (2015) Zupan et al. (2016)

### Contour detection, contour features and video masking

Prior to contour detection, each movement map was median blurred with a 15 × 15 kernel and dilated with a 3 × 3 kernel of ones to close the contours. Contours were detected using Canny edge detection (thresholds of 30 and 150). For each contour in the movement map with a contour area above 5000 pixels and that did not only consist of low activity pixels (dark grey, pixel intensity  $\frac{1}{4}$  85), the minimum area rectangle (MAR) was produced. The MAR was used to segment the movement map into contour images and from these, the following six contour features were calculated for each contour image from each movement map (Eqs. (1) e (6)):

$$\text{Aspect ratio} = \frac{\text{MAR}_{\text{length}}}{\text{MAR}_{\text{width}}} \quad (1)$$

$$\text{Extent} = \frac{\text{Contour}_{\text{area}}}{\text{MAR}_{\text{area}}} \quad (2)$$

$$\text{Mean intensity} = \frac{\sum_1^{\text{Contour area}} \text{Intensity}}{\text{Contour area}} \quad (3)$$

$$\text{Prop}_{\text{Low activity}} = \frac{\text{No. low activity pixels}}{\text{Contour area}} \quad (4)$$

$$\text{Prop}_{\text{High activity}} = \frac{\text{No. high activity pixels}}{\text{Contour area}} \quad (5)$$

$$\text{Prop}_{\text{Overlap}} = \frac{\text{No. overlap pixels}}{\text{Contour area}} \quad (6)$$

The contour area and MAR area was the number of pixels counted within each contour and MAR, respectively. See Fig. 1 for image schematics of a contour, MAR, and MAR length and width. These contour features were chosen as they are all relative measures and may not depend on the size of the pig or the distance between the camera and the pigs; thus, making the image features invariant from scaling, rotating transformations, and expecting better generalisation capabilities. The MAR was also used to segment the 1-sec raw videos into contour videos. For each 1-sec video sequence, the result was a number of contour videos paired with the calculated contour features; in the following referred to as contours. The contour features were used as input to the contour features classifiers as described in section 2.3.5, whilst the contour videos were used as the input to the contour videos deep learning classifiers described in section 2.3.6.

### Labelling of contours

Locomotor play is a solitary play type, meaning the pig plays alone, although the performance may trigger play in penmates. As described in Table 2, locomotor play may also involve the playing pig bumping into other pigs. Thus, it was decided to include only contours with one pig playing and contours with one pig bumping into other pigs as a result of locomotor play. Which contours that fulfilled these criteria was determined by watching the contour videos for the periods labelled as locomotor play as explained in section 2.2. It was decided to classify these play contours from other solitary pig behaviour based on the levels of activity, i.e., standing, walking, and running, as described in Table 2. From the pool of constructed contours, 7000 contours were semi-randomly chosen for labelling the other solitary pig behaviours as described in Table 3. The chosen contours were labelled according to the ethogram in Table 2 by the labeller observing the contour videos. The result was 2438 standing contours, 792 walking contours and 62 running contours, together with 855 locomotor play contours.

### Training, validation, and test data

Of the data available, five pens and eight days were used to train the play detection algorithm, whilst one pen and two days (not seen by the algorithm) were used as test data for internal validation of the developed play detection algorithm (see Table 1). The training data was further divided into training and validation data with an approximately 75:25 split. This split was

done separately for the contour features (section 2.3.5) and contour videos classification (section 2.3.6) and separately for each of the four data combinations explained below. As the focus of this study was to classify locomotor play from other solitary behaviours, and not to classify the three other behaviours from each other, it was decided to not use multiclass classification. Instead, a step-wise binary classification was done in four steps resulting in four data combinations: (1) Standing: classification of locomotor play from standing, (2) Walking: classification of locomotor play from walking, (3) NoRunning: classification of locomotor play from standing and walking, (4) WithRunning: classification of locomotor play from standing, walking, and running. Number of contours included for each behaviour within each data combination can be seen in Table 5 and Table 6. For each data combination, the behaviours not being locomotor play were labelled 'Other'. For training the classifiers, the number of contours was balanced between the labels of 'Play' and 'Other', whilst the test data always included 218 play contours and 465 standing, 152 walking and 21 running contours, depending on the data combination.

### **Contour features classification**

For each of the four data combinations (Standing, Walking, NoRunning, WithRunning), the contour features together with their respective label (Other, Play) were used as inputs to five different classifiers: K-Nearest Neighbour, Logistic Regression, Naive Bayes, Support Vector Machines, and Random Forest. Four-fold cross validation was used for hyperparameter tuning of each classifier on the training data and the levels of tuning of each hyperparameter can be seen in Table 4. The hyperparameter combination with the highest cross-validation accuracy was chosen for training each of the five classifiers and the classifier with the highest training accuracy was used for internal validation on the test data (see Table 5). The number of contours in the training and test data for each data combination can be seen in Table 5.

### **Contour videos classification**

The contour videos together with their respective label (Other, Play) were used as inputs to a deep learning classifier being trained by extracting both spatial (EfficientNet-B0 convolutional neural network; CNN; Tan & Le (2019)) and temporal (long short-term memory, LSTM) features. In this context, the GMM and contour detection replaced tracking-by-detection in the production of the contour videos. The contour videos were used instead of the full-frame videos to include only the relevant information for the classification, being the playing pig. Prior to training and testing of the deep learning classifier, the contour videos were processed to a frame resolution of 224 \_ 224 pixels with 3 colour channels, independent of their original resolution, to match the input demand of the EfficientNet-B0 architecture. To increase data variation, contour videos in the training data were augmented by rotating each frame of each contour video 90, 180 and 270°, resulting in four times as many data than used for the contour features classification. The CNN model extracted spatial features frame by frame (15 frames per video) for each contour video with an input layer of 224 \_ 224 \_ 3 pixels, a 1000-dimensional feature vector as output layer and the EfficientNet-B0 architecture and scaling in between, with pretrained ImageNet weights. For the CNN model, stochastic gradient descent with a learning rate of 0.0001 was used as optimiser and binary cross entropy as the loss function. Multiple state-of-the-art EfficientNet architectures were tested (B0, B3, and B7). EfficientNet is a combined CNN architecture and scaling method that uniformly scales network

**Table 3 – Number of contours chosen for labelling of the other behaviours divided between the 10 combinations of pen and day, based on rules ensuring variation in the contours to get contours with standing, walking, and running pigs.**

No. contours	No. contours per pen and day	Rule
1000	100	$\text{Prop}_{\text{High activity}} < 0.10$
2000	200	$0.10 \geq \text{Prop}_{\text{High activity}} < 0.30$
2000	200	$0.30 \geq \text{Prop}_{\text{High activity}} < 0.60$
2000	200	$0.60 \geq \text{Prop}_{\text{High activity}}$

**Table 4 – Levels of hyperparameters tested and chosen for each of five classifiers: K-Nearest Neighbour (KNN), Logistic Regression (Logit), Naive Bayes (NaiveB), Support Vector Machines (SVM), and Random Forest (RF), and the four data combinations (Standing, Walking, NoRunning, WithRunning). The hyperparameter combination with highest 4-fold cross validation accuracy were chosen and used for training each of the five classifiers.**

Classifier	Hyperparameter	Levels tested	Standing	Walking	NoRunning	WithRunning
KNN	n_neighbours	[1, 3, 5, 10, 20]	10	10	20	20
	weights	[uniform, distance]	distance	distance	uniform	distance
	metric	[euclidean, manhattan, minkowski]	Euclidean	Euclidean	Euclidean	Euclidean
Logit	solver	[newton-cg, lbfgs, liblinear]	liblinear	liblinear	liblinear	liblinear
	penalty	[l2]	l2	l2	l2	l2
	C	[0.01 to 100 * 10]	100	100	100	100
NaiveB	var_smoothing	[1e-9 to 1 * 10]	0.001	0.001	0.1	0.001
SVM	kernel	[poly, rbf, sigmoid, linear]	poly	poly	linear	linear
	C	[0.01, 0.1, 1.0, 10, 50]	10	10	10	10
	gamma	[scale, auto]	scale	scale	scale	scale
RF	n_estimators	[1, 10, 50, 100]	100	50	100	10
	criterion	[gini, entropy]	gini	entropy	entropy	entropy
	max_features	[sqrt, log2]	log2	log2	log2	sqrt

**Table 5 – Performance metrics of the contour features classifiers with a classification threshold of 0.5, on each of the four data combinations (Standing, Walking, NoRunning, WithRunning). Only the classifier with the highest training accuracy within each data combination is presented.**

Training data	Standing	Walking	NoRunning	WithRunning
No. play contours	637	637	637	637
No. standing contours	650	–	320	300
No. walking contours	–	640	320	300
No. running contours	–	–	–	41
Model	SVM	Logit	Logit	Logit
Accuracy	0.919	0.850	0.847	0.834
Test data	Standing	Walking	NoRunning	WithRunning
No. play contours	218	218	218	218
No. other contours <sup>a</sup>	465	152	617	638
Recall <sup>b</sup>	0.779	0.463	0.766	0.766
Precision <sup>b</sup>	0.817	0.944	0.746	0.699
f1 score <sup>b</sup>	0.798	0.622	0.756	0.731
Specificity	0.918	0.961	0.908	0.887
TP <sup>c</sup>	170	101	167	166
FN <sup>c</sup>	48	117	51	52
FP <sup>c</sup>	38	6	57	72
TN <sup>c</sup>	427	146	560	566

<sup>a</sup> Other contours include the behaviours: standing, walking and/or running, depending on the test data.

<sup>b</sup> Recall, precision and f1 score for the play class.

<sup>c</sup> TP: True positives – number of play contours predicted as play contours; FN: False negatives – number of play contours predicted as other contours; FP: False positives – number of other contours predicted as play contours; TN: True negatives – number of other contours predicted as other contours.

width, depth, and resolution (Tan & Le, 2019). The EfficientNet-B0 architecture with the least number of parameters resulted in better performance for the current classification challenge and was therefore chosen for reporting. The output vector of the CNN model for each frame was used as input to the LSTM model. The output of the LSTM model was a 2-dimensional vector resulting from a SoftMax activation function and representing the classification of the two labels: Other, Play. For the LSTM model, RMSprop with a learning rate of 0.00005 was used as optimiser and binary cross entropy as the loss function. For each of the four data combinations (Standing, Walking, NoRunning, WithRunning), the best combination of number of epochs [20, 50, 70, 100] and batch size [8, 16, 32, 64] was evaluated based on training and validation accuracy/loss curves to get the highest validation accuracy while avoiding under- and overfitting. The model with the best combination of number of epochs and batch size for each data combination was used to internally validate the contour videos classifier on the test data. The number of contour videos in the training and test data for each data combination can be seen in Table 6.

### Combining classifiers

To potentially increase the performance of the play detection algorithm, the two types of classifiers (contour features and contour videos) were combined. The contour features classifiers trained and selected for each of the four data combinations (Standing, Walking, NoRunning, WithRunning) were applied to the WithRunning test data with a classification threshold optimised for a recall of at least 0.98 (see Table 7).

**Table 6 – Performance metrics of the contour videos classifiers (CNN + LSTM deep learning algorithm) on each of the four data combinations (Standing, Walking, NoRunning, WithRunning).**

Training data	Standing	Walking	NoRunning	WithRunning	
No. play videos	2548	2548	2548	2548	
No. standing videos	2600	–	1250	1200	
No. walking videos	–	2560	1250	1200	
No. running videos	–	–	–	164	
Epochs	70	20	50	20	
Batch size	64	16	8	8	
Validation loss	0.094	0.161	0.146	0.197	
Validation accuracy	0.970	0.938	0.950	0.921	
Test data	Standing	Walking	NoRunning	WithRunning	WithRunning
No. play videos	218	218	218	218	218
No. other videos <sup>a</sup>	465	152	617	638	638
Test time (s/video) <sup>b</sup>	0.763	0.776	0.799	0.758	0.826
Recall <sup>c</sup>	0.991	0.931	0.936	0.936	0.954
Precision <sup>c</sup>	0.896	0.956	0.919	0.879	0.852
f1 score <sup>c</sup>	0.941	0.942	0.927	0.907	0.900
Specificity	0.946	0.934	0.971	0.956	0.944
TP <sup>d</sup>	216	203	204	204	208
FN <sup>d</sup>	2	15	14	14	10
FP <sup>d</sup>	25	10	18	28	36
TN <sup>d</sup>	440	142	599	610	602

<sup>a</sup> Other videos include the behaviours: standing, walking and/or running, depending on the test data.

<sup>b</sup> Video of 1 s duration.

<sup>c</sup> Recall, precision and f1 score for the play class.

<sup>d</sup> TP: True positives – number of play videos predicted as play videos; FN: False negatives – number of play videos predicted as other videos; FP: False positives – number of other videos predicted as play videos; TN: True negatives – number of other videos predicted as other videos.

The result was a low number of false negatives but a high number of false positives; meaning that few locomotor play events were lost, whilst a high number of other behaviour events were retained. The respective contour videos classifiers for the four data combinations were then applied to the true and false positives (the events classified as play by the contour features classifiers) for a second round of classification. Finally, the two classification rounds were summed resulting in the total number of true and false positives and negatives. Besides increasing the performance of the play detection algorithm, this combination strategy was also chosen to lower the number of contour videos to be classified by the deep learning classifier and thereby reduce the computational power needed.

### Performance metrics

As training of all classifiers was carried out on a balanced data set of contours with respect to the two labels: Other and Play, Accuracy was used as the performance metric as defined in Eq. (7). As the test data set was unbalanced, the performance metrics used in the internal validation were Recall, Precision, f1 score and Specificity as defined in Eqs. 8e11.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{f1 score} = \frac{2 * (\text{Recall} * \text{Precision})}{\text{Recall} + \text{Precision}} \quad (10)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (11)$$

Where TP is the number of true positives meaning the number of correctly classified locomotor play events; TN is the number of true negatives meaning the number of correctly classified other behaviour events; FP is the number of false positives meaning the number of other behaviour events incorrectly classified as locomotor play events; and FN is the number of false negatives meaning the number of locomotor play events incorrectly classified as other behaviour events. Accuracy represents the proportion of correct classifications of both locomotor play and other behaviour events. Recall and Specificity represent the proportion of correctly classified locomotor play events and other behaviour events, respectively. Precision represents the proportion of play classifications that actually were locomotor play events, while the f1 score combines the Recall and Precision into one metric.

## Results and discussion

### Contour features classification

The performance of the contour features classifiers can be seen in Table 5. The best performance was obtained when classifying locomotor play from standing with a training accuracy of 0.92, and an internal validation f1 score and specificity of 0.80 and 0.92, respectively. This was expected, as standing contours differed from play contours especially when considering

aspect ratio (mean  $\pm$  SD; Standing:  $2.00 \pm 0.5$ ; Play:  $2.99 \pm 1.2$ ), mean intensity (Standing:  $133 \pm 28$ ; Play:  $163 \pm 17$ ) and the proportion of low activity pixels (Standing:  $0.44 \pm 0.29$ ; Play:  $0.12 \pm 0.12$ ); differences less pronounced between locomotor play, walking, and running contours. When including walking and running, the performance decreased to a training accuracy of 0.83. During internal validation on the WithRunning test data, 76% percent of locomotor play events were correctly classified, while 70% of the play classifications represented locomotor play events. Furthermore, 89% of the other behaviour events were correctly classified. Interestingly, allowing the classifier to train on contour features of running, did not improve the performance, although this could be due to the low number of running contours in the training data ( $n = 41$ ).

**Table 7 – Internal validation performance metrics when combining the contour features classifier (first round) and the contour videos classifier (second round) into a final classification on the WithRunning test data.**

First classification round	Standing	Walking	NoRunning	WithRunning
Classification threshold <sup>a</sup>	0.05	0.05	0.06	0.12
Included events				
Play (TP)	209	214	210	208
Other (FP)	351	391	358	296
Excluded events				
Play (FN)	9	4	8	10
Other (TN)	287	247	280	342
% Other	45	39	44	54
Sum of the two classification rounds				
Test time (s/video) <sup>b</sup>	0.499	0.540	0.506	0.452
Recall <sup>c</sup>	0.936	0.922	0.890	0.890
Precision <sup>c</sup>	0.703	0.874	0.898	0.878
f1 score <sup>c</sup>	0.803	0.897	0.894	0.884
Specificity <sup>d</sup>	0.865	0.955	0.966	0.958
TP <sup>e</sup>	204	201	194	194
FN <sup>e</sup>	14	17	24	24
FP <sup>e</sup>	86	29	22	27
TN <sup>e</sup>	552	609	616	611

<sup>a</sup> Classification threshold resulting in a recall of at least 0.98.

<sup>b</sup> Video of 1 s duration.

<sup>c</sup> Recall, precision and f1 score for the play class obtained by summing the false negatives of the first and second classification round.

<sup>d</sup> Specificity obtained by summing the true negatives of the first and second classification round.

<sup>e</sup> TP: True positives – number of play videos predicted as play videos; FN: False negatives – number of play videos predicted as other videos; FP: False positives – number of other videos predicted as play videos; TN: True negatives – number of other videos predicted as other videos.

## Contour videos classification

Training curves of loss and accuracy for each of the four contour videos classifiers (Standing, Walking, NoRunning, WithRunning) can be seen in Fig. 2, illustrating that neither underfitting nor overfitting occurred with the chosen number of epochs and batch sizes (see Table 6). The performance of the contour videos classifiers can be seen in Table 6. A similar pattern as with the contour features classifiers was seen with the best performance obtained when classifying locomotor play from standing. Again, this was expected, as standing was defined as the pig keeping a leg still within the 1 s video, whereas pigs performing at least two forward hops were labelled as performing locomotor play. Internal validation on the WithRunning test data showed the best performance when using the NoRunning classifier. A similar performance could probably be obtained with the WithRunning classifier, if more running events were included in the training data, thereby allowing the classifier to learn more subtle differences between running and locomotor play events. Ninety-four percent of locomotor play events were correctly classified, while 88% of the play classifications represented locomotor play events. Furthermore, 96% of the other behaviour events were correctly classified.

Deep learning algorithms can also be defined as data-based models since they are characterised as being highly adapted to the data they are trained on, meaning that they can suffer difficulties when generalising unseen data. In the current study, internal validation showed satisfying generalisation abilities. This means that the developed classifier performed well on data from the same environment but including new animals. This may be explained by that the contour videos which zoomed in on a single pig, only included the relevant information, and that the internal validation data was obtained from an unseen pen of the same design and from pigs of similar age and weight. However, this situation may not be the same for external validation where the classifier would be tested on data including e.g., an unknown environment, other lighting conditions and other ages/weights of the pigs. It was hoped that the contour features classifier could provide such generalisation abilities, but unfortunately it did not perform satisfactorily in the internal validation and had too many false classifications (see Fig. 4). To the authors knowledge, no previous studies have attempted to automatically detect locomotor play in pigs, although there have been attempts in calves utilising accelerometer data (Gladden, Cuthbert, Ellis, & McKeegan, 2020; Rushen & de Passill\_e, 2012). Other recent studies within automatic behaviour classification in pigs have utilised similar deep learning methods as the current one. This include the detection of e.g., object manipulation (Chen, Zhu, Oczak, et al., 2020), aggression (Chen, Zhu, Steibel, et al., 2020), and tail-biting behaviour (Liu et al., 2020). These studies obtained training accuracies of 97%, 97% and 96%, respectively, whilst only the study of Chen, Zhu, Oczak, et al. (2020) conducted internal validation obtaining an accuracy of 97%. Although, the current play detection algorithm had a worse performance than the above-mentioned studies, it does represent a first attempt at classifying a challenging task and it did obtain a validation accuracy of 0.95 with a test recall of 0.94 and a test specificity of 0.96. Overall, the current study provides a proof of concept that locomotor play can be classified from other behaviours using automatic deep learning methods. Future continuation of the work is needed to make the play detection algorithm operate in a real-life setting (see section 3.5).

### **Comparing classifiers**

As the inclusion of the running events in the training of the classifiers seemed to lower their performance, it was decided to compare the NoRunning classifiers when applied to the WithRunning test data. The performance metrics of the contour features classifier, contour videos classifier and their combined classification can be seen in Fig. 3. The contour videos classifier greatly outperformed the contour features classifier with a higher recall, higher precision, and higher specificity. As locomotor play is a short-lasting, complex behaviour and is similar to (for example) aggression, it is not surprising that more complex deep learning methods, taking both spatial and temporal features into consideration, are needed to detect locomotor play. This is unlike longer-lasting and location specific behaviours such as drinking, feeding and posture detection where simpler computer-vision techniques may suffice.

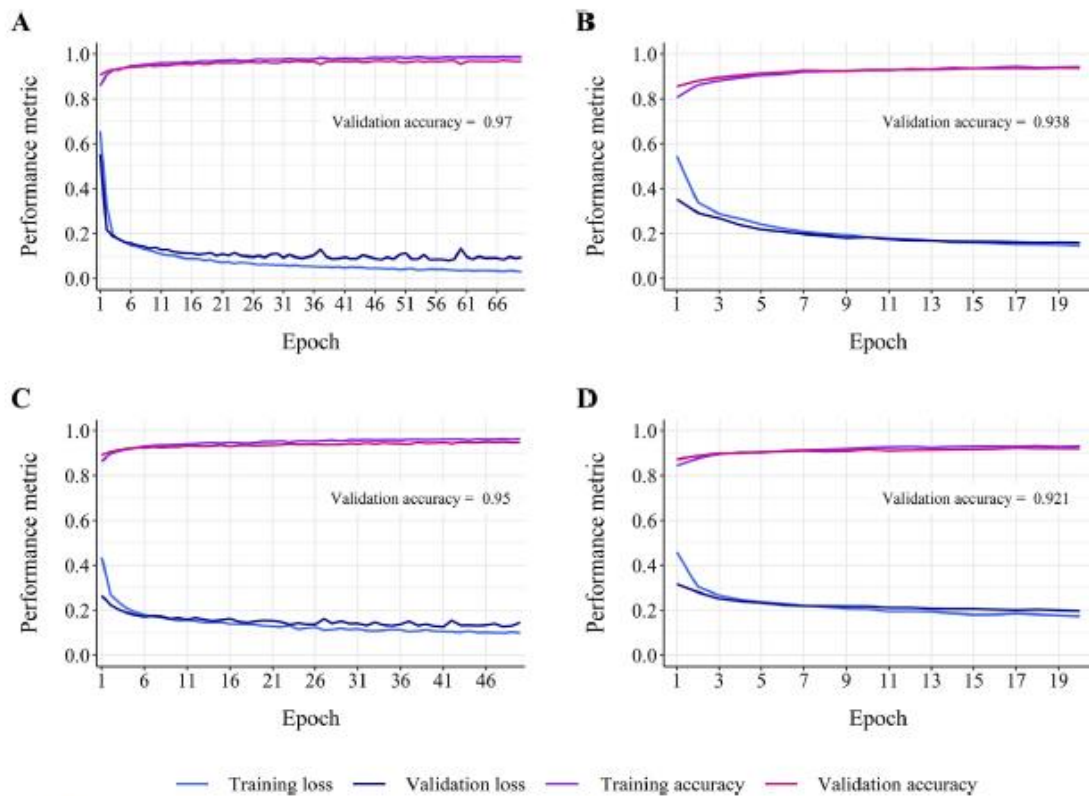


Fig. 2 – Training curves of each of the four contour videos classifiers (CNN + LSTM): A - Standing, B - Walking, C - NoRunning, D - WithRunning.

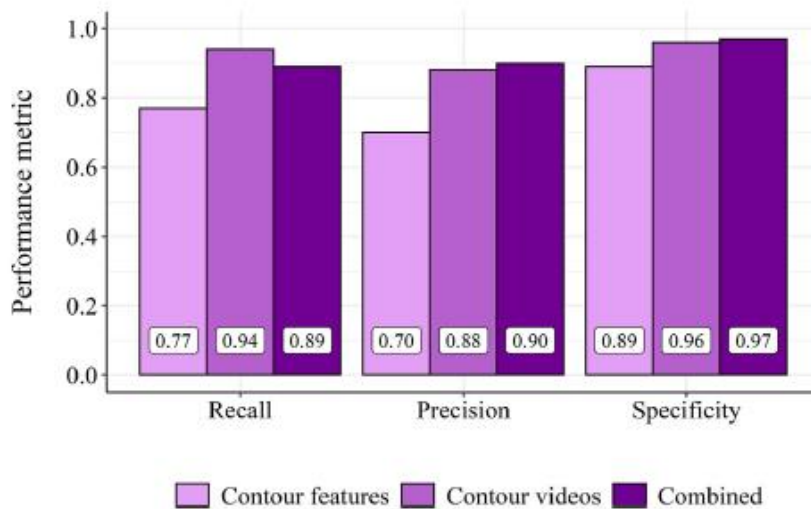
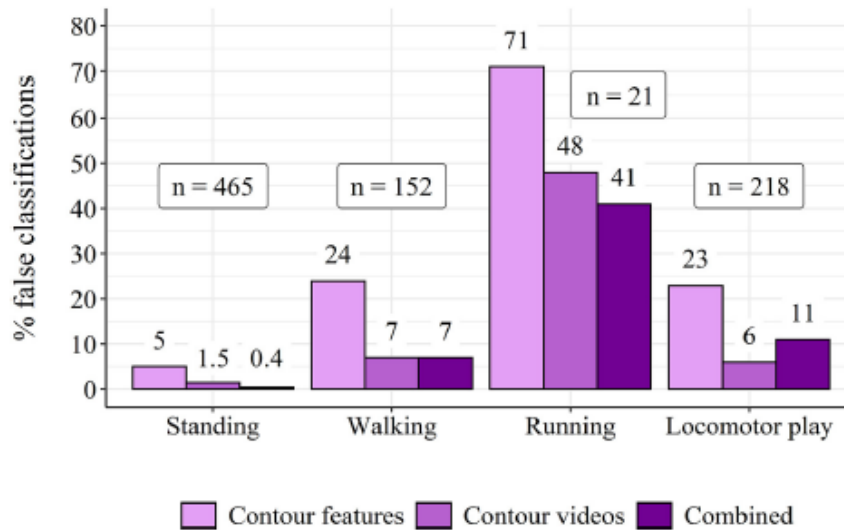


Fig. 3 – Recall, precision and specificity of the contour features classifier, contour videos classifier (CNN + LSTM) and their sequentially combined classification when trained on the NoRunning data combination and internally validated on the WithRunning test data.



**Fig. 4 – Percentage of false classifications of standing, walking, running, and locomotor play events (n = number of events) for the contour features classifier, contour videos classifier (CNN + LSTM) and their sequentially combined classification when trained on the NoRunning data combination and internally validated on the WithRunning test data.**

The higher specificity and precision indicate a lower percentage of false classifications for the other behaviours as can also be seen in Fig. 4. The specificity is an especially important metric in this classification task and should be prioritised (Dominiak & Kristensen, 2017). Locomotor play, being the minority behaviour, is performed much less than many other behaviours in young pigs. Thus, it is important to correctly classify the other behaviours (standing, walking, running) to minimise the number of false positives that otherwise could result in a large overestimation of the duration of locomotor play. In the current case, the percentage of false classifications decreased for all the behaviours when turning to deep learning methods (see Fig. 4). The decrease in false classifications for standing events from 5% to 1.5% is especially important. In this small-scale investigation, it may seem like a small difference. However, when taking into consideration that most behaviour performed by the young pigs will fall into the standing category, this difference will have a huge impact on the number of false positives in a real-life setting and thereby on the estimation of locomotor play duration. Although both classifiers struggled with the classification of locomotor play from running, the percentage of false classifications of running decreased from 71 to 48% when turning to deep learning methods. To distinguish, for example, the difference between running from locomotor play can be very difficult even for the human observer and therefore it is not surprising that this was the area where the play detection algorithm struggled the most, independent of methodology used.

### Combining classifiers

The two classifiers (contour features and contour videos) were sequentially combined to potentially improve performance even more. Detailed performance results of the combined classification can be seen in Table 7. Again, inclusion of the running contours in the training data seems to decrease the overall performance. For the NoRunning classifier, the combined classification detected 89% of locomotor play events correctly, while 90% of the play classifications represented locomotor play events. Furthermore, 97% of the other behaviour events were correctly classified. The combined method led to lower recall than the sole contour videos classifier (Fig. 3) and thereby a higher percentage of false locomotor play classifications (Fig. 4). On the contrary, it led

## Contour features classification

The performance of the contour features classifiers can be seen in Table 5. The best performance was obtained when classifying locomotor play from standing with a training accuracy of 0.92, and an internal validation *f1 score* and specificity of 0.80 and 0.92, respectively. This was expected, as standing contours differed from play contours especially when considering aspect ratio (mean  $\pm$  SD; Standing:  $2.00 \pm 0.5$ ; Play:  $2.99 \pm 1.2$ ), mean intensity (Standing:  $133 \pm 28$ ; Play:  $163 \pm 17$ ) and the to a slightly lower percentage of false standing classifications; a decrease that could be very important in a real-life setting, as discussed in section 3.3. The combining of classifiers also had the purpose of lowering the computational power needed to run the play detection algorithm. The NoRunning combined classifier resulted in 44% of the other behaviour contours being correctly classified already in the first round of classification and thus, excluded from the second round (see Table 7). This number will probably be higher in a real-life setting with an even higher proportion of the other behaviours belonging to the standing category. The combined classifier resulted in a 36% reduction in time used on classifying the test data (contour videos classifier: 11min; combined classifier, second round: 7min; 856 contour videos; see Tables 6 and 7, NoRunning classifier onWithRunning test data). The fact that the combined method has potential to significantly reduce the computational power needed, whilst retaining equal ability to classify the other behaviours correctly (comparable specificity and precision), may compensate for the method's lower ability to classify locomotor play correctly. The current study provides evidence of the power of combining a less computational heavy method with a deep learning method to lower the computational power needed in the classification and detection of complex behaviours.

## Future perspectives

Only few locomotor play periods were not detectable as contours in the movement maps created by the GMM. Problem scenarios included, for example, brown pigs in shadowed areas in the pen, producing images with low contrast to the background, and pigs being partly covered by the inventory and thus, having a contour with a too small area to be included. Another issue with the use of contour detection for behavioural classifications is a combination of the housing environment and the behavioural characteristics of locomotor play. Pigs are in general housed in a relatively small space and in groups of varying sizes. Furthermore, locomotor play is characterised by the pigs covering a large part of the pen, sometimes even bumping into each other. As a result, play has even been found to be contagious (Held & Spinka, 2011), meaning that multiple pigs often show locomotor play at the same time. Unfortunately, these factors result in contours that often include multiple pigs, either multiple pigs playing or one/few pigs playing together with other pigs performing other behaviour. In the current study, only 44% of the labelled locomotor play periods resulted in a contour with one pig playing alone. Thus, the use of contour detection for segmentation of individual pigs to obtain the contour features and the contour videos is not realistic in a real-life setting since too many contours will include multiple pigs, thereby making it impossible to count the number of pigs playing in a certain 1 s period.

Liu et al. (2020) developed a tracking-by-detection algorithm to segment pair-wise interactions between pigs from the larger frame, as part of a tail biting detection algorithm. A similar method could be used to segment individual pigs prior to locomotor play detection. The method could further be improved by using a rotated bounding box to detect and track the pigs to produce segmented videos similar to the contour videos of the current study. Such work has already been initiated in pigs within the authors' research team and may be

particularly important to avoid overlapping images between pigs. However, the detection and tracking of pigs is challenged by the fast movements (blurring in frames) of locomotor play (Wang, Larsen, Liu, & Norton, 2022; Wang, Larsen, Liu, Winters, et al., 2022) and thus the detection and tracking part of the algorithm also needs further development. Automatic detection of play behaviour in pigs will most likely be used as a measure to assess the on-farm welfare of young and growing pigs and to document welfare labels on meat products. This means that even though individual monitoring is always preferable, it may not be necessary in this case. Thus, the developed detection and tracking algorithm may not need to provide the correct identification to a single pig, but merely be able to follow an individual pig throughout the 1 s period of the video, or the length of the play event, in order to segment the pig from the rest of the video. This would make it possible to count the number of pigs playing in each 1 s period and thereby to monitor the duration of locomotor play on pen level.

Whether a play detection algorithm is used for research purposes to create data on play behaviour or it is used for onfarm monitoring, the algorithm is currently challenged by the computational power required and thus the time needed to run the algorithm. This is similar for other algorithms designed for other behavioural events. Future work needs to focus on the effectiveness of such algorithms to enable them to run in real-time on several pens of pigs at the same time. When the algorithms can examine large datasets then their potential will be elucidated, resulting in true value creation by improving our knowledge and implementing behaviour detection for animal welfare assessment and disease prediction.

## **Conclusion**

The current study presents a first attempt on developing an algorithm for automatic detection of locomotor play in pigs. When using potentially generalisable contour features and standard machine learning methods to classify locomotor play from other solitary behaviours, *Recall*, *Precision*, and *Specificity* of 77%, 70% and 89%, respectively, were obtained during internal validation. Contour videos and deep learning methods (CNN + LSTM) increased *Recall*, *Precision*, and *Specificity* to 94%, 88% and 96%, respectively. When combining the two classification methods, an almost similar performance was retained, whilst 44% of the other behaviour contours were correctly classified without the need of deep learning methods and the combination thereby decreasing the computational power needed to run the algorithm. Overall, the current study provides a proof of concept that locomotor play can be automatically detected in young pigs. However, the algorithm is currently challenged by the computational power required and future work needs to focus on the effectiveness of such algorithms to enable them to run in real-time on several pens of pigs at the same time.

## **Data statement**

Data supporting the analyses can be found on Mendeley Data (Larsen, Wang, Willems, Liu, & Norton, 2023).

## **CRediT authorship contribution statement**

Mona Lilian Vestbjerg Larsen: Project administration, Resources, Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualisation, Writing – original draft, Writing - review & editing. Meiqing Wang: Conceptualization, Methodology, Formal analysis, Writing – review & editing. Sam Willems: Conceptualisation, Methodology, Writing – review & editing. Dong Liu: Conceptualisation, Methodology, Writing – review & editing. Tomas Norton: Project administration, Resources, Conceptualisation, Methodology, Supervision, Writing e review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

This study was conducted as part of the AutoPlayPig project funded by the Marie Skłodowska-Curie grant agreement No. 842555 and the ClearFarm project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862919, while the data collection was funded by the Ministry of Food, Agriculture and Fisheries of Denmark under the program "Veterinary agreement III" (Veterinærforlig III, AU6). The authors send great thanks to the staff at the Pig Research Unit, Aarhus University, Denmark, and to PhD student Jeanet F. M. Winters, assisting technicians Carsten K. Christensen and Dines T. Bolt, post-doctoral colleague Guilherme Amorim Franchi, and student helper Ida H. Kristoffersen.

## References

- S.M. Brown *et al.*  
Evidence for litter differences in play behaviour in pre-weaned pigs  
Applied Animal Behaviour Science  
(2015)
- C. Chen *et al.*  
A computer vision approach for recognition of the engagement of pigs with different enrichment objects  
Computers and Electronics in Agriculture  
(2020)
- C. Chen *et al.*  
Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory  
Computers and Electronics in Agriculture  
(2020)
- K.N. Dominiak *et al.*  
Prioritizing alarms from sensor-based detection models in livestock production - a review on model performance and alarm reducing methods  
Computers and Electronics in Agriculture  
(2017)
- T.M. Donaldson *et al.*  
Effects of early play experience on play behaviour of piglets after weaning  
Applied Animal Behaviour Science  
(2002)
- G.A. Franchi *et al.*  
Play behaviour positively relates to weight gain, feeding behaviour and drinking behaviour in weaner pigs (*Sus scrofa*)  
Applied Animal Behaviour Science  
(2023)
- H. Gan *et al.*  
Automated detection and analysis of social behaviors among preweaning piglets using key point-based spatial and temporal features  
Computers and Electronics in Agriculture  
(2021)
- S.D.E. Held *et al.*

- Animal play and animal welfare  
Animal Behaviour  
(2011)
- D.B. Jensen *et al.*  
A multivariate dynamic linear model for early warnings of diarrhea and pen fouling in slaughter pigs  
Computers and Electronics in Agriculture  
(2017)
  - D. Liu *et al.*  
A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs  
Biosystems Engineering  
(2020)
  - J.E. Martin *et al.*  
The influence of neonatal environment on piglet play behaviour and post-weaning social and cognitive development  
Applied Animal Behaviour Science  
(2015)
  - R.C. Newberry *et al.*  
Playful behaviour of piglets  
Behavioural Processes  
(1988)
  - M. Ocepek *et al.*  
Which types of rooting material give weaner pigs most pleasure?  
Applied Animal Behaviour Science  
(2020)
  - E.A. O'Connor *et al.*  
The impact of chronic environmental stressors on growing pigs, *Sus scrofa* (Part 1): Stress physiology, production and play behaviour  
Animal  
(2010)
  - M. Oostindjer *et al.*  
Effects of environmental enrichment and loose housing of lactating sows on piglet behaviour before and after weaning  
Applied Animal Behaviour Science  
(2011)
  - D.S. Rodriguez-Baena *et al.*  
Identifying livestock behavior patterns based on accelerometer dataset  
Journal of Computational Science  
(2020)