Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Deep learning-based model classifies thermal conditions in dairy cows using infrared thermography

Verônica M. Pacheco [a], Rafael V. Sousa [a], Edson J.S. Sardinha [a], Alex V.S. Rodrigues [a], Tami M. Brown-Brandl [b], Luciane S. Martello [a,*]

[a] Department of Biosystems Engineering, Faculty of Animal Science and Food Engineering (FZEA), University of São Paulo (USP), Av. Duque de Caxias Norte, 225 Pirassununga, SP, 13635-900, Brazil

[b] Biological Systems Engineering Department, University of Nebraska-Lincoln, 1400 R Street, Lincoln, NE, 68588, USA

ARTICLE INFO

Article history:
Received 24 November 2021
Received in revised form
14 June 2022
Accepted 1 July 2022

Keywords:
Convolutional Neural Network
Thermography
Non-Invasive Measurement
Precision Livestock Farming
Thermal Comfort

Infrared thermography is a technique that has been utilized to assess the thermal status of animals. This study proposes a convolutional neural network (CNN)-based method for the individual classification of the thermal condition of dairy cows using thermal images of specific body surface regions. The experiment was carried out with 26 lactating cows (Holstein) during summer and winter (40 days). Meteorological data were collected every 30 min and rectal temperature (t_{rectal}), respiratory rate (RR), and surface temperature (t_{sf}) were measured three times a day (5 a.m., 1 p.m., and 7 p.m.). t_{sf} was correlated with RR and t_{rectal} and selected as input data of the models. Thus, thermal images were labelled according to the RR and t_{rectal} categories and employed in the investigation of four CNN-based models constructed by supervised learning and cross-validation protocols. Performance was assessed by confusion matrix metrics (accuracy, precision, recall, and F1-score) comparing the predicted labels and true labels. The best results were obtained with the models that used forehead images. The model using images labelled according to three RR thermal levels has an accuracy of 76%, and the model labelled according to three t_{rectal} thermal levels has an accuracy of 71%. The method based on deep learning allowed us to generate a computational classifier that considers not only the temperature intensities from thermographic images but also their distribution profile to identify patterns referring to thermal conditions in dairy cows.

© 2022 IAgrE. Published by Elsevier Ltd. All rights reserved.

* Corresponding author.
E-mail addresses: veronica.pacheco@usp.br (V.M. Pacheco), rafael.sousa@usp.br (R.V. Sousa), sardinha@usp.br (E.J.S. Sardinha), alex.rodrigues@usp.br (A.V.S. Rodrigues), tami.brownbrandl@unl.edu (T.M. Brown-Brandl), martello@usp.br (L.S. Martello).
https://doi.org/10.1016/j.biosystemseng.2022.07.001
1537-5110/© 2022 IAgrE. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Thermal stress evaluation has been increasingly performed to maintain the welfare of dairy cattle, as its impact on the health and productivity of animals has already been consolidated. The infrared thermography technique has become a useful tool for assessing thermal stress by monitoring the body surface temperature (t_{sf}) in dairy cattle (Sellier et al., 2014).

If the internal temperature of animals is affected by environmental factors, thermoregulatory mechanisms are activated to generate or dissipate heat, seeking to recover the internal thermal balance (homeothermy). Thermal stress occurs because of the imbalance in these heat exchanges that in unfavourable environmental conditions require animals to accelerate thermoregulatory activities. Under these conditions, cattle may not express their genetic potential by using the energy available to maintain homeothermy, affecting their productive and reproductive performance (Allen et al., 2015).

Cows subjected to high air temperatures tend to decrease matter consumption, which directly affects the production and composition of milk. Additionally, thermal stress affects reproductive activity as it negatively influences natural mating behaviour, hormone production, and heat (Polsky & von Keyserlingk, 2017).

Aside from the advantage of being able to quickly and noninvasively collect a thermoregulatory response, thermal images still require time and specialized staff to interpret the results. An automatic way to interpret this type of image can contribute to the generation of real-time information and more efficient reports, despite the need to expand the data exchange capacity between two systems (Van Hertem et al., 2017). In this context, a convolutional neural network (CNN), which is a deep learning (DL) technique, has been applied to detect patterns in images.

Advances in CNN knowledge have stimulated the use of the technique in livestock research to solve problems involving image classification (Misimi et al., 2017), including the prediction of body condition score (BCC) in dairy cows (Rodríguez Alvarez et al., 2018), individual pig recognition systems (Hansen et al., 2018), prediction of intramuscular fat in live pigs through ultrasound image analysis (Kvam & Kongsro, 2017) and posture recognition systems of lactating sows using Kinect (Zheng et al., 2018). Nonetheless, CNNs have not yet been widely employed to interpret thermal images, which are commonly used to monitor heat exchange between animals and the environment (Pacheco et al., 2020; Ricci et al., 2019; Salles et al., 2016). CNNs can be useful for extracting patterns and colour distribution (temperature distribution) to predict the thermal condition of dairy cows. In addition, this technique enables the use of raw images to predict heat stress in dairy cows without requiring additional on-farm expertise, significantly reducing the workload of processing, standardizing and pipeline analysis of the images.

To date, no work has developed a CNN model that only used thermal images to assess the thermal condition level of cows. Therefore, the objective of this work is to propose an innovative method using a CNN-based model to classify thermal images obtained from dairy cattle using two image databases: a database labelled according to three respiratory rate (RR) levels and a database labelled according to three rectal temperature (t_{rectal}) levels.

2. Materials and methods

The investigated computational models were based on CNNs and applied thermal images of different body regions of lactating cows (Holstein) as input. For this purpose, the thermal images were labelled according to thermal levels defined for the RR and t_{rectal} readings to build an image database for model training. The experiment was carried out according to the Institutional Animal Care and Use Committee Guidelines of the Faculty of Animal Science and Food Engineering of the University of São Paulo (protocol 4,768,290,118). Data collection and modelling are detailed below.

2.1. Experimental design and data collection

The experiments were carried out during the winter and summer seasons at the facilities of the Faculty of Animal Science and Food Engineering (FZEA) of the University of São Paulo (USP) in Pirassununga, SP, Brazil, located at 21°57′02″S, 47°27′50″W, at a mean elevation of 630 m above sea level. The climate of the region is Köeppen Cwa with two well-defined seasons (rainy summer and dry winter with rare occurrences of frost). The city has an average annual air temperature of 21.5 °C. In winter, the mean air temperature, relative humidity, and dew point were 19.17 ± 4.97 °C, 70.81% ± 18.73%, and 13.00 ± 3.04 °C, respectively. In summer, the parameters were 29.21 ± 6.20 °C, 62.59% ± 20.51%, and 20.33 ± 1.79 °C, respectively.

The adult cows are housed in a free-stall barn with a cement tile roof (uninsulated), wooden structures separating bays, a rough cement floor with a slope of 10%, and a ceiling height of 3.2 m (total covered area of 200 m^2, 16 × 13 m, southeast orientation, with an outside pasture area). The building has 20 free-stall bays with sand litter. The feeding area is covered and located inside the building and is dimensioned based on an extension of 80 cm per animal. The water troughs are located at the ends and centre of the free-stall barn. The feed distribution (fodder) and milking were performed twice a day at 7 a.m. and 2 p.m. The animals were fed corn silage (30%), Tifton hay (10%), and concentrate (60%) twice a day at 7 a.m. and 3 p.m., which contained 76% total digestible nutrients and 18% crude protein. RR and t_{rectal} were collected with data from an experiment carried out for 20 days in winter and 20 days in summer. In both seasons, RR, t_{rectal} and thermal images (in this order) were collected to evaluate body surface temperatures (t_{sf}) of different regions of the animals.

The physiological measurements of t_{rectal}, RR, and t_{sf} of all cows were conducted for 20 days in winter and 20 days in summer with all cows at 5 a.m., 1 p.m., and 7 p.m. The t_{rectal} was manually recorded with a digital thermometer (VMDT01, Viomed, China). The RR was determined through a time count of every ten flank movements (breaths), and then breaths per minute were calculated. t_{sf} was measured with a thermographic camera (875–2, Testo SE & Co. KGaA, Germany) with an emissivity of 0.98, approximately 2 m from each of the four

animal body regions: forehead, ocular area, rib, and flank. The weather data were monitored with a data logger (HOBO U12, Onset Computer Corporation, USA), which was fixed to the centre of the pens at a height of 3 m. The weather variables were automatically recorded every 30 min throughout both the summer period and winter period and accordingly matched the physiological measurements.

The thermal images were interpreted with the software IRsoft 4.4 (Testo SE & Co. KGaA, Germany) to determine the mean temperatures of the forehead ($t_{sf\text{-}fr}$), rib ($t_{sf\text{-}r}$), and flank ($t_{sf\text{-}fl}$) and the maximal temperature of the ocular area ($t_{sf\text{-}o}$). The reflective temperature in the software was set according to the average air temperature during the time of collection and the emissivity was set to 0.95. For each body region, a specific smaller area defined a subarea in each image (Fig. 1).

2.2. Data analysis and construction of the thermal image database

The SAS Corr procedure (SAS Institute Inc., Cary, NC) was selected to determine the best Pearson correlation coefficient among physiological data (t_{sf} by thermography, RR, and t_{rectal}) to choose which body regions would be most suited for the CNN model. Subsequently, the images of the selected body regions ($t_{sf\text{-}fr}$ and $t_{sf\text{-}o}$) were preprocessed using the software IRsoft and scripts developed in Python 3.5.4 rc1 and MATLAB R2015a to improve and organise the database according to the following steps: (1) saving temperature matrices (IRSoft); (2) reconstruction of the thermal images; (3) regions of interest (RoIs) defined by a rectangular area of the forehead (RoI_f) and a circle of the ocular area (RoI_o) was selected in the reconstructed images; (4) determination of the temperature range used to reconstruct the thermal image database; (5) data augmentation operations; and (7) organization of the image database labelled according to RR and t_{rectal} labelling classes.

Using IRSoft software, the images were selected as temperature matrices of 160 × 120 pixels, reflecting the temperature profile of each region of the selected body surface ($t_{sf\text{-}fr}$ and $t_{sf\text{-}o}$). In total, 1867 images of the RoI_f and 1865 images of the RoI_o were saved. Thermal images were reconstructed so that each specific region of each image was cropped (MATLAB R2015a script), generating new temperature matrices. The cutout sizes are listed as follows: 32 × 48 pixels for the RoI_f and 24 × 24 pixels for the RoI_o. From these RoIs, a general histogram was developed for each region to determine the ideal temperature range for the uniform reconstruction of the RoIs from the thermal images (Fig. 2). The temperatures were converted to the *Jet* colourmap, and the palette was selected from the *matplotlib* library (Python). The palettes (colourmaps), including the jet, have 256 different colours in which the temperatures can be represented. To obtain a colour image with a good representation of the temperature matrix, a different colour (or tone) is selected for each tenth of a degree of temperature variation. In the case of the RoI_f, as the histogram was broader, the range of 14–39 °C was selected, resulting in images with 250 different colours. In the case of the RoI_o, as the histogram was narrower, the range of 24–39 °C was utilized, resulting in images with 150 different colours.

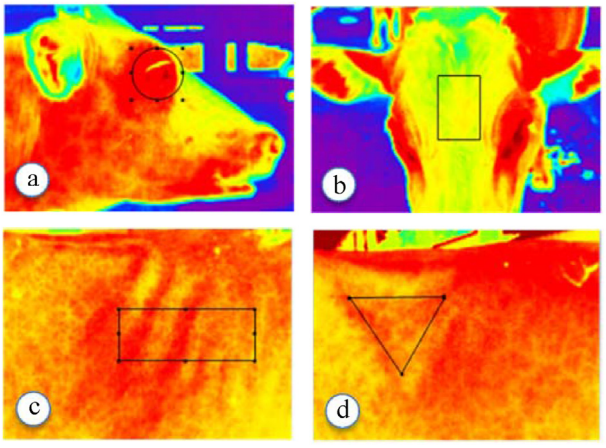

Fig. 1 – Thermal images of the specific sub-areas for each animal body region: (a) ocular area, (b) forehead, (c) rib, and (d) flank, respectively.

Translation and mirroring techniques have also been applied to the original images to enlarge the database using Python script. Translation consists of applying a crop filter that is smaller than the original image at different positions of the image so that new images are formed from the cropped pieces of the original image. The mirroring technique consists of making a 180° rotation in the original image to form a second image with the inverse of the original image. Thus, 10 samples with 20 × 20 pixels (jpg) were obtained from each cropped thermal image (jpg) through horizontal mirroring combined with 5 translations (Fig. 3).

The total thermal image database that was constructed was labelled according to three RR labelling classes and three t_{rectal} labelling classes (Table 1). The criteria for the limits for the RR and t_{rectal} labelling classes were created to establish a balanced database in each class according to available data and previous studies (Burfeind et al., 2012; Kadzere et al., 2002), but the proposed method allows adjustment of these limits before running new models according to the need for age and lineage race. Thus, it was possible to work with similar amounts of images in each class of RR and t_{rectal}, which is important for the learning process of the CNN-based models. Thus, two image databases were manually created from thermal images, hereinafter referred to as "true label", with each database having its thermal images divided into 3 RR or t_{rectal} labelling classes.

2.3. Deep learning-based modelling and evaluation

The CNN-based models that use thermal images were setup, trained, and modified, if necessary, to obtain accurate responses and the best model for automatic classification of thermal images. At the end of this work, four classifier models were developed. Two classifiers employed thermographic RoI_f: the first classifier was trained with respiration rate-based labelling ($CNN_{F\text{-}RR}$), and the second classifier was trained with rectal temperature-based labelling ($CNN_{F\text{-}RT}$). Similarly, two more models were built with thermographic RoI_o: the first model was trained with respiration rate-based labelling ($CNN_{O\text{-}RR}$), and the second model was trained with rectal temperature-based labelling ($CNN_{O\text{-}RT}$).

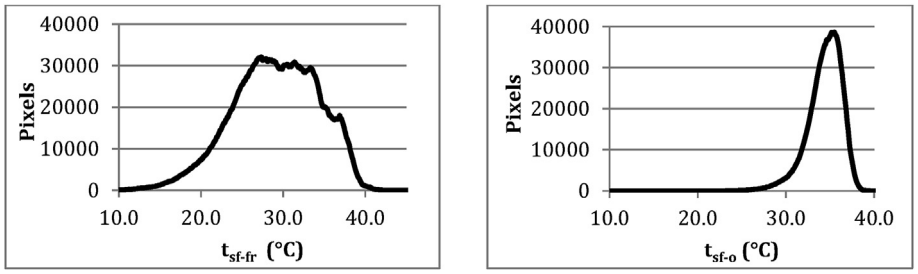

Fig. 2 – Temperature distribution histograms from the temperature maps of regions of interest of the forehead ($t_{sf\text{-}fr}$) (a) and ocular area ($t_{sf\text{-}o}$) (b).

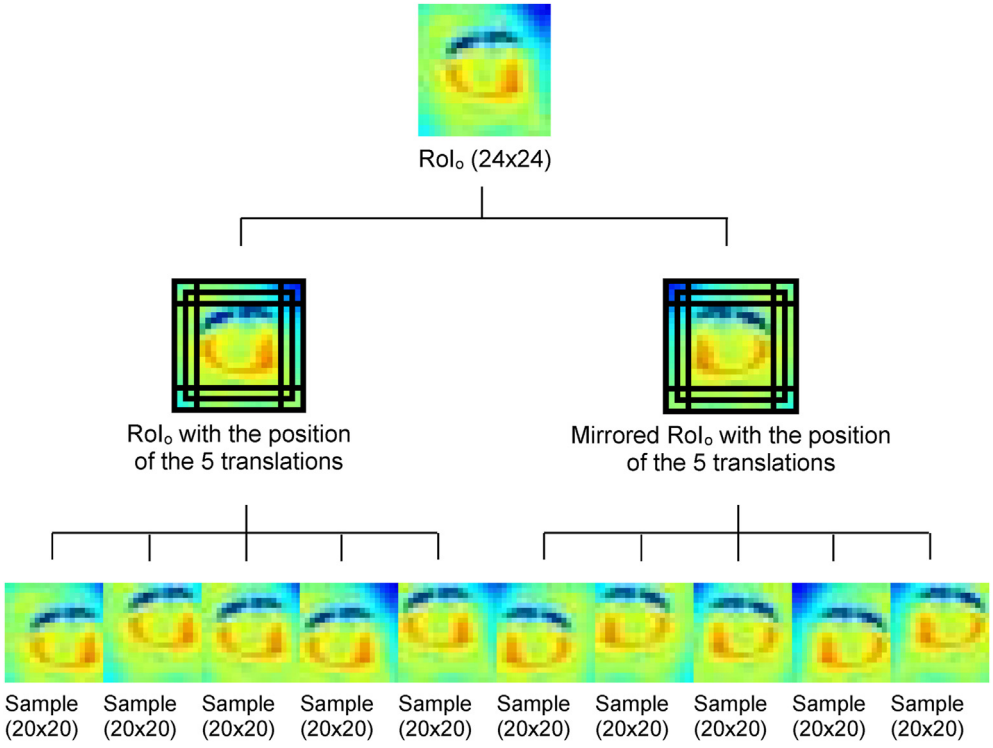

Fig. 3 – Example of the mirroring and translation techniques to increase the database of the region of interest of the ocular area RoI_o.

The CNN architecture was built according to the sequential model and uses TensorFlow of the Keras 2.0.6 library for Python, using TensorFlow as the back-end. According to the cross-validation protocol, the RoIs images were divided into 80% for training, 10% for validation, and 10% for testing. The models were built from the training data and their evolution followed through the validation data. The trained models were compared with the test data during a process that was repeated 3 times, preserving the model with better accuracy. A fine-tuning process based on the best accuracy model made it possible to define values for the following hyperparameters: type (convolution and pooling), size and number of filters, step, activation function ('relu') and number of layers until the final models were obtained (Fig. 4). The results reflect the best accuracy obtained for each model ($CNN_{F\text{-}RR}$, $CNN_{F\text{-}RT}$, $CNN_{O\text{-}RR}$ and $CNN_{O\text{-}RT}$).

The performance of the trained models was evaluated using the confusion matrices by comparing the predicted labels obtained from the test phase of the classifiers with the true labels. Through the confusion matrix, it is possible to evaluate the efficiency of the classifiers in terms of accuracy, precision, recall, and F1 score, as described by Sokolova and Lapalme (2009; Eqs. (1)–(4)). General accuracy is an important metric that cannot be evaluated in isolation. As the accuracy observes the general sum for the correct answers, it can count on the addition of erroneously classified examples, or in the case of unbalanced classes, it can represent the correct answers accumulated in just one class. Therefore, it is important to observe other metrics, such as precision and recall. Precision indicates the classifier's ability to avoid classifying incorrect data such as correct answers, and recall shows the classifier's ability to obtain all

Table 1 – Respiratory rate (RR) and rectal temperature (t_{rectal}) labelling classes considered for the distribution of the thermal imaging database.

RR labeling class (breaths/min)	RoI_f (Number of images)	RoI_o (Number of images)
<40	6500	6490
41–60	6370	6380
>60	5610	5610
Total	18,480	18,480
t_{rectal} labeling class (°C)	RoI_f (Number of images)	RoI_o (Number of images)
<38.4	6740	6730
38.5–39.0	6640	6650
>39.0	5130	5130
Total	18,510	18,510

[a]The difference in the number of samples between the databases is due to the difference in the number of images between the region of interest of the forehead (RoI_f) and the region of interest of the ocular area (RoI_o) and the existence or not of the RR and/or t_{rectal} value associated with a given image.

the correct samples for each class (Sokolova & Lapalme, 2009). Thus, if the precision is low, the model is giving examples from other classes to the observed class. On the other hand, if the recall is low, the classifier encounters difficulty obtaining correct examples. To concurrently evaluate these two metrics, the F1 score is calculated as the harmonic average between precision and recall. An ideal model has an F1 score value similar to that of accuracy. All metrics were calculated based on the values identified in the confusion matrix as true positives (t_p), true negatives (t_n), false-positives (f_p) and false negatives (f_n). Values identified as t_p are those correctly identified as belonging to the class, and t_n are the values correctly identified as not belonging to the analysed class. The f_p values are those misidentified as belonging to the class and the f_n values misidentified as not belonging to the analysed class (Sokolova & Lapalme, 2009).

$$Accuracy = \frac{\sum_i^l t_p}{n} \quad (1)$$

$$Precision = \frac{t_p}{t_p + f_p} \quad (2)$$

$$Recall = \frac{t_p}{t_p + f_n} \quad (3)$$

$$F1 - Score = \frac{\sum_i^l \frac{2 * Precision_i * Recall_i}{Precision_i * Recall_i}}{classes} \quad (4)$$

where n is total number of images classified, and l is the number of classes.

3. Results

Table 2 presents the descriptive data from the animals (RR, t_{rectal}, and t_{sf}) and from the environment (DBT and WBT) obtained during the entire period of the experiment. The DBT (37.65 °C) and the maximum RT (41.1 °C) observed indicated that the animals were exposed to thermal stress.

The mean ($t_{sf\text{-}fr}$, $t_{sf\text{-}r}$, and $t_{sf\text{-}fl}$) and maximum ($t_{sf\text{-}o}$) values were correlated with physiological data (RR and t_{rectal}) (Table 3). Among all body surface areas, $t_{sf\text{-}fr}$ was more correlated with the RR (0.70), and $t_{sf\text{-}o}$ was the best correlated with t_{rectal} (0.61); therefore, the thermal images of these specific body regions were selected as inputs for CNN-based computational models. Therefore, considering the available database, 18,480 thermal images from RoI_f were selected to build $CNN_{F\text{-}RR}$ and $CNN_{F\text{-}RT}$, and 18,510 images from RoI_o were utilized in the construction of $CNN_{O\text{-}RR}$ and $CNN_{O\text{-}RT}$.

Although $t_{sf\text{-}fr}$ was better correlated with RR and $t_{sf\text{-}o}$ was better correlated with t_{rectal}, all body regions showed median values of correlation with these variables (between 0.55 and 0.70).

3.1. *Forehead thermographic classifier with respiration rate-based labelling*

The $CNN_{F\text{-}RR}$ confusion matrix is shown in Table 4. The correct classifications accounted for an accuracy of 76.3% distributed in the following classes: <40 breaths/min (26.9%), 41–60 breaths/min (23.7%), and >61 breaths/min (25.8%). To obtain

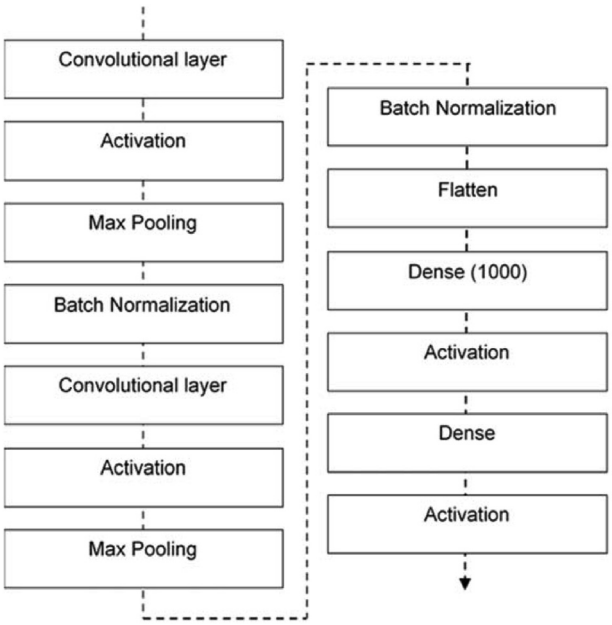

Fig. 4 – Layer's structure of the developed model.

Table 2 – Descriptive statistics.

Data	Variables		Mean	Standard Deviation	Min	Max
Physiology	RR (breaths/min)		52.83	20.55	14.19	126.05
	t_{rectal} (°C)		38.74	0.73	36.40	41.10
Infrared Thermography (°C)	$t_{sf\text{-}fr}$	Ave	27.85	5.48	12.70	41.10
		Min	24.81	6.40	7,00	39.70
		Max	32.05	3.65	20.40	43.00
	$t_{sf\text{-}fl}$	Ave	32.01	3.58	15.50	38.90
		Min	30.40	4.07	12.40	37.70
		Max	33.40	4.07	12.40	37.70
	$t_{sf\text{-}r}$	Ave	32.14	3.42	17.50	38.80
		Min	30.22	3.78	15.10	38.10
		Max	33.67	3.07	19.20	40,00
	$t_{sf\text{-}o}$	Max	36.65	3.78	15.10	38.10
Temperature	DBT (°C)		23.45	7.43	7.71	37.65
	WBT (°C)		18.36	4.46	6.60	25.70

*Physiological data: respiration rate; t_{rectal}: rectal temperature; $t_{sf\text{-}fr}$: forehead surface temperature; $t_{sf\text{-}o}$: surface temperature of ocular area; $t_{sf\text{-}r}$: rib surface temperature; $t_{sf\text{-}fl}$: flank surface temperature. Weather data: dry bulb temperature (DBT); wet bulb temperature (WBT). Parameters: average (Ave); minimum (Min); maximum (Max).

Table 3 – Pearson's correlation coefficient relating physiological variables.

Variable	t_{rectal}	$t_{sf\text{-}fr}$	$t_{sf\text{-}o}$	$t_{sf\text{-}r}$	$t_{sf\text{-}fl}$
RR	0,72	0,70	0,66	0,68	0,66
t_{rectal}	–	0,55	0,61	0,56	0,55

[a]RR: respiration rate;: rectal temperature; $t_{sf\text{-}fr}$: mean forehead surface temperature; $t_{sf\text{-}o}$: maximal surface temperature of ocular area; $t_{sf\text{-}r}$: mean rib surface temperature; $t_{sf\text{-}fl}$: mean flank surface temperature.

these accuracy values in the classes, we divide the number of correct answers in the class by the total number of images. $CNN_{F\text{-}RR}$ presented medium to high values of precision (mean = 76.7%) and recall (mean = 76.7%) and an F1-score of 76.6%. The lowest values of precision and recall were in the 41–60 breaths/min class (67.3% and 68.8%, respectively).

In the <41 breaths/min class, 496 images were correctly classified by $CNN_{F\text{-}RR}$ (precision = 80.0% and recall = 76.3%). Of the 650 images belonging to this class, the model could not obtain 154 images, of which 136 images were classified as belonging to the 41–60 breaths/min class and 18 images were classified as belonging to the >60 breaths/min class. Considering an applied situation, the model would consider 154 cows with an RR higher than the real RR. Regarding the images that the model considered as belonging to the <41 breaths/min class (620 images), it incorrectly classified 124 images: 117 images were incorrectly classified as belonging to the 41–60 breaths/min class and 7 images were incorrectly classified as belonging to the >60 breaths/min class. In this case, the model would consider 117 cows with an RR lower than the real RR in an applied situation.

In relation to the central class (41–60 breaths/min), 438 images were correctly classified (precision = 67.3% and recall = 68.8%). Of the total images belonging to this class (637 images), the model predicted 199 images as belonging to other classes (117 in the <41 breaths/min class and 82 in the >60 breaths/min class). In an applied situation, 117 cows would be classified with an RR below the real RR and 82 with an RR above the real RR. Additionally, of the 651 images that the model classified as belonging to the central class (41–60 breaths/min), 213 images belonged to other classes. Thus, 136 cows would be considered to have an RR lower than the real RR, and 77 cows would be considered to have an RR higher than the real RR.

In the last class (>60 breaths/min), 477 images were correctly classified (precision = 82.7% and recall = 85%). Of the 561 images that were tested, the model incorrectly classified 84 images: 77 images in the 41–60 breaths/min class, and 7 images in the <41 breaths/min class. On the other hand, of the 577 images that the model considered for this class, 100 images belonged to the other classes (82 images belonged to the 41–60 breaths/min class and 18 images belonged to the <41 breaths/min class).

Table 4 – Confusion matrix between the predicted label by $CNN_{F\text{-}RR}$ and the true label, obtained from the respiratory rate (RR) labeling.

True label – RR (breaths/min)	Predicted label– RR (breaths/min)			Recall
	<41	41–60	>60	
< 41	**496**	136	18	76.3%
41–60	117	**438**	82	68.8%
> 60	7	77	**477**	85.0%
Precision	80.0%	67.3%	82.7%	**76.3%**

[a]Bold cells indicate the correct predictions in each class.

3.2. Forehead thermographic classifier with rectal temperature-based labelling

A confusion matrix was also created to evaluate the efficiency of $CNN_{F\text{-}RT}$ (Table 5). The accuracy of the model was 70.5%, and the correct answers were 27.8% in the <38.5 °C class, 20.8% in the 38.5–39.0 °C class and 21.9% at > 39.1 °C class. Precision (mean = 70.8%) and recall (mean = 71.1%) presented medium values, and as with the previous model, the results were lower in the central class (38.5–39.0 °C), with values of 65.7% and 58.0%, respectively. In this case, the F1-score was 70.9%.

Of the 674 images belonging to the <38.5 °C class, 514 images were correctly classified (precision = 69.4% and recall = 76.3%). In this class, the model could not obtain 160 images, of which 132 images were classified as belonging to the 38.5–39.0 °C class and 28 images were classified as belonging to the >39.0 °C class. Considering an applied situation, the model would consider 160 cows with a t_{rectal} higher than the real t_{rectal} for this true label. Regarding the images that $CNN_{F\text{-}RT}$ considered to belong to the <38.5 °C class (740 images), it incorrectly classified 226 images: 188 images were incorrectly classified as belonging to the 38.5–39.0 °C class, and 38 images were incorrectly classified as belonging to the >39.0 °C class. In this class, the model would consider 88 cows with a t_{rectal} lower than the real t_{rectal} in an applied situation.

In the central class (38.5–39.0 °C), 385 images were correctly placed (precision = 65.7% and recall = 58.0%). In a total of 664 images, $CNN_{F\text{-}RT}$ classified 279 images as belonging to other classes (188 images were classified as belonging to the <38.5 °C class and 91 images were classified as belonging to the >39.0 °C class). In this case, 188 cows would be classified with a t_{rectal} below the real t_{rectal}, and 91 cows would be classified with a t_{rectal} above the real t_{rectal}. Additionally, of the 586 images that the model classified as belonging to the central class (38.5–39.0 °C), 201 images belonged to other classes. Thus, 132 cows would be considered to have a t_{rectal} lower than the real t_{rectal}, and 69 cows would be considered to have a t_{rectal} higher than the real t_{rectal}. As previously mentioned, the middle category is the most difficult to predict, as it is a transition region between the lowest t_{rectal} region and the highest t_{rectal} region, and not all animals have the same physiological response pattern when they are submitted to the same environmental stimulus.

In the last class (>39.0 °C), 406 images were predicted in the correct class (precision = 77.3% and recall = 79.1%). Of the 513 images that were tested in this class, $CNN_{F\text{-}RT}$ classified 69 images in the 38.5–39.0 °C class and 38 images in the <39.0 °C class (107 errors). Additionally, of the 525 images that the model considered for this class, 119 images belonged to the other classes (91 images belonged to the 38.5–39.0 °C class, and 28 images belonged to the <38.5 °C class).

3.3. Ocular thermographic classifier with respiration rate-based labelling

In the confusion matrix between the $CNN_{O\text{-}RR}$ predicted labels and the true labels, it was possible to observe that $CNN_{O\text{-}RR}$ reached an accuracy of 64.0% (Table 6). The correct answers were distributed as follows: 24.3% for the <41 breaths/min class, 17.6% for the 41–60 breaths/min class and 22.1% for the >60 breaths/min class. For $CNN_{O\text{-}RR}$, the precision value and recall value were lower in the central class (41–60 breaths/min), with values of 52.1% and 50.1%, respectively. The precision value (mean = 64.5%) and recall value (mean = 39.8%) of $CNN_{O\text{-}RR}$ were below the values of $CNN_{F\text{-}RR}$. The F1-score value was 64.3%, which was similar to that of accuracy.

A total of 449 images were correctly classified as belonging to the <41 breaths/min class (precision = 65.7% and recall = 69.2%). Of the 649 images belonging to this class, 171 images were erroneously classified as belonging to the 41–60 breaths/min class and 29 images were erroneously classified as belonging to the >60 breaths/min class. In this case, the model would consider 200 cows with an RR higher than the real RR. Regarding the images that the $CNN_{O\text{-}RR}$ considers to belong to the <41 breaths/min class (683 images), it misclassified 234 images, that is, the model considered 234 cows with an RR higher than the real RR.

The central class had 325 images correctly labelled (precision = 52.1% and recall = 50.1%). Of the 638 images belonging to the central class (41–60 breaths/min), 313 images were incorrectly classified as belonging to other classes (210 images in the <41 breaths/min class and 103 images in the >60 breaths/min class). In an applied situation, 210 cows would be classified with an RR below the real RR, and 103 cows would be classified with an RR above the real RR. Additionally, of the 624 images that the model classified as belonging to the central class (41–60 breaths/min), 171 cows were classified with an RR lower than the observed RR, and 128 cows were classified as belonging to the highest class (>60 breaths/min). This class, as occurred in the $CNN_{O\text{-}RR}$, had the lowest values of precision and recall. Following the same protocol for the RoI_f, these images seem to have been influenced by the adopted

Table 5 – Confusion matrix between the predicted label by $CNN_{F\text{-}RT}$ and the true label, obtained from the rectal temperature (t_{rectal}) labeling.

True label – t_{rectal} (°C)	Predicted label – t_{rectal} (°C)			Recall
	<38.5	38.5–39.0	>39.0	
< 38.5	**514**	132	28	76.3%
38.5–39.0	188	**385**	91	58.0%
> 39.0	38	69	**406**	79.1%
Precision	69.4%	65.7%	77.3%	**70.5%**

[a]Bold cells indicate the correct predictions in each class.

Table 6 – Confusion matrix between the predicted label by $CNN_{O\text{-}RR}$ and the true label, obtained from the respiratory rate (RR) labeling.

True label – RR (breaths/min)	Predicted label– RR (breaths/min)			Recall
	<41	41–60	>60	
< 41	449	171	29	69.2%
41–60	210	325	103	50.1%
> 60	24	128	409	72.9%
Precision	65.7%	52.1%	75.6%	**64.0%**

[a]Bold cells indicate the correct predictions in each class.

temperature range (24–39 °C). In the case of RoI_o, the range was smaller, which improved the contrast.

The last class (>60 breaths/min) had 409 correctly labelled images (precision = 75.6% and recall = 72.9%). Of the 561 images belonging to this class, $CNN_{O\text{-}RR}$ could not identify 152 images, which were misclassified, and of the 541 images that $CNN_{O\text{-}RR}$ considered from this class, 132 images belonged to the other labelling classes. $CNN_{O\text{-}RR}$ represents 152 animals with an RR above 60 breaths/min classified as having a lower RR and 132 animals with a low RR classified as having an RR above 60 breaths/min (representing 8.2% and 7.1% of the total data, respectively). As with the classification of RoI_f, the images of the RoI_o were classified with the highest precision and recall values for the last category (>60 breaths/min), which is the most important for accessing thermal stress in dairy cows.

3.4. *Ocular thermographic classifier with rectal temperature-based labelling*

The accuracy of the $CNN_{O\text{-}RT}$ (Table 7) was 60.1%, and the correct answers were distributed as follows: 26.6% in the <38.5 °C class, 12.9% in the 38.5–39.0 °C class and 20.6% in the >39.0 °C class. The precision (mean = 60.0%) and recall (mean = 62.7%) presented more unbalanced results compared with the other models, and lower values were obtained for the central class (52.4% and 40.0%, respectively). For the $CNN_{O\text{-}RT}$ model, the F1-score was 61.3%, which was similar to that of accuracy.

$CNN_{O\text{-}RT}$ classified 492 corrected images from 673 images classified as belonging to the <38.5 °C class (precision = 58.1% and recall = 73.1%). Of the 847 images that the model considered to belong to this class, 355 images were incorrectly classified: 299 images were incorrectly classified as belonging to the 38.5–39.0 °C class, and 56 images were incorrectly classified as belonging to the >39.0 °C class. In this case, the model indicated 181 cows with a t_{rectal} higher than the real t_{rectal} and 355 cows with a t_{rectal} lower than the real t_{rectal}.

In the central class (38.5–39.0 °C), 238 were correctly classified (precision = 52.4% and recall = 40.0%). In this class, which had 665 test images, the $CNN_{O\text{-}RT}$ predicted 427 images as belonging to other classes (299 images in the <38.5 °C class and 128 images in the >39.0 °C class). In addition, of the 454 images that the model predicted to belong to the central class (38.5–39.0 °C), 216 images belonged to other classes. Thus, 141 cows would be considered to have a t_{rectal} higher than the real t_{rectal}, and 75 cows would be considered to have a t_{rectal} lower than the real t_{rectal}. The adopted temperature range (24–39 °C) for RoI_o was smaller and improved the contrast but did not improve the model's performance.

In the last class (>39.0 °C), 382 images were correctly classified (precision = 69.4% and recall = 74.9%) of the 513 images that were tested, and the model misclassified 131 images: 75 images in the 38.5–39.0 °C class, and 56 images in the <38.5 °C class. Of the 550 images that the model considered for this class, 168 images belonged to the other classes (128 images belonged to the 38.5–39.0 °C class and 40 images belonged to the class <38.5 °C). In this case, 168 cows with a t_{rectal} less than 39.0 °C were considered to have a higher t_{rectal}. The critical situation, which would be to classify images as belonging to the >39 °C class as lower classes and to classify the lower classes as belonging to the >39.5 °C class, would be represented by 7.1% and 9.1%, respectively, of the total data.

Table 7 – Confusion matrix between the predicted by $CNN_{O\text{-}RT}$ and the true label, obtained from the rectal temperature (t_{rectal}) labeling.

True label – t_{rectal} (°C)	Predicted label – t_{rectal} (°C)			Recall
	<38.5	38.5–39.0	>39.0	
< 38.5	**492**	141	40	73.1%
38.5–39.0	299	**238**	128	40.0%
> 39.0	56	75	**382**	74.9%
Precision	58.1%	52.4%	69.4%	**60.1%**

[a]Bold cells indicate the correct predictions in each class.

4. Discussion

In general, the CNN developed from thermography of the RoI_f presented better results than those developed with the RoI_o. The results are promising when considering the challenges in standardizing thermal images. Although several authors have verified the potential of thermography as a noninvasive tool to monitor the surface temperature of animals (Brown-Brandl et al., 2013; Church et al., 2014; Salles et al., 2016), there is a consensus among researchers regarding the difficulty of standardizing methodologies, both in the use of the equipment and regarding the best body region to capture images (Church et al., 2014; Schaefer et al., 2004). Meteorological factors and differences among the investigated animals should be considered when calibrating equipment, as they influence the radiation patterns detected by the thermographic camera (Montanholi et al., 2015).

The accuracy and F1-score showed approximate values for both $CNN_{F\text{-}RR}$ (76.3 and 76.7%, respectively) and $CNN_{F\text{-}RT}$ (70.5% and 70.9%, respectively). These results indicate that the correct answers were well distributed among the classes. The F1-scores of $CNN_{O\text{-}RR}$ (64.0% and 64.3%) and $CNN_{O\text{-}RT}$ (60.1% and 61.3%) also showed approximate values but were smaller than those of $CNN_{F\text{-}RR}$ and $CNN_{F\text{-}RT}$. In view of this context, the $CNN_{F\text{-}RR}$ and $CNN_{F\text{-}RT}$ classifiers presented better results for predicting the physiological true labelling classes when the data were submitted to the test stages.

As observed in the methodology, the temperature ranges represented by the colour maps for the images of RoI_f and RoI_o differed. With a broader histogram, to generate the RoI_f, the range of 14–39 °C was selected, resulting in images with 250 different colours. In the case of the RoI_o, with a narrower histogram, the range of 24–39 °C was selected, resulting in images with 150 different colours. The adoption of different temperature ranges and/or colour scales can affect the accuracy of the model, since as previously mentioned, they learn through the pattern of colours and shapes observed in the images. The range of colours defined to represent the temperatures of the RoI_o and the humidity present in the ocular area may have influenced the lower precision obtained for the $CNN_{O\text{-}RR}$ and $CNN_{O\text{-}RT}$ models compared with $CNN_{F\text{-}RT}$ and

$CNN_{F\text{-}RT}$. The temperature limits distributed over a wide range may have caused the images in the central temperatures to lose their shape. This region is the most difficult to predict, as it is a transition region between the lowest and highest. As these computational models detect patterns by observing shapes and colours, this may have impaired the performance of the model and contributed to the decrease in precision and recall.

Church et al. (2014) investigated the influence of solar charge, wind speed, and distance between camera and object on infrared temperature measurements. The results showed a difference of 0.56 °C ($p < 0.05$) between eye temperature data for the conditions of direct sunlight exposure or no direct sunlight exposure. Regarding wind speed, when the measurements were performed under conditions of low wind speed (7 km/h) and high wind speed (12 km/h), the temperature reduction was 0.43 °C and 0.78 °C, respectively ($p < 0.001$). Church et al. (2014) also cited that the direct solar charge on Holstein cows has a major influence on body surface temperature, indicating a difference of up to 15 °C above the light (white) parts in the dark (black) parts. This finding reinforces the difficulties encountered in the standardization of methodologies and the need to adopt certain measures to reduce variations in the image capture and analysis process, such as capturing images with the smallest possible variation in the distance between the camera and the analysed object; avoiding measurements with direct sunlight on the object; and monitoring the wind speed, relative humidity, air temperature, and surface humidity of the object.

Although these measures were taken into account, some sources of variation, such as luminosity and wind speed, were not controlled as the thermal images were captured in the field. Note that the luminosity and wind speed were sources of variation in this study since the collections were performed at different times (5 a.m., 12 p.m. and 5 p.m.) and in different seasonal periods (summer and winter). As CNN learning considers the shape and colour of images to learn patterns, the conversion of temperatures to colours influenced the learning of these models. In addition, inherent climatic factors influenced the capture of surface temperatures, and consequently, the conversion to colour scale. Thus, exploring how these sources of variation influence image quality and associated information of period and time of day can further contribute to improving the accuracy of these CNN-based models.

5. Conclusion

This research explored the potential of deep learning tools for use as a rapid interpreter of thermal images of adult dairy cows. Four CNN-based models were designed and compared using the input of thermal images labelled according to thermal conditions related to the respiration rate and rectal temperature of cows. The models have better potential to classify thermal images as belonging to three different classes of stress for the forehead and ocular regions in relation to the other regions of the body surface. The model generated for forehead images labelled from respiration rate is more accurate, although both respiration rate and rectal temperature have served as a good reference for labelling images. The need to standardise the collection of thermal images and to include characteristics of the weather in the models is identified, as these factors directly interfere in the calculation of body surface temperature and colour distribution.

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.

Acknowledgement

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES

Allen, J. D., Hall, L. W., Collier, R. J., & Smith, J. F. (2015). Effect of core body temperature, time of day, and climate conditions on behavioral patterns of lactating dairy cows experiencing mild to moderate heat stress. *Journal of Dairy Science, 98*(1), 118–127. https://doi.org/10.3168/jds.2013-7704

Brown-Brandl, T. M., Eigenberg, R. A., & Purswell, J. L. (2013). Using thermal imaging as a method of investigating thermal thresholds in finishing pigs. *Biosystems Engineering, 114*(3), 327–333. https://doi.org/10.1016/j.biosystemseng.2012.11.015

Burfeind, O., Suthar, V. S., & Heuwieser, W. (2012). Effect of heat stress on body temperature in healthy early postpartum dairy cows. *Theriogenology, 78*(9), 2031–2038. https://doi.org/10.1016/j.theriogenology.2012.07.024

Church, J. S., Hegadoren, P. R., Paetkau, M. J., Miller, C. C., Regev-Shoshani, G., Schaefer, A. L., & Schwartzkopf-Genswein, K. S. (2014). Influence of environmental factors on infrared eye temperature measurements in cattle. *Research in Veterinary Science, 96*(1), 220–226. https://doi.org/10.1016/j.rvsc.2013.11.006

Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., & Grieve, B. (2018). Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry, 98*, 145–152. https://doi.org/10.1016/j.compind.2018.02.016

Kadzere, C. T., Murphy, M. R., Silanikove, N., & Maltz, E. (2002). Heat stress in lactating dairy cows: A review. *Livestock Production Science, 77*, 59–91. https://doi.org/10.1016/S0301-6226(01)00330-X

Kvam, J., & Kongsro, J. (2017). In vivo prediction of intramuscular fat using ultrasound and deep learning. *Computers and Electronics in Agriculture, 142*, 521–523. https://doi.org/10.1016/j.compag.2017.11.020. September.

Misimi, E., Øye, E. R., Sture, Ø., & Mathiassen, J. R. (2017). Robust classification approach for segmentation of blood defects in cod fillets based on deep convolutional neural networks and support vector machines and calculation of gripper vectors for robotic processing. *Computers and Electronics in Agriculture, 139*, 138–152. https://doi.org/10.1016/j.compag.2017.05.021

Montanholi, Y. R., Lim, M., Macdonald, A., Smith, B. A., Goldhawk, C., Schwartzkopf-Genswein, K., & Miller, S. P. (2015). Technological, environmental and biological factors: Referent variance values for infrared imaging of the bovine.

Journal of Animal Science and Biotechnology, 6(1), 1–16. https://doi.org/10.1186/s40104-015-0027-y

Pacheco, V. M., Sousa, R. V. de, Rodrigues, A. V. da S., Sardinha, E. J. de S., & Martello, L. S. (2020). Thermal imaging combined with predictive machine learning based model for the development of thermal stress level classifiers. *Livestock Science, 241*, Article 104244. https://doi.org/10.1016/j.livsci.2020.104244. October 2019.

Polsky, L., & von Keyserlingk, M. A. G. (2017). Invited review: Effects of heat stress on dairy cattle welfare. *Journal of Dairy Science, 100*(11), 8645–8657. https://doi.org/10.3168/jds.2017-12651

Ricci, G. Dela, Silva-Miranda, K. O. da, & Titto, C. G. (2019). Infrared thermography as a non-invasive method for the evaluation of heat stress in pigs kept in pens free of cages in the maternity. *Computers and Electronics in Agriculture, 157*, 403–409. https://doi.org/10.1016/j.compag.2019.01.017. January.

Rodríguez Alvarez, J., Arroqui, M., Mangudo, P., Toloza, J., Jatip, D., Rodríguez, J. M., Teyseyre, A., Sanz, C., Zunino, A., Machado, C., & Mateos, C. (2018). Body condition estimation on cows from depth images using Convolutional Neural Networks. *Computers and Electronics in Agriculture, 155*, 12–22. https://doi.org/10.1016/j.compag.2018.09.039. December 2017.

Salles, M. S. V., da Silva, S. C., Salles, F. A., Roma, L. C., El Faro, L., Bustos Mac Lean, P. A., Lins de Oliveira, C. E., & Martello, L. S. (2016). Mapping the body surface temperature of cattle by infrared thermography. *Journal of Thermal Biology, 62*, 63–69. https://doi.org/10.1016/j.jtherbio.2016.10.003. October.

Schaefer, A. L., Cook, N., Tessaro, S. V., Deregt, D., Desroches, G., Dubeski, P. L., Tong, A. K. W., & Godson, D. L. (2004). *Early detection and prediction of infection using infrared thermography* 1.

Sellier, N., Guettier, E., & Staub, C. (2014). A review of methods to measure animal body temperature in precision farming. *American Journal of Agricultural Science and Technology*. https://doi.org/10.7726/ajast.2014.1008. October.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Van Hertem, T., Rooijakkers, L., Berckmans, D., Peña Fernández, A., Norton, T., Berckmans, D., & Vranken, E. (2017). Appropriate data visualisation is key to Precision Livestock Farming acceptance. *Computers and Electronics in Agriculture, 138*, 1–10. https://doi.org/10.1016/j.compag.2017.04.003

Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., & Xue, Y. (2018). Automatic recognition of lactating sow postures from depth images by deep learning detector. *Computers and Electronics in Agriculture, 147*, 51–63. https://doi.org/10.1016/j.compag.2018.01.023. August 2017.