Journal of Animal Science, 2023, **101**, 1–15
https://doi.org/10.1093/jas/skad347
Advance access publication 9 October 2023
Animal Behavior and Cognition

A computer vision image differential approach for automatic detection of aggressive behavior in pigs using deep learning

Jasmine Fraser,[†] Harry Aricibasi,[‡] Dan Tulpan,[†,1] and Renée Bergeron[†]

[†]Department of Animal Biosciences, Ontario Agricultural College, University of Guelph, 50 Stone Road East, Guelph, ON, Canada N1G2W1
[‡]Department of Biomedical Sciences, Ontario Veterinary College, University of Guelph, 50 Stone Road East, Guelph, ON, Canada N1G2W1
[1]Corresponding author. dtulpan@uoguelph.ca

Abstract

Pig aggression is a major problem facing the industry as it negatively affects both the welfare and the productivity of group-housed pigs. This study aimed to use a supervised deep learning (DL) approach based on a convolutional neural network (CNN) and image differential to automatically detect aggressive behaviors in pairs of pigs. Different pairs of unfamiliar piglets ($N = 32$) were placed into one of the two observation pens for 3 d, where they were video recorded each day for 1 h following mixing, resulting in 16 h of video recordings of which 1.25 h were selected for modeling. Four different approaches based on the number of frames skipped (1, 5, or 10 for Diff1, Diff5, and Diff10, respectively) and the amalgamation of multiple image differences into one (blended) were used to create four different datasets. Two CNN models were tested, with architectures based on the Visual Geometry Group (VGG) VGG-16 model architecture, consisting of convolutional layers, max-pooling layers, dense layers, and dropout layers. While both models had similar architectures, the second CNN model included stacked convolutional layers. Nine different sigmoid activation function thresholds between 0.1 and 1.0 were evaluated and a 0.5 threshold was selected to be used for testing. The stacked CNN model correctly predicted aggressive behaviors with the highest testing accuracy (0.79), precision (0.81), recall (0.77), and area under the curve (0.86) values. When analyzing the model recall for behavior subtypes prediction, mounting and mobile non-aggressive behaviors were the hardest to classify (recall = 0.63 and 0.75), while head biting, immobile, and parallel pressing were easy to classify (recall = 0.95, 0.94, and 0.91). Runtimes were also analyzed with the blended dataset, taking four times less time to train and validate than the Diff1, Diff5, and Diff10 datasets. Preprocessing time was reduced by up to 2.3 times in the blended dataset compared to the other datasets and, when combined with testing runtimes, it satisfied the requirements for real-time systems capable of detecting aggressive behavior in pairs of pigs. Overall, these results show that using a CNN and image differential-based deep learning approach can be an effective and computationally efficient technique to automatically detect aggressive behaviors in pigs.

Lay Summary

Aggressive behavior in pigs is a major concern for the swine industry that negatively affects animal welfare. This study aims to provide an efficient automatic solution based on computer vision and supervised deep learning models able to distinguish between aggressive and non-aggressive behavior of pigs using video recordings.

Key words: aggressive behavior, computer vision, deep learning, image analysis, pig behavior, video recording

Abbreviations: AUC, area under the curve; CNN, convolutional neural network; DL, deep learning; FN, false negatives; FP, false positives; ICC, intraclass correlation coefficient; LSTM, long short-term memory; ML, machine learning; ReLU, rectified linear unit; ROC, receiver operating characteristic; SVM, support vector machines; TN, true negatives; TP, true positives; VGG, visual geometry group

Introduction

Pig's social structures are based on a dominance hierarchy, formed through a series of agonistic interactions, which determine the dominant-submissive relationships (Meese and Ewbank, 1973). In stable social groups, aggressive interactions are much lower as pigs can regulate situations through "avoidance order" (Jensen, 1982). However, in commercial settings, it is common practice for pigs to be regrouped and mixed with unacquainted pigs resulting in an unstable social group. Each time pigs are regrouped, a new hierarchy must be formed resulting in an influx of aggressive interactions. While these interactions are necessary to establish the group hierarchy, it is often the case that serious injuries occur and it would be desirable to have the capability to monitor this behavior and test the efficacy of management practices. For example, socializing piglets prior to mixing is a potential avenue to reduce post-mixing aggression (Fels et al., 2021). Additionally, pigs are exposed to increased competition through intensive feeding and housing systems resulting in a further increase in levels of aggression and injuries (Stukenborg et al., 2011). Aggression is a major concern in the swine industry as it can have negative effects on both the welfare of the animals and their productivity (Arey and Edwards, 1998). Thus, the ability to detect aggression in farm settings could be a first step towards developing enhanced management practices, and efficient intervention solutions to prevent fighting.

Received July 19, 2023 Accepted October 6, 2023.

As a result, there has been a significant increase in research investigating the capabilities of artificial intelligence approaches such as machine and deep learning (ML, DL) that utilize computer vision techniques and digital imagery as a potential option for quicker and more accurate detection of these undesirable behaviors.

One promising approach for automated behavior detection, is the use of machine learning. Machine learning is a form of artificial intelligence that attempts to mimic human's ability to learn and make predictions. The process involves fitting predictive models to data through the recognition of patterns within the data (Greener et al., 2021). DL, a subset of machine learning, uses algorithms based on a hierarchical learning process and artificial neural networks to detect patterns in data, which can then be used to make predictions (Bengio, 2009). However, what makes DL so powerful, is its ability to automatically detect and extract features without having to specify in the algorithm what to look for, which is typically required in conventional machine learning (Olden et al., 2008). With previous research showing that DL can perform close to or even better than humans in tasks such as image recognition and machine translation (He et al., 2016; Wu et al., 2016), it is a promising technique for the automatic detection of behaviors. The use of DL architectures has shown to be especially useful for biological image classification problems, such as disease detection and diagnosis (Suk et al., 2014; Le et al., 2019; Bi et al., 2020).

The ability to handle large and complex data have made machine and DL popular approaches for behavior detection in animals. Valletta et al. (2017) provide an in-depth overview of ML techniques pertinent to the study of animal behavior. The authors present a concise guide on the rationale behind unsupervised and supervised learning and illustrate the application of these methods by developing three data analytical workflows to convert datasets into useful biological knowledge (Valletta et al., 2017). Viazzi et al. (2014) attempted to detect aggressive behavior continuously and automatically in group-housed piglets using image processing. Information pertaining to the pig's motion was initially extracted, through Motion History Images. Two features, the mean intensity, and the occupation index were extracted, and using a Linear Discrimination Analysis, the interactions were classified as aggressive or not (Viazzi et al., 2014). Following this, Oczak et al. (2014) also aimed to test a method for the automatic detection of aggressive behavior in pairs of pigs, using an activity index and a multilayer feed-forward network (Oczak et al., 2014). Five features were calculated, such as the average, maximum, minimum, sum, and variance of the activity index, which were then used to train a multilayer feed-forward network to classify high (neck biting, ear biting, and body biting) and medium-level aggression (head and body knocking, parallel, and inverse parallel pressing).

Lee et al. (2016) tested a method, which utilized kinetic depth sensors to extract the maximum, minimum, average, and standard deviation of velocity as well as the distance between pigs. Hierarchical support vector machines then used the extracted features to detect aggression. The ability to detect high (neck biting, ear biting, and body biting) and medium (head and body knocking, parallel, and inverse parallel pressing) levels of aggression was further tested by Chen et al. (2017) using the connective area and adhesion index to separate two aggressive pigs from the remaining five pigs in the group (Chen et al., 2017). In this method, the two pigs engaged in an aggressive interaction were considered as one rectangular unit to calculate acceleration. Based on acceleration, recognition rules for medium and high aggression were designed. Similarly, Chen et al. (2018) also used connective area and adhesion index to locate the aggressive pigs, and once again both pigs engaging in an aggressive interaction were considered as one rectangle, used for feature extraction (Chen et al., 2018). Feature points (the head and four kink points) of aggressive pigs were selected and using the motion from these points, kinetic energy was calculated. The kinetic energy difference between frames was used as an additional feature. Thresholds were determined using hierarchical clustering, trained on the features extracted, and then used to determine recognition rules. One year later, Chen et al. (2019) developed a depth-based image analysis technique using a Motion Shape Index calculated from frame-to-frame distance statistics and applied it to detection of aggressive behaviors in group-housed pigs (Chen et al., 2019). They used support vector machine classifiers to detect aggression and reported results with accuracies, sensitivities, specificities, and precisions above 96%. Most recently, Chen et al. (2020) tested a method based on a type of DL method called convolutional neural network (CNN) and Long Short-Term Memory (LSTM) to recognize aggression (Chen et al., 2020). This method directly processed video episodes, unlike previous work, which processed individual frames. It used the CNN architecture visual geometry group (VGG)-16 to extract spatial features, which were then inputted into the LSTM to extract temporal features. A prediction function through a fully connected layer was then used to determine if the episode was aggressive.

While many of these methods showed strong results, it is important to note that the accuracy of these methods is greatly affected by the dataset. For instance, sedentary behavior is a very consistent and highly predominant behavior, involving little to no movement. Therefore, using a dataset with large amounts of sedentary behavior may produce a model with seemingly high results, but these results may not accurately reflect the models' ability to detect more complex behaviors. For example, the study conducted by Viazzi et al. (2014) reports the inclusion of 150 aggressive and 150 non-aggressive behavior episodes, nevertheless, the average duration of non-aggressive episodes is almost twice larger (25.8 to 30.0 s) than that of aggressive episodes (17.6 s) leading to the conclusion that their dataset was unbalanced. In many cases, little information is provided on the datasets used, making it difficult to examine the validity and practicality of the results. For example, the study conducted by Viazzi et al. (2014) refers to aggressive behavior as fighting without providing details about specific behaviors included in this generic category, while the study conducted by Lee et al. (2016) only mentions about two types of aggressive behavior (head-to-head knocking and chasing) without further details about other types of aggressive behaviors that might have been included in their work. In contrast, the studies of Chen et al. (2017, 2019, and 2020) provide a detailed description of the ethogram used to classify the seven observed aggressive behaviors. In addition, many of the above studies use a combination of two or more feature extraction methods and DL models programmed in different languages, which are important to obtain quality results, but can be very costly computation-wise and not fully automated. For example, the study conducted by Viazzi

et al. (2014) extracts motion history images and performs post-processing calculations leading to two features (mean intensity and occupation index) using Matlab, while the modeling using linear discriminant analysis is performed in LPSS. Furthermore, there is no additional information suggesting that the two tasks were included in an automatic computational pipeline, but rather performed independently.

Therefore, this paper aims to find and test a computationally less intensive solution based on the fully automatic detection of aggression in pairs of pigs using a CNN and an imaging differential approach, all integrated with a standalone computational pipeline that could pave the way to the most sought-after solution for this problem, i.e., detecting pig aggression in real-time from live videos. Moreover, the manuscript focuses on detecting aggressive behavior in pairs of pigs rather than larger groups, thus reducing the confounding factors associated with mixed behaviors in larger groups, and testing if the implementation of a real-time solution for practical applications is feasible for this rather simplified scenario. While the detection of aggressive behavior in pairs of pigs is less practical than in larger groups, the models proposed in this study rely on a methodology based on image differences able to capture motion and proximity among pigs and do not rely on individual pig detections or segmentations thus rendering it applicable and scalable to larger groups of pigs.

Materials and Methods

Data collection and protocols

All experimental procedures followed the guidelines outlined by the Canadian Council of Animal Care and were approved by the University of Guelph Animal Care Committee. The study was conducted at the Arkell Swine Research Station at the University of Guelph (Guelph, ON, Canada) from June to August 2022.

Animals and housing

In total, 32 Landrace × Yorkshire × Duroc crossbred piglets ranging from ages 6 to 8 wk were used, with an average weight of 12.4 kg. All pigs were the same color (white) with no visible markings. The pens were 1.95 m wide and 4.37 m long with 1.02 m matte white plastic walls and fully slatted concrete floors. The pens were illuminated with incandescent lights. Each pen contained a 0.89 m wide 3-hole Crystal Springs Wet/Dry Feeder and a bowl drinker.

Equipment and data acquisition

Sixteen tests were conducted in a 3-d period (five tests on the first and third day and six tests on the second day). During each test, a pair of randomly selected unfamiliar piglets were placed in one of the two observation pens (five to six pairs per pen per day in total) and were recorded continuously for 1 h following mixing. In total, we used 16 h of video recordings including 5 h on the first and third days and 6 h on the second day. We have opted for a 1-h recording period per pair of pigs due to the increased occurrence of aggressive behaviors during the period following their initial mixing. When unfamiliar piglets are mixed and their behavior is observed during the initial hour post-mixing, we found that, on average, there were four instances of aggressive interactions per hour. This frequency is significantly higher than what would naturally occur in a group of pigs that are more familiar with each other. To provide a point of reference, a prior study determined that the number of fights observed in a group of piglets ranged from 0.4 to 5 events per hour during the 5 h following mixing, whereas it ranged from 0.2 to 1.4 events per hour when observations were made 24 h later and extended over a 6-h period (Mei et al., 2016).

The videos used for this trial were recorded using a top-down oriented Ro-main RS-CCPOE280IR4-DH (Ro-main Agro-technological Products and Solutions, ST-Lambert-de-Lauzon, QC) camera, which was mounted 3.15 m above and 1.40 m from the back of the pen. The videos were recorded at a resolution of 1080 p and a frame rate of 25 frames per second (fps). The camera recorded continuously for an hour and footage was saved to the server as 10-min clips (six clips per pair). The computer processor was Intel (R) Core (TM) i7-8650U CPU @ 1.90 GHz with 16 GB of RAM. The operating system used was Microsoft Windows 10 Pro Edition. The graphic card was NVIDIA GTX 1060. The scripts used in this study were developed using Python 3.9, Keras 2.9.0, and Tensorflow 2.9.1.

Datasets preparation

Behavior definitions

The videos were all manually labeled second by second by a single observer, using an ethogram based on the work of Jensen (1982), with the addition of mounting and chasing behavior, based on previous work (O'Connell et al., 2003; Fàbrega et al., 2013). The videos were labeled first using the

Table 1. Labeled pig behaviors based on Jensen (1982), O'Connell et al. (2003), and Fàbrega et al. (2013). The table includes the type of behavior, a binary class representing aggressive or non-aggressive, and a description for each behavior type

Behavior	Binary class	Description
Head-to-head biting	Aggressive	Biting the head area of the receiver
Parallel pressing	Aggressive	Pressing of shoulders against each other, facing the same direction
Inverse parallel pressing	Aggressive	Pressing of shoulders against each other, facing the opposite direction
Head-to-head knock	Aggressive	Striking the head of the receiver with the snout
Head-to-body knock	Aggressive	Striking the body of the receiver with the snout
Chasing	Aggressive	Actively pursuing another pig before or immediately following an aggressive behavior
Mounting	Non-aggressive	Placing hooves on the back of another pig, with or without a thrusting movement
Mobile	Non-aggressive	One or more pig is moving in a non-aggressive manner
Immobile	Non-aggressive	Both pigs are sedentary

specific behaviors described in Table 1, then further labeled into a binary classification of either aggressive or non-aggressive. Mounting behavior was labeled separately from non-aggressive mobile as it is the target behavior for future research and results from this trial will be used as preliminary data for future trials.

Intra-observer reliability

Intra-observer reliability was estimated for the duration of behaviors during five 10-min clips, by calculating the intra-class correlation coefficient using PROC MIXED in SAS. It was found that the intra-observer reliability for the duration of behaviors was 99.99%.

Dataset design

Based on the statistical analysis performed by Chen et al. (2020), which determined that the minimum duration of aggressive behaviors is 2 s, we automatically extracted 27,975 2-s video clips from the recorded videos. Of the total 27,975 2-s video clips, 1,123 (4%) made up the aggressive class, and the remaining 26,852 (96%) made up the non-aggressive class. To avoid distorted results due to the highly imbalanced dataset as previously reported by Chen et al. (2020), we performed down-sampling on the majority class where 4% of the non-aggressive clips were randomly selected and combined with all aggressive clips, which were included in the final dataset. This process resulted in roughly 50% of the dataset containing aggressive behavior clips (1,123 clips) and 50% containing non-aggressive behavior clips (1,124 clips). The clips were then divided into training, validation, and testing through random stratified sampling, resulting in a 50%, 25%, and 25%, split respectively as shown in Table 2. The stratified sampling was employed to ensure that a roughly equal representation of behaviors is present in the training, validation, and testing datasets.

Each video clip was separated into consecutive frames and further post-processed into frame differences (with or without frame skipping) resulting in four datasets. To mitigate errors from interference occurring outside the pen area caused by people and pig motion in neighboring pens or walkways, all frames were cropped from the original 1,280 × 720 pixels to 1,088 × 612 pixels by removing bands of boundary pixels, and then further resized and rescaled to 244 × 244 pixels (Figure 1), which was the input size of each individual frame.

Since pigs' behaviors include mobile and immobile situations where the aggressive episodes are typically mobile, the patterns and amount of motion in a video recording could be associated with a specific type of behavior. Thus, it is essential to properly detect motion and investigate if there is an optimal choice for the frame difference construction process. Four approaches, shown in Figure 2, were considered based on the number of frames skipped (1, 5, or 10) and the amalgamation of multiple image differences into one image (blended). Therefore, the first dataset *(Diff1)* consisted of differential images representing pairs of consecutive frame differences. For example, if we have two consecutive frames (previous, current) where only one pixel-size object has changed position due to movement and the remaining pixel values representing the background intensities remained unchanged, the resulting frame difference obtained by subtracting the intensity pixels values of the previous frame from the ones in the current frame (consider the absolute values) will only contain two positive non-zero pixel values, while the remaining ones will equal to zero (Figure 3). The second dataset (*Diff5*) consisted of differential images representing frame differences that were five frames apart. The third dataset (*Diff10*) consisted of differential images representing frame differences that were 10 frames apart. The fourth dataset (*Blended*) consisted of additive superpositions of disjoint consecutive frame differences (e.g., frame 2 minus (—) frame 1 superposed on frame 4—frame 3, frame 6—frame 5, …, and on frame 50—frame 49) that spanned a 2-s video clip. Since each pixel has intensity values in the range of 0 to 255, the maximum value of the pixels in the superposed frame differences were capped at 255. We have also reset to zero the bottom 20% of the pixel values to remove unnecessary noise caused by accumulated pixel intensities, and thus increasing the contrast for the pixels representing pig motion.

Using a blended image differential approach was of particular interest as it captured the entirety of motion occurring within the 2-s video clips. Since DL uses pattern recognition to learn and make predictions (Bengio, 2009), using a blended method allows for more information on both the directional movement, as well the intensity of movement related to a specific behavior. Figure 4 illustrates how much variation occurs between the different types of aggressive behaviors. For

Table 2. Summary of datasets containing 2-s video clips used to train, validate, and test the models. Data are separated into training, validation, and testing with a 50/25/25 percent split respectively using a randomized stratified sampling approach

Behavior	Binary class	Clip count		
		Training	Validation	Testing
Head biting	Aggressive	173	86	86
Parallel pressing	Aggressive	88	44	43
Inverse parallel pressing	Aggressive	13	7	6
Chasing	Aggressive	4	2	2
Head-to-head knocking	Aggressive	271	135	135
Head-to-body knocking	Aggressive	14	7	7
Mounting	Non-aggressive	70	35	35
Mobile	Non-aggressive	423	211	211
Immobile	Non-aggressive	70	35	34

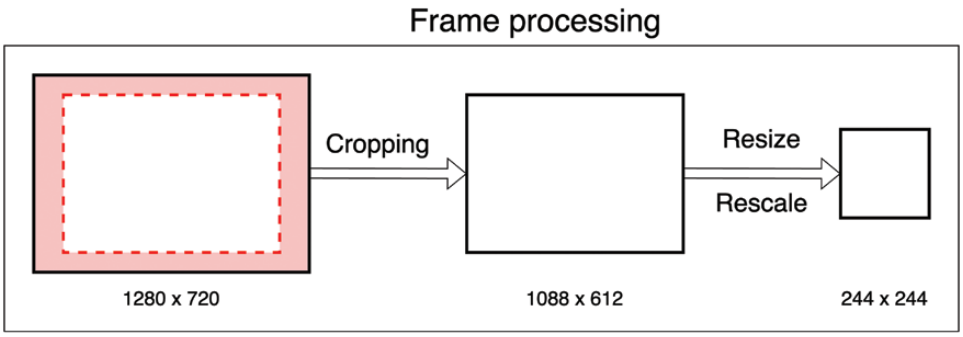

Figure 1. Frame cropping and rescaling. The cropping and rescaling process reduces a 1280 × 720 pixels original frame to a 244 × 244 pixel image.

example, chasing behavior can be characterized by intense motion, usually in a linear direction. Comparatively, immobile behavior (Figure 5) shows very little if any movement. Pigs may move around slightly while lying down, which accounts for the little motion shown, but it is almost always only one pig moving. Since aggression is largely characterized by high-intensity movement, the non-aggressive mobile movement can be challenging to classify. In cases where the two pigs are not in proximity, it is expected to be easier to classify the behavior as non-aggressive. However, when the pigs are in proximity and moving (e.g., paired exploratory behaviors), the image can look very similar to one representing lower-intensity aggressive behaviors, such as head-to-head knocking or head-to-body knocking.

Deep learning algorithm

Based on its successful application in various computer vision-based projects, a deep CNN architecture (Neubauer, 1998) with multiple convolutional layers based on a Visual Geometry Group model architecture (Simonyan and Zisserman, 2014) was used in this study. The original VGG-16 architecture was adapted to fit the needs of the project. Since the dataset used for this project was very small when compared to the dataset used by Simonyan and Zisserman (2014), including only grayscale images, the modifications implemented here aimed to improve the speed and overall efficiency of the model. The learning rate was set to 0.00001 (the final learning rate of VGG-16), and the filters were reduced by 75%. The kernel and pooling size were the same as VGG-16 at 3 × 3 and 2 × 2, respectively. The last layer used a sigmoid activation function, rather than a Softmax function, because for this project we wanted to be able to test if using a different threshold than the default 0.5 had a significant impact on the model prediction quality. The threshold was compared to the output score between 0 and 1 to determine whether a behavior was aggressive or not. Since aggressive behavior can range in motion intensity, testing different thresholds to find the optimal value to improve recall and accuracy without increasing loss significantly was very important.

The first CNN architecture consisted of five convolutional layers combined with five max-pooling down-sampling layers, three dense layers, and two dropout layers as described in Figure 6 (A). The first ten layers were used to reduce the input dimension from 224 × 224 pixels to a latent representation of 14 × 14 pixels and extract the main features for the pigs' motion pattern detection. This model had 3.7 million parameters compared to the significantly larger 138 million parameters of VGG-16. The dense layer included a sigmoid activation function. Using the Adam optimizer (Kingma and Ba, 2014) and the binary cross-entropy loss, the network applied a sigmoid activation for the last layer to output a score between 0 and 1 that can be further compared with a predefined threshold to determine if a motion pattern represents aggressive behavior or not. In this study, the threshold for aggressive behavior was set at 0.5.

The second CNN architecture had the same architecture as the first model but included 13 stacked convolutional layers (Figure 6B). The purpose of these convolutional layers is to learn low-level features. Hence, stacking the layers allows subsequent layers to learn combinations of low-level features to form higher-level features (Bengio, 2009). The convolutional layers after the max-pooling layers work in a similar way but the input size is reduced for performance. Therefore, the stacking allows for a combination of learning at each stage before downsizing. All convolutional layers were implemented with a Rectified Linear Unit (ReLU) activation function, with zero padding and a stride parameter of 1. Except for the layer sizes, which were scaled down to speed up computation, no explicit hyper-parameter optimization was needed since we initialized our code with the already optimized VGG-16 model architectures. Moreover, based on preliminary trials where we compared the performance of the DL models with and without stacking trained on the Diff1, Diff5, Diff10, and Blended datasets, we decided to apply the stacked model only on the Blended dataset, due to its reduced size which is roughly 50 times smaller than the Diff1 dataset.

Model evaluation

The models were built and compared using five different scenarios based on three k-difference frames (k = 1, 5, and 10), the blended 50-frame differences, and the blended 50-frame differences with stacked convolutional layers. The average and standard deviation of loss, accuracy, precision, recall, AUC, and runtime corresponding to training, validation, and testing experiments were calculated using a repeated hold-out validation approach like the one used for developing the VGG-16 model. In our study, each training, validation, and testing subset triplet was obtained by applying random stratified subsampling on the full collection of 2-s video recordings. This procedure was repeated three times. The stratification at video recording level of the subsampling process is needed to avoid situations where frames from the same video recording are included in the training, validation, or testing sets, leading to over-inflate model evaluations.

Accuracy, precision, and recall were calculated using Equations 1, 2 and 3

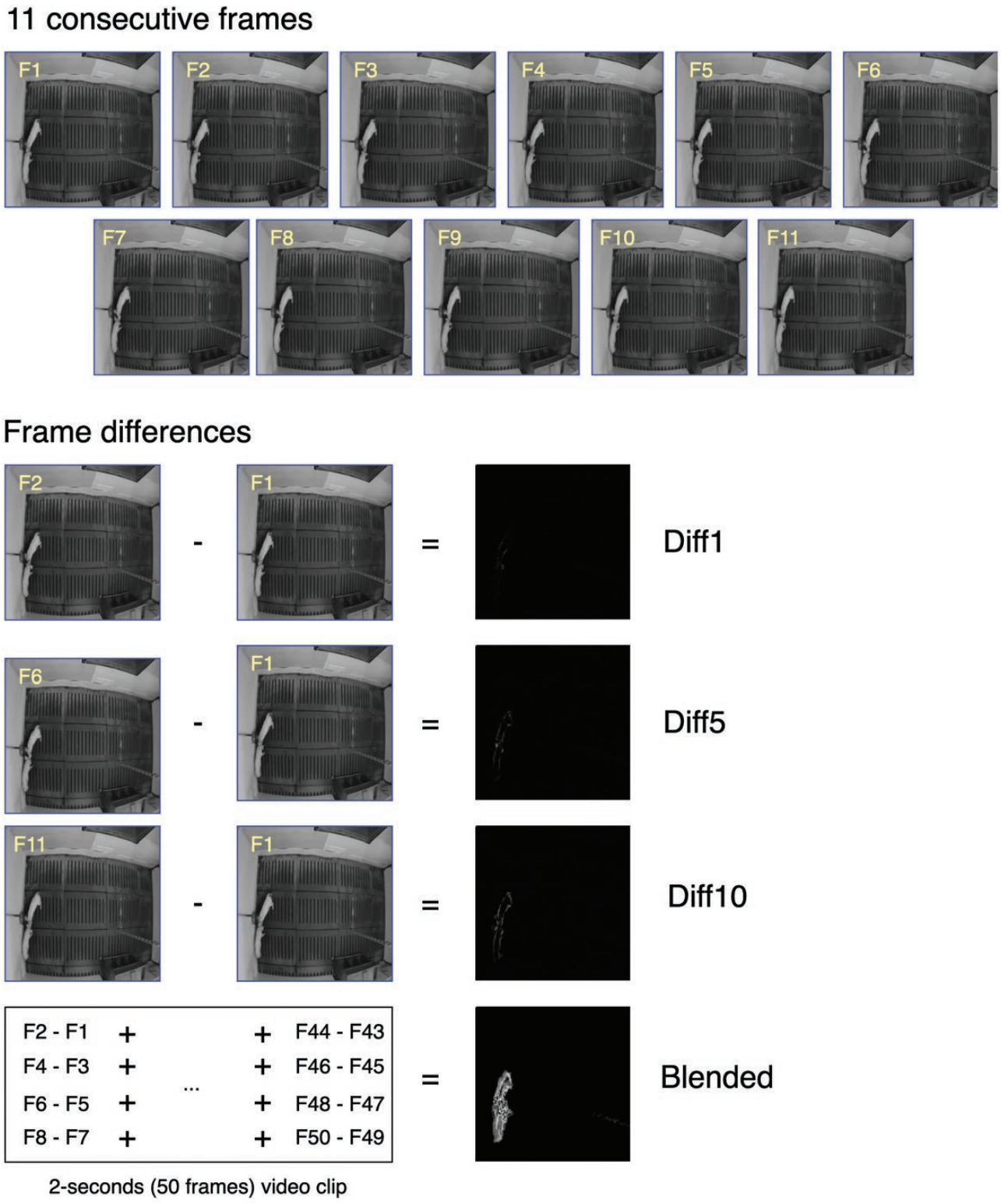

Figure 2. Examples of frame differences included in the four datasets used in this study. The top part includes 11 consecutive frames from a video recording and the bottom part provides details on how the image differential approach was applied to obtain four different types of binary images.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Where true positives (TP) represent aggressive episodes that were correctly classified as aggressive, true negatives (TN) represent non-aggressive episodes that were also correctly

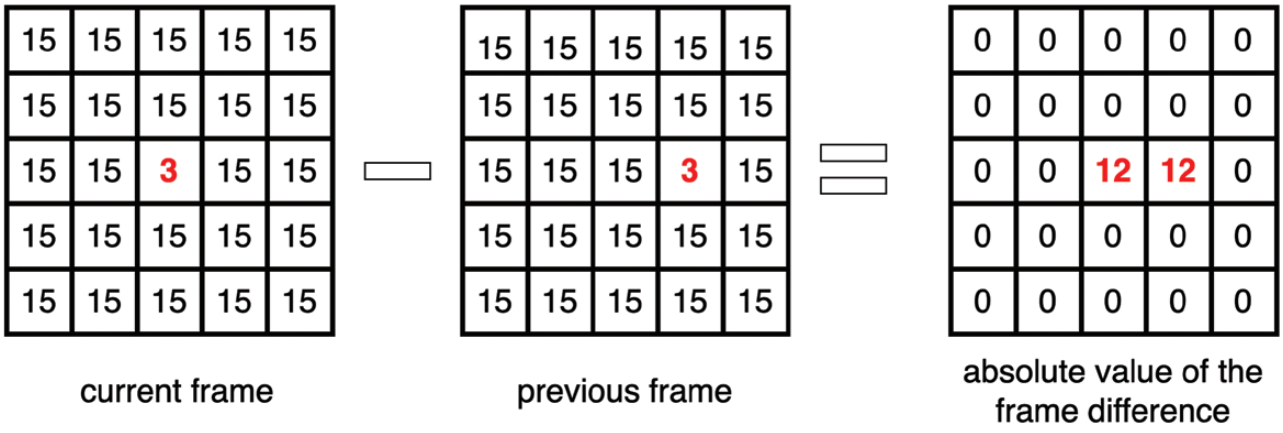

Figure 3. Example of frame difference calculations. The figure shows how to obtain the absolute value of pixel intensities by subtracting pixel intensities for corresponding (x, y) values from two consecutive 5 × 5 frames (current, previous) and recording the absolute values in the resulting frame.

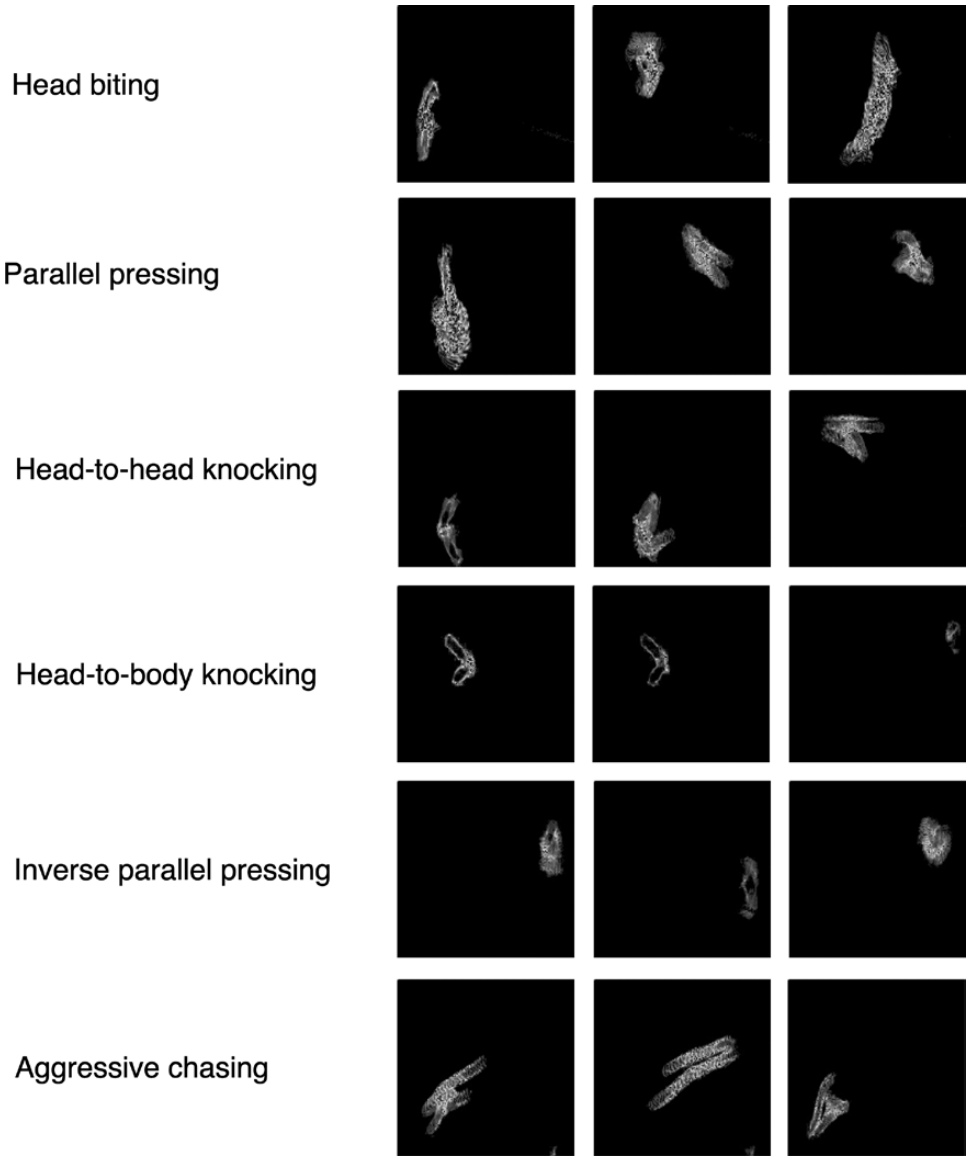

Figure 4. Blended representation of six types of aggressive pig behaviors. Six types of aggressive behavior are listed, and three examples of binary images are provided for each type.

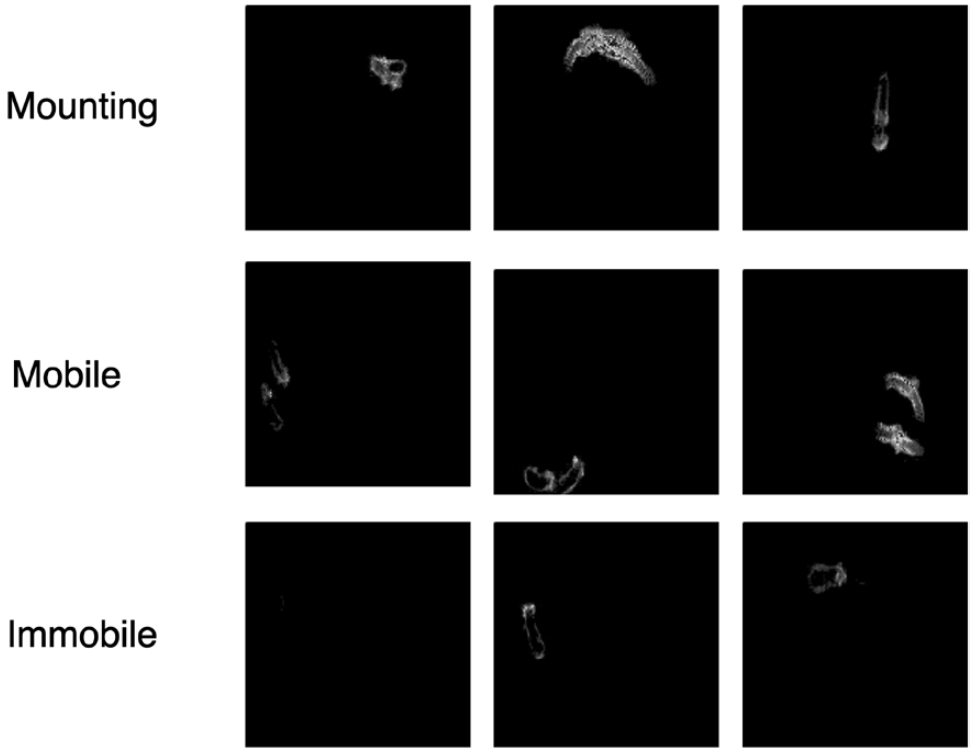

Figure 5. Blended representation of three types of non-aggressive pig behaviors. The figure includes three examples of binary images for each type of non-aggressive behavior.

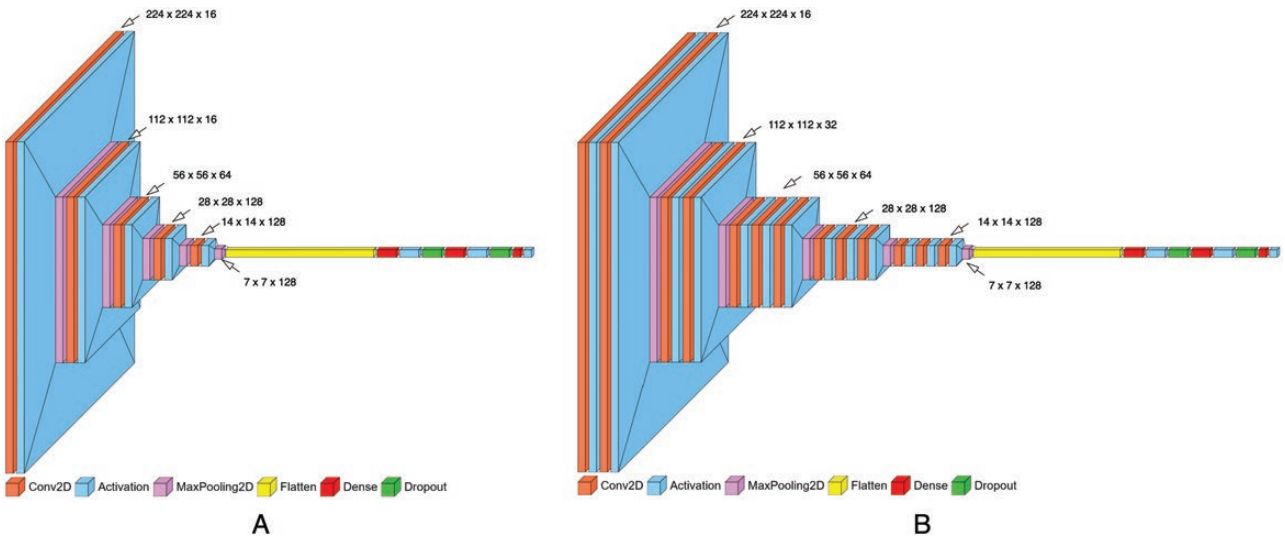

Figure 6. The architectures of the first and second convolutional neural network models: without (A) and with (B) stacked convolutional layers. The model architectures include six types of layers. The total number of parameters for the first model (A) without stacked convolutional layers is 3,719,681, while the total number of parameters for the second model (B), which includes stacked convolutional layers is 4,395,441.

classified as non-aggressive, false positives (FP) represent non-aggressive episodes that were incorrectly classified as aggressive, and false negatives (FN) represent aggressive episodes that were incorrectly classified as non-aggressive.

A fourth measure, area under the curve (AUC), was also calculated. AUC measures the area under a receiver operating characteristic curve, which shows the performance of a classification model using true positive and false positive rates and measures the model's capability to distinguish between classes.

In addition, to estimate overfitting, a binary cross-entropy function was used as the loss function described in (Equation 4)

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (4)$$

Where y represents the class (1 for aggressive and 0 for non-aggressive) and p(y) represents the predicted probability of the data point being aggressive for all N data points.

This function is used to evaluate the model by returning high values for incorrect predictions and low values for correct predictions. This is achieved by adding log[−p(y)] to the loss for each point $y = 1$ and adding a log[1−p(y)] for every $y = 0$.

The classification results based on sub-class behaviors were reported based on checking if the model correctly classified the instances, and the percentages reported in the manuscript include individual values representing how many instances of each sub-class were classified correctly divided by the total number of predictions for that sub-class.

Results and Discussion

Model performance evaluation

A summary of all the results obtained in this study are provided in Table 3.

The average training, validation, and testing loss decreased when larger frame differences datasets were considered, and the lowest values were consistently obtained with the Blended stacked CNN model (Table 3A). The same pattern could be observed when we compared the models using accuracy (Table 3B), precision (Table 3C), recall (Table 3D), and AUC (Table 3E). The highest accuracy value was 0.80 for the stacked CNN model applied to the blended dataset, while the other models produced slightly lower accuracy scores, with the lowest value (0.76) corresponding to the Diff1 dataset. These results are slightly lower than those presented in previous studies, which reported accuracies of over 0.90 (Chen et al., 2017, 2020). However, comparing results directly is difficult as the datasets vary significantly between projects. The highest precision and recall values (0.81 and 0.78, respectively) were obtained with models built on the blended dataset and they were comparable to the accuracy results, while the AUC values were slightly elevated (0.84 to 0.88). For this study, which focuses on identifying aggressive episodes in pig video recordings, recall is perhaps more important than precision since the former measures how many aggressive episodes the model missed while the latter focuses on the total number of correctly predicted episodes over the total number of incorrect and correct predictions. Thus, mislabeling an episode as aggressive (equivalent to a false alarm), while there is no pig aggression, has a lower impact on pigs' health and wellbeing compared to missing an aggressive event that could damage one or all subjects involved in the aggression. The relatively high AUC values suggest that the models can effectively distinguish between aggressive and non-aggressive pig behaviors using the four datasets, with the highest results obtained using the Stacked CNN model applied to the Blended dataset.

Overfitting analysis

As is often the case, deep neural networks are prone to overfitting and unpredictable behavior and therefore require additional computational efforts to detect and address the problem (Greener et al., 2021).

In this work, we performed an overfitting analysis of our models by repeating the training and validation of the CNN models three times. Figure 7 shows the training and validation loss curves for the CNN model without stacked convolutional layers for 100 epochs (Figure 7A) and 1,000 epochs (Figure 7B). While the training process continues to produce increasingly lower loss values, the validation curves diverge from the training curves after 50 to 70 epochs (the green area) and eventually start to plateau. This is followed by a

Table 3. Average and standard deviation training, validation, and testing results for the five models. For each model, we report the loss, accuracy, precision, recall, and area under the curve averaged over three repeated experiments. Bolded values represent the best results obtained with the models employed in this study

Task	Diff1	Diff5	Diff10	Blended	Blended stacked
*A. Mean **loss** ± std.*					
Training	0.50 ± 0.01	0.47 ± 0.01	0.47 ± 0.02	0.48 ± 0.01	**0.44** ± 0.01
Validation	0.51 ± 0.02	0.49 ± 0.01	0.48 ± 0.02	0.49 ± 0.03	**0.47** ± 0.04
Testing	0.51 ± 0.02	0.50 ± 0.01	0.49 ± 0.01	0.49 ± 0.03	**0.47** ± 0.03
*B. Mean **accuracy** ± std.*					
Training	0.76 ± 0.01	0.78 ± 0.01	0.79 ± 0.01	0.79 ± 0.01	**0.80** ± 0.01
Validation	0.76 ± 0.01	0.77 ± 0.01	0.77 ± 0.01	0.77 ± 0.01	**0.79** ± 0.01
Testing	0.76 ± 0.02	0.76 ± 0.01	0.77 ± 0.01	0.78 ± 0.03	**0.79** ± 0.02
*C. Mean **precision** ± std.*					
Training	0.78 ± 0.01	0.78 ± 0.01	0.80 ± 0.01	0.79 ± 0.01	**0.81** ± 0.01
Validation	0.76 ± 0.01	0.78 ± 0.01	0.78 ± 0.01	**0.80** ± 0.02	**0.80** ± 0.01
Testing	0.77 ± 0.01	0.78 ± 0.01	0.78 ± 0.01	**0.81** ± 0.03	**0.81** ± 0.03
*D. Mean **recall** ± std.*					
Training	0.73 ± 0.01	0.77 ± 0.01	0.77 ± 0.01	**0.78** ± 0.01	**0.78** ± 0.01
Validation	0.75 ± 0.02	0.75 ± 0.01	0.76 ± 0.01	0.74 ± 0.02	**0.77** ± 0.03
Testing	0.74 ± 0.03	0.74 ± 0.01	0.75 ± 0.01	0.73 ± 0.02	**0.77** ± 0.03
*E. Mean **AUC** ± std.*					
Training	0.84 ± 0.01	0.86 ± 0.01	0.87 ± 0.01	0.86 ± 0.01	**0.88** ± 0.01
Validation	0.83 ± 0.01	0.85 ± 0.01	0.85 ± 0.01	0.85 ± 0.02	**0.86** ± 0.02
Testing	0.84 ± 0.01	0.85 ± 0.01	0.85 ± 0.01	0.85 ± 0.03	**0.86** ± 0.02

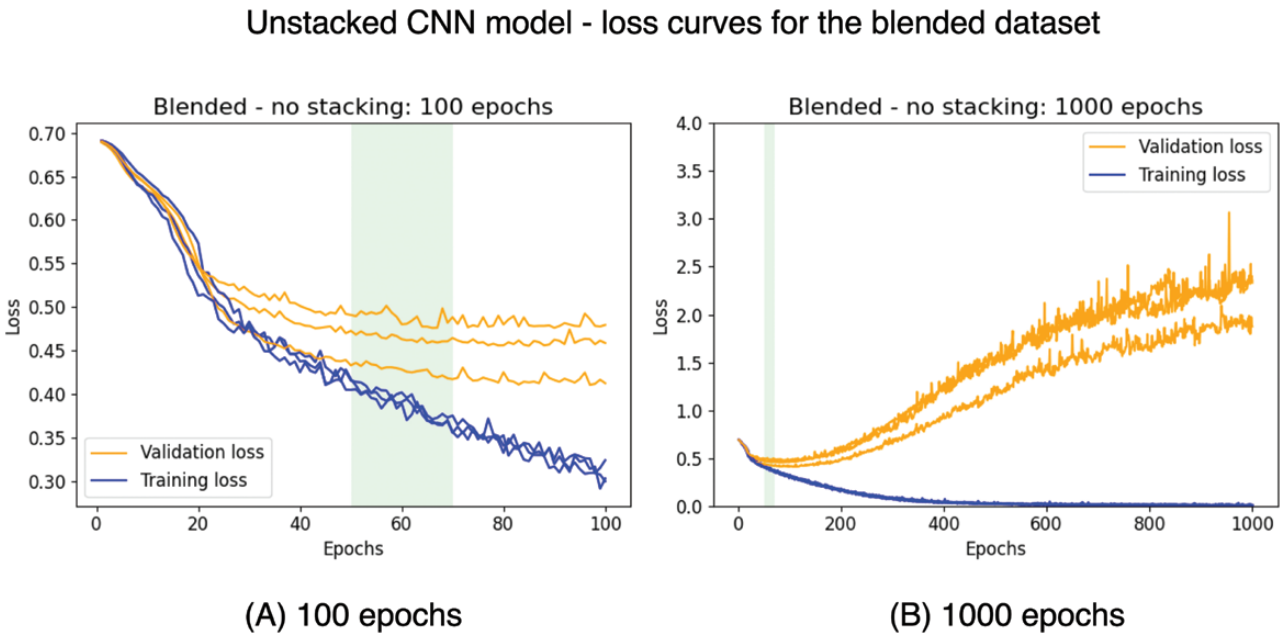

Figure 7. Training and validation loss for the convolutional neural network model with unstacked convolutional layers trained on the blended dataset for 100 and 1000 epochs. The training and validation process was repeated three times leading to three training and three validation curves. The highlighted vertical region represents the range of epochs where the two types of curves start to significantly diverge marking the start of the overfitting phenomenon. (A) The models were trained for 100 epochs. (B) The models were trained for 1000 epochs.

rise in the validation loss values and a continuous decrease in the training loss values, which are depicted in Figure 7B. This divergence region represents the start of the overfitting phenomenon, and it is recommended to either stop the model training process or apply more intense model regularization. The same pattern can be observed in Figure 8, which depicts the training and validation curves for the CNN model with stacked convolutional layers. We also note the increased ruggedness of the loss curves for the stacked CNN model, which is due to the significantly larger number of parameters (compared to the unstacked model) and the use of the same-size dataset. For the unstacked CNN model built on the Diff1 dataset, the loss curves start to diverge earlier after four to six epochs due to a significantly higher number of datapoints (frame differences) presented to the network in each epoch (Figure 9). The same phenomenon was observed for the loss curves corresponding to the CNN models built on the Diff5 and Diff10 datasets (data not included). We note that in this work, all models have been regularized by including dropout layers in their network architectures.

Prediction of individual behaviors

One of the most interesting results in terms of showing the capabilities of this model was the individual behavior recall results. While the main task of the project was to distinguish between aggressive and non-aggressive behaviors, it is often very important to analyze the results and identify the behavioral subtypes that are harder to predict correctly. As previously discussed, recall is a very important metric as the goal of this project was to test the capabilities of detecting aggressive behavior. Compared to the results shown in Table 3, the average recall value was above 0.80 for aggressive behaviors (Table 4) while the average recall value for non-aggressive behaviors was slightly lower (0.77). The low recall values (0.63 and 0.75) of both the mounting and mobile non-aggressive behaviors indicate that these behaviors are difficult to classify. Mounting behavior made up a relatively small portion of the dataset (6.2%) and can be associated with little movement if the receiver pig does not react, or fast-paced movement if the receiver pig does react, making it a very hard behavior to classify. Chen et al. (2020) had similar misclassifications of mounting behavior because of the displacement created by the receiver pig. For this project, the goal was to identify aggressive behaviors, so mis-classifying non-aggressive behaviors is not of great concern. However, Table 4 also illustrates the problem with identifying low-level aggression, such as head-to-head knocking or head-to-body knocking. While these behaviors have been previously characterized as high to medium-level aggression (Lee et al., 2016; Chen et al., 2017), our observations showed that these behaviors could often be characterized by short, low-intensity movements, making them difficult to classify correctly. The standard deviations for both chasing and head-to-body knocking were significantly higher than the other behaviors, which may be in part due to the limited data on these two behaviors. Another confounding factor that can hypothetically affect the quality of the predictions of aggressive behaviors based on image differentials is the presence of high-speed movement scenarios representing non-aggressive behaviors such as scampering or play-fighting, where pairs or groups of pigs will be in close proximity and perform fast, non-aggressive movements. These scenarios might be easier to label correctly by more complex classification approaches based on non-motion-oriented data. Currently, we are not aware of any successful classification model capable of distinguishing among such scenarios.

Optimal sigmoid function thresholds

We explored the impact of using different sigmoid activation function thresholds in the range of 0.1 to 0.9 (with increments of 0.1) for the final layer on the prediction performance of the

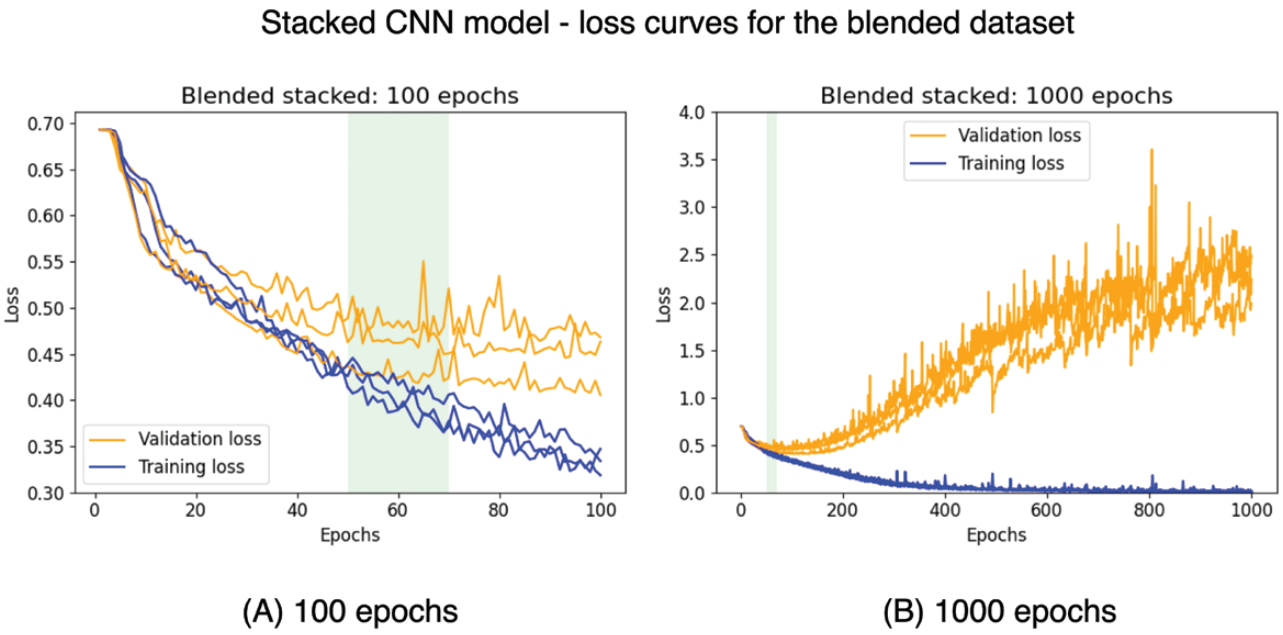

Figure 8. Training and validation loss for the convolutional neural network model with stacked convolutional layers trained on the blended dataset. The training and validation process was repeated three times leading to three training and three validation curves. The highlighted vertical region represents the range of epochs where the two types of curves start to significantly diverge marking the start of the overfitting phenomenon.

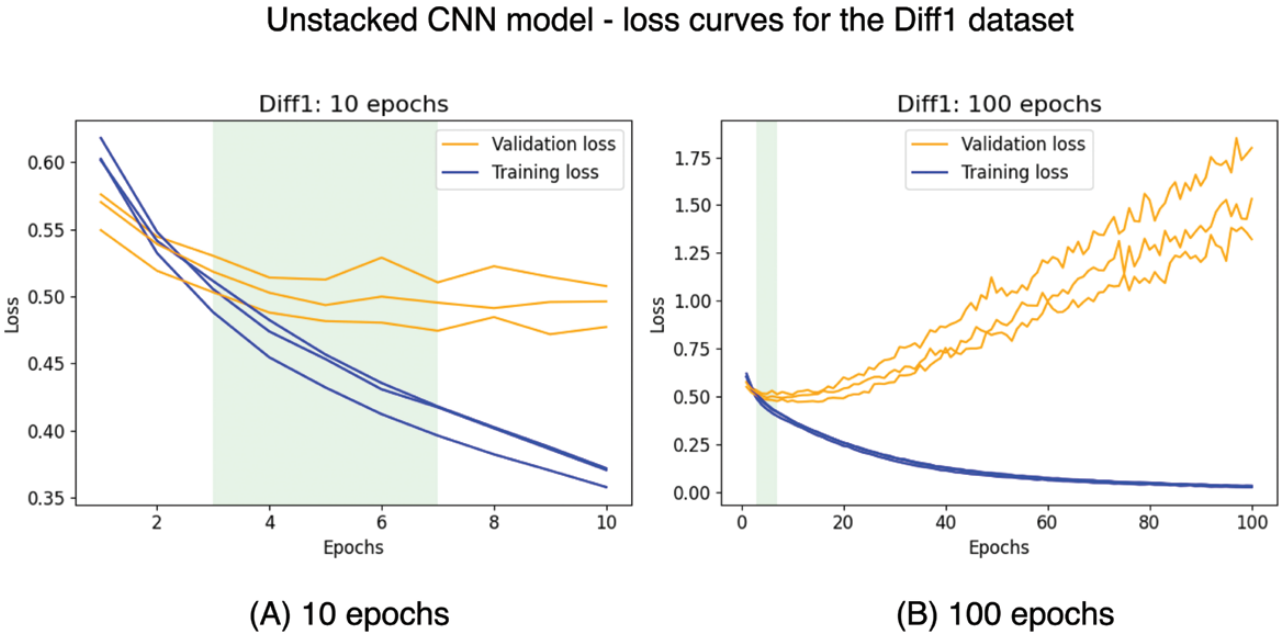

Figure 9. Training and validation loss for the convolutional neural network model with unstacked convolutional layers trained on the Diff1 dataset for 10 and 100 epochs. The training and validation process was repeated three times leading to three training and three validation curves. The highlighted vertical region represents the range of epochs where the two types of curves start to significantly diverge marking the start of the overfitting phenomenon.

five models (Figure 10). The Diff 1, Diff5, and Diff10 models show similar performance patterns with respect to choosing optimal sigmoid function thresholds with optimal values for accuracy and F1-score in the range of 0.2 to 0.5. while the Blended and the Blended-stacked models showed better performance for higher thresholds in the range of 0.4 to 0.8. All five models have increasingly higher precision for increasingly higher threshold values while their recall decreases at a much steeper rate (more accentuated for Diff1, Diff5, and Diff10) as the threshold values increase. The Diff 1, Diff5, and Diff10 models are more sensitive to movement, using small differences in intensity to make predictions. Therefore, when the threshold is increased, only movements with high intensity remain visible, which could account for the steeper decline in recall. Comparatively, for the blended models less intense movement will add up and still be used to make predictions. As a result, less false negatives will occur as the threshold increases, further resulting in a higher recall. To maintain a good performance

Table 4. Prediction results for individual sub-class behaviors obtained with the stacked blended model. The results include average recall and standard deviation averaged over three repeated experiments as well as, the average recall for aggressive behaviors, non-aggressive behaviors, and the overall weighted average

Behavior	Binary class	Recall (↓)
Head biting	Aggressive	0.95 ± 3.61
Immobile	Non-aggressive	0.94 ± 0.58
Parallel pressing	Aggressive	0.91 ± 2.52
Inverse parallel pressing	Aggressive	0.89 ± 9.81
Chasing	Aggressive	0.84 ± 28.29
Mobile	Non-aggressive	0.75 ± 2.08
Head-to-head knocking	Aggressive	0.72 ± 8.54
Mounting	Non-aggressive	0.63 ± 3.00
Head-to-body knocking	Aggressive	0.57 ± 28.50
Summary		
Average recall of aggressive behaviors		0.81
Average recall of non-aggressive behaviors		0.77
Weighted average recall		0.79

balance for all models in this study, our research results are reported for a sigmoid function threshold of 0.5.

Computational efficiency

Table 5 includes a summary of the training-validation, testing, and data preprocessing runtimes. The models using the blended dataset, without stacked convolutional layers, require up to 4 times less time to train and validate when compared to the Diff1, Diff5, and Diff10 models. This is mainly due to a significantly reduced size of the datasets (up to 50 times less data) since each blended frame is created using 50 single frames. The reduced dataset does lower the runtime for the stacked model using the blended dataset, compared to the Diff1, Diff5, and Diff10 models, but the addition of extra layers results in it having double the runtime of the unstacked model using the Blended dataset.

Similarly, the testing time of the models trained and validated on the blended dataset is up to 24 times shorter compared to the ones trained on the Diff1 dataset, while the dataset preprocessing takes up to 2.3 times less time than for the Diff1 dataset (Figure 11). From a data preprocessing perspective, the most efficient models are the ones using the blended approach, which are up to 2.3 times more efficient than the Diff1, Diff5, and Diff10 models. In a practical implementation on farm where data preprocessing and testing are bundled in a software system, the combined average preprocessing and testing time for a single 2-s video requires between 0.34 s (Blended) and 0.89 s (Diff1) suggesting that real-time capabilities can be achieved using our solution. Nevertheless, our solution was applied and tested only on pairs of animals and thus additional computational costs need to be considered when applied to larger pens consisting of larger groups of pigs.

Challenges and limitations

While we were particularly careful with the experimental design and setup for this study, we have noticed limitations and challenges that are worth mentioning and could be potentially useful for future studies. One limitation related to using image differentials is the potential for error because of artifacts occurring in the data. While the conditions in these experiments were highly controlled, fluctuations in illumination, pest interference, and movement of enrichment devices (e.g., chain) could not be controlled. As previously discussed, frames were cropped to eliminate interferences from outside the pen. Threshold settings for pixel intensities for the blended dataset from 0 to 51 were set to 0 to further eliminate any noise from within the pen. Despite these strategies to mitigate errors, there were still artifacts (e.g., visible floorboards) that occurred as shown in Figure 12. These artifacts create the illusion of movement potentially resulting in false classifications. Such situations and pen settings could significantly affect the final results when DL models are deployed in commercial farms, and therefore, alternative modeling approaches would be desired that are more flexible and more robust to various environmental factors. Moreover, our study focused on pairs of pigs, which potentially limits the diversity of aggressive and non-aggressive behaviors that could be encountered in larger groups of pigs where the number of interactions among individuals increases.

Another significant limitation of the study is the relatively low number of specific pig aggressive sub-behaviors, such as chasing compared to other more prevalent aggressive sub-behaviors (head biting) present in our dataset. The lack of sufficient examples of a specific sub-behavior could lead to a decreased prediction accuracy for the models. Moreover, our study includes a relatively low number of images since we apply a supervised DL approach that requires significant efforts to watch hours of video recordings and manually annotate them second by second. Nevertheless, when compared to dataset sizes from other studies that seem to suffer from the same limitations, our dataset, which consists of 16 h of total recording of which 4,494 s were included in the analyses, is of comparable size. For example, the Chen et al. (2017) study selected 2,057 s of video recordings from 60 h of videos, while studies conducted by the same group in 2019 and 2020 selected 10,434 and 9,600/2,400 s (with/without augmentation) of video recordings out of 48 and 24 h of total recordings.

Conclusions

Being able to identify aggressive behavior automatically and accurately in pigs is an important step towards mitigating the issue of aggression. In this study, we tested an image differential approach with supervised DL to detect aggression in pairs of pigs, with a focus on creating a faster model that requires fewer computational needs. We tested four approaches to determine the optimal frame differential construction (Diff1, Diff5, Diff10, and blended) and then we used two CNN architectures to classify the behavior. The results showed that all four approaches were able to classify aggressive behavior with a relatively high degree of accuracy, with the stacked CNN model on the blended dataset producing the best results and requiring the least training time. Nine different sigmoid activation function thresholds were tested to determine optimal threshold. All models saw an increase in precision and a decrease in recall at higher thresholds. The Diff1, Diff5, and Diff10 models produced the best F1-score and accuracy results at lower

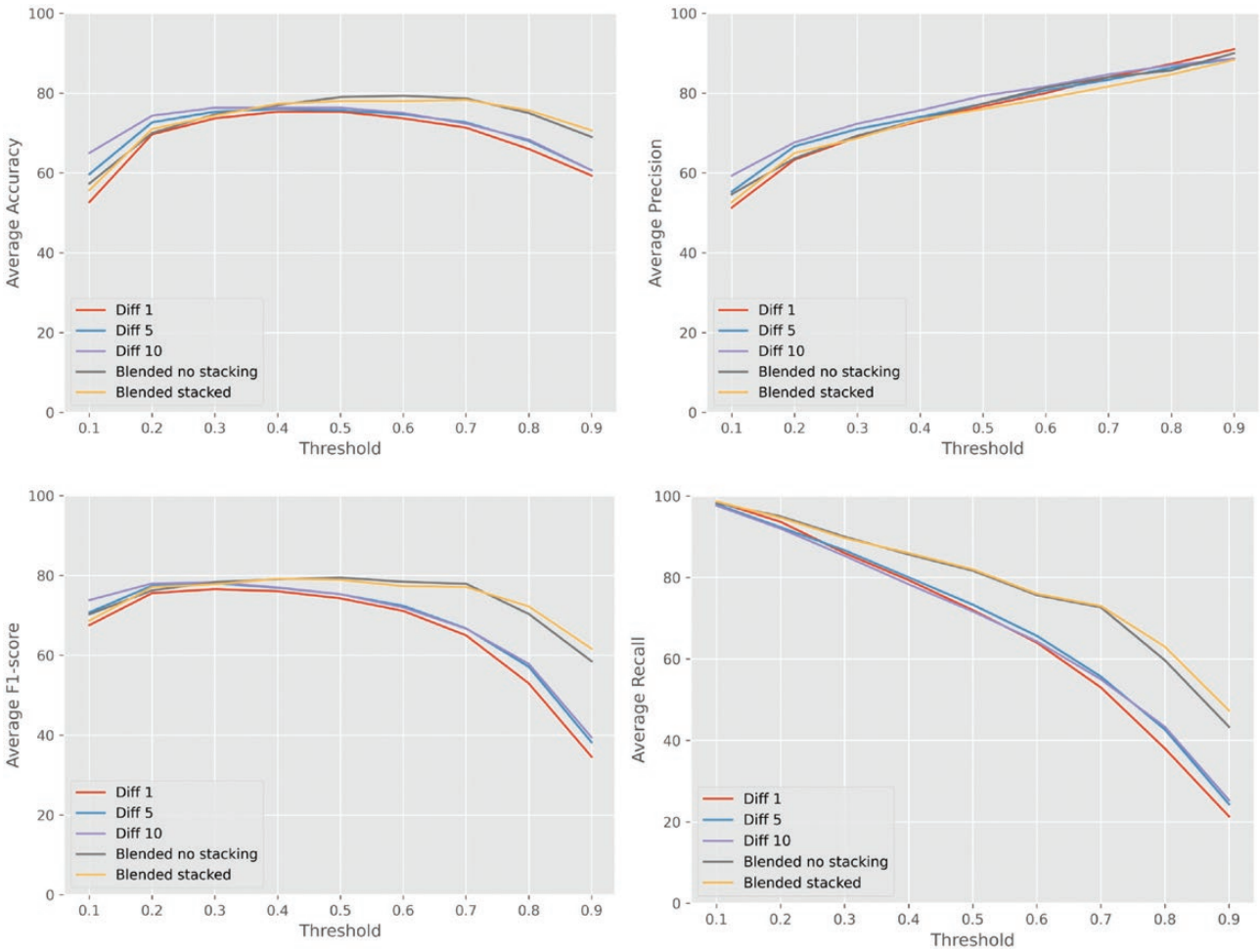

Figure 10. The impact of using various sigmoid activation function thresholds on the prediction performance of the five models. Each plot includes five curves representing either the accuracy, precision, recall or F1-score corresponding to a combination of convolutional neural network model architectures (with or without stacking), and datasets (diff1, diff5, diff5, and blended).

Table 5. Average and standard deviation training-validation, testing, and data preprocessing runtimes for the five models. Each result was averaged over three repeated experiments. Bolded values represent the best results obtained with the models employed in this study.

	Training-validation run time (s)	Testing run time (s)	Data preprocessing run time (s)
Diff1	1,382.33 ± 110.53	142.67 ± 19.63	859.33 ± 54.24
Diff5	1,259.67 ± 112.07	134.67 ± 11.93	686.00 ± 104.68
Diff10	1,207.00 ± 158.79	126.33 ± 1.53	656.33 ± 40.02
Blended	**396.67** ± 45.39	6.67 ± 0.58	**373.33** ± 118.49
Blended stacked	893.00 ± 56.56	**6.00** ± 0.00	**373.33** ± 118.49

thresholds. Comparatively, the blended models produced the best F1-score and accuracy at higher thresholds. To ensure the best performance balance for all models, a 0.5 threshold was used. Runtimes for the blended dataset without stacked CNNs, took four times less time to train and validate than the Diff1, Diff5, and Diff10. Furthermore, preprocessing took 2.3 times less time for the blended approach, and the dataset was 24 times smaller than the other non-blended datasets. Two significant limitations of this study were the small dataset and the use of pairs of pigs, which resulted in a low availability of certain behaviors such as chasing or mounting. Future work with larger datasets and groups of pigs could lead to improved results and will eventually show the scalability of our method. Overall, this study showed that using CNN DL models and an image differential approach can produce meaningful results faster and require less computational needs when applied to video recordings for pairs of pigs.

Acknowledgments

The authors would like to thank the Arkell Swine Research Facility personnel for their help and dedication and University of Guelph for continuously supporting our efforts. This work was a part of the project funded by the Canada First Research

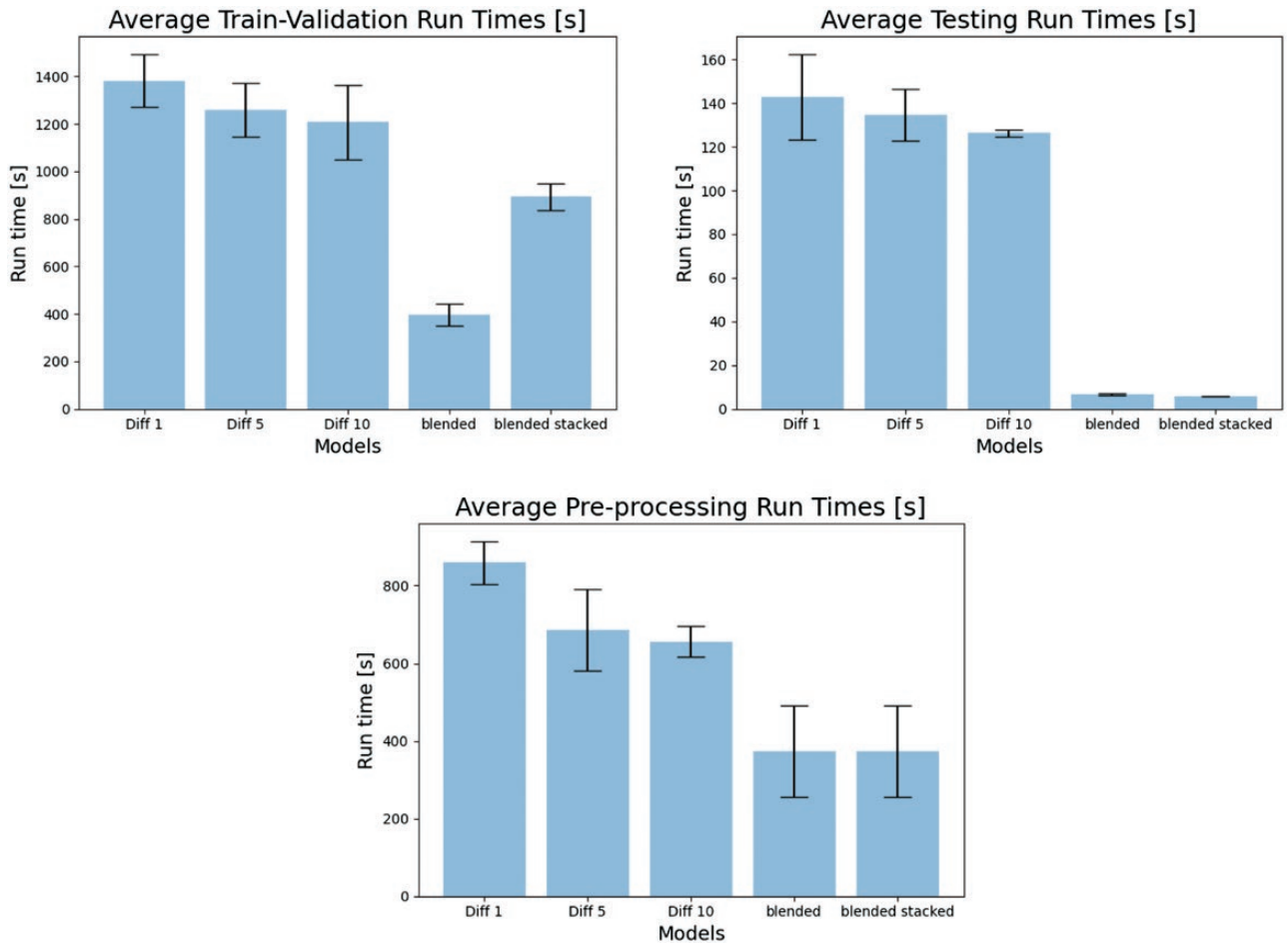

Figure 11. Summary of runtimes for the five models. The figure includes the average and standard deviation training, validation, testing, and data preprocessing runtimes averaged over three repeated experiments.

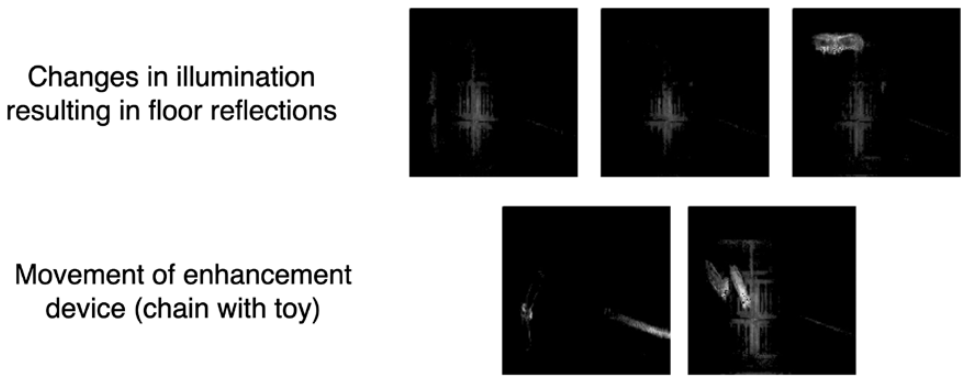

Figure 12. Artifacts caused by changes in the pig housing environment. The changes depicted in this figure were caused by sudden changes in illumination or movement of enhancement devices.

Excellence Fund (CFREF grant number: 499091) and the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Supplement (grant number: 401790).

CRediT

Jasmine Fraser: Investigation, Data curation, Formal analysis, Project administration, Writing—original draft and edits. **Harry Aricibasi:** Software, Data curation, Investigation, Methodology, Formal analysis, Project administration. **Renee Bergeron:** Conceptualization, Writing—review & editing, Resources, Funding acquisition, Supervision. **Dan Tulpan:** Conceptualization, Writing—review & editing, Resources, Funding acquisition, Supervision.

Conflict of Interests Statement

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.

References

Arey, D. S., and S. A. Edwards. 1998. Factors influencing aggression between sows after mixing and the consequences for welfare and production. *Livest. Prod. Sci.* 56:61–70. doi: 10.1016/s0301-6226(98)00144-4

Bengio, Y. 2009. Learning deep architectures for AI. foundations and trends in machine learning. 2:1–127. doi: 10.1561/2200000006

Bi, X., X. Zhao, H. Huang, D. Chen, and Y. Ma. 2020. Functional brain network classification for alzheimer's disease detection with deep features and extreme learning machine. *Cognit.. Comput.*. 12:513–527. doi: 10.1007/s12559-019-09688-2

Chen, C., W. Zhu, C. Ma, Y. Guo, W. Huang, and C. Ruan. 2017. Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs. *Comput. Electron. Agric.* 142:380–387. doi: 10.1016/J.COMPAG.2017.09.013

Chen, C., W. Zhu, Y. Guo, C. Ma, W. Huang, and C. Ruan. 2018. A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs. *Livest. Sci.* 218:70–78. doi: 10.1016/J.LIVSCI.2018.10.013

Chen, C., W. Zhu, D. Liu, J. Steibel, J. Siegford, K. Wurtz, J. Han, and T. Norton. 2019. Detection of aggressive behaviours in pigs using a RealSence depth sensor. *Comput. Electron. Agric.* 166:105003. doi: 10.1016/J.COMPAG.2019.105003

Chen, C., W. Zhu, J. Steibel, J. Siegford, K. Wurtz, J. Han, and T. Norton. 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. *Comput. Electron. Agric.* 169:105166. doi:10.1016/J.COMPAG.2019.105166

Fàbrega, E., X. Puigvert, J. Soler, J. Tibau, and A. Dalmau. 2013. Effect of on farm mixing and slaughter strategy on behaviour, welfare and productivity in Duroc finished entire male pigs. *Appl. Anim. Behav. Sci.* 143:31–39. doi: 10.1016/j.applanim.2012.11.006

Fels, M., L. Schrey, S. Rauterberg, and N. Kemper. 2021. Early socialisation in group lactation system reduces post-weaning aggression in piglets. *Vet. Rec.* 189:e830. doi: 10.1002/vetr.830

Greener, J. G., S. M. Kandathil, L. Moffat, and D. T. Jones. 2021. 2021 A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* 23:40–55. doi: 10.1038/s41580-021-00407-0

He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016-December:770–778, Las Vegas (NV). doi: 10.1109/CVPR.2016.90

Jensen, P. 1982. An analysis of agonistic interaction patterns in group-housed dry sows — aggression regulation through an "avoidance order". *Appl. Anim. Ethol.* 9:47–61. doi: 10.1016/0304-3762(82)90165-1

Kingma, D. P., and J. L. Ba. 2014. Adam: a method for stochastic optimization. 3rd international conference on learning representations, ICLR 2015 - Conference Track Proceedings, San Diego (CA). doi: 10.48550/arxiv.1412.6980

Le, E. P. V., Y. Wang, Y. Huang, S. Hickman, and F. J. Gilbert. 2019. Artificial intelligence in breast imaging. *Clin. Radiol.* 74:357–366. doi: 10.1016/j.crad.2019.02.006

Lee, J., L. Jin, D. Park, and Y. Chung. 2016. Automatic recognition of aggressive behavior in pigs using a Kinect depth sensor. *Sensors (Basel)*. 16:631–643. doi: 10.3390/s16050631

Meese, G. B., and R. Ewbank. 1973. The establishment and nature of the dominance hierarchy in the domesticated pig. *Anim. Behav.* 21:326–334. doi: 10.1016/s0003-3472(73)80074-0

Mei, H., B. Yang, J. Luo, and L. Gan. 2016. The effect of mixing levels on aggression at weaning in piglets. *Appl. Anim. Behav. Sci.* 179:32–38. doi: 10.1016/J.APPLANIM.2016.03.009

Neubauer, C. 1998. Evaluation of convolutional neural networks for visual recognition. *IEEE Trans. Neural Netw.* 9:685–696. doi:10.1109/72.701181

O'Connell, N. E., V. E. Beattie, and B. W. Moss. 2003. Influence of social status on the welfare of sows in static and dynamic groups. *Anim. Welf.* 12:239–249. doi: 10.1017/S0962728600025665

Oczak, M., S. Viazzi, G. Ismayilova, L. T. Sonoda, N. Roulston, M. Fels, C. Bahr, J. Hartung, M. Guarino, D. Berckmans, et al. 2014. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. *Biosyst. Eng.* 119:89–97. doi: 10.1016/j.biosystemseng.2014.01.005

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: a primer for ecologists. *Q. Rev. Biol.* 83:171–193. doi: 10.1086/587826

Simonyan, K., and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. 3rd international conference on learning representations, ICLR 2015 - Conference Track Proceedings. doi: 10.48550/arxiv.1409.1556

Stukenborg, A., I. Traulsen, B. Puppe, U. Presuhn, and J. Krieter. 2011. Agonistic behaviour after mixing in pigs under commercial farm conditions. *Appl. Anim. Behav. Sci.* 129:28–35. doi: 10.1016/j.applanim.2010.10.004

Suk, H. I., S. W. Lee, and D. Shen; Alzheimer's Disease Neuroimaging Initiative. 2014. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI Diagnosis. *Neuroimage* 101:569–582. doi: 10.1016/j.neuroimage.2014.06.077

Valletta, J. J., C. Torney, M. Kings, A. Thornton, and J. Madden. 2017. Applications of machine learning in animal behaviour studies. *Anim. Behav.* 124:203–220. doi: 10.1016/j.anbehav.2016.12.005

Viazzi, S., G. Ismayilova, M. Oczak, L. T. Sonoda, M. Fels, M. Guarino, E. Vranken, J. Hartung, C. Bahr, and D. Berckmans. 2014. Image feature extraction for classification of aggressive interactions among pigs. *Comput. Electron. Agric.* 104:57–62. doi: 10.1016/j.compag.2014.03.010

Wu, Y., M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al. 2016. Google's neural machine translation system: bridging the gap between human and machine translation. *ArXiv preprint. arXiv:1609.08144*.