

Article

Automatic Scoring System for Monitoring Foot Pad Dermatitis in Broilers

Helen Louton ^{1,*}, Shana Bergmann ², Andre Pillar ², Michael Erhard ², Jenny Stracke ³, Birgit Spindler ⁴, Paul Schmidt ⁵, Jan Schulte-Landwehr ⁶ and Angela Schwarzer ²

- ¹ Animal Health and Animal Welfare, Faculty of Agricultural and Environmental Sciences, University of Rostock, Justus-von-Liebig-Weg 6b, D-18059 Rostock, Germany
- ² Chair of Animal Welfare, Ethology, Animal Hygiene and Animal Husbandry, Department of Veterinary Sciences, Faculty of Veterinary Medicine, LMU Munich, Veterinaerstr. 13/R, D-80539 Munich, Germany; s.bergmann@lmu.de (S.B.); a.pillar@tierhyg.vetmed.uni-muenchen.de (A.P.); m.erhard@tierhyg.vetmed.uni-muenchen.de (M.E.); a.schwarzer@tierhyg.vetmed.uni-muenchen.de (A.S.)
- ³ Institute of Animal Science, Farm Animal Ethology, University of Bonn, Endenicher Allee 15, D-53115 Bonn, Germany; jenny.stracke@itw.uni-bonn.de
- ⁴ Institute for Animal Hygiene, Animal Welfare and Farm Animal Behavior, University of Veterinary Medicine Hannover, Foundation, Bischofsholer Damm 15, D-30173 Hannover, Germany; Birgit.Spindler@tiho-hannover.de
- ⁵ Statistical Consulting for Science and Research, Große Seestraße 8, D-13086 Berlin, Germany; paul.schmidt.mail@gmail.com
- ⁶ CLK GmbH, Zur Steinkuhle 3, D-48341 Altenberge, Germany; schulte-landwehr@clkgmbh.de
- * Correspondence: helen.louton@uni-rostock.de; Tel.: 49-381-498-3360



Citation: Louton, H.; Bergmann, S.; Pillar, A.; Erhard, M.; Stracke, J.; Spindler, B.; Schmidt, P.; Schulte-Landwehr, J.; Schwarzer, A. Automatic Scoring System for Monitoring Foot Pad Dermatitis in Broilers. *Agriculture* **2022**, *12*, 221. <https://doi.org/10.3390/agriculture12020221>

Academic Editor: Claudia Arcidiacono

Received: 13 January 2022

Accepted: 1 February 2022

Published: 3 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The assessment of foot pad dermatitis at slaughter is a suitable method to assess and monitor the welfare of broilers. The goals of this study were to define and validate a camera-based score that could identify macroscopic lesions of the foot pads, to identify errors, and to assess possible external factors that could influence the assessment. In the first phase 200 feet of broilers and in the second phase 500 feet were collected at slaughter, assessed visually, hung back into the evisceration line, and assessed by an automatic system. The camera score cut-off values were defined in the first (=calibration) phase. In the second (=validation) phase, the performance of diagnosis for these cut-off values was evaluated, and possible errors in the assessment of reference surface area and foot pad lesions were analyzed. The results showed that, in particular, Macro Scores 0, 2, and 3 could be identified with sufficiently high sensitivity. For Macro Score 1, the sensitivity of diagnosis was not sufficiently high in the two evaluated software versions. The current automatic assessment systems at slaughter could be adjusted to the cut-off values in order to classify foot pad dermatitis lesions. Furthermore, software updates can enhance the performance measures and lower the probability of errors.

Keywords: automatic assessment; foot pad lesion; precision livestock farming; welfare assessment; welfare indicator

1. Introduction

The assessment of animal-based indicators, such as foot pad dermatitis (FPD), is considered a suitable method to assess and monitor the welfare of broilers at farm level [1]. Saraiva et al. [1] stated that FPD is the most frequently observed welfare indicator at slaughter, in addition to hock burn and dirty feathers. FPD causes inflammatory and often necrotic lesions of the foot and toe pads in broilers [2]. The authors assume that FPD is potentially painful [3]. However, FPD is not only a welfare problem, but can also cause economic losses due to carcass condemnation and low growth rate [2,4].

FPD can occur as a mild lesion with discoloration of the skin and hyperkeratosis or, in severe cases, it can cause swelling, erosions, and ulcerations [4]. The identified risk factors

for the onset of FPD are nutritional aspects [4], seasonal effects, the age at slaughter, low daily weight gain [5], and the type of litter [6,7]. However, the most important risk factor is wet litter, especially if it is combined with chemical agents, such as ammonia [5,8]. The stocking density does not seem to be a major risk factor for the development of FPD [5]. De Jong et al. [9] observed that wet litter had a negative influence not only on FPD, but also on overall health and welfare parameters, such as gait score, hock burn, cleanliness, growth, feed intake, performance, and carcass condemnation or rejection.

Different possible scoring systems are used for the assessment of FPD. They vary in the number of score categories and the approach of the assessment. In most of the scoring systems, three [10–12], four [13,14] or five [15–17] categories are used for the classification of FPD. In some systems, especially the size of the FPD, lesions are considered [16], whereas other systems consider the depth or severity of FPD [13,14]. In Germany, officials at slaughter can use a three- or four-point scoring system to evaluate FPD in broilers [18,19]. Piller et al. [16] recommended a three-category classification of FPD for the identification of ulcers, whereby lesions larger than 0.5 cm in diameter have a high probability of consideration as a “deep” lesion.

Even though FPD is a suitable indicator for the assessment of broiler health and welfare [20,21], the authors state that welfare monitoring of broilers cannot be performed on the basis of only one type of lesion (e.g., FPD). Welfare monitoring is a multicriterial approach, and in addition to FPD, also hock burn, breast burn, skin scratches, and breast blisters should be assessed to draw conclusions regarding welfare problems [21]. Assessing animal welfare indicators is time consuming, and the automatic assessment offers a novel method to monitor welfare on a large scale [15].

According to a survey of poultry slaughterhouses in the German-speaking territory, FPD is currently monitored in general [22]. In slaughterhouses, 100 feet are examined by a veterinary official or additional camera systems are installed to monitor the foot pad health [15]. According to the “QS Qualitätssicherung,” a system to assure the quality of food in Germany, all of the slaughterhouses that slaughter more than 4000 broilers per hour need an automatic system to evaluate the foot pad health of all the flock feet [23]. The sample size of 100 feet in manual assessment is also recommended by other authors [23–25]. The authors of different studies agree that an official monitoring can reduce the prevalence of FPD. Directors of poultry slaughterhouses stated that not only the implementation of bonus–malus systems, but also the fear of consequences has increased the pressure to improve the care to prevent FPD [22]. In Denmark, an official monitoring of FPD in the year 2002 led to a distinct reduction in FPD by 2005 [5]. In approximately one third of the member states of the European Union, which account for 32% of the broiler production, FPD is monitored [26].

Automatic assessment systems to evaluate the health of animals at slaughter are promising and are already used under conditions of practice [27]. The documentation of sensitivity, specificity, and replicability, as well as the implementation of frequent verification and calibration are claimed to be inherent features of quality assurance systems. However, the exact procedures are not specifically defined [23]. Even though threshold values need to be chosen carefully and should be validated to provide reliable results, there is no scientific literature considering these values. This study was performed to fill the gap of missing data in the literature considering threshold values and validated camera systems, as well as possible errors that occur in automatic assessment systems and how these could be avoided.

The general goal of this study was to improve and validate an automatic assessment system for FPD in broilers and to generate threshold values for relevant lesions. For this purpose, the specific goals were (1) to define a camera-based score that could detect macroscopic alterations of foot pads, (2) to validate this camera-based scoring system for the classification of foot pad lesions in broilers, (3) to identify errors, and (4) to assess possible external factors that might influence the identification of reference surface areas and FPD lesions.

2. Animals, Materials, and Methods

2.1. Animals and Materials

The feet (200 in the first phase, 500 in the second phase) of Ross 308 broilers of an age of 36–42 fattening days were collected at a slaughterhouse in Bavaria, Germany.

2.2. Inter-Observer Reliability Test

The inter-observer reliability was tested using the prevalence-adjusted and bias-adjusted kappa (PABAK) calculation by Byrt et al. [28]. For this test, 250 feet of Ross 308 broilers were evaluated by five observers. Piller et al. [16], who validated the used macroscopic assessment scheme histologically, already presented the results of the inter-observer reliability in their article.

2.3. Visual Assessment

2.3.1. Phase 1: Calibration

Two hundred feet (40 feet of each macroscopic score) of Ross 308 broilers were initially collected from the slaughter line for an individual assessment. The collection was conducted on 10 days from 6 February 2018 to 26 April 2018. On each day, between 7 and 32 samples were collected. An assessment scheme, revised according to the Welfare Quality[®] assessment protocol for poultry [10] and validated by Piller et al. [16], was used for the visual assessment (“macro score”) of FPD ranging from Macro Score 0 (no alteration) to Macro Score 4 (severe alteration). Macro Score 0 was defined as physiological skin without lesion, Macro Score 1 as single superficial lesion or several cumulated superficial lesions or deep lesion ≤ 0.5 cm, Macro Score 2 as superficial lesion >0.5 cm or deep lesion >0.5 to ≤ 1.0 cm, Macro Score 3 as deep lesion >1.0 cm, and Macro Score 4 as lesion on the foot pad and one or more deep lesions on the toe. The camera system only generates information on the foot pad and not on the toes. However, for Macro Score 4, contrary to Macro Score 3, the toe pads must have lesions in addition to the foot pad lesions. Therefore, for this study, Macro Score 4 was allocated to Macro Score 3. After the visual assessment, a photographic documentation of each foot was performed using a Sony Cyber-shot DSC-RX100 digital camera (Sony Europe Limited, Surrey, UK).

2.3.2. Phase 2: Validation

In the second phase of the project, 500 feet of Ross 308 broilers were collected from the slaughter line for an individual assessment (consisting of 85 samples of Macro Score 0, 86 of Macro Score 1, 159 of Macro Score 2, and 170 of Macro Scores 3 and 4). The sample collection was performed from 26 July to 11 December 2019. On 7 days of sampling, between 60 and 126 samples were taken per day. In this phase, the feet were similarly assessed using the same macro scores as in the calibration phase, and a photograph of each foot was taken.

2.4. Assessment by the Camera System

Following the described procedure of visual assessment and photographic documentation, each foot was individually hung back into the evisceration line and a picture of the foot, and if present of lesions, was taken by the camera system that was used by the slaughterhouse. For the automatic assessment, a 1.3-MP color camera (IDS Imaging, Obersulm, Germany) was used with “Chicken Check” software (CLK GmbH, Altenberge, Germany) and in combination with Halcon software (MVTec Software GmbH, Munich, Germany) for image processing. With this procedure, for each of the 200 (calibration phase) and 500 feet (validation phase), the size of the lesion in percent (%) measured by the camera system was evaluated (“camera score”). Two different software versions were used for each picture taken by the automatic assessment system to evaluate possible optimizations in terms of the detection of reference surface areas of the feet and the detection of lesions. In the following, these versions will be denoted as “original” and “updated” software. For the updated camera score, the camera detection limits were adapted and the threshold for the detection of differences in contrasts was lowered.

Once per sampling during the automatic camera assessment, the light intensity in lux (lx) (LMT Pocket-Lux 2B, LMT Lichtmesstechnik GmbH, Berlin, Germany), the relative humidity in percent (%), and the temperature in degrees Celsius (°C) (Testo 410-1 Flügelrad-Anemometer, Testo North America, West Chester, PA, USA) were measured at the height of the feet.

2.4.1. Phase 1: Calibration

In the first (=calibration) phase, camera scores were allocated for each macro score of foot pad lesions. Threshold values of camera scores for the respective macro scores (0–3) were defined, and the performance measures of the respective diagnosis were evaluated.

2.4.2. Phase 2: Validation

In the second phase, the threshold values generated in the calibration phase were validated in 500 feet. Additionally, errors in the assessment of the reference surface area of the foot pad and in the assessment of lesions were evaluated by the retrospective assessment of the pictures taken. In addition, possible impacts of climatic circumstances (temperature, humidity, and light intensity) were analyzed.

2.5. Statistical Methods

For all of the analyses, the statistical programming language R [29] was used. Prior to the experiment, we performed a power analysis to estimate the necessary sample size for assigning the macro score (categorized into three categories: 0, 1–2, 3–4) to the size of a lesion (categorized into three categories using two cut-off values). Multiple simulated data sets were generated assuming three different sets of assigned (conditional) probabilities (optimistic, neutral, and pessimistic). Multinomial regression models were used to analyze the simulated data sets in order to reveal the conditional probabilities. In this setting, power was defined as the probability that the true conditional probability lies within the estimated confidence interval. Given the optimistic and neutral conditional probability scenarios, it was estimated that 30 samples per macroscopic score would be sufficient for a power above 0.8. Given the pessimistic scenario, it was estimated that 100 samples would be necessary to reach this power. For this reason, we chose a sample size of 40 samples per macro score in the calibration phase. Another power analysis was conducted to estimate the required sample size for the validation of the scores. Here, different hypothetical sensitivity values (optimistic: 0.85–0.95, neutral: 0.70–0.80, pessimistic: 0.60–0.65) were considered. It was estimated that 45 samples were necessary to reach a sufficient power of estimation in an optimistic or a neutral scenario. Therefore, 500 samples were used, assuming an equal distribution among the macro scores.

To predict the conditional probabilities of macroscopic findings given specific camera scores, multinomial logistic regression models for categorical data were used. Following this analysis, the macro score categories with the highest probability according to the fitted models were used to determine the camera score cut-off values for macro score categories. To evaluate the performance for the classification of these cut-off values, performance measures were used as presented in Louton et al. [30]. Values close to 1.0 were in favor.

To measure the effect of software versions (original and updated) on the error assessment for reference surface areas and lesions. Similarly, multinomial logistic regression models for categorical data were used. The information of software versions was used as a predictor in the model, thus estimating the conditional probabilities (risk) of the different error types given the software version. To analyze the relationship between the two error types, another multinomial logistic regression model for categorical data was used. In this model, the conditional probabilities (risk) in the error assessment for reference surface areas were estimated given the error assessment for lesions.

Finally, to measure the effects of temperature, humidity, and light intensity on the error assessment for reference surface areas and lesions, further multinomial logistic regression models were used. The results are presented as estimated risks and their corresponding

95% uncertainty intervals. Comparisons are presented by risk ratios, relative risks, their corresponding 95% uncertainty intervals, and p -values.

3. Results and Discussion

The results of the inter-observer reliability were presented by Piller et al. [16]. The inter-observer reliability showed an average PABAK value of 0.88 with a range of 0.86–0.89, and thus represents an almost perfect inter-observer reliability [31].

3.1. Calibration of the Camera Scores, Threshold Values

Initially, the camera scores for the original (Figure 1A) and updated (Figure 1B) software are presented at given macro scores. As presented in the figures, both the original and the updated camera scores increased with an increasing macro score. The visual distinction, especially between the appearance of Macro Scores 0, 2, and 3 seems to be possible. Macro Score 4 was allocated to Macro Score 3 as described in the Materials and Methods Section, resulting in a macro score system ranging from 0 to 3 in further analysis. In the descriptive presentation, it is evident that the camera scores of the two combined macro scores were similar. Table 1 depicts a comparison of the visual and camera scores and the corresponding visual and camera pictures.

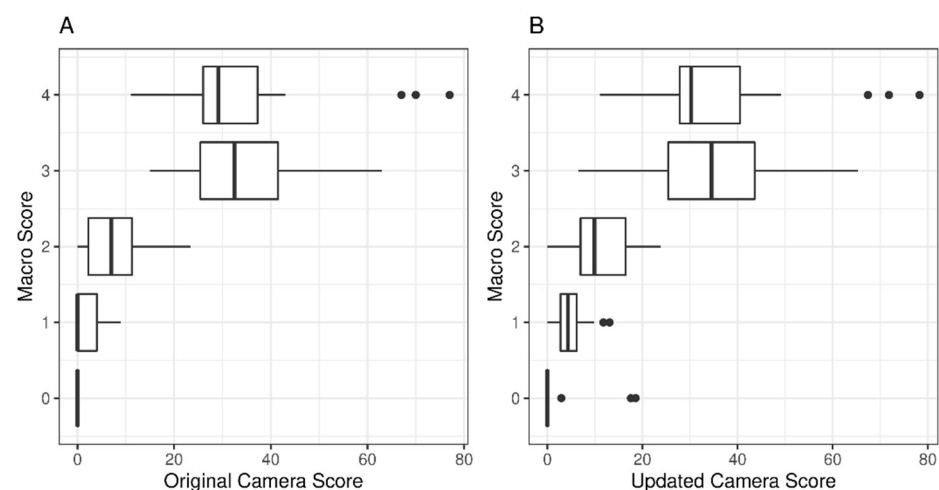




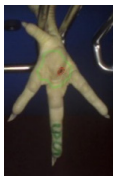





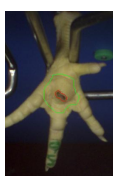
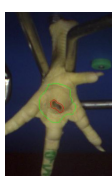


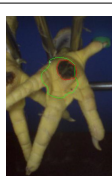

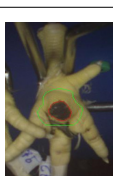



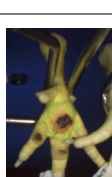


Figure 1. Presentation of the observed camera scores. (A) “Original” camera score; (B) “updated” camera score) at given macro scores (0–4) of foot pad dermatitis lesions (according to Piller et al. [16]) during the calibration phase ($n = 200$). Camera scores represent the percentage of the size of the alteration in relation to the reference surface area.

Table 1. Examples of assessed feet with the macro score system according to Piller et al. [16] and the respective camera scores and corresponding visual presentation. Macro Score 0 was defined as physiological skin without lesion, Macro Score 1 as single superficial lesion or several cumulated superficial lesions or deep lesion ≤ 0.5 cm, Macro Score 2 as superficial lesion >0.5 cm or deep lesion >0.5 to ≤ 1.0 cm, Macro Score 3 as deep lesion >1.0 cm, and Macro Score 4 as lesion on the foot pad and one or more deep lesions on the toe. Camera scores represent the percentage of the size of the alteration in relation to the reference surface area of the foot pad.

Macro Score	Visual Picture	Original Camera Score	Original Camera Picture	Updated Camera Score	Updated Camera Picture
0		0		0	
1		4.17		5.30	
1		3.98		5.44	
2		7.96		10.20	
3		29.67		29.01	
4		26.43		30.22	
4		23.63		24.48	

Considering the conditional probability, threshold values for camera scores were generated, at which the corresponding macro scores would have the highest probability. The threshold values for the original camera score were overall lower than the ones for the updated camera score (Table 2). Macro Score 1 was most probably at an original camera score of ≥ 0.01 and an updated camera score of ≥ 0.36 , and below these camera scores, Macro Score 0 had the highest probability. Macro Score 2 had a threshold value of ≥ 1.45 with the original camera score and of ≥ 4.01 with the updated camera score. Original camera scores of ≥ 12.66 and updated camera scores of ≥ 16.68 denoted Macro Score 3. These scores are lower than the values that are applied in slaughterhouses with the increments 0–5 (Score 0), 6–20 (Score 1), 21–50 (Score 2), and 51–100 (Score 3) in scoring the severity [15]. The European Commission [26] stated that a monitoring system for FPD is only effective by setting up threshold values (“trigger levels”). If these are set too low, more reports than predicted would be generated and thereby unnecessarily overload authorities. If threshold values are set too high, welfare-relevant lesions would not be reported. The estimated threshold values described above and mentioned in Table 2 could be adapted to the applied macroscopic scoring system.

Table 2. Camera score (“original” and “updated”) threshold values for the macro scores (0–3) of foot pad dermatitis lesions (according to Piller et al. [16]; Macro Scores 3 and 4 combined) during the calibration phase ($n = 200$).

Macro Score	Minimum	Maximum
Original Camera Score		
0	0.00	0.00
1	0.01	1.44
2	1.45	12.65
3	12.66	Infinite
Updated Camera Score		
0	0.00	0.35
1	0.36	4.00
2	4.01	16.67
3	16.68	Infinite

If the threshold values as described above were applied, the performance measures (sensitivity, specificity, positive predictive value, negative predictive value) for the given macro scores varied (Table 3). The sensitivity of diagnosis of 0.00 (original camera score) and 0.22 (updated camera score) indicates that Macro Score 1 is not diagnosable with the applied threshold values. If, for this reason, Macro Score 1 was allocated to Macro Score 0, the previous high sensitivity of diagnosis for Macro Score 0 dropped to 0.78 (previously 1.00), which still seems reasonable, and the accuracy of performance increased (Table 3). Macro Score 2 (FPD lesions >0.5 cm in diameter) was diagnosable with a sensitivity of 0.52 with the original and 0.70 with the updated camera software. In this case, the update of the camera version led to an improvement. Large foot pad lesions >1.0 cm in diameter were well diagnosed with a sensitivity of 0.99 (original) or 0.96 (updated) in the calibration phase. The current study evaluated the occurrence of FPD at slaughter. Therefore, the results can only be used to improve the health of subsequent flocks. Other authors evaluated FPD monitoring systems on-farm. Dawkins et al. [32] used a camera system on-farm to monitor the movements of broiler flocks. The authors stated that their evaluated camera-based optical flow system had more predictive power for the detection of FPD than the assessment of water consumption, bodyweight or cumulative mortality, and they recommended the monitoring system as a management tool. Others recommended a non-invasive measurement of dielectric constants of foot pads on-farm [33].

Table 3. Performance measures of predicted macro scores at given camera scores (for the original and the updated camera scores) for the evaluation of foot pad dermatitis in broilers during the calibration phase of the camera system ($n = 200$). The cut-off value was previously set at specific evaluated threshold camera scores. For the original camera score, Macro Score 1 was assigned to Macro Score 0 in one calculation. Sens: Sensitivity; Spec: Specificity; PPV: Positive predictive value; NPV: Negative predictive value.

Prediction ($n = 200$ feet)					Performance			
Original Camera Score								
	<0.01	0.01 to 1.44	1.45 to 12.65	≥ 12.66	Accuracy = 0.70			
Macro Score	0	1	2	3	Sens	Spec	PPV	NPV
0	40	0	0	0	1.00	0.80	0.56	1.00
1	22	0	18	0	0.00	1.00	-	0.80
2	10	0	21	9	0.52	0.88	0.52	0.88
3	0	0	1	79	0.99	0.92	0.90	0.99
	<1.45		1.45 to 12.65	≥ 12.66	Accuracy = 0.81			
Macro Score	0 + 1		2	3	Sens	Spec	PPV	NPV
0 + 1	62		18	0	0.78	0.92	0.86	0.86
2	10		21	9	0.52	0.88	0.52	0.88
3	0		1	79	0.99	0.92	0.90	0.99
Updated Camera Score								
	<0.36	0.36 to 4.00	4.01 to 16.67	≥ 16.68	Accuracy = 0.76			
Macro Score	0	1	2	3	Sens	Spec	PPV	NPV
0	37	1	0	2	0.92	0.94	0.79	0.98
1	9	9	22	0	0.22	0.98	0.75	0.84
2	1	2	28	9	0.70	0.84	0.53	0.92
3	0	0	3	77	0.96	0.91	0.88	0.97

3.2. Validation of Threshold Values

In the second phase of the project, the identified threshold values were validated with a sample size of 500 feet. Similar to the results of the calibration phase, the sensitivity of diagnosis of Macro Score 1 was 0.00 (original camera score) or 0.19 (updated camera score), indicating that Macro Score 1 was not diagnosable with the camera system, whereas the specificity was 1.00 (original camera score) and 0.98 (updated camera score) (Table 4). The sensitivity of diagnosis of Macro Score 2 was 0.48 (original camera score) and 0.72 (updated camera score) if the threshold values were set as mentioned in Table 2. The specificity of diagnosis of Macro Score 2 was 0.83 (original camera score) and 0.76 (updated camera score) if these threshold values were set. Macro Score 4 was allocated to Macro Score 3, and with the threshold values of 12.66 (original camera score) and 16.68 (updated camera score), these scores were well diagnosable with a sensitivity of 0.82 (original camera score) and 0.81 (updated camera score) and a specificity of 0.80 (original camera score) and 0.87 (updated camera score). Jung et al. [34] examined the automatic assessment of keel bone damage using a similar camera system as described in our study. After optimizations of the system, the authors reached a sensitivity of 95% and a specificity of 80% from the previously determined sensitivity of 28% and specificity of 66%. For the FPD Macro Scores 0 and 3, sensitivity values above 90% were reached in our study. Solely small foot pad lesions of ≤ 0.5 cm in diameter showed low detection performance measures. Louton et al. (in preparation) observed similar results for the automatic assessment of hock burn lesions. Similarly, small hock burn lesions were identified with a lower sensitivity than no lesions or

lesions >0.5 cm. Moreover, De Jong et al. [35] observed in an automatic assessment of foot pad lesions that the examined system showed very low agreement rates of their Visual Score 1 in a three-category (0–2) classification. These results are comparable to ours. Furthermore, Lund et al. [12] referred to the difficulty of inter-rater agreement if less severe lesions are evaluated. Interestingly, other authors observed for the automatic assessment systems of FPD a tendency that no camera scores could be generated for lesions with high macroscopic scores (as defined by experts on-farm or at slaughter) [36]. In another study, the visual assessment of foot pad lesions resulted in a low agreement between assessors, especially for Macro Score 2, and lesions were commonly underestimated. However, especially visual scoring of foot pad lesions of Macro Score 0 showed a high agreement between raters [37]. This finding is in agreement with the results of our automatic system, which identified no lesions (Macro Score 0) with the high agreement.

Table 4. Performance measures of predicted macro scores at given camera scores (for the original and the updated camera scores) for the evaluation of foot pad dermatitis in broilers during the validation phase of the camera system ($n = 500$). The cut-off value was previously set at specific evaluated threshold camera scores during the calibration phase. Sens: Sensitivity; Spec: Specificity; PPV: Positive predictive value; NPV: Negative predictive value.

Prediction ($n = 500$ feet)					Performance			
Original Camera Score								
	<0.01	0.01 to 1.44	1.45 to 12.65	≥ 12.66	Accuracy = 0.70			
Macro Score	0	1	2	3	Sens	Spec	PPV	NPV
0	82	0	3	0	0.96	0.81	0.51	0.99
1	43	0	40	3	0.00	1.00	-	0.83
2	18	0	77	64	0.48	0.83	0.57	0.78
3	17	0	14	139	0.82	0.80	0.67	0.89
Updated Camera Score								
	<0.36	0.36 to 4.00	4.01 to 16.67	≥ 16.68	Accuracy = 0.76			
Macro Score	0	1	2	3	Sens	Spec	PPV	NPV
0	74	4	7	0	0.87	0.93	0.73	0.97
1	19	16	50	1	0.19	0.98	0.70	0.85
2	3	1	114	41	0.72	0.76	0.58	0.85
3	5	2	25	138	0.81	0.87	0.77	0.90

3.3. Errors in Assessments of Reference Surface Area of Foot Pads and Lesions

Figure 2 depicts the errors in the assessment of the reference surface area of the foot pad with the original and the updated software in the validation phase. The results show that the updated software led to an improvement in the correct identification of the reference surface area of the foot pad (risk original software = 0.66 [0.62–0.70]; risk updated software = 0.74 [0.70–0.78]; relative risk [RR] original vs. updated software = 0.890 [0.818–0.969; $p = 0.008$). In addition, to the less shifted reference surface areas (risk original software = 0.23 [0.19–0.26]; risk updated software = 0.17 [0.14–0.20]; RR original vs. updated software = 1.334 [1.039–1.723]; $p = 0.021$) compared with the original software. Only on rare occasions, the reference surface area of the foot pad was assessed as “too large” or “completely wrong”.

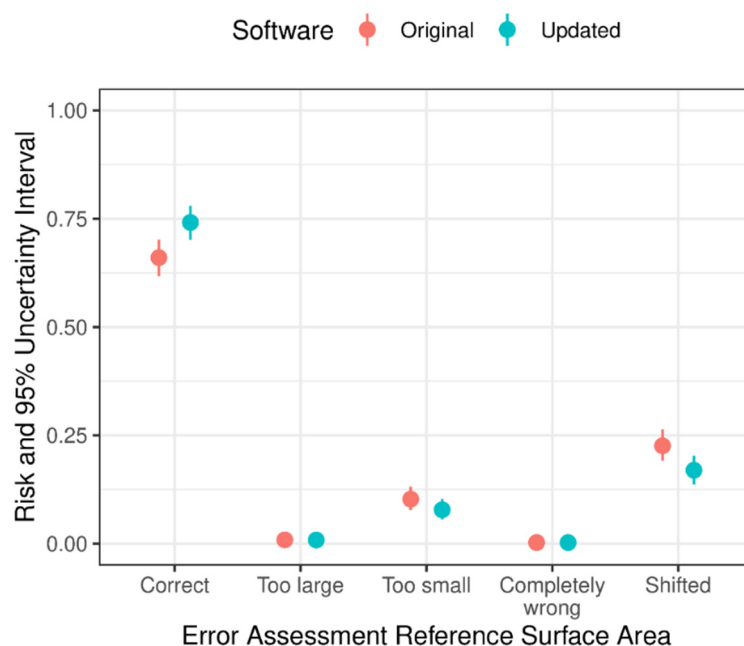


Figure 2. Estimated probabilities of occurrence of errors in the camera-based assessment of reference ranges of the foot pad surface area with the original and the updated camera software ($n = 500$; validation phase).

Considering the identification of lesions, the software versions differed in their potential of correct identification (Figure 3). The updated software differed in nearly all of the types of faulty identification from the original software and showed an improvement in most of the types. The probability of a correct identification of lesions was significantly higher if the updated software was used than if the original software was used (risk original software = 0.74 [0.70–0.77]; risk updated software = 0.80 [0.77–0.84]; RR original vs. updated software = 0.914 [0.854–0.978]; $p = 0.01$). The original software had a significantly lower risk of identifying lesions as too large than the updated software (risk original software = 0.01 [0.00–0.02]; risk updated software = 0.04 [0.03–0.06]; RR original vs. updated software = 0.132 [0.034–0.377]; $p < 0.001$). The updated software showed an improvement concerning not identifying the lesions at all (risk original software = 0.17 [0.14–0.20]; risk updated software = 0.08 [0.06–0.11]; RR original vs. updated software = 2.093 [1.487–3.045]; $p < 0.001$) or identifying them as too small, although the latter improvement was not statistically significant (risk original software = 0.09 [0.07–0.12]; risk updated software = 0.07 [0.05–0.09]; RR original vs. updated software = 1.289 [0.845–1.954]; $p = 0.224$), compared with the original software. The differences were only marginal, and the accurate estimation was probably caused by the large sample size. The results are in line with a study by Jung et al. [34], who found that optimizations of algorithms led to an improvement in the detection of keel bone damage in automatic assessments. The authors observed that the camera system generally underestimated the presence of keel bone damage. In our study, considering the detection of FPD lesions, especially if the updated software version was used, the risk of a faulty detection of FPD lesions was lower than the chance of a correct identification, which had the highest probability. However, approximately 20% of the lesions were not identified correctly. This aspect should be improved, especially if the systems are used by authorities for monitoring poultry welfare.

Furthermore, we assessed the association between the errors in the assessment of the reference surface area of the foot pad and the errors in the assessment of lesions (Figures 4 and 5). The probability of the correct identification of lesions was highest, even if the assessment of the reference surface area of the foot pad was incorrect, both with the original and with the updated camera software.

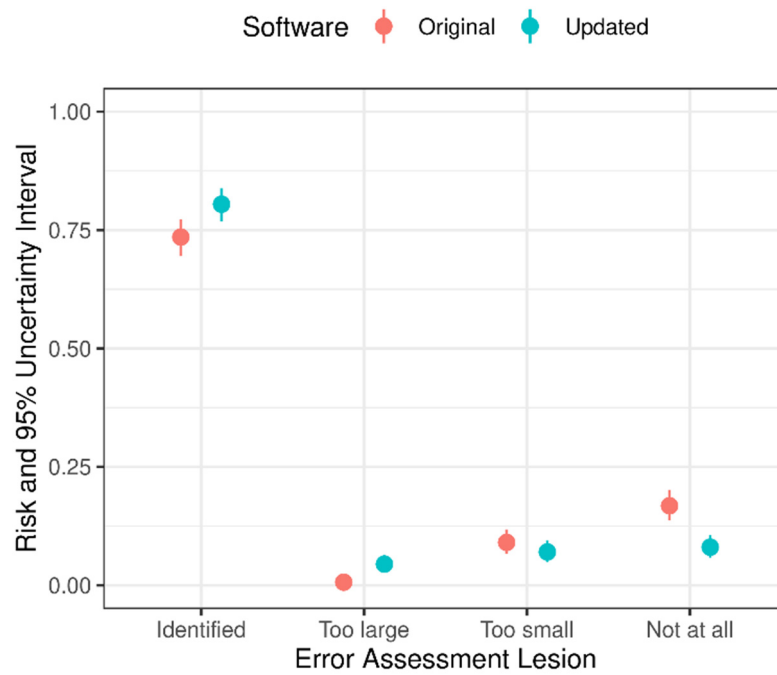


Figure 3. Estimated probabilities of occurrence of errors in the camera-based assessment of foot pad dermatitis lesions with the original and the updated camera software ($n = 500$; validation phase).

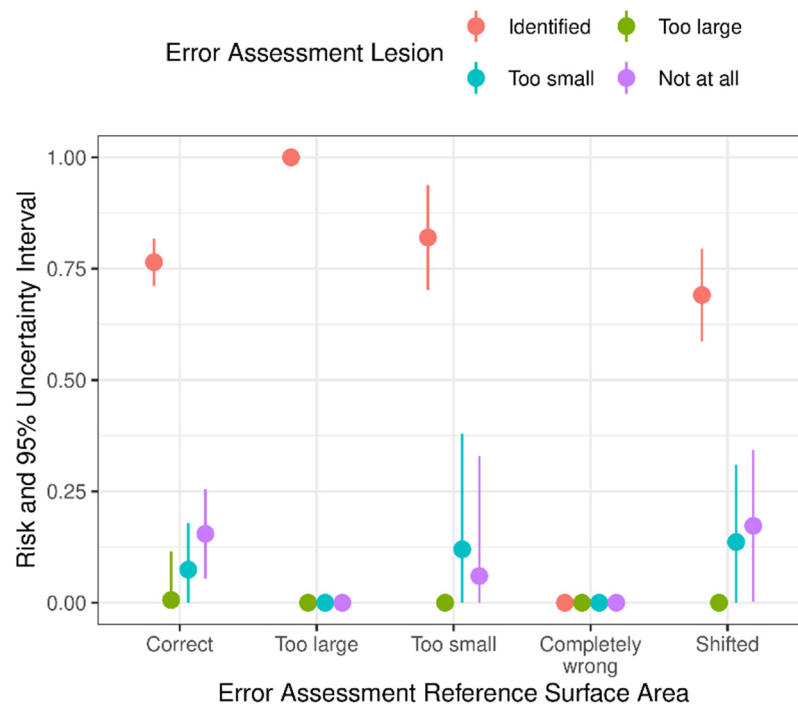


Figure 4. Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of foot pad dermatitis lesions (identified, too large, too small, not at all) by the camera system with the original software and the association to the faulty detection of the reference ranges of the foot pad surface area (correct, too large, too small, completely wrong, shifted) ($n = 500$; validation phase).

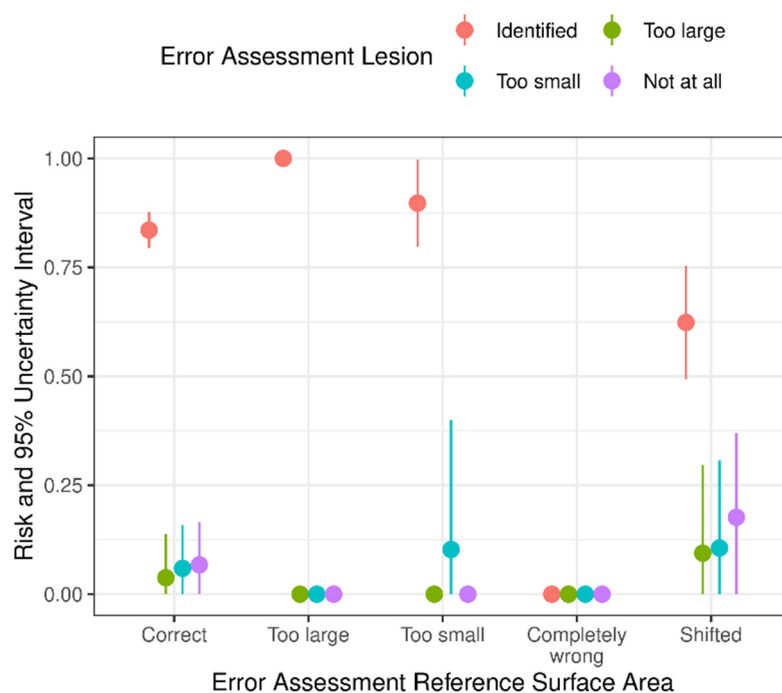


Figure 5. Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of foot pad dermatitis lesions (identified, too large, too small, not at all) by the camera system with the updated software and the association to the faulty detection of the reference ranges of the foot pad surface area (correct, too large, too small, completely wrong, shifted) ($n = 500$; validation phase).

In a study by Vanderhasselt et al. [36], 15.2% of the feet were not recognized by the systems. In addition, in 49.4% of the broilers' feet, the systems identified a lesion erroneously even if a lesion was not present. Compared with their results, the system presented in this study demonstrates a major improvement. The authors stated that the correlation between the automatic score and the expert score was improved if only data without a faulty assessment of lesions were used. As a possible explanation for this discrepancy, they pointed out that the automatic systems do not consider the depth of the lesion. However, according to Vanderhasselt et al. [36] and Heitmann et al. [13], the size of lesions correlates to their depth. Therefore, one could assume that an automatic system could identify the macro score, given a macro score that considers the size of the lesion. This aspect underlines the importance of the choice of macro score and the adaptation of software systems and algorithms to the chosen macro score. The system used in the present study assessed the size of the lesion in relation to the reference surface area of the foot pad. However, the examined software could be adapted, and the size of the lesion (e.g., 0.5 cm or 1.0 cm) rather than the percentage of a camera score could be set as the threshold.

De Jong et al. [24] defined several important criteria that should be fulfilled by the camera system. First, more than 70% of the feet should be scored by the system. With the setup of our study, we cannot draw conclusions on the overall scoring of feet on a flock basis. In our study, with the use of the updated software, the chance for a correct identification of the reference surface area of the foot pad was 74% and the chance for the correct identification of lesions was 80%. Furthermore, even if the reference surface area of the foot pad was assessed as too small, the chance of a correct identification of lesions had the highest probability. Our study does not allow the conclusion that 70% of the feet were scored, since we did not assess the feet on the flock level. Therefore, our study is not directly comparable. However, considering the identification of the reference surface area and lesions, this criterion seems to be fulfilled by the automatic system in our study. Recent numbers that were registered by the slaughterhouse in which the study

was performed show that in 2017 (98.4%), in 2018 (87.0%), in 2019 (53.4%), and in 2021 (January to October 97.9%) of all of the feet that passed the camera system were scored (51 million feet passed the camera system in 2017, 48 million in 2018, 51 million in 2019, and 42 million in 2021). The low percentage of scored feet in the year 2019 could be due to a technical issue in which an additional function was tested, which caused a high sensitivity of only the scoring feet that hung correctly in the shackles. At that time, it was considered most important that only the feet that hung correctly in the shackle would be assessed. With a new software, the scoring rate was improved in 2021. Overall, the system seems to comply with the mentioned criterion of scoring 70% of the feet if no serious disruptions occur. Nonetheless, the overall scoring rate is an important aspect to evaluate the general functioning of the system, although it does not coercively affect the quality of the results in scoring the reference surface area or the lesions. The other criterion mentioned by de Jong et al. [24] was that the agreement of the generated camera score with a golden standard should be at least 75%. In our study, this agreement was reached by the updated camera score, depending on the macro score. Regarding Macro Score 0 (no lesion) and Macro Score 3 (lesions >1.0 cm in diameter), the agreement was above 75%. However, especially small lesions ≤ 0.5 cm were not identified with sufficiently high certainty.

The statistical methods we used considered the evaluation of foot pad lesions of individual feet. This method has the advantage that visual camera-based pictures can be allocated exactly to the manual visual assessment of the respective foot. Other previously published studies evaluated automatic assessment systems that assessed the health of the feet on a flock basis, which poses the risk of incorrect allocation of individual feet. The advantageous or superior functions of the assessed camera systems are that by adaptations of the software, major improvements of the general scoring and the precision of scoring are possible. Furthermore, if needed, the software could be adapted to changes in politics and regulations, for example, if the absolute size of the lesion should be scored rather than the percentage in relation to a reference area.

3.4. Effects of Temperature, Humidity, and Light Intensity on Errors

The temperature ranged from 24.7 to 27.3 °C, the relative humidity from 60.0% to 78.7%, and the light intensity from 5200 to 7820 lx in the calibration phase. In the validation phase, the temperature ranged from 23.5 to 29.9 °C, the relative humidity from 44.9% to 94.9%, and the light intensity from 3723 to 6880 lx. The effects of temperature, humidity, and light intensity on the errors in the assessment of the reference surface area of the foot pad and on the errors in the assessment of lesions were evaluated. These climatic effects are presented in the Supplementary Material, in Figures S1 and S3–S5 for the reference surface area of the foot pad and in Figures S2 and S6–S8 for the errors in the assessment of lesions. The possible effects were rather small. The correct assessment of the reference surface area of the foot pad decreased slightly (in both software versions) with the increasing temperature and humidity. In terms of the assessment of foot pad lesions, the associations with climatic circumstances were similarly very low, and the effects were small. Increasing the humidity or light intensity increased the possibility of identifying the lesions as too small.

4. Conclusions

The automatic assessment of FPD is a hot topic. In our study, threshold values of a camera score to identify different categories of macroscopic scores were defined in a first step. Next, the performance measures of diagnosis of these macro scores were analyzed in a validation phase. In particular, the macro scores of FPD of 0 (no lesion), 2 (lesion >0.5 cm in diameter), and 3 (lesion >1.0 cm) could be identified with a sufficiently high sensitivity. The sensitivity of the diagnosis of Macro Score 1 (lesions >0 cm to 0.5 cm) was not high enough in both examined software versions (original and updated). Furthermore, we showed that the software updates can enhance the performance measures and lower the probability of errors. The automatic assessment system does not seem particularly

susceptible to influences of temperature, relative humidity or light intensity, which is an important issue for its operation in slaughterhouses. An individual assessment of feet, as performed in our study, allows an exact allocation of macro scores to the respective camera scores. However, this procedure does not allow an assessment on a flock basis. Therefore, one limitation of our study is that we cannot make a statement regarding the precision of assessment on a flock basis. However, under conditions of practice, it is not possible to assess all of the feet of one flock (e.g., 30,000 broilers) and individually allocate the macroscopic scores. For the calibration and validation of the system, this individual assessment seemed most important, and this aspect was given priority. In future studies, the overall functionality under conditions of practice at slaughter should be proven. One further limitation is that our study design does not allow any statement regarding the perception of pain in different stages of FPD. To overcome this limitation, threshold values of predefined scoring systems need to be used.

Supplementary Materials: The following materials are available online at <https://www.mdpi.com/article/10.3390/agriculture12020221/s1>. Figure S1: Possible false detection of the reference surface area of the foot pad (correct, too large, too small, shifted) with the original and the updated software depending on the temperature in degrees Celsius (left), humidity in percent (middle), and light intensity in lux (right) ($n = 500$; validation phase); Figure S2: Possible false detection of foot pad lesions (identified, too large, too small, not at all) with the original and the updated software depending on the temperature in degrees Celsius (left), humidity in percent (middle), and light intensity in lux (right) ($n = 500$; validation phase); Figure S3: Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of the reference surface area of the foot pad (correct, shifted, too large, too small) with the original and the updated software depending on the temperature in degrees Celsius ($n = 500$; validation phase); Figure S4: Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of the reference surface area of the foot pad (correct, shifted, too large, too small) with the original and the updated software depending on the humidity in percent ($n = 500$; validation phase); Figure S5: Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of the reference surface area of the foot pad (correct, shifted, too large, too small) with the original and the updated software depending on the light intensity in lux ($n = 500$; validation phase); Figure S6: Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of foot pad lesions (identified, not at all, too large, too small) with the original and the updated software depending on the temperature in degrees Celsius ($n = 500$; validation phase); Figure S7: Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of foot pad lesions (identified, not at all, too large, too small) with the original and the updated software depending on the humidity in percent ($n = 500$; validation phase); Figure S8: Estimated risk and uncertainty interval (95%) of the occurrence of errors during the assessment of foot pad lesions (identified, not at all, too large, too small) with the original and the updated software depending on the light intensity in lux ($n = 500$; validation phase).

Author Contributions: Conceptualization, H.L., A.S., M.E. and S.B.; methodology, H.L., A.S. and S.B.; software, J.S.-L.; validation, H.L. and A.P.; formal analysis, P.S.; investigation, H.L., A.P., B.S. and J.S.; resources, M.E.; data curation, P.S.; writing—original draft preparation, H.L.; writing—review and editing, A.S.; visualization, P.S.; supervision, M.E.; project administration, B.S.; funding acquisition, B.S., H.L. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision from the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program (grant number: FKZ 2817903715).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki. Any aspect of the work covered in this manuscript that has involved animals has been conducted according to the international and national guidelines for humane animal treatment, in compliance with relevant legislation and with the high ethical standards concerning animal welfare of the IAVE (International Association of Veterinary Editors) Guidelines. All of the animals used were housed according to the German Order on the Protection of Animals and the Keeping of Production Animals prior to slaughter. Sampling procedures were in accordance with

the German Animal Welfare Act, the regulations on the welfare of animals used for experiments or for other scientific purposes, and national and international regulations on the welfare of animals at slaughter (Council Regulation [EC] No. 1099/2009, of 24 September 2009, on the protection of animals at the time of killing). Ethical review and approval were waived for this study since the feet of the broilers (samples for the research) were taken after slaughter from dead animals that were primarily rejected for use in the food chain for other reasons. All of the mandatory laboratory health and safety procedures were complied with during the sampling procedures.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed are included in this article.

Acknowledgments: The authors thank Josef Bachmeier for the support and organization of the study. We thank Verena Lietze for the scientific language editing. We thank Stefan and Yeliz Wiesbeck for their valuable and dedicated support during data collection. We acknowledge financial support by Deutsche Forschungsgemeinschaft and Universität Rostock within the funding programme Open Access Publishing.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Saraiva, S.; Saraiva, C.; Stilwell, G. Feather conditions and clinical scores as indicators of broilers welfare at the slaughterhouse. *Res. Vet. Sci.* **2016**, *107*, 75–79. [[CrossRef](#)] [[PubMed](#)]
2. Shepherd, E.M.; Fairchild, B.D. Footpad dermatitis in poultry. *Poult. Sci.* **2010**, *89*, 2043–2051. [[CrossRef](#)] [[PubMed](#)]
3. Algers, B.; Berg, C. Monitoring animal welfare on commercial broiler farms in Sweden. *Acta Agric. Scand. A Anim. Sci.* **2001**, *51*, 88–92. [[CrossRef](#)]
4. Martland, M.F. Ulcerative dermatitis in broiler chickens: The effects of wet litter. *Avian Pathol.* **1985**, *14*, 353–364. [[CrossRef](#)]
5. Kyvsgaard, N.C.; Jensen, H.B.; Ambrosen, T.; Toft, N. Temporal changes and risk factors for foot-pad dermatitis in Danish broilers. *Poult. Sci.* **2013**, *92*, 26–32. [[CrossRef](#)]
6. Berk, J. Fußballendermatitis bei männlichen Broilern in Abhängigkeit von unterschiedlichen Einstreuarten. *Landbauforsch. Völkenrode* **2007**, *57*, 171–178.
7. Bilgili, S.F.; Hess, J.B.; Blake, J.P.; Macklin, K.S.; Saenmahayak, B.; Sibley, J.L. Influence of bedding material on footpad dermatitis in broiler chickens. *J. Appl. Poult. Res.* **2009**, *18*, 583–589. [[CrossRef](#)]
8. McKeegan, D. *Foot Pad Dermatitis and Hock Burn in Broilers: Risk Factors, Aetiology and Welfare Consequences*; Research Project Final Report; Faculty of Veterinary Medicine, University of Glasgow: Scotland, UK, 2010.
9. De Jong, I.C.; Gunnink, H.; van Harn, J. Wet litter not only induces footpad dermatitis but also reduces overall welfare, technical performance, and carcass yield in broiler chickens. *J. Appl. Poult. Res.* **2014**, *23*, 51–58. [[CrossRef](#)]
10. Welfare Quality®. *Welfare Quality® Assessment Protocol for Poultry (Broilers, Laying Hens)*; Welfare Quality® Consortium: Lelystad, The Netherlands, 2009.
11. Ekstrand, C.; Carpenter, T.E.; Andersson, I.; Algers, B. Prevalence and control of foot-pad dermatitis in broilers in Sweden. *Br. Poult. Sci.* **1998**, *39*, 318–324. [[CrossRef](#)]
12. Lund, V.P.; Nielsen, L.R.; Oliveira, A.R.S.; Christensen, J.P. Evaluation of the Danish footpad lesion surveillance in conventional and organic broilers: Misclassification of scoring. *Poult. Sci.* **2017**, *96*, 2018–2028. [[CrossRef](#)]
13. Heitmann, S.; Stracke, J.; Petersen, H.; Spindler, B.; Kemper, N. First approach validating a scoring system for foot-pad dermatitis in broiler chickens developed for application in practice. *Prev. Vet. Med.* **2018**, *154*, 63–70. [[CrossRef](#)] [[PubMed](#)]
14. Riber, A.B.; Rangstrup-Christensen, L.; Hansen, M.S.; Hinrichsen, L.K.; Herskin, M.S. Characterisation of footpad lesions in organic and conventional broilers. *Animal* **2020**, *14*, 119–128. [[CrossRef](#)] [[PubMed](#)]
15. Louton, H.; Erhard, M.; Wirsch, K.; Bergmann, S.; Piller, A.; Schmidt, P.; Rauch, E. Comparison of four assessment methods of foot pad dermatitis and hock burn of broilers. *Berl. Münch. Tierärztl. Wochenschr.* **2020**, *133*, 1–11. [[CrossRef](#)]
16. Piller, A.; Bergmann, S.; Schwarzer, A.; Erhard, M.; Stracke, J.; Spindler, B.; Kemper, N.; Schmidt, P.; Bachmeier, J.; Schade, B.; et al. Validation of histological and visual scoring systems for foot-pad dermatitis in broiler chickens. *Anim. Welf.* **2020**, *29*, 185–196. [[CrossRef](#)]
17. Michel, V.; Prampart, E.; Mirabito, L.; Allain, V.; Arnould, C.; Huonnic, D.; Le Bouquin, S.; Albaric, C. Histologically-validated footpad dermatitis scoring system for use in chicken processing plants. *Br. Poult. Sci.* **2012**, *53*, 275–281. [[CrossRef](#)] [[PubMed](#)]
18. *German Order on the Protection of Animals and the Keeping of Production Animals*; Amended and Promulgated on 22 August 2006, Last Changed on 5 February 2014; Bundesministerium der Justiz: Berlin, Germany, 2014.
19. Implementary Rules of the Lower Saxonian Ministry of Nutrition. Agriculture and consumer protection for §20 (4; 5). Latest Edition: 31 July 2015, from the German Order on the Protection of Animals and the Keeping of Production Animals (2006). Avail-

- able online: <https://www.voris.niedersachsen.de/jportal/?quelle=jlink&psml=bsvorisprod.psml&feed=bsvoris-vv&docid=VVND-VVND000041583#ivz5> (accessed on 16 October 2021).
20. Bessei, W. Welfare of broilers: A review. *World Poult. Sci.* **2006**, *62*, 455–466. [[CrossRef](#)]
 21. Allain, V.; Mirabito, L.; Arnould, C.; Colas, M.; Le Bouquin, S.; Lupo, C.; Michel, V. Skin lesions in broiler chickens measured at the slaughterhouse: Relationships between lesions and between their prevalence and rearing factors. *Br. Poult. Sci.* **2009**, *50*, 407–417. [[CrossRef](#)]
 22. Louton, H.; Erhard, M.; Wöhr, A.C. Acquisition of animal based welfare measures at slaughter of poultry. *Fleischwirtschaft* **2018**, *11*, 94–98.
 23. Leitfaden. *Befunddaten in der Geflügelschlachtung. Version 01.01.2022*; QS Fachgesellschaft Geflügel GmbH: Bonn, Germany, 2022.
 24. De Jong, I.; Reimert, H.; Vanderhasselt, R.; Gerritzen, M.; Gunnink, H.; Hindle, V.; Lourens, S. *Development of Methods to Monitor Foot Pad Lesions in Broiler Chickens*; Wageningen UR Livestock Research: Lelystad, The Netherlands, 2011.
 25. De Jong, I.C.; van Harn, J.; Gunnink, H.; Lourens, A.; van Riel, J.W. Measuring foot-pad lesions in commercial broiler houses. Some aspects of methodology. *Anim. Welf.* **2012**, *21*, 325–330. [[CrossRef](#)]
 26. European Commission Directorate-General for Health and Food Safety. *Use of Slaughterhouse Data to Monitor Welfare of Broilers on Farm*; Luxembourg Publications Office of the European Union: Luxembourg, 2016.
 27. Ben Sassi, N.; Averós, X.; Estevez, I. Technology and poultry welfare. *Animals* **2016**, *6*, 62. [[CrossRef](#)]
 28. Byrt, T.; Bishop, J.; Carlin, J.B. Bias, prevalence and kappa. *J. Clin. Epidemiol.* **1993**, *46*, 423–429. [[CrossRef](#)]
 29. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: <https://www.R-project.org> (accessed on 5 January 2022).
 30. Louton, H.; Piller, A.; Bergmann, S.; Erhard, M.; Stracke, J.; Spindler, B.; Kemper, N.; Schmidt, P.; Schade, B.; Boehm, B.; et al. Histologically validated scoring system for the assessment of hock burn in broilers. *Avian Pathol.* **2020**, *49*, 230–242. [[CrossRef](#)] [[PubMed](#)]
 31. Landis, R.J.; Koch, G.G. The M measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]
 32. Dawkins, M.S.; Roberts, S.J.; Cain, R.J.; Nickson, T.; Donnelly, C.A. Early warning of footpad dermatitis and hockburn in broiler chicken flocks using optical flow, bodyweight and water consumption. *Vet. Rec.* **2017**, *180*, 499. [[CrossRef](#)] [[PubMed](#)]
 33. Hoffmann, G.; Ammon, C.; Volkamer, L.; Sürle, C.; Radko, D. Sensor-based monitoring of the prevalence and severity of foot pad dermatitis in broiler chickens. *Br. Poult. Sci.* **2013**, *54*, 553–561. [[CrossRef](#)] [[PubMed](#)]
 34. Jung, L.; Nasirahmadi, A.; Schulte-Landwehr, J.; Knierim, U. Automatic assessment of keel bone damage in laying hens at the slaughter line. *Animals* **2021**, *11*, 163. [[CrossRef](#)]
 35. De Jong, I.C.; Gerritzen, M.; Reimert, H.; Fritsma, E.; Pieterse, C. Automated measurement of foot pad lesions in broiler chickens. In Proceedings of the 4th International Workshop on the Assessment of Animal Welfare at Farm & Group Level, Ghent, Belgium, 10–13 September 2008; p. 32.
 36. Vanderhasselt, R.F.; Sprenger, M.; Duchateau, L.; Tuytens, F.A.M. Automated assessment of footpad dermatitis in broiler chickens at the slaughter-line: Evaluation and correspondence with human expert scores. *Poult. Sci.* **2013**, *92*, 12–18. [[CrossRef](#)]
 37. Oliveira, A.R.S.; Lund, V.P.; Christensen, J.P.; Nielsen, L.R. Inter-rater agreement in visual assessment of footpad dermatitis in Danish broiler chickens. *Br. Poult. Sci.* **2017**, *58*, 224–229. [[CrossRef](#)]