Agronomy Research **24**(2), 76–95, 2026
https://doi.org/10.15159/AR.26.034

Development and pilot testing of the 'Lesion-Detect' system for automated identification and scoring of gross lung lesions in pigs during slaughter

T.M. Banhazi[1,2,*], M. Dunn[1,3], A. Banhazi[1], J. Reixach[4], S. Gol[4], A. Ramon-Perez[5], P. Llonch[5], V. Vujacic[6], J. Sudar[6], N. Kajari[6], J. Maselyne[7], N. Van Noten[7] and F. Tuyttens[7,8]

[1]InnoTech Vision ApS., Niels Pedersens Alle 2, 8830 Tiele, Denmark
[2]Wroclaw University of Environmental and Life Sciences, Norwida 25, 50-375 Wroclaw, Poland
[3]Leading Edge Ltd., Brisbane, PO Box 363, Brisbane, Australia
[4]Selección Batallé S.A., 17421 Riudarenes, Girona, Spain
[5]Autonomous University of Barcelona, Plaça Cívica, Bellaterra, 08193 Cerdanyola del Vallès, Catalonia, Spain
[6]Carnex d.o.o., Kulski put 26, 21460, Vrbas, Serbia
[7]ILVO, Burgemeester van Gansberghelaan 92 box 1, 9820 Merelbeke-Melle, Belgium
[8]Ghent University, Sint-Pietersnieuwstraat 25, B-9000 Ghent, Belgium
[*]Correspondence: thomas.banhazi@plfag.com

Received: January 23rd, 2026; Accepted: May 7th, 2026; Published: May 25th, 2026

Abstract. Respiratory diseases are the most significant health and welfare challenges in pig production. Enzootic pneumonia (EP) is often subclinical at farm level, making early detection difficult, yet its economic and welfare impacts are substantial. Suboptimal housing conditions are well-established risk factors that predispose pigs to respiratory disease. These conditions negatively affect animal health and welfare and are associated with increased mortality, higher medication costs, reduced feed intake, sub-optimal average daily gain, and increased organ condemnations, representing a major constraint to sustainable pig production. At slaughter, gross lung lesions associated with respiratory disease can be assessed in all carcasses, offering a valuable opportunity for herd-level monitoring. However, current Pig Health Monitoring System (PHMS) lesion scoring methods are labour-intensive, subjective, and poorly scalable. To address these limitations, we developed and pilot-tested Lesion-Detect (LD), an automated image-analysis system for identifying and scoring gross lung lesions under commercial slaughterhouse conditions. LD prototypes were deployed in two European slaughterhouses (Spain: hanging lungs; Serbia: tray-based presentation) and benchmarked against manual PHMS scoring. The system showed strong agreement with trained human assessors, with false positive and false negative rates typically within 2–3% of lung area, comparable to inter-observer variability. By delivering objective and scalable lesion scoring, LD could enable consistent respiratory health surveillance. Beyond data collection, rapid herd-level feedback supports earlier and more targeted on-farm interventions, potentially reducing antimicrobial use and improving animal welfare. With further refinement, the system may be

extended to additional carcass-level welfare indicators, strengthening its role within precision livestock farming and integrated animal welfare assessment frameworks.

Key words: precision livestock farming, enzootic pneumonia, pig health, lung lesion scoring, slaughterhouse monitoring; animal welfare assessment, image-analysis.

INTRODUCTION

Respiratory diseases are among the most persistent and damaging health challenges in pig production. Particularly enzootic pneumonia (EP), (caused predominantly by *Mycoplasma hyopneumoniae*) remains highly prevalent across Europe and globally. Respiratory diseases are often subclinical, making the on-farm diagnosis quite difficult (Murphy, T. et al., 2012), but EP exerts profound impacts on animal health, welfare, and production economics on farms.

There are multiple risk factors reported in the literature that predispose pigs (and farm workers) to respiratory diseases, mainly associated with suboptimal housing conditions (Banhazi, T. & Cargill, C., 1998; Murphy, T. et al., 2012; Banhazi, T.M., 2013; Banhazi, T. & Pisaniello, D., 2018; Banhazi, T.M. et al., 2022). It has been proven that high concentration of airborne pollutants is a major risk factor for respiratory disease development in livestock and farm workers (Curtis, S.E. et al., 1975; Donham, K.J., 1989; Crook, B. et al., 1991; Donham, K.J. et al., 1995; Stärk, K.D.C.2000; Done, S.H. et al., 2005; Murphy, T. et al., 2012). In turn, these diseases are causing substantial economic losses for pig producers because of the extra costs associated with increased mortality, veterinary care, reduced feed intake and sub-optimal average daily gain (Niemi, J.K. et al., 2004; Niemi, J.K. et al., 2010). Furthermore, respiratory disease contributes to the condemnation of carcasses and organs at slaughter, losses that are frequently borne directly by slaughterhouses (Heinonen, M. et al., 2001). Collectively, these extra cost factors represent a recurring financial burden on producers, processors, and the broader pork industry.

Respiratory diseases can be potentially diagnosed during the slaughtering process, because the gross lesions associated with these diseases are relatively readily observable (Pointon, A.M. & Banhazi, T., 1995; Holt, H.R. et al., 2011). Currently, the Pig Health Monitoring Scheme (PHMS) serves as a key international instrument for evaluating the health of pig herds at slaughterhouses (Rodriguez Carino, C. et al., 1998; Pointon, A.M. et al., 1999; Holt, H.R. et al., 2011). This surveillance framework employs trained assessors to conduct visual inspections of various respiratory diseases such as enzootic pneumonia immediately after slaughter. The PHMS scoring system has been described previously in a number of publications (Stärk, K.D.C. & Nevel, A., 2009; Holt, H.R. et al., 2011; Maes, D. et al., 2023). In summary, the simplified version of the Australian PHMS system used in this study allocates a maximum 50 points to a lung that is fully impacted by EP like lesions using the following breakdown of score allocation:

- Top lung lobes (left and right) – 10 points each if fully affected by EP like lesions
- Middle lung lobes (left right) – 10 points each if fully affected by EP like lesions
- The upper corner of the lower or large lungs lobes (left and right) – 5 points if fully affected by EP like lesions

Unfortunately, the PHMS system is expensive and inefficient due to the need for highly trained staff to permanently undertake manual data collection. Furthermore, the

inherent variability in human observations, limits both the efficiency and scalability of the PHMS approach.

Recent advances in computer vision and image analysis offer a compelling alternative (Lawrence, K.C. et al., 2001; Brosnan, T. & Sun, D.-W., 2002; Cronin, G.M. et al., 2008; Mehdizadeh, S.A. et al., 2015; Shwetha, V. et al., 2024; Iseki, H. et al., 2025). Automating these health inspections via image analysis techniques could make the PHMS system more efficient, consistent and economically viable. Beyond reducing labour demands, automated systems provide objective measurements, enabling more reliable surveillance of herd respiratory health.

Therefore, R&D aims of the aWISH project (Horizon Europe, Grant agreement ID: 101060818) include (1) creating an automated and cost-effective method of assessing the respiratory health of pigs at slaughter and (2) utilising this information to assess the general welfare status of pigs. Specifically, the objective of this self-contained study within the aWISH framework was to create a new, innovative lung lesion detection system, utilising advanced image analysis methods and artificial intelligence techniques. It is hoped that such automated lung lesion monitoring systems will enhance the ability of veterinarians to treat the remaining animals on farm in a more appropriate and timely manner, thus improving the general health, welfare status and productivity of animals. Furthermore, it is assumed that early treatment of respiratory diseases can reduce antibiotic use generally, resulting in reduced antimicrobial resistance. It is expected that by replacing manual inspection with scalable automation, Lesion-Detect (LD) has the potential to transform how respiratory health is monitored, contributing to a more sustainable, efficient, and welfare-conscious livestock industry.

MATERIALS AND METHODS

Overall LD system description

The LD system consists of a high-resolution camera unit integrated into the slaughter line, an image acquisition module, and an AI-based lesion segmentation algorithm. The algorithm identifies lung boundaries, detects lesion areas, and estimates lesion severity according to the Australian PHMS scoring system. Its design integrates with slaughterhouse workflows while withstanding the operational challenges of the slaughterhouse environment, such as dust, water exposure, and variable lighting.

Study sites

The LD system was evaluated in two commercial slaughterhouses in Europe, each representing different lung presentation methods. Trial 1 was conducted in Spain, where lungs were assessed in a hanging presentation directly on the slaughter line. A prototype LD system was installed in the Spanish slaughterhouse in late 2023. The slaughter pig processing line where the experiments took place is presented in Fig. 1 with the LD camera system visible in the middle of the picture. Trial 2 was conducted in Serbia, where lungs were presented in viscera trays. The second prototype LD system (depicted in Fig. 2) was installed in Serbia in late 2024. Both sites process pigs with average carcass weights ranging between 85–98 kg depending on breed (Large White and Duroc) and slaughter age (175–220 days). Trial 1 site used an average process chain speed of 7–10 pigs per minute, while at Trial 2 site the chain speed was around 4–5 pigs per minute.



Figure 1. The Trial 1 processing line with the camera (visible in the middle of picture) installed to provide a side view of the lungs hanging on hooks. Ultimately, the camera design will be such to provide flexibility with slaughter line integration.



Figure 2. Left: The installed camera above the viscera trays at the Trial 2 site. Right: Side view of the viscera trays.

Data collection hardware

The image capture system was built from customised off-the-shelf parts assembled into an IP67 enclosure (Fig. 3).

The main computer platform consists of a fanless Intel Celeron Jasper Lake N5105 (Mele Quieter 3Q, Mele Technologies Ltd. Shenzhen, China) with Ubuntu 20.04 operating system. Two separate USB3 cameras with different lenses were installed

inside the enclosure behind the transparent plastic lid, ensuring clear field of view as well as water and dust protection. Two See3CAM_CU20 (e-con Systems™ Inc., Riverside, USA) cameras were installed with a narrow field of view (Camera 0) and wide field of view (Camera 1) lens, with image resolution of 3,840x2,160 pixels. The recording system was a custom program written in C++ to record an image from each camera periodically (every 1 second). The captured images were stored onto a network accessible storage drive Synology Diskstation DS223 (Synology America, Bellevue, USA). The general flow for the image processing was established by the following steps: (1) read images, (2) detect lung, (3) measure the extent of lesion and (4) log results in a secure folder.



Figure 3. Lesion-Detect system developed for capturing images of lungs under commercial conditions at the Trial 1 site.

Image capture procedures at the experimental sites

In Trial 1, the initial experimental procedure was simple and involved having a person manipulating and positioning the lungs passing in front of the camera to ensure maximum visibility of lung lesions by the cameras (Fig. 4). In the later part of the LD development, such manipulation was deemed to be unnecessary, making the system more suitable for commercial application. At the Trial 2 site, slaughterhouse workers attempted to separate the lungs from the remaining viscera (as part of the normal operational procedures) and placed the lungs in a separate tray.

a) 

b) 

Figure 4. Synchronised a) Camera 0 sample image and b) Camera 1 sample image in Trial 1.

Post-processing of images captured

Images were post-processed (as described below in detail) that included automated cropping of images, segmentation and lung-score generation. First, the lungs had to be segmented from the 'cluttered' slaughterhouse background environment. Altogether, in excess of 17,000 images have been collected throughout the various experiments and portion of these collected images were used to validate the relationship between the visual assessment scoring system used during PHMS inspections (gold standard) and the images interpreted by the image analysis system (Lesion-Detect). To optimize the processing, a region of interest of 800×700 pixels in the centre of the image was extracted from the full-size images. This both reduced the complexity of the image analysis, as well as ensured that the correct sample was analysed when multiple objects were present in the full field of view (Fig. 5). In the recorded video stream, a single image of each lung is captured for the training data as it passes the middle of the frame. In this way, we achieve consistency of scale and lighting, although not orientation.

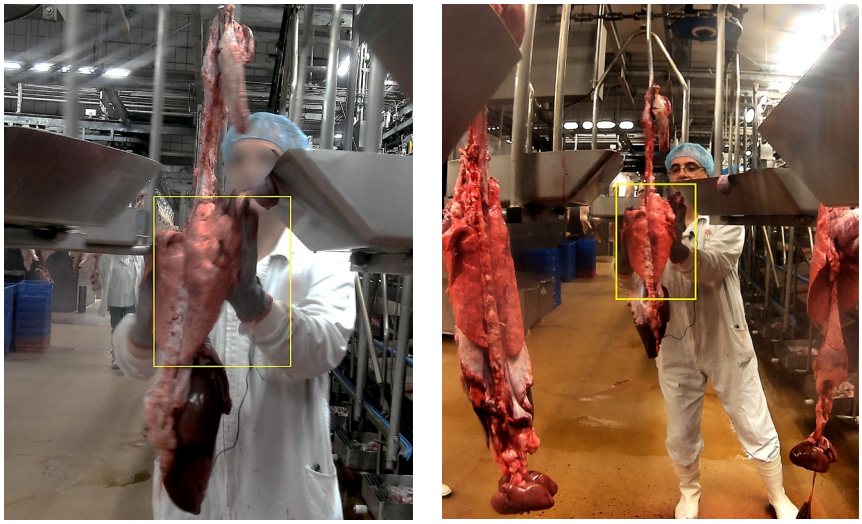

Figure 5. Automatically centralized region of interest selected by the Lesion-Detect system.

A set of 413 hanging images (from Trial 1) were manually labelled by delineating the lung region to separate it from the surrounding carcass tissue and background (Fig. 6). Manual annotation was performed using the polygon selection tool in ImageJ (https://imagej.net/ij/) by a single trained observer. For each image, the annotator placed key points along the visible outer edge of the lung to capture its boundary as accurately as possible. The visible perimeter was used as the defining criterion for inclusion, and areas with unclear margins were traced following the most continuous anatomical contour. Each completed selection was converted to a binary mask using the õCreate Maskõ function and saved with the same base filename as the original image plus an additional extension. Similarly for Trial 2 (tray images), a set of 229 images were labelled using the same procedure.

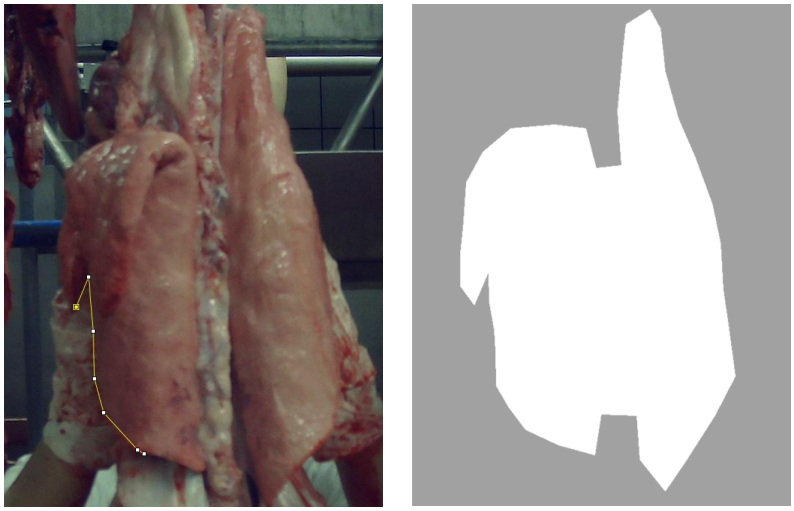

Figure 6. Manually segmented image and the binary mask that has been created.

Segmentation model development. For Trial 1, the segmentation of the lung region in hanging carcass images was performed using a DeepLabV3+ architecture implemented with the segmentation_models_pytorch library (SMP v0.3 Iakubovskii 2019). The model used (Traore, B.B. et al., 2018; Mohapatra, S. et al., 2021) follows the class of deep convolutional encoder-decoder networks developed for image segmentation tasks (Badrinarayanan, V. et al., 2017; Zhang, F. et al., 2018). Input images were RGB, and the output was a single-channel binary mask delineating the visible outer lung boundary.

The 413 manually annotated hanging-lung images were randomly split into 80% training and 20% validation subsets. To enhance robustness, small random rotations (±10°), horizontal flips, brightness/contrast jitter, and resizing to 256×256 pixels were applied synchronously to each image-mask pair. Optimisation used the Adam optimiser (learning rate = 0.001), and the best checkpoint was selected based on minimum validation loss.

Trial 2 involved tray-mounted lung images, which presented a markedly different orientation, background, and illumination from Trial 1. Consequently, a separate segmentation pipeline was developed and trained using similar encoder-decoder architectures but with preprocessing and augmentations adapted to the tray presentation. The same region-of-interest definition (the visible outer lung boundary) was maintained to ensure comparability between trials. Training again employed 80/20 split of the 229 manually segmented images, BCE + Dice loss and identical hyper-parameters. At this stage, no further detail of the segmentation process can be given due to patenting considerations (provisional patent application No. 2025906448).

Validation and quality assessment. Predicted masks were evaluated against manual ground-truth annotations using Dice similarity and Intersection-over-Union (IoU) metrics on the held-out validation set. In addition, a single trained observer performed a visual quality review, classifying each segmentation as perfect (full boundary agreement), good (< 5% boundary discrepancy), or failed (> 5%

discrepancy). These visual categories corresponded approximately to Dice ≥ 0.97, $0.90 \leq$Dice < 0.97, and Dice < 0.90, respectively. All prediction overlays were logged and archived for traceability. Example results are shown in Figs 7 and 8.

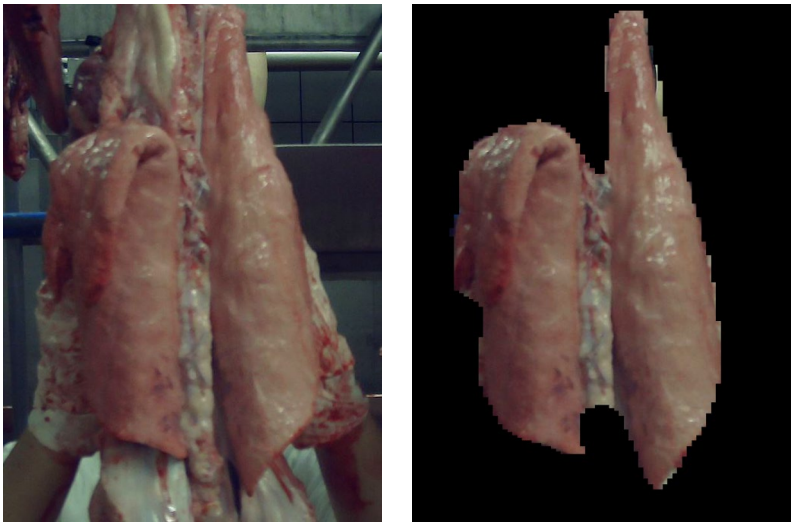

Figure 7. Automatically segmented example hanging image displayed as mask.

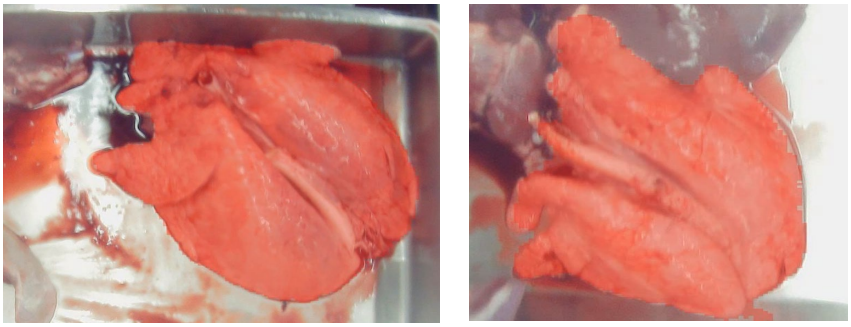

Figure 8. Automatically segmented example tray image displayed as visible overlay.

Identifying and scoring lung lesions

Ground truth annotation. Lesions were annotated in 332 images using a specialised custom-GUI tool. The custom annotation software, developed in Python, enabled efficient labelling through keystroke and mouse interactions. For each image the trained observer identified the *centrepoint* of each visible lesion within the lung region (as defined in the preceding segmentation step). From each centrepoint, a Gaussian intensity profile (heat-kernel) was generated to create the ground-truth mask: i.e., pixel values within the mask correspond to a 2D Gaussian dropping off from the lesion centre (Fig. 9). Using this approach a continuous probabilistic lesion map (rather than a strict binary mask) was produced. This style of heat-map regression (with Gaussian smoothing of point annotations) has been used as a 'heatmap regression' approach for lesion detection (Myers-Colet, C. et al., 2022).

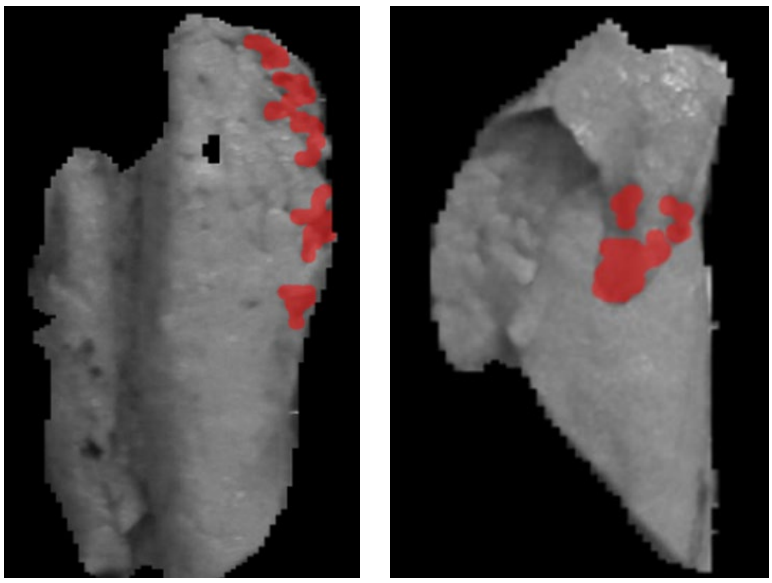

Figure 9. Manually segmented example images.

Training. A fully convolutional encoder-decoder network was trained to predict lesion-likelihood maps from the input images. The architecture followed a U-Net or DeepLabV3+ design, with a ResNet-based encoder pre-trained on ImageNet and a decoder configured for dense pixel-wise prediction. A data pool of approximately, 5700 images were used for both training and validation purposes. These images were processed and manually scored by the lead author of this article (who has 4 years of hands-on experience with lung scoring in commercial slaughterhouse environments) using the custom-GUI tool that was developed in-house and described under 2.6.1 subheadings (Fig. 9). Input images were normalised and resized to a consistent spatial resolution, and data were randomly divided into training (80%) and validation (20%) subsets.

The model was optimised using a compound loss combining binary cross-entropy and Dice similarity loss, encouraging both intensity calibration and spatial overlap with the annotated heatmaps. Standard data augmentation (geometric and photometric) was applied to improve robustness to lighting and orientation. Training continued until convergence of the validation loss, with model checkpoints selected by the lowest validation error.

Validation. Model performance was assessed on the held-out validation set. Quantitative evaluation used pixel-wise Dice similarity and Intersection-over-Union (IoU) between predicted and annotated heatmaps, as well as mean squared error (MSE) for intensity correlation. Visual inspection of overlay images confirmed that predicted lesion peaks aligned with annotated regions and were restricted to plausible anatomical locations within the lung segmentation. The system was designed to ensure that it can be ultimately related to the scoring system used in European and Australian slaughterhouses (Maes, D. et al., 2023).

Initial study to deal with occlusion and obstruction issues at the Trial 1 site

Occlusion and obstruction of parts of the lung were identified as recurring challenges in the hanging presentation in Trial 1. To assess whether partial-lung scoring could be extrapolated to full-lung scoring, 180 lungs were manually scored in Trial 1 using the PHMS system by the main author (TB) who has 4 years of hands-on experience with PHMS scoring. The difference between the left and right lung areas were compared using simple descriptive statistics.

Data analysis

System output was compared to manual scoring using the following traditional analysis methods specifically used in various classification studies (Park, B. et al., 2007; Dervić, E. et al., 2024; Deng, Y. et al., 2025; Shams, M.Y. et al., 2025; Valeris-Chacin, R. et al., 2025).

- Precision: proportion of automatically identified lesion pixels that were correct.
- Recall: proportion of manual (gold standard) lesion pixels identified by the system.
- False Positive (FP) rate: % of lung area incorrectly identified as lesion.
- False Negative (FN) rate: % of lesion area missed.
- Intersection over Union (IoU): overlap between manual and automated segmentation.

The Australian PHMS scoring system Points-to-Pixel conversion was calculated according to (Pointon, A.M. et al., 1985; Pointon, A.M. et al., 1999):

1 PHMS PE score point = (0.5×Average lung pixels)/50

Where only 0.5 lung area is assumed to be affected by lung lesions (not the large lobes) and 50 PHMS score is the maximum amount that can be given for the EP affected 0.5 lung area.

As mentioned above, segmentation quality was assessed by comparing LD's automated lesion detection capability with manually delineated lesions. Traditionally, the segmentation is expected to be executed with high level of precision, because the application of such segmentation is often used in for example medical imaging, where millimetre differences matter. Therefore, high level of precision is needed. However, the application of the selected image segmentation techniques within the LD system is aimed to be highly practical but with a greater level of error tolerance, given the relatively course-grained nature of the PHMS scoring system that is used as gold standard (Goodwin, R.F.et al., 1969). Even in the images captured, the precise border between healthy and EP affected tissue is difficult to differentiate, despite the fact an evaluator would have lot more time for precise evaluation while evaluating the images compared to when the PHMS inspection is undertaken in situ (Pointon, A.M. et al., 1996). Depending on the processing chain speed, manual evaluators would literally have only seconds to make rapid decisions about the scores they want to allocate to the various lung lobes (Stärk, K.D.C. & Nevel, A. 2009; Holt, H.R. et al., 2011). Considering all these aspects, it is proposed that it is better and more practical if all pixels identified by manual evaluators and the automatic system are compared, regardless if it is classified as FP or FN. In this manner, the project can achieve a practical precision (probably already exceeding the precision of the manual PHMS lung scoring) without unnecessarily spending excessive time and money to achieve a precision that would not

make any practical difference. In essence, the authors of this article argue that the precision has to be ultimately evaluated at the batch and not at a lesion precision level. In Fig. 10 an example of the difference between manual and automatic scoring is demonstrated.

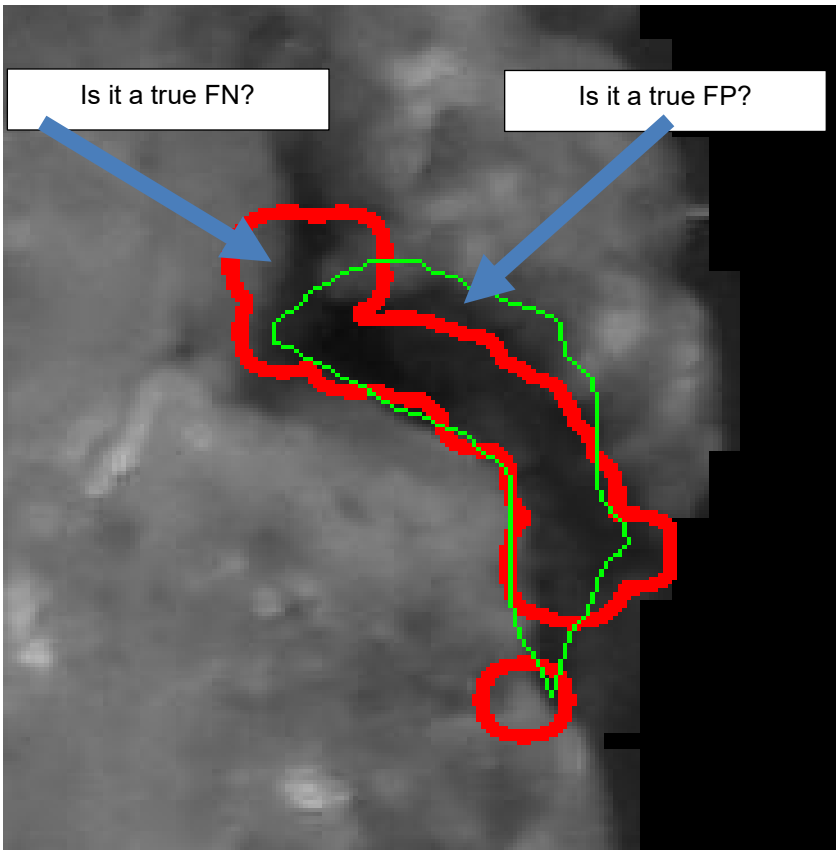

Figure 10. Segmented example image. Red = manually ID-d lesion area; Green = LD automatic ID. It can be seen from this image that classifying certain areas as False Negative (FN) & False Positive (FP) might not have a great relevance for the precision that is required for this practical application.

RESULTS AND DISCUSSIONS

The two LD prototypes evaluated under commercial slaughterhouse conditions were tested at Trial 1 site (with hanging lung presentation) and at Trial 2 site (with tray presentation). The key results are presented in tables 1–4 and figures 11–12 for average lung size, pixel-to-score conversion, segmentation accuracy, and practical considerations arising from field use.

The average lung size differed considerably between the two pilot sites due to differences in presentation and camera setup (Table 1). In Trial 1 (hanging lungs), the average lung size was ~120,000 pixels equivalent to 1,200 pixels per PHMS point.

Table 1. Key results obtained at the Trial 1 and 2 sites in terms of average lung size in pixels and pixel-to-PHMS score conversion

Site	Presentation	Images (n)	Average lung size (pixels)	Pixel-to-PHMS score conversion	FP* % of lung size	FN** % of lung size
Trial 1	hanging lungs	102	~120,000	1 PHMS score ≈ 1,200 pixels	2.54 (SD 2.41)	1.84 (SD 2.25)
Trial 2	lungs in trays	228	~206,661	1 PHMS score ≈ 2,067 pixels	3.28 (SD 2.53)	2.09 (SD 1.92)

*FP = false positive; **FN = False negatives.

In Trial 2 (tray presentation), the average size was ~206,661 pixels, equivalent to 2,067 pixels per PHMS point. The difference in lungs sizes in pixel and the resulting differences in pixel-per-PHMS point values has practical implications for calibration. The system needs to be adjusted per slaughterhouse setup to ensure consistency in score interpretation. Hanging presentation produced smaller pixel counts, potentially due to the greater distance to the camera, but the camera setup above the trays offered more controlled imaging conditions (Figs 11 and 12).

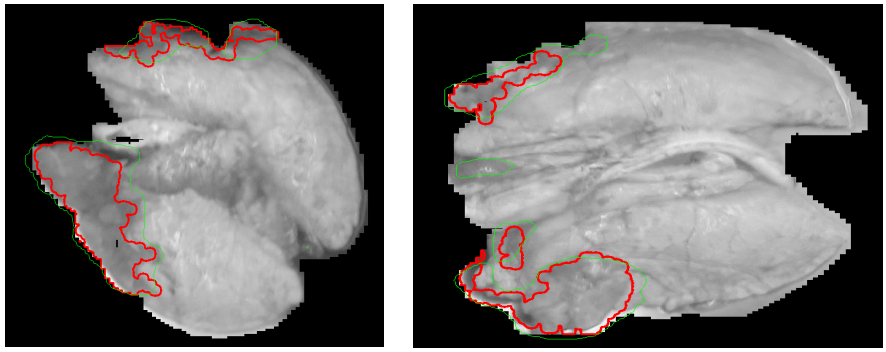

Figure 11. Examples of the overlap between automated and manual lesion identification in Trial 2. Red = manually ID-d lesion area; Green = LD automatic ID.

Precision and recall values between 0.5 and 0.6 were considered satisfactory in this context, as the objective of the system was not pixel-accurate medical diagnosis but reliable scoring of lesion extent over large anatomical regions. In production-line conditions, minor boundary uncertainty or small false detections have limited impact on the aggregate PHMS lesion score, which is averaged across thousands of pixels. For example, a precision of 0.5 at the pixel level still yields a correlation of > 0.9 with the final site-level PHMS grading, given the relative scale of the lung region (typically > 200,000 pixels). Thus, while precision and recall thresholds > 0.8 are desirable for clinical image segmentation, in this indusTrial application a balance of recall around 0.6 and precision > 0.5 provides sufficient reliability for automated scoring and trend monitoring.

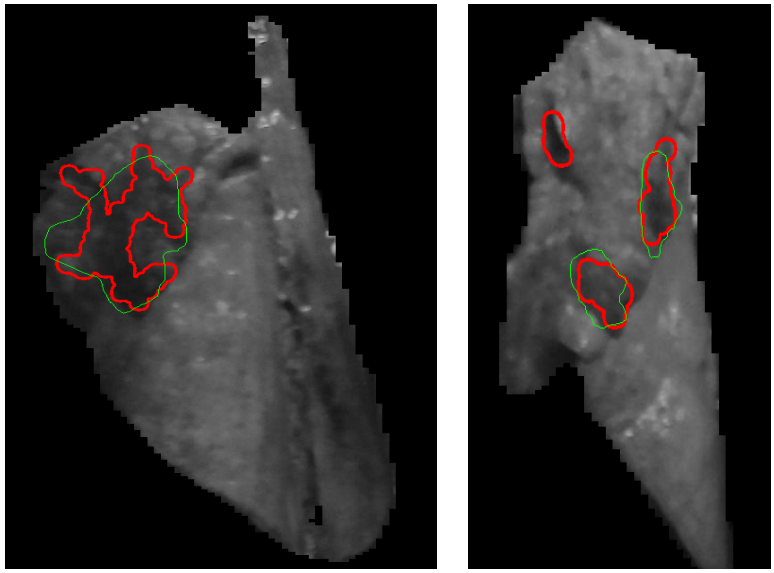

Figure 12. Examples of the overlap between automated and manual lesion identification in Trial 1. Red = manually ID-d lesion area; Green = LD automatic ID.

The initial algorithm demonstrated a high sensitivity for detecting positive cases but suffered from a high rate of false positives, as shown in Table 2. The first version of the detection algorithm had a very high recall (83%), so it was able to successfully identify 83% of all manually scored positive pixels, indicating good sensitivity to lesion presence. However, it had a very low precision (25%) as only 25% of the model's positive predictions were correct, leading to a high number of false positives.

Table 2. Key results obtained at the Trial 2 site over a number of iterations of the algorithm developed. Versions 1 and 2 compared and then versions 2 and 3 are also compared in terms of improvement in precision, recall, IoU, TP, FP, and FN detections

Metric	v1	v2	Δ (%)	Interpretation	v2	v3	Δ (%)	Interpretation
Precision	0.25	0.46	+87.6%	Fewer false positives	0.46	0.50	+8.7%	Fewer false positives
Recall	0.83	0.57	−30.3%	More missed detections	0.57	0.62	+8.7%	More true positives found
IoU	0.24	0.34	+45.6%	Substantial improvement	0.34	0.37	+8.7%	Improved overlap
TP (avg) Pixels	9,612	6,788	−29.4%	Fewer detections	6,788	7,382	+8.8%	More hits per image
FP (avg) Pixels	27,038	6,979	−74.2%	Significant reduction	6,979	6,784	−2.8%	Less noise per image
FN (avg) Pixels	1,350	4,175	+209.6%	Trade-off cost	4,175	4,454	+6.7%	Slight increase in misses

This model had a moderate IoU (25%), indicating a partial but incomplete overlap between predicted and manually scored lesions. Overall, the initial version of tray-based lesion identification algorithm frequently identifies normal tissue as abnormal, but rarely

misses actual lesions, which is important for sensitive detection. In the subsequent versions of the detection algorithm, the performance of the model was significantly improved and these improvement efforts are ongoing. By removing low-quality input images from the dataset, further improvements in performance metrics were and will be achieved. These results show that data quality directly influences the performance of the model. The improved performance metrics can be achieved by pre-selecting and automatically removing incorrect/misleading images before they are presented for evaluation. Thus, presenting cleaner, more representative input images to the LD system will help to avoid penalizing the model by feeding the system with difficult or misleading images. This will mean that not all lungs will be evaluated during the slaughtering process, but also human PHMS evaluators typically only manage to evaluate approx. 30–35% of all presented lungs, due to processing chain speed (Banhazi, 2023. Pers. com.). Obviously at a higher processing chain speed, the number of missed lung samples will be even higher. In Table 3 and Fig. 12 the key results of the study implemented in Trial 1 are presented.

Table 3. Key results obtained at the Trial 1 site in terms of precision, recall, IoU, TP, FP, and FN detections. Average, maximum and minimum values are also presented

	precision	recall	IoU	TP (avg)	FP (avg)	FN (avg)
avg	0.68	0.75	0.56	6,116	2,960	2,204
max	0.94	1.00	0.85	34,593	16,406	16,183
min	0.07	0.13	0.06	639	201	0

Segmentation accuracy was high across both pilots. In Trial 1, FP and FN rates were 2.50% and 1.84% respectively (Table 1). In Trial 2, FP and FN rates were 3.28% and 2.16% (Table 1). Contrarily to what was expected, the hanging presentation consistently yielded lower error rates probably due to improved orientation identification. Unfortunately, the tray presentations had higher rate of occlusions than expected due to incorrect positioning, but the tray presentations still produced results within acceptable error ranges (Table 1 and 4).

Table 4. Average differences in overall pixel identification by the manual and automatic system at Trial 1 and Trial 2 sites in terms of pixel differences expressed as number and% of lung area

	Pixel (trial1 $n = 102$)	% (Trial 1)	Pixel (trial2 $n = 228$)	% (Trial 2)
Automatic	8,203 (SD 6,006)	6.8	14,166 (SD 8,848)	6.9
Manual	8,320 (SD 6,601)	6.9	11,836 (SD 9,846)	5.7
Difference	117	0.1	2,330	1.2
1 PHMS score difference vs. average error	1,200 vs. 117	approx. 0.1 PHMS score equivalent	2,067 vs. 2,330	approx. 1 PHMS score equivalent

The correspondence between manual PHMS scores and LD-generated scores was assessed for both pilot sites (Table 4). The reported FP and FN rates, combined with pixel-to-score conversions, allow for a practical interpretation of system agreement. Pixel-to-score sensitivity analysis showed that the average difference between the automatically and manually identified lesion pixels corresponded to approx. 0.1 PHMS point in hanging lungs and approx. 1 point in tray presentations (Table 4). This indicates that small discrepancies had proportionally larger impacts in tray setups, highlighting

the importance of standardized lung presentation procedures (i.e. Standard Operational Procedures, SOPs).

In both pilots, the magnitude of deviation between manual and automated scoring is comparable to the inter-observer variability typically seen in slaughterhouse lesion scoring, which is often 1 to 3 PHMS points (Stärk, K.D.C. & Nevel, A. 2009; Holt, H.R. et al., 2011; Maes, D. et al., 2023). This suggests that, operationally, the LD system can replace human scoring without loss of diagnostic resolution in most cases. It is important to understand that for herd health monitoring, small differences in individual lung scores have minimal impact on herd-level prevalence or severity assessments (Voogt, A. et al., 2023; Valeris-Chacin, R. et al., 2025). The largest discrepancies are expected in cases with high lesion coverage and partial occlusion; conditions that can be mitigated by SOP improvements and selective image assessment.

Comparison of human scores and machine vision scores – hanging lung images

The manual scoring of 180 lungs showed that the average difference between left and right lungs segments was 1.72 PHMS points (Table 5), well within the typical inter-observer variability (± 2–3 points) (Hattab, J. et al., 2023). This suggests that partial images can reliably approximate whole-lung scoring. These results suggest that partial-lung images (as in occluded hanging presentations) can provide a reliable estimate of the full-lung score. However, the relationship should be validated further in controlled trials.

Table 5. Average differences in overall PHMS lung scoring (undertaken manually by the lead author) between the left and right sides of the lung samples evaluated in Trial 1

Variable	Value (*n* =180)
Average lung scores	8.26 (SD: 9.29)
Difference in lung scores (L vs. R sides) PHMS score	1.72 (SD: 2.11)
Difference in lung scores (L vs. R sides) percentage	21

Operationally, image quality and lung presentation variability were critical factors. It is assumed that SOPs for lung placement in trays will significantly improve accuracy,while quality control filters will reduce the impact of poor images. Overall, deviations between LD and manual scoring were within ±1 PHMS points (Table 4), aligning well with accepted human variability (Hattab, J. et al., 2023).

Hanging lungs were often presented at varying angles that caused segmentation errors. Unexpectedly in trays, portions of the lung were often hidden by other organs or intestines. Implementing a standardized lung placement procedure in trays will reduce variability and improve segmentation accuracy. While algorithmic improvements can and will reduce segmentation errors, it is evident that changes in slaughterhouse workflow arrangements are expected to significantly improve accuracy. SOP adoption appears to be the single most impactful operational change.

General comments and further developments

The system's practicality was demonstrated in fast-paced commercial environments and while not every image was usable, the accuracy of the system was sufficient for herd-level monitoring (Stärk, K.D.C. & Nevel, A., 2009; Holt, H.R. et al., 2011). LD needs to be viewed as a tool for scalable herd-level surveillance rather than individual diagnostic tool (Stärk, K.D.C. & Nevel, A., 2009). The differences between automated

and manual scores, even in the worst-case scenarios, were small relative to the total lesion coverage and unlikely to affect herd-level health assessments (Maes, D. et al., 2023). For herd monitoring purposes, the detection of relative changes in prevalence and severity over time is more important than exact point-for-point score matching. This means LD's real-world value lies in its consistency and throughput, enabling large sample sizes to be evaluated objectively without adding significant labour. As demonstrated above, the LD system achieved accuracy levels comparable to manual lesion scoring, with deviations well within the inter-observer variability of trained human assessors. This confirms LD as a practical replacement for manual scoring in commercial slaughterhouses. Its objectivity eliminates observer bias and provides consistent results across sites free from errors introduced by human operators (Bonicelli, L. et al., 2021a; De Luca, S. et al., 2021).

Beyond technical performance, the implications of LD are significant. By enabling early detection, LD supports welfare improvements, reduces mortality, improves feed efficiency, and lowers the incidence of carcass condemnations (Stärk, K.D.C. & Nevel, A., 2009; Bonicelli, L. et al., 2021b; Holt, H.R. et al., 2011; Maes, D. et al., 2023). It might also contribute to antimicrobial stewardship by reducing reliance on broad-spectrum antibiotics by allowing veterinarians to administer treatment early. By providing rapid, objective feedback on respiratory lesion prevalence, LD could facilitate earlier interventions at the herd level. This aligns with the goals of antimicrobial stewardship by allowing farmers and veterinarians to attend respiratory health challenges before they escalate to require broad-spectrum antibiotic treatment (Cole, D. et al., 2000; Chauvin, C. et al., 2002; Bai, H. et al., 2022). The potential welfare benefits are also significant, as earlier detection and targeted management may reduce morbidity and improve general wellbeing as well as productivity (Maes, D. et al., 2023).

Remaining challenges include handling occlusion in hanging lungs, improving segmentation outcomes, calibration across sites, automating quality control to flag unusable images for exclusion and expanding the overall datasets. In addition, the data collection hardware (LD cameras) has to be significantly ruggedised in the future. While an effort was made to design simple water and dust resistant cameras for these initial trials, the slaughterhouse environment proved to be more challenging than envisaged. During these initial experiments, one of the cameras had to be replaced due to water damage. Thus, further efforts will be made to improve the water and dust resistance design of the cameras (Bull, C.R. et al., 1996).

Future work should also explore extending LD to additional carcass-level indicators such as pleurisy, pericarditis and/or ileitis. In the context of integrated welfare assessment frameworks, LD could become part of a suite of tools providing continuous, objective monitoring of health/welfare status of various livestock species (Pessoa, J. et al., 2023; Tuyttens, F.A.M. et al., 2025).

CONCLUSIONS

The LD system provides a reliable, scalable, and objective alternative to lung lesion scoring by a human. Across two pilot sites with different lung presentation methods, the system achieved agreement levels with trained human assessors that fell within the reported normal range of inter-observer variability, while offering the benefits of high

throughput, consistency, and objectivity. With FP and FN rates of 2–3%, its accuracy is deemed sufficient for commercial herd health monitoring. By integrating LD into slaughterhouse workflows, producers can obtain timely and standardized feedback on respiratory health.

Tray presentation delivered slightly lower segmentation accuracy due to inconsistent positioning/presentation of lungs and thus increased occlusion, but with appropriate SOPs the research team is confident that these results can be improved further. It is expected that hanging presentation can also achieve more precise results in the future via selective processing and general improvement in segmentation procedures. From a herd health perspective, small deviations in individual lesion scores have negligible influence on herd-level assessments. More importantly, LD's ability to process large sample sizes quickly enables timely detection of changes in respiratory lesion prevalence or severity. This can facilitate earlier, targeted interventions, supporting antimicrobial stewardship by reducing the need for broad-spectrum antibiotic treatments and improving animal welfare outcomes.

Beyond respiratory disease surveillance, the LD approach could be adapted to monitor other carcass-level welfare indicators, forming part of a comprehensive, integrated welfare monitoring framework. Continued developments, such as dataset expansion, quality control automation, and partial *vs.* whole lung evaluation efforts, are expected to further strengthen the system's applicability and robustness across diverse operational settings. With these refinements, LD has the potential to become a cornerstone technology for data-driven health and welfare management in modern pig production, enabling continuous improvement in productivity, sustainability, and animal care.

ACKNOWLEDGEMENTS. Funded by the European Union project 'Animal Welfare Indicators at the Slaughterhouse, aWISH' (Horizon Europe under grant agreement ID: 101060818). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. We would like to thank the members of the aWISH consortium, expert panel and stakeholder advisory board members for their valuable contributions.

REFERENCES

Badrinarayanan, V., Kendall, A. & Cipolla, R. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, Article 12. https://doi.org/10.1109/tpami.2016.2644615

Bai, H., He, L.-Y., Wu, D.-L., Gao, F.-Z., Zhang, M., Zou, H.-Y., Yao, M.-S. & Ying, G.-G. 2022. Spread of airborne antibiotic resistance from animal farms to the environment: Dispersal pattern and exposure risk. *Environment International* **158**, 106927. https://doi.org/10.1016/j.envint.2021.106927

Banhazi, T. & Cargill, C. 1998. An assessment of the benefits of age-segregated rearing and all-in/all-out management in respiratory disease free herds. In *Proceedings of the 15th IPVS Congress*, Vol. **2**, 387. (Eds C. Cargill and S. McOrist). Birmingham, England: University of Birmingham.

Banhazi, T. & Pisaniello, D. 2018. The effect of air quality in livestock buildings on the occupational health of farm workers. *Air Quality and Livestock Farming* **175**.

Banhazi, T.M. 2013. Environmental and management effects associated with improved production efficiency in a respiratory disease free pig herd in australia. In *Livestock housing: Modern management to ensure optimal health and welfare of farm animals*, 297–314. (Eds A. Aland and T. Banhazi). Wageningen, The Netherlands WAP.

Banhazi, T.M., Ji, B., Rutley, D. & Phillips, C.J.C. 2022. Modelling the effects of environmental stress on weight gain in pigs. In *Practical precision livestock farming: Hands-on experiences with plf technologies in commercial and r&d settings* Vol. **1**, 1261–1266. (Eds T.M. Banhazi, V. Halas and F. Maroto-Molina). Amsterdam, The Netherlands Wageningen Academic Publishers.

Bonicelli, L., Trachtman, A.R., Rosamilia, A., Liuzzo, G., Hattab, J., Mira Alcaraz, E., Del Negro, E., Vincenzi, S., Capobianco Dondona, A., Calderara, S. & Marruchella, G. 2021a. Training convolutional neural networks to score pneumonia in slaughtered pigs. *Animals* **11**(11), 3290.

Bonicelli, L., Trachtman, A.R., Rosamilia, A., Liuzzo, G., Hattab, J., Mira Alcaraz, E., Del Negro, E., Vincenzi, S., Capobianco Dondona, A., Calderara, S. & Marruchella, G. 2021b. Training convolutional neural networks to score pneumonia in slaughtered pigs. *Animals* **11**(3290), 1–13. https://doi.org/10.3390/ani11113290

Brosnan, T. & Sun, D.-W. 2002. Inspection and grading of agricultural and food products by computer vision systems–a review. *Computers and Electronics in Agriculture* **36**(2–3), 193–213.

Bull, C.R., McFarlane, N.J.B., Zwiggelaar, R., Allen, C.J. & Mottram, T.T. 1996. Inspection of teats by colour image analysis for automatic milking systems. *Computers and Electronics in Agriculture* **15**(1), 15–26.

Chauvin, C., Beloeil, P.-A., Orand, J.-P., Sanders, P. & Madec, F. 2002. A survey of group-level antibiotic prescriptions in pig production in france. *Preventive Veterinary Medicine* **55**(2), 109–120.

Cole, D., Todd, L. & Wing, S. 2000. Concentrated swine feeding operations and public health: A review of occupational and community health effects. In *Environmental Health Perspectives*, Vol. **108**, 685.

Cronin, G.M., Borg, S.S. & Dunn, M.T. 2008. Using video image analysis to count hens in cages and reduce egg breakage on collection belts. *Australian Journal of Experimental Agriculture* **48**, 768–772.

Crook, B., Robertson, J.F., Glass, S.A., Botheroyd, E.M., Lacey, J. & Topping, M.D. 1991. Airborne dust, ammonia, microorganisms, and antigens in pig confinement houses and the respiratory health of exposed farm workers. *American IndusTrial Hygiene Association Journal* **52**(7), 271–279.

Curtis, S.E., Anderson, C.R., Simon, J., Jensen, A.H., Day, D.L. & Kelley, K.W. 1975. Effects of aerial ammonia, hydrogen sulfide and swine-house dust on rate of gain and respiratory-tract structure in swine. *Journal of animal science* **41**(3), 735–739.

De Luca, S., Zanardi, E., Alborali, G.L., Ianieri, A. & Ghidini, S. 2021. Abattoir-based measures to assess swine welfare: Analysis of the methods adopted in european slaughterhouses. *Animals* **11**(1), 226.

Deng, Y., Qu, H., Leng, A., Tang, X. & Zhai, S. 2025. Methods and challenges in computer vision-based livestock anomaly detection, a systematic review. *Biosystems Engineering* **253**, 104135. https://doi.org/10.1016/j.biosystemseng.2025.104135

Dervić, E., Matzhold, C., Egger-Danner, C., Steininger, F. & Klimek, P. 2024. Improving lameness detection in cows: A machine learning algorithm application. *Journal of dairy science* **107**(12), 11550–11562. 10. 3168/jds.2024-24730

Done, S.H., Chennells, D.J., Gresham, A.C.J., Williamson, S., Hunt, B., Taylor, L.J., Bland, V., Jones, P., Armstrong, D., White, R.P., Demmers, T.G.M., Teer, N. & Wathes, C.M.2005. Clinical and pathological responses of weaned pigs to atmospheric ammonia and dust. *Veterinary Record* **3**(157), 71–80.

Donham, K.J. 1989. Relationships of air quality and productivity in intensive swine housing. *Agriculture Practice* **10**, 15.

Donham, K.J., Reynolds, S.J., Whitten, P., Merchant, J.A., Burmeister, L. & Popendorf, W.J. 1995. Respiratory dysfunction in swine production facility workers: Dose-response relationships of environmental exposures and pulmonary function. *American Journal of Industrial Medicine* **27**(3), 405–418.

Goodwin, R.F., Hodgson, R.G., Whittlestone, P. & Woodhams, R.L. 1969. Immunity in experimentally induced enzootic pneumonia in pigs. *J. Hyg. (Lond)* **67**(2), 193–208. 10.1017/s0022172400041607

Hattab, J., Porrello, A., Romano, A., Rosamilia, A., Ghidini, S., Bernabò, N., Capobianco Dondona, A., Corradi, A. & Marruchella, G. 2023. Scoring enzootic pneumonia-like lesions in slaughtered pigs: Traditional vs. Artificial-intelligence-based methods. *Pathogens* **12**(12), 1460.

Heinonen, M., Grohn, Y.T., Saloniemi, H., Eskola, E. & Tuovinen, V.K. 2001. The effects of health classification and housing and management of feeder pigs on performance and meat inspection findings of all-in-all-out swine-finishing herds. *Preventive Veterinary Medicine* **49**(1–2), 41–54.

Holt, H.R., Alarcon, P., Velasova, M., Pfeiffer, D.U. & Wieland, B. 2011. Bpex pig health scheme: A useful monitoring system for respiratory disease control in pig farms? *BMC Veterinary Research* **7**(1). 82.10.1186/1746-6148-7-82

Iseki, H., Furukawa, E., Shimasaki, T. & Higaki, S. 2025. Automatic monitoring of activity intensity in a chicken flock using a computer vision-based background image subtraction technique: An experimental infection study with fowl adenovirus. *Smart Agricultural Technology* **10**, 100821. https://doi.org/10.1016/j.atech.2025.100821

Lawrence, K.C., Windham, W.R., Park, B. & Buhr, R.J. 2001. Hyperspectral imaging system for identification of fecal and ingesta contamination on poultry carcasses. (Ed 013076).

Maes, D., Sibila, M., Pieters, M., Haesebrouck, F., Segalés, J. & de Oliveira, L.G. 2023. Review on the methodology to assess respiratory tract lesions in pigs and their production impact. *Veterinary Research* **54**(1). 8.10.1186/s13567-023-01136-2

Mehdizadeh, S.A., Neves, D., Tscharke, M., Nääs, I. & Banhazi, T.M. 2015. Image analysis method to evaluate beak and head motion of broiler chickens during feeding. *Computers and Electronics in Agriculture* **114**, 88–95.

Mohapatra, S., Swarnkar, T. & Das, J. 2021. 2 - deep convolutional neural network in medical image processing. In *Handbook of deep learning in biomedical engineering*, 25–60 (Eds V.E. Balas, B.K. Mishra and R. Kumar). Academic Press. https://doi.org/10.1016/B978-0-12-823014-5.00006-5

Murphy, T., Cargill, C., Rutley, D. & Stott, P. 2012. Pig-shed air polluted by α-haemolytic cocci and ammonia causes subclinical disease and production losses. *Veterinary Record*: vetrec-2011-100413

Myers-Colet, C., Schroeter, J., Arnold, D. & Arbel, T. 2022. Heatmap regression for lesion detection using pointwise annotations. In *Milland@miccai* 3-12.10.1007/978-3-031-16760-7_1

Niemi, J.K., Sevón-Aimonen, M.-L., Pietola, K. & Stalder, K.J. 2010. The value of precision feeding technologies for grow-finish swine *Livestock Science* **129**, 13–23.

Park, B., Yoon, S.C., Lawrence, K.C. & Windham, W.R. 2007. Fisher linear discriminant analysis for improving fecal detection accuracy with hyperspectral images *Transactions of the ASABE* **50**(6), 2275–2283.

Pessoa, J., McAloon, C., Boyle, L., García Manzanilla, E., Norton, T. & Rodrigues da Costa, M. 2023. Value of simplified lung lesions scoring systems to inform future codes for routine meat inspection in pigs. *Porcine Health Management* **9**(1). 31.10.1186/s40813-023-00324-y

Pointon, A.M. & Banhazi, T. 1995. Evaluation of pig health monitoring scheme as an industry service. In *Proceedings of the 8-th Symposium of the International Society for Animal Hygiene*, Vol. **1**, 750–755. (Ed F. Hollo). Budapest, Hungary: Hungarian Veterinary Journal

Pointon, A.M., Davis, P.R. & Bahnson, P.B. 1999. Disease surveillance at slaughter. In *Diseases of swine*, Vol. **8**, 1111–1132. (Eds B. Straw, S. D'Allaire, W.L. Mengeling and D.J. Taylor). Ames, Iowa, USA: Iowa State University Press.

Pointon, A.M., Heap, P. & McCloud, P. 1985. Enzootic pneumonia of pigs in south australia - factors relating to incidence of disease. *Australian Veterinary Journal* **62**(3), 98–101.

Pointon, A.M., Moore, M.J. & Davies, P.R. 1996. Estimation of the sensitivity and specificity of gross lesions monitored at slaughter. In *Proceedings of the 14th IPVS Congress*, **718** Bologna, Italy.

Rodriguez Carino, C., Infante, R., Sogbe, E., Rodriguez, G., Rodriguez Carino, H., Diaz, C. & Yufa, B. 1998).Incidence of lesions in pigs at a slaughterhouse in venezuela. In *15th International Pig Veterinary Society Congress*, Vol. Vol. III: Poster Presentations, 22 Birmingham, England.

Shams, M.Y., Elmessery, W.M., Oraiath, A.A.T., Elbeltagi, A., Salem, A., Kumar, P., El-Messery, T.M., El-Hafeez, T.A., Abdelshafie, M.F., Abd El-Wahhab, G.G., El-Soaly, I.S. & Elwakeel, A.E. 2025. Automated on-site broiler live weight estimation through yolo-based segmentation. *Smart Agricultural Technology* **10**, 100828. https://doi.org/10.1016/j.atech.2025.100828

Shwetha, V., Maddodi, B.S., Vijaya, L., Abhinav, K. & Sakshi, S. 2024. Latest trend and challenges in machine learning– and deep learning–based computational techniques in poultry health and disease management: A review. *Journal of Computer Networks and Communications* 2024.10.1155/2024/8674250

Stärk, K.D.C. 2000. Epidemiological investigation of the influence of environmental risk factors on respiratory diseases in swine--a literature review. *The Veterinary Journal* **159**(1), 37–56.

Stärk, K.D.C. & Nevel, A. 2009. Strengths, weaknesses, opportunities and threats of the pig health monitoring systems used in england. *Veterinary Record* **165**(16), 461–465. https://doi.org/10.1136/vr.165.16.461

Traore, B.B., Kamsu-Foguem, B. & Tangara, F. 2018. Deep convolution neural network for image recognition. *Ecological Informatics* **48**, 257–268. https://doi.org/10.1016/j.ecoinf.2018.10.002

Tuyttens, F.A.M., Lawrence, A.B. & Mullan, S. 2025. A framework for a comprehensive animal welfare label: Scientific, logistic, and ethical challenges. *Animal Frontiers* **15**(2), 61–68. 10.1093/af/vfaf003

Valeris-Chacin, R., Garcia-Morante, B., Sibila, M., Canturri, A., Ballarà Rodriguez, I., Bernal Orozco, I., Jordà Casadevall, R., Muñoz, P. & Pieters, M. 2025. Scoring of swine lung images: A comparison between a computer vision system and human evaluators. *Veterinary Research* **56**(1). 9.10.1186/s13567-024-01432-5

Voogt, A., Schrijver, R., Temürhan, M., Bongers, J. & Sijm, D. 2023. Opportunities for regulatory authorities to assess animal-based measures at the slaughterhouse using sensor technology and artificial intelligence: A review. *Animals* **13**, 3028. 10.3390/ani13193028

Wathes, C.M., Demmers, T.G.M., Teer, N., White, R.P., Taylor, L.L., Bland, V., Jones, P., Armstrong, D., Gresham, A.C.J., Hartung, J., Chennells, D.J. & Done, S.H. 2004. Production responses of weaned pigs after chronic exposure to airborne dust and ammonia. *Animal Science* **78**(1), 87–97.

Zhang, F., Cai, N., Wu, J., Cen, G., Wang, H. & Chen, X. 2018. Image denoising method based on a deep convolution neural network. *IET Image Processing* **12**(4), 485–493. https://doi.org/10.1049/iet-ipr.2017.0389