Food Control 179 (2026) 111581



Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont





Post-mortem detection of unhealthy livers and hearts in chickens using deep learning, logistic regression and Computed Tomography (CT) scanning

Kacper Libera [a], Effrosyni Kritsi [a], Dirk Schut [b], Louis van Steijn [c], Lourens Heres [a], Len Lipman [a,*]

[a] Institute for Risk Assessment Sciences (IRAS), Utrecht University, Yalelaan 2, 3584 CM, Utrecht, the Netherlands
[b] Computational Imaging Group, Centrum Wiskunde & Informatica (CWI), Science Park 123, 1098 XG, Amsterdam, the Netherlands
[c] Meyn Food Processing Technology B.V., Westeinde 6, 1511MA, Oostzaan, the Netherlands

ARTICLE INFO

Keywords:
Image classification
PM inspection
Food safety
Edible offal
X-rays

ABSTRACT

Diseased chicken organs, including livers and hearts, are frequently observed during post-mortem inspection in poultry slaughterhouses. It is crucial to accurately identify and discard these organs, since they are unfit for human consumption. The current method of inspection is based on visual human examination and this procedure has limitations that negatively affect its reliability e.g. subjectivism and working under time pressure. It implies that new technologies should be investigated.

Computed Tomography (CT) scanning can detect pathologically changed organ tissues, because the radiodensity of the organs is altered. Therefore, the aims of this study were to compare the radiodensity between healthy and diseased hearts and livers and then develop different classifiers to identify diseased organs.

264 chicken hearts and 252 livers were collected from a slaughterhouse including healthy and diseased samples. All the organs were CT scanned in a veterinary clinic. Two logistic regression models (Log_Reg_Hearts and Log_Reg_Livers) and four deep learning models were developed including deep and shallow neural networks (NN_Deep_Hearts/Livers and NN_Shallow_Hearts/Livers) to classify these organs.

Deep learning models for hearts (accuracy 0.91 for NN_Shallow_Hearts; 0.92 for NN_Deep_Hearts) outperformed logistic regression model (0.82, Log_Reg_Hearts). There was no difference in accuracy for the liver models (0.78 for NN_Shallow_Livers; 0.75 for NN_Deep_Livers; 0.79 for Log_Reg_Livers).

Our study confirms that diseased chicken hearts and livers can be automatically and accurately detected using CT scans classified by logistic regression/deep learning models. Overall, CT scanning has potential to increase the safety of poultry edible organs and streamline the workflow in the slaughterhouse.

1. Introduction

According to European Union's (EU) food safety regulations every carcass needs to be individually checked by an official veterinarian and/or a meat inspector during post-mortem (PM) inspection (Regulation of European Commission (EC) no 2017/625, 2017; Commission Implementing Regulation EU 2019/627, 2019). The course of post-mortem inspection differs per farm animal species, but mainly consist of thorough visual examination, palpation and in some cases incision of certain groups of lymph nodes, organs and muscles. It is important to highlight that in PM inspection attention is paid not only to the carcass, but there are also many checks that need to be done to the offals, which are defined as fresh meat other than that of the carcass, including viscera (internal organs) and blood (Regulation of European Commission (EC) no 853/2004, 2004). The justification for that is that in many countries offals are used for human consumption, so the safety of offals should also be assured. For example, in 2023 EU slaughterhouses exported almost 100,000 tons of chicken offals to third countries (AVEC, 2024). Heart, liver and gizzard (stomach) form the chicken edible offals, culinarily termed as giblets. Giblets are consumed daily in many countries

* Corresponding author.
E-mail addresses: k.b.libera@uu.nl (K. Libera), e.kritsi@uu.nl (E. Kritsi), dirk.schut@cwi.nl (D. Schut), lvansteijn@meyn.com (L. van Steijn), l.heres1@uu.nl (L. Heres), l.j.a.lipman@uu.nl (L. Lipman).

https://doi.org/10.1016/j.foodcont.2025.111581
Received 13 May 2025; Received in revised form 11 July 2025; Accepted 17 July 2025
Available online 17 July 2025
0956-7135/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

worldwide and have been described as a good and affordable source of essential nutrients (Seong et al., 2015). Moreover, the organs are a reliable indicator of the general health status of the slaughtered animal. Many diseases can only be diagnosed post-mortem based on the examination of the organs, because the clinical signs are not present or they are nonspecific. For example in the course of infection with fowl adenovirus in chickens (inclusion body hepatitis and hepatitis-hydropericardium syndrome) the gross hepatic lesions may include: yellowish or pale discoloration, pinpoint hemorrhage and swelling, while the only clinical finding might be the sudden increase in mortality (El-Shall et al., 2022).

The poultry slaughterhouses stand out from other types of slaughter plants due to automatization and robotization at most of the processing steps including removal of internal organs (evisceration). In modern slaughterhouses the organs are automatically removed by a highly specialized machine (eviscerator), which brings out the package of organs from the carcass and places it on a special bowl/rack for the PM examination (Lichtner et al., 2024). Then meat inspectors need to visually examine the packet of organs and determine if there are no lesions/contaminations making the carcass and/or organs unfit for human consumption (Regulation of European Commission (EC) no 2017/625, 2017).

Among the viscera inspected, special attention is usually paid to hearts and livers, because these organs are part of chicken edible offal, so they need to meet food safety standards (Regulation of European Commission (EC) no 853/2004, 2004). Furthermore, common infectious health problems in broiler chickens are affecting the heart and liver. Consequently, inflammation of the heart, more specifically the pericardium (pericarditis), and inflammation of the liver (hepatitis) are both significant reasons of chicken carcass rejections worldwide (Salines et al., 2017). For example, in the study by Buzdugan et al. (Buzdugan et al., 2020) the rejection rates per 1000 were 1.5 for diseased livers and 0.3 for diseased hearts (including total and partial rejections). According to another research hepatitis was the 3rd most significant reason for rejection affecting 0.17 % of chickens slaughtered (Forseth et al., 2023).

During evisceration there is a risk that organs can be mechanically damaged and, if they contain pathogens, the bacteria can spread on other organs and/or carcasses creating extra food safety/quality hazards. This contamination can be transmitted through the direct spill of the intestinal content or via airborne transmission (Lichtner et al., 2024). Therefore it would be highly advantageous if the organs can be examined before evisceration. Moreover, there is practically no need to eviscerate the carcasses that will be discarded anyways. To tackle this challenge, a non-destructive evaluation technique would be required. The technique should be able to penetrate through the tissues in order to measure internal structures, which makes all technologies based on visible or near infrared light not applicable. One potentially suitable technique is Computed Tomography (CT) scanning, which can create a 3D image of any object of interest based on X-rays. X-rays penetrate different tissues to various extents (e.g. bone vs. blood) and these differences can be measured and seen in CT scans as gray-scale (Hermena & Young, 2025). To objectively measure the opacity of any object to X-rays, the Hounsfield scale (Hounsfield units, HU) has been introduced, which is a quantitative scale for describing radiodensity. As a reference, pure water has a HU of 0 and air is −1000 HU. For example in horses adipose tissue can vary from −30 to −80 HU, soft tissues +30 to +220, while bone can be close to +2000 to +3000 (Kimberlin, 2016).

Over the course of a disease the structure of an organ can change due to e.g. injury, accumulation of inflammation tissue, lesions, neoplasia etc., which would consequently influence its radiodensity (Hoey et al., 2016). Thus, there is a possibility to distinguish between healthy and diseased organs based on their radiodensity. This characteristic is frequently used in human medical imaging where for every organ there is described range of Hounsfield units for healthy and diseased tissues (DenOtter & Schubert, 2025). There are also some reference values described in veterinary medical imaging for dogs and horses (Thrall, 2017).

Given the differences in radiodensity between healthy and unhealthy organs, a binary classifier may be developed to detect unhealthy organs. Classification methods can be divided into traditional or neural network-based methods. Traditional methods rely on features that are derived from the data, and each feature is designed by hand. These features are combined using a relatively simple decision model (linear, quadratic, nearest-neighbor …). The benefits of this approach are that is relatively fast to train, explainable and robust to overfitting. The main downsides are that designing features requires domain knowledge, and that designing features and decision models to efficiently combine data is challenging (e.g. from multiple pixels, timepoints, or sensors). Neural networks are decision models with a very large amount of parameters, making them capable of modeling complex relations (Hornik et al., 1989). Instead of relying on handcrafted features, neural networks are typically applied directly to raw input data. Neural networks have shown remarkable accuracy in modeling all kinds of complex relationships with relatively little guidance. However, they have the downside of requiring large datasets and a lot of computational resources for training. Moreover, they are prone to overfitting and the inner reasoning which led to the decision of the model may be unclear. For image classification, Convolutional Neural Networks (CNNs) are a popular type of neural network and CNNs have shown very promising results in classifying organs as healthy or diseased (Shatnawi et al., 2025).

If an accurate classifier can be developed, it could be incorporated into the setup in the slaughterhouse to assist, and potentially even replace, meat inspectors performing PM inspection at the slaughter line. That might lead to more accurate, quicker and more robust PM inspection at a lower cost due to reduced labor costs and the possibility to increase the speed of the slaughter line.

However, to the best of our knowledge CT scanning has not been investigated for post-mortem inspection of chickens. In this paper we show it has the potential to increase the safety of poultry products and streamline the workflow in the slaughterhouse by identifying pathologically changed organs. Therefore, the aims of this study were:

- to compare the mean radiodensity of healthy and diseased chicken hearts and livers and develop logistic regression model based on this feature
- to develop AI classification models (neural networks) based on CT scans to distinguish between healthy and diseased organs
- to compare the performance of logistic regression, a shallow neural network, and a deep neural network on this task

2. Materials and methods

2.1. Sample collection and study design

In total 264 chicken hearts and 252 livers (chicken broiler hybrid: Ross 308) have been collected from a slaughterhouse located in the Netherlands, which uses a CO_2 chamber for stunning and an automatic system for bleeding. The samples were collected during 6 sampling days from February 2024 to March 2025 b y a slaughterhouse employee just after evisceration. The samples were transported from the slaughterhouse to the laboratory in a cooling box (<3 °C). All organs (264 hearts and 252 livers = 516) were CT scanned. The brief explanation of study design is given in Fig. 1.

2.2. Organs processing and categorization (ground truth)

All the organs have been categorized into one of two categories: healthy or diseased. For clarity, 4 unique labels were developed:

1. HEAL_H – healthy hearts
2. DIS_H – diseased hearts
3. HEAL_L – healthy livers

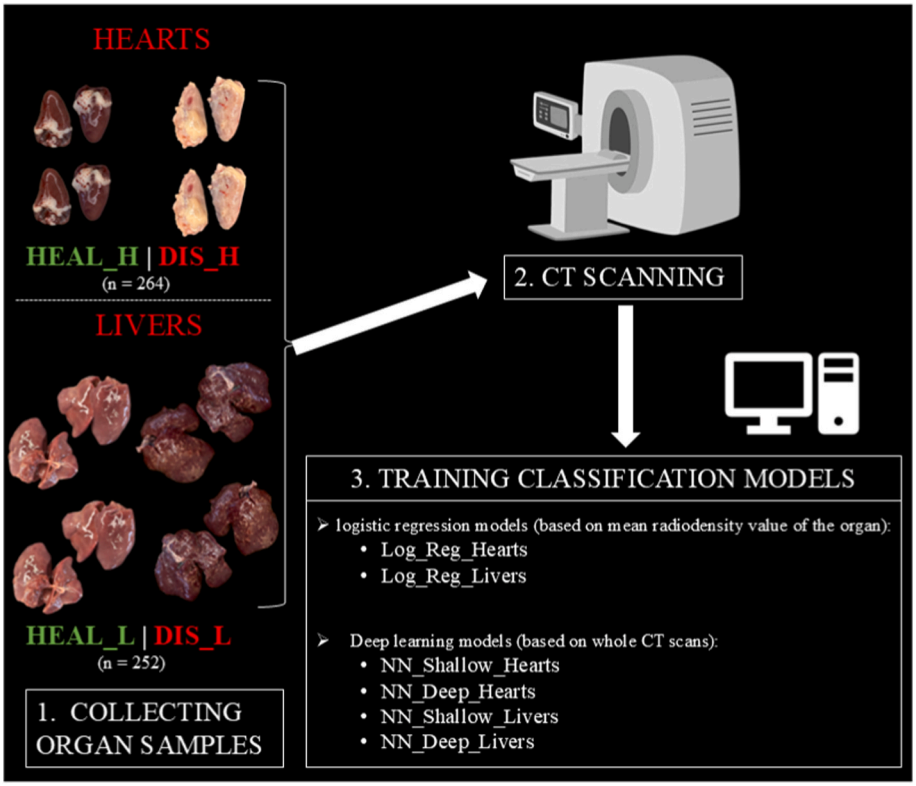


Fig. 1. Graphical explanation of the study design. HEAL_H, healthy hearts; DIS_H, diseased hearts; HEAL_L, healthy livers; DIS_L, diseased livers; CT scanning, Computed Tomography scanning.

4. DIS_L – diseased livers

The slaughterhouse's employee (guided by the meat inspector) was asked to gather equal numbers of HEAL_H/DIS_H and HEAL_L/DIS_L on every sampling day and put it in different plastic bags. Nevertheless, every sample was re-examined in the lab then classified to one of the mentioned categories. Briefly, all samples were categorized as healthy or diseased for each organ type by two veterinarians with food safety background. In case of disagreement, the final decision was made after following a second evaluation of the specific organ. If after the re-evaluation the two veterinarians still had different opinions, the sample was assigned to the diseased class (DIS_L/DIS_H). This was agreed beforehand in order to minimize the risk of underestimating the diseased categories and is in line with EU regulation stating that in case of any doubts official veterinarian can declare meat (and offal) as unfit for human consumption (Commission Implementing Regulation EU 2019/627, 2019). Ideally, some questionable samples could have been subjected to histological examination to confirm\exclude the presence of pathology. However, it was beyond the scope of the current research.

Table 1
The classification criteria for healthy and diseased chicken hearts and livers. For both types of organs two categories were applied: HEAL_H and DIS_H for hearts and HEAL_L and DIS_L for livers.

Hearts	HEAL_H • transparent and thin pericardium; no visible fibrous deposits and other lesions	DIS_H • visible inflammation of heart membrane pericardium including visible fibrous deposits and thickened pericardium
Livers	HEAL_L • dark brown, uniform in color, no lesions	DIS_L • swollen liver containing multiple pale or hemorrhagic foci, spots on surface, visible fibrous deposits

HEAL_H, healthy hearts; DIS_H, diseased hearts; HEAL_L, healthy livers; DIS_L, diseased livers.

The classification criteria used by two veterinarians can be found in Table 1.

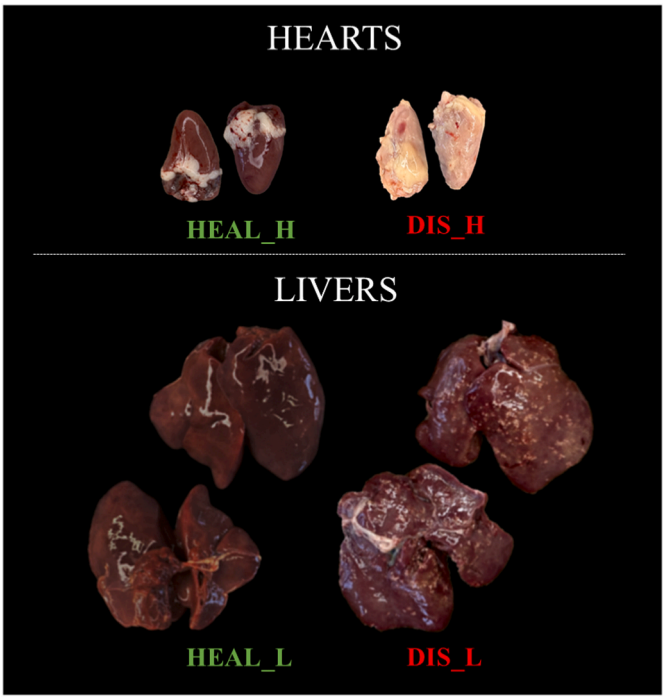


Fig. 2. Photographs of typical samples from the 4 different groups used in the study: healthy hearts (HEAL_H), diseased hearts (DIS_H), healthy livers (HEAL_L) and diseased livers (DIS_L). Note: in DIS_H the heart is covered by yellowish tissue, which is inflamed and thickened pericardium; in DIS_L the liver is covered with yellowish spots and its color is not uniform.

Additionally, photographs of some selected samples from each group are given in Fig. 2.

2.3. CT scanning and CT scan processing

All samples were scanned using a SOMATOM Definition AS CT scanner (Siemens Healthineers, the Hague, the Netherlands) at the veterinary hospital of Utrecht University, the Netherlands. The scanner's software presets were utilized to suggest the optimal combination of voltage and current based on sample size, thickness, and density. The parameters were set to 120 kV, 80–150 mA, and a slice thickness of 2.0 mm. The samples were kept overnight in the lab fridge (<3 °C) and scanned on the day following the day of slaughter. For time-efficiency reasons the organs were not scanned individually but were scanned in randomly allocated batches of 6–9 livers and 13–15 hearts.

The CT scans were saved as DICOM files. Since every DICOM file represented a batch of either 6–9 livers or 13–15 hearts, the individual organs needed to be extracted and cropped. All organs (n = 516) have been manually cropped using 3D Slicer (version 5.6.2) and then the background was removed (set to −1000 HU). The volume of every organ was masked using a threshold (from 0 to 300 HU) and a smoothing function (opening, remove extrusions) was applied. Then the value of mean radiodensity of the whole organ was gathered using Python code (calc_MGV.py in GitHub repository provided at the end of the article).

2.4. CT scans of liver and hearts – input data for models' development

The CT datasets consisted of 264 and 252 CT scans for hearts and livers, respectively. Table 2 provides an overview including the class balance between healthy and diseased organs.

2.5. Statistical analysis of the organs' mean radiodensity

The statistical analysis of the organs' mean radiodensity has been conducted separately for livers and hearts. The normality test (Kolmogorov-Smirnov's test with Lilliefors correction) and the test of equality of variances (Levene's test) have been performed. The differences between groups have been investigated using Welch's *t*-test (unequal variances *t*-test for independent samples) for hearts and livers. All the statistical calculations were performed in SPSS (version: 29.0.2.0, IBM) and the level of significance was set to p-value = 0.05.

2.6. Overview of the classifiers (logistic regression models and deep learning models)

Three different classifiers were used for each organ: a logistic regression classifier, a shallow CNN and a deep CNN (3D Resnet18). The logistic regression classifier used the mean radiodensity as its only input feature, and the neural networks used the cropped organ image resized to 128x128x128 as input. An overview of the differences between the classifiers is provided in Table 3. The depth is defined as the number of times a non-linear activation function (i.e. ReLu, sigmoid or softmax) was applied to a linear combination of the data (i.e. linear or convolution layer).

Table 2
Overview of the number of samples included during model development.

	Condition	Number of samples	Class balance
Livers	HEAL_L	121	HEAL_L: DIS_L
	DIS_L	131	0.48 : 0.52
	Total	252	
Hearts	HEAL_H	142	HEAL_H: DIS_H
	DIS_H	122	0.54 : 0.46
	Total	264	

HEAL_L, healthy livers, DIS_L; diseased livers; HEAL_H, healthy hearts; DIS_H; diseased hearts.

Table 3
Comparison of the different classifiers studied in this paper.

Classifier	Depth	# of trainable parameters	Input
Logistic regression (Log_Reg_Livers & Log_Reg_Hearts)	1	2	mean radiodensity
Shallow CNN (NN_Shallow_Livers & NN_Shallow_Hearts)	5	311,554	128x128x128 volume
3D Resnet18 (NN_Deep_Livers & NN_Deep_Hearts)	18	33,161,026	128x128x128 volume

CNN, convolutional neural network; NN_Shallow_Livers, model to detect unhealthy livers based on CT scans using shallow neural network; NN_Deep_Livers, model to detect unhealthy livers based on CT scans using deep neural network; Log_Reg_Livers; logistic regression model to detect unhealthy livers based on radiodensity; NN_Shallow_Hearts, model to detect unhealthy hearts based on CT scans using shallow neural network; NN_Deep_Hearts, model to detect unhealthy hearts based on CT scans using deep neural network; Log_Reg_Hearts, logistic regression model to detect unhealthy hearts based on radiodensity.

In this study two logistic regression models were created:

1. Model to detect unhealthy livers based on their mean radiodensity (Log_Reg_Livers)
2. Model to detect unhealthy hearts based on their mean radiodensity (Log_Reg_Hearts)

Furthermore, four neural network-based models were developed:

1. Model to detect unhealthy livers based on the CT scans (NN_Shallow_Livers) using a relatively simple (shallow) neural network
2. Model to detect unhealthy livers based on the CT scans (NN_Deep_Livers) using a more complex (deep) neural network
3. Model to detect unhealthy hearts based on the CT scans (NN_Shallow_Hearts) using a relatively simple (shallow) neural network
4. Model to detect unhealthy hearts based on the CT scans (NN_Deep_Hearts) using a more complex (deep) neural network.

2.7. Development of classifiers for automatic detection

For the logistic regression classifier, the implementation from the scikit-learn library was used with the default parameters (an inverse regularization strength of 1). The shallow CNN was implemented in PyTorch (Paszke et al., 2017), and it was based on the architecture described in (Zunair et al., 2020), but with the number of features halved in all layers, and with the first block of convolution, max pooling and batch normalization removed. Therefore, there were a total of 13 layers. The feature extraction block consisted of three convolutional layers, three maxpooling layers, and three batch normalization layers. The classification block consisted of a flatten layer, two fully connected layers, and a dropout layer. The details regarding this architecture can be found in a Github repository linked to the article (script train_nn.py). The 3D Resnet18 was also implemented in PyTorch, and the architecture was converted from its original 2D implementation (He et al., 2016; Wightman et al., 2021) to 3D using the timm_3D library (Solovyev et al., 2022). Briefly, there was one initial convolutional layer, then 4 residual blocks (each of them consisted of 4 convolutional layers), and then a final fully connected layer at the end (18 layers in total). For details, please see the original paper (He et al., 2016).

2.7.1. Cross-validation

Cross validation was applied to make efficient use of the available data, and to show the variations in classification results caused by the random nature of both the data and the neural network training. 5 cross-validation folds were used and each fold was split into 10 % test data, 30

% validation data, and 60 % training data. Consequently, in every fold there were around 151 liver samples and 158 heart samples used for training and 76 liver samples and 79 heart samples used for validation. The test data was unique in each fold (around 25 liver samples and 26 heart samples) and stratified sampling was used over both the label (diseased or healthy) and the scanning day (6 different scanning days) to ensure that each test set was an accurate representation of the dataset. In each fold, the remaining 90 % of the data was assigned to either training or validation data using stratified sampling.

2.7.2. Regularization and hyperparameters

Several regularization approaches were applied to avoid overfitting in the neural network-based classifiers: data augmentation, weight decay (Loshchilov & Hutter, 2018), dropout (Srivastava et al., 2014), droppath/stochastic depth (Huang et al., 2016) (only on the 3D ResNet18), and early stopping based on the validation set. The data augmentation was implemented using the TorchIO library (Pérez-García et al., 2021) and had the following hyperparameters: the probability of applying an affine transform, and the maximum translation distance, scaling factor, and degrees of rotation along each axis (separately) of such a transform; the probability of applying a Gaussian blur kernel, and its maximum kernel width; the probability of adding Gaussian noise and its maximum standard deviation; the probability of flipping the volume along each axis (separately). Moreover, there were hyperparameters for the learning rate, weight decay and batch size of the optimizer, and for the probability of dropout and droppath.

2.7.3. Training and hyperparameter optimization

The PyTorch Lightning library (Falcon, 2019) was used to train the neural networks. The optimizer was the AdamW optimizer (Loshchilov & Hutter, 2018) and the cost function was the binary cross entropy (BCE). The accuracy and the BCE on the validation set were calculated at the end of every epoch. An epoch was considered better than a previous one if the accuracy improved, or if the accuracy remained the same and the BCE improved. The training was stopped when the best epoch had not changed for 200 epochs, and the network weights of the best epoch were used as the final result.

The hyperparameters were optimized using 50 iterations of the Tree-structured Parzen Estimator (Bergstra et al., 2011) implemented in the Optuna library (Akiba et al., 2019). In each iteration a new neural network was trained, and the cost function was the same combination of accuracy and BCE over the validation set that was used for early stopping. A separate hyperparameter optimization was performed for each fold of the cross validation.

2.8. Performance evaluation

All the models (logistic regression and deep learning models) have been cross-validated with 5 different splits (5-fold cross validation) and the final performance presented in the results sections is the average value on the test dataset across all the splits.

Several evaluation metrics have been used to assess both models performance on the test data set including:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

True positive, (TP); a sample, which was correctly predicted as DIS_L or DIS_H.

True negative, (TN); a sample, which was correctly predicted as HEAL_L or HEAL_H.

False positive, (FP); a sample, which was incorrectly predicted as DIS_L or DIS_H, but in fact it was HEAL_L or HEAL_H.

False negative, (FN); a sample, which was incorrectly predicted as HEAL_L or HEAL_H, but in fact it was DIS_L or DIS_H.

Area under Curve (AUC) = area underneath the ROC curve (receiver operating characteristic curve plotting Recall vs. FPR).

The class DIS_L/DIS_H was considered as a positive class (1), while HEAL_L/HEAL_H as a negative (0).

3. Results

3.1. Statistical analysis

The results of statistical analysis are given in Table 4. In both types of organs the diseased hearts/livers have significantly lower radiodensity values comparing to healthy hearts/livers.

Additionally, Fig. 3 presents the mean radiodensity distribution in the form of histograms. The gray bars indicate the overlap between healthy and diseased groups.

3.2. Logistic regression models (Log_Reg_Livers and Log_Reg_Hearts) and deep learning models (NN_Shallow_Livers/Hearts and NN_Deep_Livers/Hearts) performance

The performance of all the models is given in Table 5. The deep learning models outperformed logistic regression models in hearts, but not in livers. The performance of shallow and deep neural networks was comparable in both the hearts and the livers.

4. Discussion

The aim of this study was to investigate the effects of diseases on the radiodensity of hearts and livers in poultry, and to develop classifiers to enable automatic detection of these diseases. Two logistic regression and four deep learning models were evaluated based on the mean radiodensity or full CT scans, respectively. The deep learning models (NN_Deep_Hearts and NN_Shallow_Hearts) outperformed the logistic regression model (Log_Reg_Hearts) in hearts, but not in livers.

Our top performing models for every organ type reached accuracy of 92 % (NN_Deep_Hearts) for hearts and 79 % (Log_Reg_Livers) for livers on the test dataset.

4.1. The pathologies of chicken organs

For many years, broiler chickens have been selected in breeding programs to stimulate the growth rate and to maximize the carcass yield, particularly the weight of the chicken breast (Hartcher & Lum, 2020). This led to a very efficient and short fattening period on modern poultry

Table 4
The differences in mean radiodensity (expressed in Hounsfield units) of healthy (HEAL_L) and diseased (DIS_L) livers and healthy (HEAL_H) and diseased (DIS_H) hearts. The values are expressed as mean ± standard deviation.

	HEAL_L, n = 121	DIS_L, n = 131	p-value
Livers, n = 252	161.47 ± 5.61	151.68 ± 7.47	<0.01
	HEAL_H, n = 142	DIS_H, n = 122	p-value
Hearts, n = 264	163.37 ± 6.70	149.99 ± 9.56	<0.01

HEAL_L, healthy livers; DIS_L, diseased livers; HEAL_H, healthy hearts; DIS_H, diseased hearts; p-values were determined by an unequal variances *t*-test for independent samples (Welch's *t*-test).

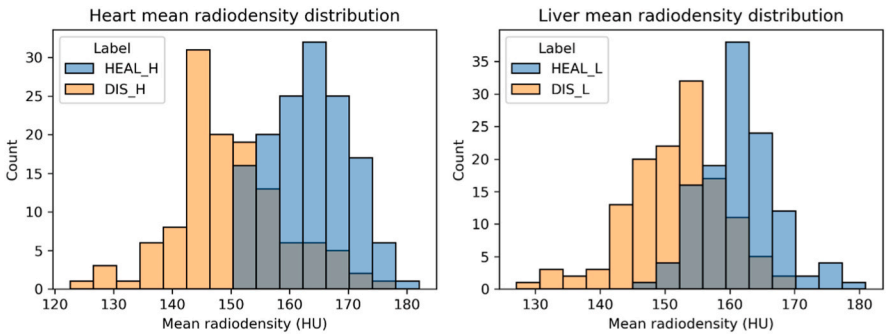


Fig. 3. The mean radiodensity distribution in healthy and diseased hearts (HEAL_H and DIS_H) and livers (HEAL_L and DIS_L). The gray bars indicate overlap between healthy and diseased groups.

Table 5
The average results on the test dataset after cross-validation of the logistic regression models (Log_Reg_Livers and Log_Reg_Hearts) and deep learning models (NN_Shallow_Livers/Hearts and NN_Deep_Livers/Hearts). The values are expressed as mean (± standard deviation).

	Accuracy	Precision	Recall (sensitivity)	Specificity	AUC
Livers					
NN_Shallow_Livers	0.7785 (± 0.0795)	0.7898 (± 0.1272)	0.8000 (± 0.0877)	0.7551 (± 0.1529)	0.8754 (± 0.1057)
NN_Deep_Livers	0.7471 (± 0.1031)	0.7737 (± 0.1404)	0.7385 (± 0.1595)	0.7538 (± 0.1770)	0.8776 (± 0.0720)
Log_Reg_Livers	0.7948 (± 0.0944)	0.8333 (± 0.1053)	0.7846 (± 0.1837)	0.8333 (± 0.1274)	0.9018 (± 0.0597)
Hearts					
NN_Shallow_Hearts	0.9088 (± 0.0221)	0.9086 (± 0.0616)	0.9013 (± 0.0916)	0.9162 (± 0.0576)	0.9766 (± 0.0153)
NN_Deep_Hearts	0.9168 (± 0.0485)	0.9405 (± 0.0937)	0.8846 (± 0.0753)	0.9457 (± 0.0871)	0.9685 (± 0.0210)
Log_Reg_Hearts	0.8245 (± 0.0853)	0.8275 (± 0.1239)	0.8013 (± 0.1135)	0.8438 (± 0.1382)	0.87525 (± 0.0782)

NN_Shallow_Livers, model to detect unhealthy livers based on CT scans using shallow neural network; NN_Deep_Livers, model to detect unhealthy livers based on CT scans using deep neural network; Log_Reg_Livers; logistic regression model to detect unhealthy livers based on radiodensity; NN_Shallow_Hearts, model to detect unhealthy hearts based on CT scans using shallow neural network; NN_Deep_Hearts, model to detect unhealthy hearts based on CT scans using deep neural network; Log_Reg_Hearts, logistic regression model to detect unhealthy hearts based on radiodensity; AUC, area under the curve.

farms, where chickens can grow from 50 g (one-day) chick to almost 3000 g of live slaughter weight within less than 6 weeks (40 days) (Souza da Silva et al., 2021). Obviously, this genetic advance was accompanied by improvements and intensification in feeding, management and veterinary care (Korver, 2023). However, these changes created many new challenges for poultry production, including maintaining a high health status of all the birds.

Because attention was mainly paid to external traits (such as muscle content), the internal organs have remained underdeveloped (Tallentire et al., 2016). That created extra pressure on the key internal organs such as liver and heart, because they need to supply more nutrients, oxygen and detoxification to meet the raised demands of the body developing at very high rate.

It is a well-established pathophysiological phenomenon that if any organ is functioning on an increased level of activity, it is more prone to health problems including infectious diseases (Marsland et al., 2002). Another risk factor for infectious disease occurrence is a high stock density frequently exceeding multiple birds on every 1 m^2 of the chicken barn and this is the production standard for many fast-growing broilers (Horne, 2020). Therefore, inflammation of the internal organs including liver and pericardium/heart are significant health problems on every broiler farm and are also important reasons for the carcass rejection during PM inspections at the slaughterhouse.

Pericarditis can result from a number of etiologies, some of which are harmful to human health including pathogenic Escherichia coli, Salmonella sp., and Pseuodmonas aeuroginosa. In literature there are already 37 potential causes of pericarditis in poultry identified (Coburn & Munday, 2004). It is also important to highlight that pericarditis commonly occurs as part of a polyserositis following septicemia, with colisepticemia (septicemia caused by Escherichia coli) as the most prevalent type being subsequently one of the major bacterial infections in the poultry worldwide (Panth, 2019).

According to Olkowski et al. (Olkowski et al., 2003), typical lesions found in hearts with pericardium include: excessive accumulation of serous fluid in the pericardial space (a.), locally extensive or focal adhesions between parietal and visceral components of the pericardium (b.), focal or locally extensive fibrous deposits on visceral pericardium that may or may not have been associated with adhesions (c.), and thickened pericardium (d.). In our study at least one of the lesions described above (a.-d.) were found in all the pericarditis samples (DIS_H).

Liver inflammation in chickens has mainly viral etiology (Yugo et al., 2016). Hepatitis is predominantly caused by fowl adenoviruses (FAdV) causing two diseases in young broilers: inclusion body hepatitis and hepatitis hydropericardium syndrome. These two conditions are characterized by non-specific clinical signs, but often a sudden increase in mortality is observed (El-Shall et al., 2022). Similarly, as pericarditis hepatitis might be a part of a generalized infection caused by Escherichia coli (colibacillosis), which also might be harmful to humans (Panth et al., 2019). According to El-Shall et al. (El-Shall et al., 2022), the livers

in birds diagnosed with hepatitis are pale, friable, enlarged with minute white foci and, in certain cases, petechial or ecchymotic hemorrhages (a.). Additionally, visible fibrous deposits might be found if hepatitis is a part of generalized infections (b.) (Panth et al., 2019). As expected in our dataset livers categorized as hepatitis (DIS_L) were identified with the lesions described in either of two scenarios (a.-b.).

Theoretically the lesions should be also visible during manual (qualitative) CT scan analysis, but we were not able to see them during CT scans analysis. That is because of their relatively small size and severity and/or resolution of our CT scans.

We also considered whether scanning the organs in batches (rather than individual scanning) can influence the quality of a CT scan and consequently influence further analysis. We applied batch scanning to reduce the time of scanning. In other words, with a single CT scanner run, we were able to get CT scans of multiple organs. It is important to mention that during CT scanning, we always kept the organs on the table in such a way that we were eliminating any physical contact between them. Additionally, all the individual organs have been carefully cropped and double-checked by two authors of the study. Each organ was given an unique ID and number, which allowed for sample identification and tracking. In the beginning, we also investigated if the cropped organs had the same properties (e.g., size, volume, mean radiodensity) compared to the same organs in a CT scan of the organ batch, and in all cases, the results were the same. Thus, scanning organs in batches did not influence image quality or subsequent analysis.

Regardless of the lesions found in any of these organs, this is the sufficient base for the official veterinarian or meat inspector to declare the viscera and corresponding carcass unfit for human consumption (Regulation of European Commission (EC) no 2017/625, 2017). Obviously, it would be ideal to describe, analyze and record organs lesions in details (such as in dissecting room setting), but unfortunately it is practically impossible due to time pressure during in-line PM inspection.

4.2. Current method of inspection and its limitations

The current method of inspection involves visual examination of the chicken viscera. After evisceration the organs are placed on a special bowl under or rack behind the carcass from which they were removed from. The carcasses with the corresponding packet of organs are transferred on a conveyor line to the post-mortem inspection station, where they are inspected by official veterinarians or official auxiliary (Regulation of European Commission (EC) no 2017/625, 2017). However, it is important to note that the speed of the line often exceeds 180 chickens per minute (Pacholewicz et al., 2019). That means that during 1 s 3 chickens should be inspected, which puts extreme time pressure on the inspectors, potentially negatively influencing their performance. Another limitation of any human visual inspection is subjectivism, susceptibility for committing mistakes, fatigue and different performance throughout the whole work shift (e.g. beginning vs. end of the shift). For example, in the study by Arzoomand et al. (Arzoomand et al., 2019), a large variation in assessments among official veterinarians and auxiliaries was observed when inspecting pig carcasses. The inter-rater variability among inspectors was also highlighted by the European Food Safety Authority (EFSA Panel on Biological Hazards EFSA Panel on Contaminants in the Food Chain & EFSA Panel on Animal Health and Welfare, 2012). All in all, this highlights that there is a need for an objective and robust method of organ inspection.

4.3. Statistical analysis and logistic regression models (Log_Reg_Hearts and Log_Reg_Livers)

In both types of organs we tested if there are statistically significant differences between mean radiodensity of healthy and diseased groups. In both livers and hearts, we observed the same pattern and reported that the mean radiodensity of the healthy organs (HEAL_H/HEAL_L) was significantly higher than in diseased organs (DIS_H/DIS_L). The explanation for this difference might be that in the course of any organ inflammation the healthy tissue is being damaged and is being partially replaced by connective/fat tissue, which has lower radiodensity (Hermena & Young, 2025). This is accompanied by the occurrence of specific lesions e.g. spots on the liver surface or fibrous deposits on the pericardium surface. These two mechanisms combined are probably causing the lower mean radiodensity of the affected organs.

According to our knowledge there are no studies investigating CT scanning of different organs in poultry. However, in another animal species, for example in dogs, there are studies describing the application of CT scanning and differences in radiodensity in detecting and characterizing liver lesions. The results show that lesions (or affected livers) have lower radiodensity than healthy organs (Burti et al., 2021). There is also another qualitative study suggesting that differences in Hounsfield units are a useful predictor of pleural effusion (accumulation of excessive fluid in the pleural space) type in dogs (Woods et al., 2018).

It is also important to note that the numerical differences between groups were higher in hearts comparing to livers and this can be partially explained the in the most cases in hearts the whole organ was pathologically changed, while in liver the lesions were more focal or dispersed. It is also worth noting that variation of radiodensity (expressed as standard deviation) was greater in diseased groups in both types of organs. It can be explained that in diseased organs there might be pathologically changed segments of the organ, while some other parts might be not affected. The type of pathology and its extent can also vary between samples. This results in higher disuniformity of the radiodensity within the organ and subsequently increases the variation (expressed as standard deviation).

Based on these significant differences in radiodensity we decided to develop two logistic regression models (Log_Reg_Hearts and Log_Reg_Livers). Logistic regression is considered as a traditional machine learning technique, where hand-crafted features (numbers) are used to create a linear classifier. In many deep learning studies logistic regression models are considered as 'baseline' classifier to which deep learning (AI) models are compared to (Song et al., 2021). Thus, the same strategy was applied in our research.

Our logistic regression models reached accuracy of 82 % for hearts (Log_Reg_Hearts) and 79 % for livers (Log_Reg_Livers) and these models used a single input feature (mean radiodensity of the organ expressed in Hounsfield units) retrieved from every CT scan. We choose this approach, because in the course of diseases many organs in animals and humans are changing their radiodensity and this is an important feature for diagnosis based on CT scanning. One can also come up with other features that can be extracted from CT scan e.g. the size or shape of the organ. However, we decided not to use these features for logistic regression models, because we think that these features might by significantly influenced by the size of the bird (size and shape) or cannot be directly linked and explained by lesions found in affected organs. Moreover, by resizing every CT scan to uniform shape (128x128x128) some information regarding the size and shape might be lost during data processing.

4.4. Deep learning models (NN_Shallow_Livers/Hearts and NN_Deep_Livers/Hearts) performance

Both our neural network-based models performed at an accuracy around 91 % for hearts and 76 % for livers. We did not observe significant differences between the shallow and deep neural networks despite the roughly 100x difference in number of parameters. In general, the models for hearts outperformed the ones for livers, which was in accordance with our expectations, because the lesions of pericarditis are more pronounced compared to hepatitis, so probably easier to detect. Furthermore, our results show that neural network-based models in hearts can extract more information from the whole CT scan than just the mean radiodensity value and use it in training to achieve better performance. On the other hand, in livers the neural network-based

models perform worse than logistic regression, which suggests that the mean radiodensity of the organs is the key feature for detecting diseased livers. It implies that in some cases relatively simple models (logistic regression) can outperform more complicated methods (neural networks). Similar findings were reported in a study on cows reproduction (Bates & Saldias, 2019) or in the review comparing logistic regression and neural networks in healthcare research (Issitt et al., 2022).

Nevertheless, anatomical variability, sample size, and the form of disease manifestation should be considered as other explanations of the fact that deep learning models outperformed logistic regression for hearts but not for livers. Chicken hearts are smaller and more uniform than livers, and pericarditis is usually heavily demonstrated in the whole organ. Consequently, deep learning models are capable of extracting more features, which indirectly could lead to better performance of deep learning models in hearts. However, it is not excluded that if the resolution of the CT scans would be increased it may be possible to gather more information or see new patterns that could be used for training more accurate neural networks.

Moreover, to get better understanding of the decision-making process made by neural networks explainable AI techniques could be applied e.g. GradCam (Selvaraju et al., 2020), but it was beyond the scope of the current study.

For our best performing models (Log_Reg_Livers for livers and NN_Deep_Hearts for hearts) the accuracy is considered as promising taking under consideration that we have well-balanced classes. The ratio of healthy to diseased organs in our dataset was already close to 50:50 for both livers and hearts (e.g. 48 % healthy livers, 52 % diseased livers), so the dataset is considered relatively balanced. Additionally, we ensured that each fold in the cross validation preserved the healthy/diseased ratio by using stratified sampling (i.e. to keep relatively the same proportion of classes across training, validation, and test sets). Consequently there was no need to handle class imbalance.

The quality of positive prediction (precision) reached value of 94 % in hearts and 83 % in livers, which is also satisfactory result. On the other hand, the values of recall (ability to identify positive instances from all the actual positive instances) reached 78 % and 88 % for Log_Reg_Livers and NN_Deep_Hearts. Not only should the performance of the model be discussed, but also the impact of misclassifications (false negative/false positive). In our study a false negative sample would be a diseased liver/heart classified as a healthy organ. This type of misclassification might lead to a situation when diseased organs are approved and considered as fit for human consumption, which could lead to food safety risks. For example, there was a study describing an outbreak of campylobacteriosis due to consumption of chicken liver dishes (pathe) (Wensley et al., 2020).

On the other hand, a false positive outcome would classify an organ as diseased, but actually it was healthy. These organs would be unnecessarily discarded, leading to increased food waste and subsequently reducing the profitability for the slaughterhouse operator (Malher et al., 2015). Nevertheless, there are still some possibilities to improve the accuracy of the classification. For example, some samples can be manually re-examined if the model is uncertain. Alternatively, the categorization of all the samples predicted by the model as positive (diseased) can be re-confirmed by human inspectors. Taking under consideration the real-life prevalence of carcasses found with diseased organs (several %), it should not involve much human labor, therefore a highly reduced workload is expected overall.

4.5. Target CT scanning setup in the slaughterhouse

It is important to note that the key benefit of CT scanning is to examine the organs when they are still in the carcass, because at that point they are not visible to the human eye or a camera. Therefore, organ classification models should also be tested in the setup where the organs are situated inside the carcass. Because of that, this study should be considered as a proof-of-concept focusing more on different methodological approaches than on a direct application. However, classifying organs separately has two benefits that are valuable at this early stage of research. Firstly, because of the class imbalance and the impossibility of checking the organ status before evisceration, many carcasses would need to be acquired and scanned to have enough positive samples, which would result in higher costs. Secondly, scanning organs separately leads to a more controlled scanning setup, allowing the research to focus precisely on the detection of organ diseases without being affected by other factors such as scanning artifacts or overfitting on other parts of the carcass.

With a single CT scan a lot of information is gathered, not only limited to the health status of the organs, but also referring to animal welfare (e.g. broken wings) or size/shape of the carcass that might be useful for further processing (Libera et al., 2024). Therefore, in our ideal target setup in the slaughterhouse every carcass should be CT scanned just before evisceration. The data regarding health status of the organs should be processed in real-time and fed to a decision-making AI model, so the carcasses with unhealthy organs are moved to another conveyor line and then discarded immediately, consequently skipping the unnecessary evisceration. The expected outcome of this solution would be a lower prevalence of affected carcasses and an improved workflow in terms of time and energy. In other words, some carcasses will not be eviscerated, which saves time and energy for the process of evisceration itself but also for meat inspectors working on post-mortem inspection station.

The data on animal welfare and carcass size/shape are collected at the same time, but can be processed later, because they will be used in further processing or final assessment of the slaughtered batch (flock). Anyways, all the data collected with CT scanning can be also send back to the poultry farmers, so they also get extra information about the health/welfare status within the flock with the ultimate goal to introduce improvements for the next flock e.g. changing housing conditions/density, vaccination strategy or changing the catching/transportation company.

Alternatively, only a pre-selected and representative part of the batch can be examined by CT scanning and AI and then compared to the other part of the batch regularly checked by official veterinarians/meat inspectors. We hypothesize that the detection rate would be higher in the group examined by CT scanning and AI compared to the group inspected by humans. This strategy might be helpful to compare the outcomes of AI model and human inspector and might act as an intermediate step before fully replacing human inspectors.

4.6. Strengths and limitations of the study

One of the strengths of this study is the collection of a unique dataset consisting of CT scans of diseased and healthy chicken hearts and livers, so we did not use any pre-existing data. It is important to highlight that we collected samples on multiple days (so multiple chicken flocks), which provides extra variability and subsequently confirms the robustness of our models. Further, we thoroughly analyzed every sample and used them to create two types of classifiers. By creating two types of classifiers on the same dataset we compared traditional machine learning (logistic regression) with neural networks-based learning. This approach allowed us to benchmark and show that neural networks can outperform more traditional techniques.

The future experiments should focus on further trying to improve the accuracy of the models and testing their performance on the organs before evisceration in the real carcasses or 3D models simulating carcasses. Other internal organs should be also inspected using CT scanning e.g. respiratory tract, gizzard or crop. This might facilitate to detect pathologies other than hepatitis/pericarditis.

Among the limitations of the study the limited number of samples should be considered. This is a consequence of independent sample collection comparing to using pre-existing (open) datasets, where thousands of CT scans were collected over multiple years. However,

during training our deep learning models we applied data augmentation, which is proven to be an effective solution to increase variability of the input data in relatively small datasets. It is also important to note that the difference of mean radiodensity between healthy and diseased organs is relatively small, which might also make the classifier sensitive to CT scanning artifacts (e.g. beam hardening) or changes of the scanner/scanning protocol. Another shortcoming of our study is that CT scanning is a quite complex, slow and expensive technique to be implemented in the industrial setting as of now. For instance, based on a rough estimation the price of a new CT scanner comparable to the one used in our study can range from 100,000 to 1,000,000 USD, but the cost of daily operation (mainly electricity) should not exceed several dozen USD (Libera et al., 2025). It is also important to note that CT scanning is based on X-rays, which are harmful to humans at high exposure, so the occupational health must also be carefully considered and radiological protection to be installed. Integrating CT scanning into the existing slaughterhouse workflow remains another to hurdle overcome before its industrial use.

However, the potential benefits of applying CT scanning in slaughterhouses (more accurate post-mortem/animal welfare inspection combined with reduced labor costs) might justify its implementation. Moreover, the rapid developments in CT scanning technology, along with the modernization of food safety regulations, will hopefully reduce the time required before its application in real-life modern poultry slaughterhouses.

5. Conclusions

Our study confirms that diseased chicken hearts and livers can be automatically detected using the combination of CT scanning and logistic regression/deep learning models. The deep learning models outperformed logistic regression models on hearts but not on livers. There were no significant differences between the performance of shallow vs. deep neural networks. All in all, CT scanning has potential to increase the safety of poultry edible organs and streamline the workflow in the slaughterhouse.

CRediT authorship contribution statement

Kacper Libera: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Effrosyni Kritsi:** Writing – review & editing, Methodology, Investigation, Data curation. **Dirk Schut:** Writing – review & editing, Methodology, Investigation, Data curation. **Louis van Steijn:** Writing – review & editing, Funding acquisition. **Lourens Heres:** Writing – review & editing, Supervision, Investigation. **Len Lipman:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data and code availability

The data used for this study are available upon reasonable request.

Code is available at: https://github.com/D1rk123/ct_chicken_organ_classification.

Declaration of competing interest

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

Acknowledgments

The authors would like to thank the slaughterhouse staff and Monique Jacobs from the diagnostic imaging department at the university veterinary hospital in Utrecht. This study was supported by a grant from Dutch Research Council (NWO), the project title: Universal Three-dimensional Passport for process Individualization in Agriculture (UTOPIA); grant no: ENWSS.2018.003.

Since this study did not involve any living animals, no ethical statement was required. The samples were obtained from the chickens routinely slaughtered in a commercial slaughterhouse in the Netherlands between February 2024 and March 2025.

Data availability

Data will be made available on request.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623–2631). New York, USA: Association for Computing Machinery. https://doi.org/10.1145/3292500.3330701.

Arzoomand, N., Vågsholm, I., Niskanen, R., Johansson, A., & Comin, A. (2019). Flexible distribution of tasks in meat inspection – a pilot study. *Food Control, 102*, 166–172. https://doi.org/10.1016/j.foodcont.2019.03.010

AVEC. (2024). Annual report 2024. Available from https://avec-poultry.eu/wp-content/uploads/2024/08/07082-AVEC-annual-report-2024_WEB.pdf. (Accessed 16 April 2025).

Bates, A. J., & Saldias, B. (2019). A comparison of machine learning and logistic regression in modelling the association of body condition score and submission rate. *Preventive Veterinary Medicine, 171*, Article 104765. https://doi.org/10.1016/j.prevetmed.2019.104765

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in neural information processing systems* (pp. 2546–2554). Curran Associates.

Burti, S., Zotti, A., Bonsembiante, F., Contiero, B., & Banzato, T. (2021). Diagnostic accuracy of delayed phase post contrast computed tomographic images in the diagnosis of focal liver lesions in dogs: 69 cases. *Frontiers in Veterinary Science, 8*, Article 611556. https://doi.org/10.3389/fvets.2021.611556

Buzdugan, S. N., Chang, Y. M., Huntington, B., Rushton, J., Guitian, J., Alarcon, P., & Blake, D. P. (2020). Identification of production chain risk factors for slaughterhouse condemnation of broiler chickens'. *Preventive Veterinary Medicine, 181*, Article 105036. https://doi.org/10.1016/j.prevetmed.2020.105036

Coburn, H., & Munday, D. (2004). A mini review of the causes and incidence of pericarditis in poultry. *Centre for Risk Analysis and Epidemiology, Veterinary Laboratories Agency, Weybridge, UK*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=680965b987b8561fe468db1f505f3ddce37504a0.

DenOtter, T. D., & Schubert, J. (2025). Hounsfield unit. In *StatPearls*. Treasure Island (FL): StatPearls Publishing.

EFSA Panel on Biological Hazards (BIOHAZ), EFSA Panel on Contaminants in the Food Chain (CONTAM), & EFSA Panel on Animal Health and Welfare (AHAW). (2012). Scientific opinion on the public health hazards to be covered by inspection of meat (poultry). *EFSA Journal, 10*(6). https://doi.org/10.2903/j.efsa.2012.2741. Article 2741.

El-Shall, N. A., El-Hamid, H. S. A., Elkady, M. F., Ellakany, H. F., Elbestawy, A. R., Gado, A. R., Geneedy, A. M., Hasan, M. E., Jaremko, M., Selim, S., El-Tarabily, K. A., & El-Hack, M. E. A. (2022). Epidemiology, pathology, prevention, and control strategies of inclusion body hepatitis and hepatitis-hydropericardium syndrome in poultry: A comprehensive review. *Frontiers in Veterinary Science, 9*, Article 963199. https://doi.org/10.3389/fvets.2022.963199

Falcon, W. (2019). PyTorch lightning team. *GitHub*. https://github.com/PyTorchLightning/pytorch-lightning.

Forseth, M., Moe, R. O., Kittelsen, K., Skjerve, E., & Toftaker, I. (2023). Comparison of carcass condemnation causes in two broiler hybrids differing in growth rates. *Scientific Reports, 13*, Article 4195. https://doi.org/10.1038/s41598-023-31422-0

Hartcher, K. M., & Lum, H. K. (2020). Genetic selection of broilers and welfare consequences: A review. *World's Poultry Science Journal, 76*(1), 154–167. https://doi.org/10.1080/00439339.2019.1680025

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Hermena, S., & Young, M. (2025). CT-Scan image production procedures. In *StatPearls*. StatPearls Publishing.

Hoey, E. T. D., Shahid, M., & Watkin, R. W. (2016). Computed tomography and magnetic resonance imaging evaluation of pericardial disease. *Quantitative Imaging in Medicine and Surgery, 6*(3), 274–284. https://doi.org/10.21037/qims.2016.01.03

Horne, P. L. M. (2020). *Economics of broiler production systems in The Netherlands: Economic aspects within the Greenwell sustainability assessment model*. Wageningen Economic Research.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, K. Q. (2016). Deep networks with stochastic depth. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer vision – ECCV 2016* (pp. 646–661). Springer. https://doi.org/10.1007/978-3-319-46493-0_39.

Implementing Regulation (EU) 2019. 627 of 15 March 2019 laying down uniform practical arrangements for the performance of official controls on products of animal origin intended for human consumption in accordance with Regulation (EU) 2017/625 of the European Parliament and of the Council and amending Commission Regulation (EC) No 2074/2005 as regards official controls. Retrieved July 10, 2025, from https://eur-lex.europa.eu/eli/reg_impl/2019/627/oj/eng.

Issitt, R. W., Cortina-Borja, M., Bryant, W., Bowyer, S., Taylor, A. M., & Sebire, N. (2022). Classification performance of neural networks versus logistic regression models: Evidence from healthcare practice. *Cureus, 14*(2), Article e22443. https://doi.org/10.7759/cureus.22443

Kimberlin, L. (2016). *Atlas of clinical imaging and anatomy of the equine head*. John Wiley & Sons.

Korver, D. R. (2023). Review: Current challenges in poultry nutrition, health, and welfare. *Animal, 17*(2), Article 100755. https://doi.org/10.1016/j.animal.2023.100755

Libera, K., Schut, D., Kritsi, E., van Steijn, L., Dallman, T., & Lipman, L. (2025). Detecting wing fractures in chickens using deep learning, photographs and computed tomography scanning. *Poultry Science, 104*(8), Article 105264. https://doi.org/10.1016/j.psj.2025.105264

Libera, K., Valadian, R., Vararattanavech, P., Dasari, S. N., Dallman, T. J., Weerts, E., & Lipman, L. (2024). Inspection of chicken wings and legs for animal welfare monitoring using X-ray computed tomography, visual examination, and histopathology. *Poultry Science, 103*(1), Article 103403. https://doi.org/10.1016/j.psj.2023.103403

Lichtner, E., Cetin, Y. E., & Kriegel, M. (2024). Poultry evisceration and cross-contamination – effectiveness of local air extraction and mechanical barriers. *Journal of Food Engineering, 381*, Article 112164. https://doi.org/10.1016/j.jfoodeng.2024.112164

Loshchilov, I., & Hutter, F. (2018). Decoupled weight decay regularization. *International conference on learning representations (ICLR)*. https://doi.org/10.48550/arXiv.1711.05101

Malher, X., Coudurier, B., & Redlingshöfer, B. (2015). Food losses and waste in the poultry production chain: From farm to retail. *XVI European Symposium on the Quality of Eggs and egg products, may 2015*. France: Nantes.

Marsland, A. L., Bachen, E. A., Cohen, S., Rabin, B., & Manuck, S. B. (2002). Stress, immune reactivity and susceptibility to infectious disease. *Physiology & Behavior, 77*(4–5), 711–716. https://doi.org/10.1016/s0031-9384(02)00923-x

Olkowski, A. A., Wojnarowicz, C., Rathgeber, B. M., Abbott, J. A., & Classen, H. L. (2003). Lesions of the pericardium and their significance in the aetiology of heart failure in broiler chickens. *Research in Veterinary Science, 74*(3), 203–211. https://doi.org/10.1016/s0034-5288(03)00004-3

Pacholewicz, E., Luning, P. A., Smits, C. B., Koene, M. G. J., & Hartog, M. (2019). *Developing nudges as an intervention to improve compliance of food handlers to procedures during broiler processing*. & Research: Wageningen University. https://doi.org/10.18174/574413

Panth, Y. (2019). Colibacillosis in poultry: A review. *Journal of Agriculture and Natural Resources, 2*(1), 301–311. https://doi.org/10.3126/janr.v2i1.26094

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in PyTorch. In *31st conference on neural information processing systems (NeurIPS 2017)*. Long Beach, CA, USA.

Pérez-García, F., Sparks, R., & Ourselin, S. (2021). TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine, 208*, Article 106236. https://doi.org/10.1016/j.cmpb.2021.106236

Regulation (EC) No 853/2004 of the European Parliament and of the Council of 29 April 2004 laying down specific hygiene rules for food of animal origin. Retrieved April 16, 2025, from https://eur-lex.europa.eu/eli/reg/2004/853/oj/eng.

Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products. Retrieved April 16, 2025, from https://eur-lex.europa.eu/eli/reg/2017/625/oj/eng.

Salines, M., Allain, V., Roul, H., Magras, C., & Le Bouquin, S. (2017). Rates of and reasons for condemnation of poultry carcases: Harmonised methodology at the slaughterhouse. *The Veterinary Record, 180*(20). https://doi.org/10.1136/vr.104000, 516–516.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision, 128*, 336–359. https://doi.org/10.1007/s11263-019-01228-7

Seong, P. N., Cho, S. H., Park, K. M., Kang, G. H., Park, B. Y., Moon, S. S., & Ba, H. V. (2015). Characterization of chicken by-products by means of proximate and nutritional compositions. *Korean Journal for Food Science of Animal Resources, 35*(2), 179–188. https://doi.org/10.5851/kosfa.2015.35.2.179

Shatnawi, M. Q., Abuein, Q., & Al-Quraan, R. (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images. *Intelligence-Based Medicine, 11*, Article 100188. https://doi.org/10.1016/j.ibmed.2024.100188

Solovyev, R., Kalinin, A. A., & Gabruseva, T. (2022). 3D convolutional neural networks for stalled brain capillary detection. *Computers in Biology and Medicine, 141*, Article 105089. https://doi.org/10.1016/j.compbiomed.2021.105089

Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *International Journal of Medical Informatics, 151*, Article 104484. https://doi.org/10.1016/j.ijmedinf.2021.104484

Souza da Silva, C., Molenaar, R., Giersberg, M. F., Rodenburg, T. B., van Riel, J. W., De Baere, K., Van Dosselaer, I., Kemp, B., van den Brand, H., & de Jong, I. C. (2021). Day-old chicken quality and performance of broiler chickens from 3 different hatching systems. *Poultry Science, 100*(3), Article 100953. https://doi.org/10.1016/j.psj.2020.12.050

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.

Tallentire, C. W., Leinonen, I., & Kyriazakis, I. (2016). Breeding for efficiency in the broiler chicken: A review. *Agronomy for Sustainable Development, 36*, Article 66. https://doi.org/10.1007/s13593-016-0398-2

Thrall, E. D. (2017). *Textbook of veterinary diagnostic radiology* (7th ed.). Saunders.

Wensley, A., Padfield, S., & Hughes, G. J. (2020). An outbreak of campylobacteriosis at a hotel in England: The ongoing risk due to consumption of chicken liver dishes. *Epidemiology and Infection, 148*, e32. https://doi.org/10.1017/S095026882000028X

Wightman, R., Touvron, H., & Jégou, H. (2021). ResNet strikes back: An improved training procedure in timm. https://doi.org/10.48550/arXiv.2110.00476.

Woods, S. J., Spriet, M., Safra, N., Cissell, D. D., & Borjesson, D. L. (2018). Hounsfield units are a useful predictor of pleural effusion cytological type in dogs but not in cats. *Veterinary Radiology & Ultrasound, 59*(4), 405–441. https://doi.org/10.1111/vru.12618

Yugo, D. M., Hauck, R., Shivaprasad, H. L., & Meng, X. J. (2016). Hepatitis virus infections in poultry. *Avian Diseases, 60*(3), 576–588. https://doi.org/10.1637/11229-070515-Review.1

Zunair, H., Rahman, A., Mohammed, N., & Cohen, P. C. (2020). Uniformizing techniques to process CT scans with 3D CNNs for tuberculosis prediction. In *MICCAI 2020 international workshop on PRedictive intelligence in MEdicine (PRIME)* (pp. 172–183). Springer. https://doi.org/10.1007/978-3-030-59354-4_15.