Research Note: Is infrared thermography an appropriate method for early detection and objective quantification of plumage damage in white and brown feathered laying hens?

Ruben Schreiter *,[1] and Markus Freick *,†

*ZAFT e.V., Centre for Applied Research and Technology, Dresden, Germany; and †HTW Dresden – University of Applied Sciences, Dresden, Germany

ABSTRACT For the standardized assessment of plumage damage in laying hens, imaging techniques can be used in addition to visual plumage scoring (**PS**). In this study, the diagnostic accuracy of infrared thermography (**IRT**) was analyzed in white-feathered (**WL**) and brown-feathered laying hens (**BL**) with PS as a reference. In 28 flocks, a 3-level PS and IRT were performed 8 times for the dorsal neck, back, and belly plumage. A total of 3,600 records for WL and 7,600 records for BL were available for both PS and IRT in each of the 3 body regions. Receiver operating characteristic (**ROC**) analyses were used to investigate diagnostic accuracy. High-accuracy detection was found for severe plumage damage on the back (WL—sensitivity: 96.0%, specificity: 98.3%; BL—sensitivity: 96.1%, specificity: 98.9%) and belly plumage (WL—sensitivity: 96.3%, specificity: 95.7%; BL—sensitivity: 95.3%, specificity: 97.2%), but insufficient accuracy for distinguishing between intact plumage and moderate plumage damage in all 3 body regions (sensitivity: 31.7−71.5%; specificity: 70.4−98.1%). The area under the curve (**AUC**) of the ROC graphs differed significantly between BL and WL for the back and belly plumage ($P \leq 0.05$). We concluded that IRT is a suitable tool to objectively detect severe plumage damage but not for early detection of incipient plumage loss.

Key words: plumage loss, feather pecking, ROC analysis

2022 Poultry Science 101:102022
https://doi.org/10.1016/j.psj.2022.102022

INTRODUCTION

Severe feather pecking (**SFP**) is the most important multifactorial behavioral disorder in laying hens that has detrimental effects on animal welfare, biological performance, and economic success (Spindler et al., 2016). SFP can be detected indirectly by plumage assessment (Welfare Quality®, 2009).

Plumage scoring (**PS**) assesses plumage condition and is usually performed by observers using a defined scheme to assign the completeness of a hen's plumage to a score reflecting the extent of plumage loss (Welfare Quality®, 2009). The weaknesses in this visual assessment result from the subjective bias of the observer and associated variability depending on the examiner's qualifications, experience, availability (time capacity), and motivation (Döhring et al., 2020). Less time required for PS procedures would also allow them to be performed more frequently in practice. Therefore, there is an increasing demand for imaging diagnostics to assess plumage conditions objectively and quantifiably. Infrared thermography (**IRT**) is used to detect the temperature of plumage with a thermal imaging camera and identify featherless areas based on higher temperatures (Zhao et al., 2013). Cook et al. (2006) showed in brown-feathered brown-egg laying hens' (**BL**) that IRT temperatures differ in a corresponding way with differing PS. Zhao et al. (2013) determined characteristic temperature ranges for individual scores in white-feathered white-egg laying hens (**WL**). In further research on WL by Pichova et al. (2017), these results were supplemented by parameters to identify featherless areas by calculating the proportion of pixels with a temperature above the threshold of 33.5°C.

Since there is no evidence-based knowledge on the diagnostic accuracy of IRT for plumage assessment in the field, our study aimed to determine the sensitivity and specificity for this technique and evaluate its suitability as a method for objective quantification and early detection of plumage loss. Furthermore, we hypothesized potential differences regarding the most appropriate IRT image parameters between BL and WL.

Received February 26, 2022.
Accepted June 14, 2022.

[1]Corresponding author: ruben.schreiter@htw-dresden.de

MATERIALS AND METHODS

Animals and Study Design

Animals from 28 flocks (n = 9 WL flocks and n = 19 BL flocks) on 11 laying hen farms in Saxony/Germany were included in this field study. Hens were kept in a barn (n = 21 flocks) or free-ranging systems (n = 7 flocks) with a median flock size of 12,357 hens (1.-3. quartile: 5,671−16,996) according to the legal requirements of the EU and Germany. Fifty hens per flock and date were randomly selected eight times in the period from placement and the subsequent 50 wk of life to assess plumage condition by PS and IRT. Each individual's, dorsal neck, back, and belly body regions were assessed because these areas are particularly suitable for indirectly quantifying the occurrence of SFP (Bilcik and Keeling, 1999). In total, 3,600 WL and 7,600 BL PS and IRT records for each of the 3 body regions were available.

Plumage Scoring and Infrared Thermography

PS was performed by the same observer using a 3-score system (0 = intact plumage, 1 = moderate plumage damage with one or more featherless areas ≤5 cm; 2 = severe plumage damage with one or more featherless areas >5 cm) according to Welfare Quality® (2009). Intraobserver reliability was assessed at 3 time points during the study, each with 50 hens.

IRT images of the hen's surface were photographed using the FLIR E5 thermal imaging camera (FLIR Systems Inc., Wilsonville, OR) with the approach and camera settings described by Zhao et al. (2013) (emissivity 0.98, humidity 60%, temperature 22°C, resolution 360 × 240 px). At a horizontal distance of 0.8 m from the camera, one image of each hen's dorsal neck, back, and belly was photographed in front of a white wall without solar reflection. The hens were fixed upside down (neck, back; legs and tail fixed in hand) or backward (belly; legs down fixed in hand, breast resting on other hand) by a trained staff member during imaging. The actual room temperature was recorded using a thermohygrometer (Klima Logger TFA; TFA-Dostmann, Wertheim, DE).In the IR images, a surface temperature for each pixel of the recorded animal was stored and analyzed using camera-specific software (FLIR −Tools + ResearchIR−Standard, FLIR Systems Inc.). For this purpose, freehand software tools were used to mark the relevant image area according to the regions defined at PS, and the temperature values of the pixels were exported to a Microsoft Excel spreadsheet (2013 version, Microsoft Corporation, Redmond, WA). The following parameters were calculated separately for each animal and body region according to Pichova et al. (2017): 1) Arithmetic mean of the temperature for all pixels (mean temperature, **MT**); 2) temperature difference between MT and the recorded ambient temperature (**ΔT**), and 3) the relative proportion of featherless areas (**FL%**), where FL% represents the percentage of pixels above a defined temperature threshold out of the total number of pixels for the body region. Based on the findings of Pichova et al. (2017), temperatures of 32.0°C, 32.5°C, 33.0°C, 33.5°C, 34.0°C, and 34.5°C were used as thresholds for defining a featherless region. Several temperature thresholds were used to assess their suitability in terms of diagnostic performance.

Statistical Analyses

Microsoft Excel (version 2013, Microsoft Corporation) was used for data collection, processing, and creating selected diagrams. SAS 9.4 for Windows (SAS Institute Inc., Cary, NC) was used to analyze the data separately by hybrid type (BL vs. WL) and identify potential differences in the diagnostic value of IRT between BL and WL.

A concordance analysis was performed to quantify the degree of agreement in integument scores. For this purpose, the prevalence-adjusted and bias-adjusted kappa (**PABAK**) was calculated as a characteristic of the intraobserver reliability according to Gunnarsson et al. (2000).

The Kolmogorov−Smirnov test was used to check the normal distribution of the data. For MT and ΔT, which were normally distributed, an F-test was performed to compare the IRT values between the plumage scores. FL% data were not normally distributed; therefore, the Kruskal−Wallis test was used to compare the plumage scores. Correlation coefficients for MT, ΔT, and Fl% to PS were calculated using Spearman's rank correlation.

Receiver operating characteristic (**ROC**) analyses were performed for the FL% of 6 different thresholds, MT, and ΔT as metrically scaled outcomes and the corresponding plumage scores' nominally scaled outcome to investigate the diagnostic value of IRT. Therefore, plumage scores 0 vs. 1, 0 vs. 2, 1 vs. 2, and 0/1 vs. 2 were compared using different approaches. The area under the curves (**AUC**) of the ROC graphs was calculated to measure test goodness. The Youden Index determined the cut-offs for an optimal ratio of sensitivity and specificity. Testing for differences in the AUC of each parameter between hybrid types was performed using the Mann−Whitney U test (du Prel et al., 2010).

Despite multiple testing of the data, we established no correction for the family-wise error rate because this study was considered an explorative approach (Victor et al., 2010). *P*-values of <0.05 were considered as remarkably low.

RESULTS

Plumage scoring for the flocks was performed at 17.6 ± 0.8, 23.3 ± 1.5, 28.1 ± 1.5, 32.5 ± 1.7, 41.2 ± 1.3, 50.1 ± 2.2, 58.8 ± 2.0, and 66.1 ± 2.7 wk of age (mean ± standard deviation [SD]). PABAK values of 0.90 for dorsal neck plumage, 0.94 for back plumage, and 0.91 for belly plumage indicated very good intraobserver

Table 1. Best-fit parameters of infrared thermography images to assess plumage damage in laying hens identified by receiver operating characteristic (ROC) analyses with the associated area under the curve (AUC), cutpoint at maximum Youden index, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV).

Body region/hybrid type	Best-fit parameter	AUC (SE)	Cutpoint	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
Dorsal neck plumage							
Brown-feathered hens							
score 0 vs. 1	ΔT	0.6863 (0.0062)	4.557	60.2	70.4	75.6	53.6
score 1 vs. 2	FL% >32.0°C	0.8846 (0.0101)	16.541	80.2	97.1	96.6	82.6
score 0/1 vs. 2	ΔT	0.9176 (0.0064)	6.319	85.7	84.2	99.0	25.2
score 0 vs. 2	FL% >32.0°C	0.9310 (0.0078)	2.647	86.5	99.2	98.7	90.9
White-feathered hens							
score 0 vs. 1	FL% >32.0°C	0.6710 (0.0104)	0.207	36.0	97.9	87.1	79.7
score 1 vs. 2	FL% >33.0°C	0.8968 (0.0125)	10.953	78.5	97.1	88.4	94.2
score 0/1 vs. 2	FL% >32.0°C	0.9333 (0.0091)	2.115	87.9	96.5	98.6	73.3
score 0 vs. 2	FL% >32.0°C	0.9455 (0.0086)	0.833	89.4	98.5	98.6	88.6
Back plumage							
Brown-feathered hens							
score 0 vs. 1	FL% >32.0°C	0.6477 (0.0064)	0.057	31.7	97.5	83.2	78.2
score 1 vs. 2	FL% >32.5°C	0.9348 (0.0047)	6.771	90.7	93.3	87.0	95.3
score 0/1 vs. 2	FL% >32.0°C	0.9712 (0.0023)	6.797	91.5	97.9	97.2	93.5
score 0 vs. 2	FL% >32.0°C	0.9818 (0.0020)	1.301	96.1	98.9	98.3	97.5
White-feathered hens							
score 0 vs. 1	FL% >32.0°C	0.7243 (0.0110)	0.042	46.8	98.1	88.5	85.7
score 1 vs. 2	FL% >32.0°C	0.9013 (0.0099)	6.475	88.5	88.5	86.2	90.4
score 0/1 vs. 2	FL% >32.0°C	0.9612 (0.0046)	1.803	95.0	91.4	98.7	72.4
score 0 vs. 2	FL% >32.0°C	0.9756 (0.0041)	0.212	96.0	98.3	98.8	94.3
Belly plumage							
Brown-feathered hens							
score 0 vs. 1	FL% >32.5°C	0.7339 (0.0084)	0.631	56.1	85.5	90.9	43.0
score 1 vs. 2	FL% >33.5°C	0.9297 (0.0052)	10.353	92.4	88.5	88.9	92.0
score 0/1 vs. 2	FL% >32.5°C	0.9719 (0.0025)	17.706	90.6	97.3	97.8	88.6
score 0 vs. 2	FL% >32.5°C	0.9807 (0.0023)	7.372	95.3	97.2	98.7	90.7
White-feathered hens							
score 0 vs. 1	FL% >32.0°C	0.7816 (0.0137)	0.794	71.5	76.8	93.3	37.4
score 1 vs. 2	FL% >33.0°C	0.9091 (0.0098)	15.28	81.2	87.2	64.1	94.2
score 0/1 vs. 2	FL% >33.0°C	0.9746 (0.0026)	5.826	94.9	92.4	97.8	84.0
score 0 vs. 2	FL% >33.0°C	0.9862 (0.0021)	4.568	96.3	95.7	98.1	91.8

FL%, proportion of featherless areas (i.e., percentage of pixels above a defined threshold temperature out of the total number of pixels for the body region); SE, standard error of the mean; ΔT, temperature difference in °C between the mean temperature (i.e., arithmetic mean of the temperature of all pixels in the region of interest) and the recorded ambient temperature.

A visual plumage scoring system according to Welfare Quality® (2009) was used as the reference method.

reliability. The MT, ΔT, and FL% of each threshold differed significantly between the 3 plumage scores (Supplementary Table 1; $P < 0.001$). The MT for both hybrid types was 22.15 ± 4.25°C (mean ± SD) for score 0, 22.32 ± 4.65°C for score 1, and 27.27 ± 4.07°C for score 2 in the dorsal neck plumage; 21.90 ± 3.87°C for score 0, 22.90 ± 5.05°C for score 1, and 29.83 ± 5.44°C for score 2 in the back plumage; 23.77 ± 3.81°C for score 0, 25.11 ± 4.66°C for score 1, and 31.59 ± 4.87°C for score 2 in the belly plumage, respectively. The highest correlation coefficients between PS and FL% >32.0°C were calculated for the back plumage (BL and WL: $r_s = 0.82$), followed by belly plumage (BL: $r_s = 0.71$, WL: $r_s = 0.80$) and dorsal neck plumage (BL: $r_s = 0.55$; WL: $r_s = 0.66$).

Table 1 summarizes the AUC, cutpoints, sensitivities, specificities, negative predictive values (**NPV**), and positive predictive values (**PPV**) determined in the ROC analyses. ROC curves for IRT parameters with the highest AUC are shown in Figure 1. We obtained the highest diagnostic accuracy by comparing intact plumage (score 0) to severe plumage damage (score 2) for the back plumage in BL and belly plumage in WL, respectively.

In certain body regions and score comparisons, the best-fit parameters differed between WL and BL (Table 1). When comparing AUC between the hybrid types, there were no significant differences in FL% for the dorsal neck plumage ($P \geq 0.091$); however, they were evident for back plumage ($P \leq 0.043$) except for the comparison between scores 1 vs. 2 ($P \geq 0.133$). For belly plumage, differences in AUC between the hybrid types were detected for scores 0 vs. 1, 0 vs. 2, and 1 vs. 2 ($P \leq 0.045$), but not for scores 0/1 vs. 2 ($P \geq 0.309$).

DISCUSSION

The study showed that IRT is a suitable tool for assessing plumage damage in laying hens in certain cases. However', the correlations to PS and diagnostic accuracy varied between body regions and hybrid types. In accordance with Pichova et al. (2017), the correlations between PS and IRT were low to moderate for the dorsal neck plumage but high for the back and belly plumage. A stronger correlation in the back plumage than the belly plumage was shown in our study, which contrasts with the results of Pichova et al. (2017). A varying density of plumage may be the reason for differences in the diagnostic accuracy of IRT between hybrid types and body regions (Cangar et al., 2008), but may also be due to differing amounts of subcutaneous fat tissue. Furthermore, possible causes for the differences

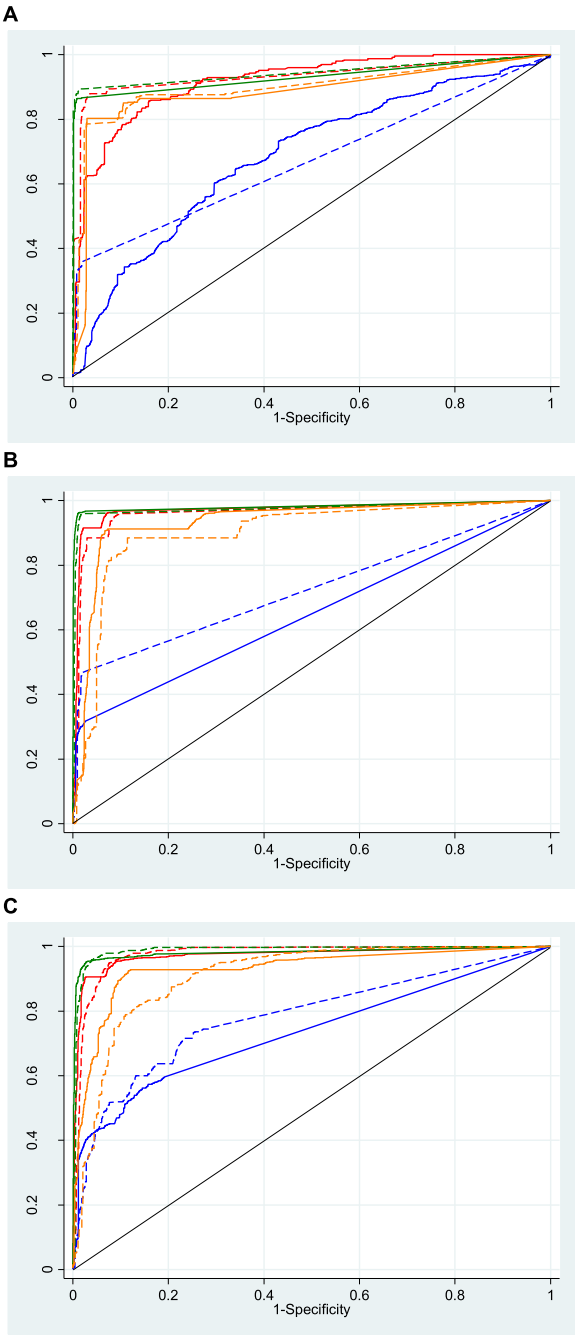

Figure 1. Receiver operating characteristic (ROC) curves for identifying plumage damage in laying hens using infrared thermography (IRT) for three different body regions: (A) dorsal neck, (B) back, and (C) belly. A visual plumage scoring system according to Welfare Quality® (2009) was used as the reference method. ROC curves for brown-feathered (solid lines) and white-feathered hens (dashed lines) for the best-fit parameters of the infrared images (i.e., the maximum area under the curves, see Table 1) comparing score 0 (intact plumage) vs. score 1 (mild plumage damage; blue lines), score 1 vs. score 2 (severe plumage damage; orange lines), scores 0 and 1 vs. score 2 (red lines), and score 0 vs. score 2 (green lines).

found between the hybrid types may be divergent reflective properties of the different colored plumage. Regarding MT as an effect of plumage cover, the observed values of our study conform to Cook et al.'s results (2006) and are below the temperatures determined by Zhao et al. (2013); the latter may be related to the ambient temperature during IRT recording. The best-fit parameter from IRT to describe plumage damage was the FL% in 22 of 24 variants. In Pichova et al.'s (2017) study, this parameter showed a higher correlation with PS than ΔT for back and belly plumage. By contrast, Pichavo et al. (2017) found the highest correlation with ΔT for dorsal neck plumage; furthermore, the dorsal neck was the only body region where partly ΔT (2 of 8 variants) proved to be the most appropriate parameter. One possible reason for this difference is that the hen can move its anatomically flexible neck during fixation into different positions, changing the total area of IRT measurement.

ROC analyses generated new findings regarding the diagnostic accuracy of IRT. The AUC and its associated sensitivities and specificities for identifying severe plumage damage from intact plumage showed with the best-fit parameter high accuracy with a sensitivity and specificity of over 95% for the back and belly plumage. It also displayed high accuracy in differentiating between moderate and severe plumage damage. By contrast, the detection of initial, moderate plumage damage (i.e., scores 0 vs. 1) showed low accuracy, which is insufficient for applying IRT as an early detection system.

CONCLUSIONS

We investigated the diagnostic accuracy of IRT for assessing the plumage of BL and WL in our study. Our results showed that this imaging technique is suitable for the objective quantification of plumage damage with differences between body regions. However, it is not a suitable tool for the early detection of incipient plumage damage. Furthermore, the AUC values and best-fit parameters of IRT differ between white- and brown-feathered laying hens, which should be considered in further studies and in developing automatic systems for detecting plumage damage.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Silke Lange (www.p-wert.de) for performing statistical analyses of the study. Special thanks go to all participating layer farmers as well as Adriano Krätzschmar for their support and constructive cooperation during data collection.

This study was funded by the Saxon State Office for Agriculture, Environment and Geology - EIP-Agri (identification number 332019017501 LWC). Responsible for the implementation of EAFRD funding in Saxony is the State Ministry for Energy, Climate Protection, Environment and Agriculture, Funding Strategy Unit, EAFRD Managing Authority). This article is funded by the Open Access Publication Fund of Hochschule für Technik und Wirtschaft-Dresden University of Applied Sciences and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)−432908064.

Ethical statement: The study was reviewed by the Country Directorate of Saxony/Germany as the responsible animal ethics committee and not classified as animal experiment (reference DD24.1-5131/277/42).

DISCLOSURES

The authors have no conflicts of interest to report.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.psj.2022.102022.

REFERENCES

Bilcik, B., and L. Keeling. 1999. Changes in feather condition in relation to feather pecking and aggressive behaviour in laying hens. Br. Poult. Sci. 40:444–451.

Cangar, O., J. M. Aerts, J. Buyse, and D. Berckmans. 2008. Quantification of the spatial distribution of surface temperatures of broilers. Poult, Sci. 87:2493–2499.

Cook, N. J., A. B. Smykot, D. E. Holm, G. Fasenko, and J. S. Church. 2006. Assessing feather cover of laying hens by infrared thermography. J. Appl. Poult. Res 15:274–279.

Döhring, S., L. Jung, and R. Andersson. 2020. Plumage damage in laying hens – automated detection in practical test - technical note. Eur. Poult. Sci. 84.

Du Prel, J.-B., B. Röhrig, G. Hommel, and M. Blettner. 2010. Auswahl statistischer testverfahren. Deutsches Ärzteblatt Int. 107:343–348.

Gunnarsson, S., B. Algers, and J. Svedberg. 2000. Description and evaluation of a scoring system of clinical health in laying hens. Laying Hens in Loose Housing Systems. S Gunnarsson, ed. Swedish University of Agricultural Sciences, Uppsala Doctoral thesis.

Pichova, K., B. Bilcik, and L. Kost. 2017. Assessment of the effect of housing on feather damage in laying hens using IR thermography. Animal 11:661–669.

Spindler, B., M. F. Giersberg, R. Andersson, and N. Kemper. 2016. Keeping laying hens with untrimmed beaks – a review of the status quo in practice and science. Züchtungskunde 88:475–493.

Victor, A., A. Elsäßer, G. Hommel, and M. Blettner. 2010. Judging a plethora of p-values: how to contend with the problem of multiple testing. Part 10 of a series on evaluation of scientific publications. Deutsches Ärzteblatt Int. 107:50–56.

Welfare Quality®. 2009. Welfare Quality® Assessment Protocol for Poultry (Broilers, Laying Hens). Welfare Quality® Consortium, Lelystad, the Netherlands.

Zhao, Y., H. Xin, and B. Bong. 2013. Use of infrared thermography to assess laying-hen feather coverage. Poult. Sci. 92:295–302.