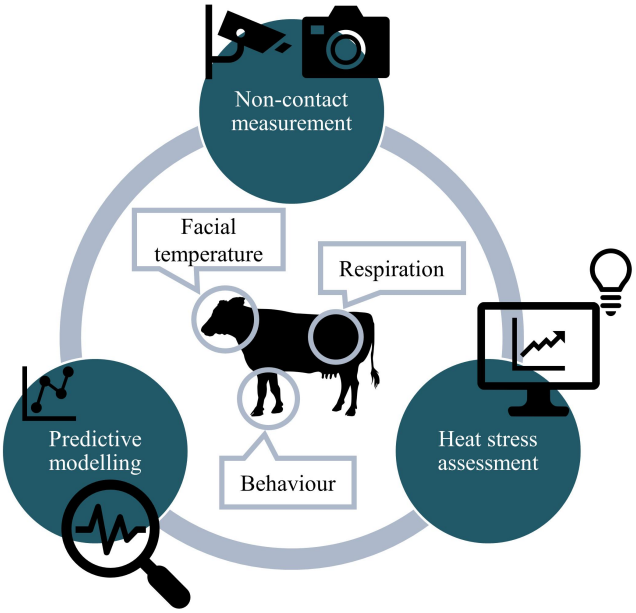
LIÈGE université
Gembloux
Agro-Bio Tech

中国农业科学院
CHINESE ACADEMY OF AGRICULTURAL SCIENCES

Advancing heat stress detection in dairy cows through machine learning and computer vision



Hang Shu

Promoters: Prof. Jérôme Bindelle

Prof. Wensheng Wang

2024

FRENCH COMMUNITY OF BELGIUM
UNIVERSITY OF LIÈGE – GEMBLOUX AGRO-BIO TECH
BELGIUM

Advancing heat stress detection in dairy cows through machine learning and computer vision

Hang Shu

Original essay for graduation as a doctor in agricultural sciences and biological engineering

Promoters: Prof. Jérôme Bindelle
Prof. Wensheng Wang
Civil year: 2024

This document is written in British English. The figures are original and designed by the author; the author must be cited in case of reproducing the results.

Abstract

Heat stress detection in dairy cows has long been connected with production loss. However, the reduction in milk yield lags behind the exposure to heat stress events for about two days. Other stress responses, such as physiological and behavioural changes, are well documented to be activated by dairy cows in the earlier stage of heat stress compared with production loss. Among all candidate indicators, body surface temperatures (BST), respiration rate (RR), and relevant behaviours have been concluded to be the most appropriate indicators due to their high feasibility of acquisition and early response. Vision-based methods are promising for accurate measurements while adhering to animal welfare principles. Meanwhile, predictive models show a non-invasive alternative to obtain these data and can provide useful insights with their interpretations. Thus, this thesis aimed to provide non-invasive solutions to the detection of heat stress in dairy cows by using artificial intelligence techniques. The detailed research content and relevant conclusions are as follows:

An automated tool based on improved UNet was proposed to collect facial BST from five facial landmarks (i.e., eyes, muzzle, nostrils, ears, and horns) on cattle infrared images. The baseline UNet model was improved by replacing the traditional convolutional layers in the decoder with Ghost modules and adding efficient channel attention modules. The improved UNet outperformed other comparable models with the highest mean Intersection of Union of 80.76% and a slightly slower but still good inference speed of 32.7 frames per second (FPS). Agreement analysis reveals small to negligible differences between the temperatures obtained automatically in the area of eyes and ears and the ground truth.

A vision-based method was proposed to measure RR for multiple dairy cows lying on free stalls. The proposed method involved various computer vision tasks (i.e., instance segmentation, object detection, object tracking, video stabilisation, and optical flow) to obtain respiration-related signals and finally utilised Fast Fourier Transform to extract RR. The results show that the measured RR had a Pearson correlation coefficient of 0.945, a root mean square error (RMSE) of 5.24 breaths per minute (bpm), and an intraclass correlation coefficient of 0.98 compared with visual observation. The average processing time and FPS on 55 test video clips (mean ± standard deviation duration of 16 ± 4 s) was 8.2 s and 64, respectively.

A deep learning-based model was proposed to recognise cow behaviours (i.e., drinking, eating, lying, standing-in, and standing-out) that are known to be influenced by heat stress. The YOLOv5s model was selected due to its ability to compress the weight size while maintaining accuracy. It had a mean average precision of 0.985 and an inference speed of 73 FPS. Further validation demonstrates the excellent capacity of the proposed model in measuring herd-level behavioural indicators, with an intraclass correlation coefficient of 0.97 compared with manual observation.

Critical thresholds were determined by using piecewise regression models with environmental indicators as the predictors and animal-based indicators as the outcomes. An ambient temperature (Ta) threshold was determined at 26.1 °C when

the automated measured mean eye temperature reached 35.3 °C. A Ta threshold of 23.6 °C and a temperature-humidity index (THI) threshold of 72 were determined when the automated measured RR reached 61.1 and 60.4 bpm, respectively. In addition, the test dairy herd began to change their standing and lying behaviour at the earliest Ta of 23.8 °C or THI of 68.5.

Four machine learning algorithms were used to predict RR, vaginal temperature (VT), and eye temperature (ET) from 13 predictor variables from three dimensions: production, cow-related, and environmental factors. The artificial neural networks yielded the lowest RMSE for predicting RR (13.24 bpm), VT (0.30 °C), and ET (0.29 °C). The results interpreted with partial dependence plots and Local Interpretable Model-agnostic Explanations show that P.M. measurements and winter calving contributed most to high RR and VT predictions, whereas lying posture, high Ta, and low wind speed contributed most to high ET predictions.

Based on these results, an integrative application of all the proposed measurement, prediction, and assessment methods has been suggested, wherein RGB and infrared cameras are used to measure animal-based indicators, and critical thresholds, along with model interpretation, are used to assess the heat stress state of dairy cows. This strategy ensures timely and thorough cooling of cows in all areas of the dairy farm, thereby minimising the negative impact of heat stress to the greatest extent.

Keywords: precision livestock farming, dairy cows, heat stress, thermal comfort, decision support, artificial intelligence, animal welfare

Résumé

La détection du stress thermique chez les vaches laitières a longtemps été associée à des pertes de production. Cependant, la réduction de la production laitière survient environ deux jours après l'exposition aux événements de stress thermique. D'autres réponses au stress, telles que les changements physiologiques et comportementaux, sont bien documentées comme étant activées par les vaches laitières au stade précoce du stress thermique par rapport aux pertes de production. Parmi tous les indicateurs potentiels, les températures de surface corporelle (TSC), le taux de respiration (TR) et les comportements pertinents ont été conclus comme étant les indicateurs les plus appropriés en raison de leur grande faisabilité d'acquisition et de leur réponse précoce. Les méthodes basées sur la vision sont prometteuses pour des mesures précises tout en respectant les principes de bien-être animal. Parallèlement, les modèles prédictifs offrent une alternative non invasive pour obtenir ces données et peuvent fournir des informations utiles avec leurs interprétations. Ainsi, cette thèse visait à fournir des solutions non invasives pour la détection du stress thermique chez les vaches laitières en utilisant des techniques d'intelligence artificielle. Le contenu de la recherche détaillée et les conclusions pertinentes sont les suivants :

Un outil automatisé basé sur un UNet amélioré a été proposé pour collecter la TSC faciale à partir de cinq points de repère faciaux (c'est-à-dire les yeux, le museau, les narines, les oreilles et les cornes) sur des images infrarouges de bovins. Le modèle UNet de base a été amélioré en remplaçant les couches convolutionnelles traditionnelles dans le décodeur par des modules Ghost et en ajoutant un modules d'attention aux canaux efficace. L'UNet amélioré a surpassé d'autres modèles comparables avec la plus haute intersection moyenne de l'union de 80,76 % et une vitesse d'inférence légèrement plus lente mais toujours bonne de 32,7 images par seconde (IPS). L'analyse de l'accord révèle de petites à négligeables différences entre les températures obtenues automatiquement dans la zone des yeux et des oreilles et la vérité au sol.

Une méthode basée sur la vision a été proposée pour mesurer le TR de plusieurs vaches laitières couchées sur des stalles libres. La méthode proposée impliquait diverses tâches de vision par ordinateur (c'est-à-dire segmentation d'instances, détection d'objets, suivi d'objets, stabilisation vidéo et flux optique) pour obtenir des signaux liés à la respiration, puis utilisait la transformation de Fourier rapide pour extraire le TR. Les résultats montrent que le TR mesuré avait un coefficient de corrélation de Pearson de 0,945, une erreur quadratique moyenne (RMSE) de 5,24 respirations par minute (rpm) et un coefficient de corrélation intraclasse de 0,98 par rapport à l'observation visuelle. Le temps de traitement moyen et le IPS sur 55 clips vidéo de test (durée moyenne ± écart type de 16 ± 4 s) étaient respectivement de 8.2 s et 64.

Un modèle basé sur l'apprentissage profond a été proposé pour reconnaître les comportements des vaches (c'est-à-dire boire, manger, se coucher, se tenir debout et sortir) qui sont connus pour être influencés par le stress thermique. Le modèle

YOLOv5s a été sélectionné en raison de sa capacité à comprimer la taille des poids tout en maintenant la précision. Il avait une précision moyenne moyenne de 0,985 et une vitesse d'inférence de 73 IPS. Une validation ultérieure démontre l'excellente capacité du modèle proposé à mesurer les indicateurs comportementaux au niveau du troupeau, avec un coefficient de corrélation intraclasse de 0,97 par rapport à l'observation manuelle.

Des seuils critiques ont été déterminés en utilisant des modèles de régression segmentée avec des indicateurs environnementaux comme prédicteurs et des indicateurs basés sur les animaux comme résultats. Un seuil précoce de température ambiante (Ta) a été déterminé à 26,1 °C lorsque la température moyenne des yeux mesurée automatiquement atteignait 35,3 °C. Un seuil Ta de 23,6 °C et un seuil d'indice température-humidité (ITH) de 72 ont été déterminés lorsque le TR mesuré automatiquement a atteint 61,1 et 60,4 rpm, respectivement. Le troupeau laitier de test a commencé à modifier son comportement debout et couché au Ta le plus précoce de 23,8 °C ou à l'ITH de 68,5.

Quatre algorithmes d'apprentissage automatique ont été utilisés pour prédire le TR, la température vaginale (TV) et la température des yeux (TY) à partir de 13 variables prédictives provenant de trois dimensions : production, liées aux vaches et environnementales. Les réseaux de neurones artificiels ont donné le RMSE le plus bas pour la prédiction du TR (13,24 rpm), de la TV (0,30 °C) et de la TY (0,29 °C). Les résultats interprétés avec des tracés de dépendance partielle et des explications interprétables de modèle agnostique local montrent que les mesures PM et la mise bas en hiver contribuaient le plus aux prévisions élevées de TR et de TV, tandis que la posture couchée, la haute Ta et la faible vitesse du vent contribuaient le plus aux prévisions élevées de TY.

Sur la base de ces résultats, une application intégrée de toutes les méthodes de mesure, de prédiction et d'évaluation proposées a été suggérée, dans laquelle des caméras RVB et infrarouges sont utilisées pour mesurer les indicateurs basés sur les animaux, et des seuils critiques, ainsi que des interprétations de modèle, sont utilisés pour évaluer l'état de stress thermique des vaches laitières. Cette stratégie assure un refroidissement opportun et complet des vaches dans toutes les zones de la ferme laitière, minimisant ainsi au maximum l'impact négatif du stress thermique.

Mots-clés: élevage de précision, vaches laitières, stress thermique, confort thermique, support à la décision, intelligence artificielle, bien-être animal

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my promoters, Professor **Jérôme Bindelle** (Precision livestock and nutrition unit, University of Liege – Gembloux Agro-Bio Tech, Belgium) and Professor **Wensheng Wang** (Agricultural Information Institute, Chinese Academy of Agricultural Sciences, China) for their patient guidance, erudite knowledge, and valuable suggestions from the beginning to the end of this Ph.D. project.

I gratefully acknowledge the financial scholarship from the **China Scholarship Council,** the funding from **Inner Mongolia Autonomous Region Science and Technology Major Project (2020ZD0004)**, the support provided for studying and researching facilities from the **University of Liege – Gembloux Agro-Bio Tech** (Belgium) and the **Chinese Academy of Agricultural Sciences** (China), the assistance in data collection, project communication, and research guidance from Professor **Xianhong Gu** (State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, China) and Doctor **Leifeng Guo** (Agricultural Information Institute, Chinese Academy of Agricultural Sciences, China), as well as the aid in providing experimental sites and animals from **Yinxiangweiye dairy farm**. Undoubtably, their support has been instrumental in shaping the course of my academic journey.

I would like to wholeheartedly express my profound gratitude to my cherished family, with a special mention of **my parents**, **my parents-in-law**, **my grandparents**, and **my beloved wife**. Their unwavering belief in my abilities, understanding, and the immeasurable sacrifices they have made alongside me have always been my rock and inspiration. I extend my sincerest wishes for success and happiness to my wife as she embarks on her Ph.D. study. I am certain that she will excel in her academic endeavours and make valuable contributions to her chosen field.

My warm thanks are also dedicated to **my experimental cows**, who played an indispensable role in this project. I would like to emphasise that this appreciation extends to every cow involved, including **the memorable cow numbered 4971**, despite the playful incidents where you tried to bump into me throughout the entire study. These remarkable friends served as the subjects of my study, patiently enduring the experimental procedures and providing crucial data for analysis. Their cooperation and participation were instrumental in gaining insights and drawing meaningful conclusions.

The last but not the least, I would like to acknowledge that, apart from my own efforts, the success of this thesis relies largely on the encouragement and guidelines I have received from many others. I would like to take this opportunity to express my gratitude to all those who have generously offered their help and provided their best advice throughout this journey, including, but not limited to, **my friends and colleagues**.

Table of contents

Abstract 3
Résumé 5
Acknowledgements 7
Table of contents 8
List of figures 14
List of tables 20
List of acronyms 22
Chapter 1 25
Measurement, prediction, and assessment of heat stress in dairy cattle: a review 25
1. Heat stress and on-farm mitigation strategies 28
1.1. Heat stress 28
1.2. State-of-the-art solutions for heat stress mitigation 28
1.2.1. Environmental regulation 29
1.2.2. Nutrition management 30
1.2.3. Genetic selection for thermotolerance 31
2. Detection of heat stress using animal-based indicators 32
3. Measurement of heat stress in dairy cows 34
3.1. Core body temperatures 34
3.1.1. Rectal temperature 37
3.1.2. Vaginal temperature 37
3.1.3. Reticulorumen temperature 38
3.1.4. Tympanic temperature 38
3.1.5. Subcutaneous temperature 39
3.1.6. Milk temperature 39
3.2. Body surface temperatures 40
3.3. Respiratory dynamics 41
3.3.1. Respiration rate 41
3.3.2. Panting score 45

3.4. Behavioural changes 46
4. Prediction of heat stress in dairy cows 48
4.1. Empirical models 48
4.2. Mechanistic models 52
5. Assessment of heat stress in dairy cows 55
5.1. Development of critical thresholds 55
5.2. Sensitivity of animal-based indicators in response to heat stress 60
5.3. Selection of thermal indices for specific environments 62
6. Future development of heat stress detection methods 64
6.1. Better acquisition of animal-based indicators by PLF techniques 64
6.2. IoT technologies help with precision management 66
7. Conclusion 67
Chapter 2 69
Problematic, research aim, thesis outline, and experimental design 69
1. Problematic 71
2. Research aim 71
3. Thesis outline 71
4. Experimental design 72
Context – Chapter 3 74
Chapter 3 75
Improved UNet for facial temperature analysis in dairy cow infrared images 75
1. Introduction 78
2. Materials and methods 79
2.1. Data acquisition 79
2.2. Data pre-processing 80
2.3. Segmentation network architectures 81
2.3.1. UNet model 81
2.3.2. Improved UNet model 81
2.4. Segmentation network training 85
2.5. Ablation and comparison studies 85

2.6. External validation 86
2.7. Performance evaluation 86
2.8. Data analysis 87
2.8.1. Agreement with ground truth 87
2.8.2. Threshold development 87
3. Results and discussion 88
3.1. Overview of the datasets 88
3.2. Results of training and ablation study 88
3.3. Segmentation results of the improved UNet 90
3.4. Comparison with other semantic segmentation models 91
3.5. Comparison of the generalisability of the two splitting strategies 94
3.6. Agreement of automated and manual methods with ground truth 95
3.7. Response of FT to heat stress 98
3.8. Limitations and perspectives 100
4. Conclusions 102
Context – Chapter 4 104
Chapter 4 105
Non-contact respiration rate measurement of multiple cows using computer vision 105
1. Introduction 108
2. Materials and methods 109
2.1. Experimental design 109
2.2. Data preparation 110
2.3. Muti-target RR measurement for lying cows based on deep learning and optical flow method 111
2.3.1. Architecture, training, and evaluation of deep learning networks 112
2.3.2. Optical flow 113
2.3.3. Object tracking and stabilisation 114
2.3.4. Optical flow calculation at abdominal areas 115
2.3.5. Signal processing for extracting RR 116
2.4. Performance analysis 116

2.5. Threshold development for heat stress 118
3. Results and discussion 118
3.1. Performance of cow segmentation and flank detection 118
3.2. Performance of RR measurement based on optical flow method 120
3.3. Factors that might have affected the results 123
3.4. Response of RR to heat stress 125
3.5. Limitations and perspectives 126
4. Conclusions 127
Context – Chapter 5 128
Chapter 5 129
Determining heat stress in a dairy herd via automated behaviour recognition 129
1. Introduction 132
2. Materials and methods 133
2.1. Experimental design 133
2.1.1. Housing, animals, and management 133
2.1.2. Behavioural and environmental measurement 134
2.2. Development of behaviour recognition model 134
2.2.1. Data preparation 134
2.2.2. Deep learning algorithm and transfer learning 135
2.2.3. Performance evaluation 136
2.3. Behavioural indicator calculation and heat stress determination 137
2.3.1. Behaviour recognition 137
2.3.2. Detection filtering 137
2.3.3. Behavioural parameters 138
2.3.4. Threshold development 138
3. Results and discussion 139
3.1. Performance of behaviour recognition 139
3.2. Behavioural pattern under heat stress conditions 141
3.3. Critical threshold for determining the onset of heat stress 145
3.4. Strength and limitations 149

4. Conclusions 151
Context – Chapter 6 152
Chapter 6 153
Predicting physiological responses of dairy cows using machine learning 153
1. Introduction 156
2. Materials and methods 157
2.1. Location, facilities, and animals 157
2.2. Variables 158
2.3. Data collection 159
2.4. Data processing 160
2.5. Predictive modelling 161
2.5.1. Random forests 161
2.5.2. Gradient boosting machines 161
2.5.3. Artificial neural networks 162
2.5.4. Regularised linear regression 162
2.6. Recognition of heat stress state 162
3. Results and discussion 163
3.1. Data cleaning and descriptive statistics 164
3.2. Model performance 166
3.3. Model interpretation 168
3.4. Recognition of heat stress state 172
3.5. Limitations and perspectives 174
4. Conclusions 177
Chapter 7 179
General discussion and conclusion 179
1. General discussion 181
1.1. The effect of data-splitting strategies on model performance 181
1.2. Choice of individual and herd-level heat stress detection 184
1.3. A promising integration of the proposed measurement methods 185
1.3.1. Role of predictive models in this integrative system 187

1.3.2. Heat stress assessment based on comprehensive information 188
1.3.3. Pros and cons from a sustainable agriculture perspective 188
2. Conclusions 189
References 191
Appendices 227
Appendix 1. Supplementary materials 229
Appendix 2. List of publications 237
Accepted publications (peer reviewed) 237
Presentations at international conferences (peer reviewed) 238

List of figures

Figure 1-1: Relationship between environmental stimuli and induced animal response. 28

Figure 1-2: Detection of heat stress and mitigations that could be taken. 33

Figure 1-3: Measurement, prediction, and assessment of heat stress in dairy cows. 65

Figure 2-1: Technical route of the project 'Advancing heat stress detection in dairy cows through machine learning and computer vision'. 72

Figure 3-1: Examples of cattle infrared images and their ground-truth annotations for facial landmarks. 80

Figure 3-2: Architecture of the baseline UNet model and the improved UNet model by replacing convolutions in the decoder with Ghost and GhostECA modules (shown in dashed boxes, where C and C' represent the input and output channel numbers, respectively, and C_=C'/4). 82

Figure 3-3: Architecture of the efficient channel attention (ECA) module. 85

Figure 3-4: Distribution of temperature-humidity index (THI) during photography in (a) image-and (b) cow-level datasets, summarised by training (including validation and before augmentation) and testing sets, respectively. 88

Figure 3-5: Loss and mean Intersection over Union (mIoU) curve of the proposed architecture trained using the cow-level dataset. The dashed lines show the epoch (153) with the lowest validation loss (0.137). 89

Figure 3-6: Detailed segmentation results on the cow-level testing set (n = 93). (a) Predictions on some example images against varying thermal conditions classified by temperature-humidity index (THI), shown at camera angles ranging from -90° to 90°. (b) Per-class Intersection over Union (IoU), Recall, and Precision. 91

Figure 3-7: Segmentation results of different semantic segmentation models on the cow-level testing set (n = 93). (a) FCN-VGG16; (b) PSPNet-MobileNetV2; (c) PSPNet-ResNet50; (d) DeepLabV3+-MobileNetV2; (e) DeepLabV3+-Xception; (f) UNet-VGG16; (g) UNet-ResNet50; (h) SegFormer-B5; (i) Proposed. 93

Figure 3-8: Per-class Intersection over Union of the improved models, trained separately using the image- and cow-level datasets, on the (a) testing and (b) external testing sets, respectively. 94

Figure 3-9: Predictions of two example images from the external testing set with cows wearing two ear tags. 95

Figure 3-10: Overview of the (a) maximum and (b) mean temperatures of ground-truth annotation, predicted segmentation by the proposed model, and manual collection on the cow-level testing set (n = 93). 96

Figure 3-11: Bland-Altman plots showing the agreement between the predicted segmentation of the improved UNet model and ground-truth annotation on the cow-level testing set (n = 93) in terms of the maximum and mean temperatures (Tmax and Tmean) of five facial landmarks. The solid and dashed lines represent mean difference and 95% limits of agreement, respectively. Datapoints are coloured by temperature-humidity index (THI). 96

Figure 3-12: Bland-Altman plots showing the agreement between the predicted segmentation of the baseline UNet model and ground-truth annotation on the cow-level testing set (n = 93) in terms of the maximum and mean temperatures (Tmax and Tmean) of five facial landmarks. The solid and dashed lines represent mean difference and 95% limits of agreement, respectively. Datapoints are coloured by temperature-humidity index (THI). 97

Figure 3-13: Bland-Altman plots showing the agreement between manual collection and ground-truth annotation on the cow-level testing set (n = 93) in terms of the maximum and mean temperatures (Tmax and Tmean) of five facial landmarks. The solid and dashed lines represent mean difference and 95% limits of agreement, respectively. Datapoints are coloured by temperature-humidity index (THI). 98

Figure 3-14: Automated measurements of the maximum temperature (Tmax) and mean temperature (Tmean) of eyes and ears (n = 1000) and their fitted profiles from linear regression (dashed line) and piecewise regression (solid line) models with (a) ambient temperature and (b) temperature-humidity index as the predictor, respectively. The breakpoints are marked as a black triangle above the x-axis. 99

Figure 4-1: Example annotations for cow instances (top) and flank objects (bottom). Polygons in red and green represent standing and lying cow instances, respectively. Rectangles in yellow and blue represent the flank objects on the hind leg side and the opposite side, respectively. 110

Figure 4-2: Pipeline for muti-target respiration rate (RR) measurement for lying cows based on deep learning and optical flow method. Lucas-Kanade = LK. 112

Figure 4-3: Three regions of interest to compute sparse optical flows: (a) cow bounding box, (b) cow mask, and (c) cow mask overlapped by flank bounding boxes. The schematic of the regions of interest and the computed optical flows are shown in the top and bottom rows, respectively. 115

Figure 4-4: Description of the dataset for the development of respiration rate measurement. The 55 video clips are sorted by datetime (mm-dd hh:mm). 117

Figure 4-5: Example predictions of cow segmentation (top) and flank detection (bottom) on the pen-level testing set. Predictions in red and green represent standing and lying cow instances, respectively. Predictions in yellow and blue represent the flank objects on the hind leg side and the opposite side, respectively. 120

Figure 4-6: Statistical analysis between automated and manual respiration rate measurements (breaths per minute (bpm), n = 180). Automated measurements relied

on optical flow- and colour-based methods. (a) Linear regression plot, where the Pearson correlation coefficient of 0.945 and 0.787, respectively, and the root mean square error of 5.24 and 9.79 bpm, respectively. (b) Bland-Altman plot, where the solid lines represent the mean difference: 1.08 and 1.58 bpm, respectively, and the dashed lines represent 95% limits of agreement: -8.97 to 11.13 bpm and -17.35 to 20.51 bpm, respectively. 121

Figure 4-7: Inference speed of the automated respiration rate measurements relied on optical flow- and colour-based methods. (a) Processing time of the 55 clips sorted by datetime (mm-dd hh:mm). (b) Relationship between the number of available cows to be processed and processing time. (c) Relationship between the number of available cows to be processed and average frames per second. 122

Figure 4-8: The effect of flank colour on measured respiration rate (RR, breaths per minute (bpm)). (a) Bland-Altman plot between automated measurements and manual observations. (b) Bar plot of one-way analysis of variance. No texture 1 = No texture but with clear sand or ribs (n = 34), No texture 2 = No texture and no clear reference (n = 7), With textures = with clear textures (n = 139). 123

Figure 4-9: The effect of flank occlusion on measured respiration rate (RR, breaths per minute (bpm)). (a) Bland-Altman plot between automated measurements and manual observations. (b) Bar plot of one-way analysis of variance. Partially occluded = Flank partially occluded by cubicles or pillars (n = 52), No occlusion = Flank not occluded (n = 128). 124

Figure 4-10: Visualisation of all cow bounding boxes in the last video frame, coloured by the difference between automated measured and manual observed respiration rate (breaths per minute, bpm). 124

Figure 4-11: Automated measurements of respiration rate (RR; breaths per minute (bpm), n = 180) and their fitted profiles from linear regression (dashed line) and piecewise regression (solid line) models with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictors. The breakpoints are marked as black triangles above the x-axis. 125

Figure 5-1: Schematic of detection filtering taking one camera for example. The predefined red polygon represents the area of interest and only the bounding boxes with their lower centres in the area are kept. 137

Figure 5-2: Comparison of training loss and validation mean average precision (mAP) for three YOLOv5 architectures. 140

Figure 5-3: Detailed performance of the YOLOv5s model. (a) F1 confidence curve with the black vertical dashed line indicating the best F1 score at a confidence threshold of 0.696. (b) Confusion matrix (normalised by column) with confidence threshold set to 0.696. (c) Example results of behavioural recognition with confidence threshold set to 0.696. 141

Figure 5-4: Comparison of herd-level behavioural parameters measured half-hourly by manual and automated methods from 10:30 to 15:00 h during the three-day experiment. Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; Standing-in% = percentage of cows standing-in; Standing-out% = percentage of cows standing-out; CI = comfort index; SUI = stall-use index; CSI = cow stress index. 142

Figure 5-5: Overall variation of indoor ambient temperature (Ta, ℃), relative humidity (RH, %), and temperature-humidity index (THI) during the 10-day experiment with a measurement interval of 10 min. 143

Figure 5-6: Herd-level behaviour distribution measured half-hourly by the proposed automated method, as well as the corresponding ambient temperature (Ta, ℃) and temperature-humidity index (THI), averaged during the 10-day experiment. 143

Figure 5-7: Herd-level behavioural indicators measured half-hourly by the proposed automated method, as well as the corresponding ambient temperature (Ta, ℃) and temperature-humidity index (THI), averaged during the 10-day experiment. Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index. 144

Figure 5-8: Spearman's rank correlation coefficients between herd-level behavioural indicators measured half-hourly by the proposed automated method from 10:30 to 15:00 h during the 10-day experiment and ambient temperature (Ta, ℃) and temperature-humidity index (THI). Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index. 145

Figure 5-9: Automated measurements of herd-level behavioural indicators and their fitted profiles from linear regression (green) and piecewise regression (red) models with (a) ambient temperature and (b) temperature-humidity index as the predictor, respectively. Breakpoints are marked as a black triangle above the x-axis. Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index. 146

Figure 5-10: Automated measurements of the percentage of cows lying and their fitted profile from advanced piecewise regression with ambient temperature as the predictor. Random effects of time of day (h) were introduced for intercept, slope difference, and breakpoint. Breakpoints are marked as a black triangle above the x-axis. 148

Figure 5-11: Automated measurements of stall-use index and their fitted profile from advanced piecewise regression with ambient temperature as the predictor. Random effects of time of day (h) were introduced for intercept, slope difference, and breakpoint. Breakpoints are marked as a black triangle above the x-axis. 149

Figure 5-12: Automated measurements of cow stress index and their fitted profile from advanced piecewise regression with ambient temperature as the predictor. Random effects of time of day (h) were introduced for intercept, slope difference, and breakpoint. Breakpoints are marked as a black triangle above the x-axis. 150

Figure 6-1: Flow chart showing the strategic plan of the present study. RF = random forests; GBM = gradient boosting machines; ANN = artificial neural networks; RLR = regularised linear regression; RR = respiration rate; VT = vaginal temperature; ET = eye temperature. 159

Figure 6-2: Daily patterns of (a) ambient temperature (Ta), (b) relative humidity (RH), and (c) temperature-humidity index (THI). Zones in green, purple, and yellow represent three experimental phases, respectively. Ta_mean, Ta_max, and Ta_min are the mean, maximum, and minimum of daily Ta, respectively; RH_mean, RH_max, and RH_min are the mean, maximum, and minimum of daily RH, respectively; THI_mean, THI_max, and THI_min are the mean, maximum, and minimum for daily THI, respectively. 164

Figure 6-3: Measured and predicted (a) respiration rate (RR), (b) vaginal temperature (VT), and (c) eye temperature (ET) from the overall best models (artificial neural networks) on the testing sets. The data points in green, purple, and yellow belong to three experimental phases, respectively. The red lines represent the linear regression. The dotted lines represent the diagonal line with a slope of 1. RMSE = root mean square error; R^2 = coefficient of determination. 167

Figure 6-4: Partial dependence plots of the overall best models (artificial neural networks) on the testing sets showing the effect of production (a-c), cow-related (d-i), and environmental factors (j-m) on respiration rate (RR), vaginal temperature (VT), and eye temperature (ET). The 95% confidence intervals for continuous and categorical variables are shown with dotted lines and error bars, respectively. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day). 170

Figure 6-5: Local interpretation heatmap of the overall best respiration rate model (an artificial neural network) showing the influence of different predictor variables on the prediction of 426 observations of the testing set. The top five influential predictor variables that best explained each observation were used for plotting. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day). 172

Figure 6-6: Local interpretation heatmap of the overall best vaginal temperature model (an artificial neural network) showing the influence of different predictor variables on the prediction of 235 observations of the testing set. The top five influential predictor variables that best explained each observation were used for plotting. DMY1D = daily milk yield of the day before the test day (kg/day);

DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day). 173

Figure 6-7: Local interpretation heatmap of the overall best eye temperature model (an artificial neural network) showing the influence of different predictor variables on the prediction of 288 observations of the testing set. The top five influential predictor variables that best explained each observation were used for plotting. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day). 174

Figure 6-8: Local interpretation heatmaps of the overall best models (artificial neural networks) plotting the top five influential predictor variables of five observations with the highest prediction selected from the testing set of (a) respiration rate, (b) vaginal temperature, and (c) eye temperature. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day). 175

Figure 6-9: Local interpretation heatmaps of the overall best models (artificial neural networks) plotting the top five influential predictor variables of five observations with the lowest prediction selected from the testing set of (a) respiration rate, (b) vaginal temperature, and (c) eye temperature. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day). 176

Figure 7-1: A systematic application of the proposed measurement methods, predictive models, and assessment. (a) Vision-based cow behaviour and respiration rate measurement informing the initiation and cessation of cooling measures; (b) infrared thermography-based facial temperature measurement at the selection gate before milking, allocating heat-stressed cows to (c) an intensive cooling session, and non-heat-stressed cows to (d) the milking parlour. 186

List of tables

Table 1-1: Summary of the state-of-the-art measurement methods for dairy cattle body temperatures. 35

Table 1-2: Summary of the state-of-the-art measurement methods for dairy cattle respiration-related data. 42

Table 1-3: Summary of the state-of-the-art measurement methods for dairy cattle behavioural indicators. 47

Table 1-4: Summary of the recent empirical models for predicting dairy cattle body temperatures and respiratory dynamics 50

Table 1-5: Summary of the recent mechanistic models for predicting dairy cattle body temperatures and respiratory dynamics. 53

Table 1-6: Summary of heat stress thresholds in dairy cattle based on body temperatures, respiratory dynamics, and behavioural changes. 57

Table 1-7: Summary of the thermal indices mentioned in this review. 63

Table 3-1: Overview of the datasets split using two strategies. 81

Table 3-2: Structure of the improved UNet. 83

Table 3-3: Performance of ablation study trained using the cow-level dataset (n = 93). 89

Table 3-4: Performance of comparison study trained using the cow-level dataset (n = 93). 92

Table 3-5: mean Intersection over Union of the improved UNet models, trained separately using the (a) image- and (b) cow-level datasets, on the testing and external testing sets, respectively. 94

Table 3-6: Parameter estimates (mean ± standard error) of the piecewise regression models predicting eye and ear temperatures with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively (n = 1000). 100

Table 4-1: Overview of the cow segmentation datasets split using two strategies. 111

Table 4-2: Overview of the flank detection datasets split using two strategies. .111

Table 4-3: Performance of the three YOLOv5-seg architectures for segmenting cow instances, each trained separately using the image- and pen-level datasets. ... 118

Table 4-4: Performance of the three YOLOv5 architectures for detecting flank objects, each trained separately using the image- and pen-level datasets. 119

Table 4-5: Parameter estimates (mean ± standard error) of the piecewise regression models predicting respiration rate with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively (n = 180). 125

Table 5-1: Definition of the target cow behaviours. 135

Table 5-2: Overview of the behaviour recognition dataset. Images were extracted using 6-min scan sampling from 10:30 to 15:00 h during the three-day experiment. 135

Table 5-3: Performance comparison of three YOLOv5 architectures. 140

Table 5-4: Parameter estimates (mean ± standard error) of basic piecewise regression models with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively. Behavioural indicators were measured half-hourly by the proposed automated method from 11:00 to 14:30 h during the 10-day experiment. [a] 147

Table 6-1: Summary of the cows at the beginning of each phase. 158

Table 6-2: Summary of the datasets for predicting respiration rate (RR), vaginal temperature (VT), and eye temperature (ET). Continuous variables are summarised as mean ± standard deviation, categorical variables are summarised as n (%). 165

Table 6-3: Performance of four candidate algorithms in predicting respiration rate (RR, breaths per min), vaginal temperature (VT, °C), and eye temperature (ET, °C) on the training sets, 5-fold cross-validation, and testing sets. 166

Table 6-4: Performance of the overall best models (artificial neural networks) in predicting respiration rate (RR, breaths per min), vaginal temperature (VT, °C), and eye temperature (ET, °C) on the testing sets summarised by three experimental phases. 167

Table 6-5: Performance of four candidate algorithms in predicting respiration rate (RR, breaths per min), vaginal temperature (VT, °C), and eye temperature (ET, °C) on the cow-level training sets, 5-fold cross-validation, and testing sets. 168

Table 6-6: Performance of the proposed classifiers, temperature-humidity index (THI) classifiers, adjusted temperature-humidity index (THIadj) classifier, and equivalent temperature index for cattle (ETIC) classifier in recognising heat stress state on the testing sets based on measured respiration rate (RR) and vaginal temperature (VT), respectively. 177

List of acronyms

AIM:	age in months
ANN:	artificial neural networks
bpm:	breaths per minute
BST:	body surface temperatures
CBT:	core body temperatures
CI:	comfort index
CNN:	convolutional neural networks
CSI:	cow stress index
DIM:	days in milk
DMY:	daily milk yield
drinking%:	percentage of cows drinking
eating%:	percentage of cows eating
ECA:	efficient channel attention
ET:	eye temperature
ETIC:	equivalent temperature index for cattle
FLOPs:	floating-point operations
FN:	false negative
FP:	false positive
FPS:	frames per second
FT:	facial temperatures
GBM:	gradient boosting machines
GF:	Gunnar-Farneback
GhostECA:	Ghost with efficient channel attention
IoT:	internet of things
IoU:	Intersection over Union
IRT:	infrared thermography
LIME:	Local Interpretable Model-agnostic Explanations
LK:	Lucas-Kanade
lying%:	percentage of cows lying

mIoU:	mean Intersection over Union
ML:	machine learning
mPA:	mean pixel accuracy
MT:	milk temperature
PLF:	precision livestock farming
Pr:	precision
PS:	panting score
R:	recall
R^2:	coefficient of determination
RET:	reticulorumen temperature
RF:	random forests
RFID:	radiofrequency identification
RH:	relative humidity
RMSE:	root mean square error
RoI:	regions of interest
RR:	respiration rate
RT:	rectal temperature
SR:	solar radiation
ST:	subcutaneous temperatures
SUI:	stall-use index
Ta:	ambient temperature
THI:	temperature-humidity index
THIadj:	adjusted temperature-humidity index
TN:	true negative
TP:	true positive
TT:	tympanic temperature
VT:	vaginal temperature
WS:	wind speed

Chapter 1

Measurement, prediction, and assessment of heat stress in dairy cattle: a review

'*All models are wrong, but some are useful.*'

George Box

Adapted from:

Shu, H., Wang, W., Guo, L., Bindelle, J. *Recent advances on early detection of heat strain in dairy cows using animal-based indicators: A review*. Animals, 2021. https://doi.org/10.3390/ani11040980

Abstract

In pursuit of precision livestock farming, the automated measurement for heat stress response has been more and more valued. Efforts have been made recently to use more sensitive animal-based indicators in the hope of improving decision-making regarding heat abatement in dairy farms. The present review focuses on the recent efforts in developing detection methods of heat stress in dairy cows using animal-based indicators including body temperatures, respiratory dynamics, and behavioural changes. For every candidate animal-based indicator, state-of-the-art data acquisition techniques, including both direct measurements and predictive models, as well as their developed thresholds were summarised. Body surface temperature, respiration rate, and relevant behaviours were concluded to be the best indicators due to their ease of acquisition and sensitivity to heat stressors. Future studies should continue to evaluate thermal indices and customise thresholds at specific farm locations due to inconsistencies in environmental conditions and animal information. Wearable and indwelling devices on the market are now available for real-time measurement of basic behaviours and core body temperature while vision-based non-contact methods are gaining popularity due to their non-invasive measurement. The development of measurement methods can accelerate the development of data-driven predictive models by allowing access to more relevant parameters, and these models in turn can guide the measurement with their interpretations. By utilising internet of things technologies, a comprehensive strategy based on both animal- and environment-based indicators is expected to increase the precision of detecting heat stress in dairy cows, allowing for more accurately identifying high-risk animals, implementing targeted interventions, and ultimately minimising the impact of heat stress.

Keywords: stress response, detection, precision livestock farming, thermal comfort, animal welfare

1. Heat stress and on-farm mitigation strategies

1.1. Heat stress

Thermoregulation, defined as the process to keep the balance between heat production and heat loss, is important for homeotherms to maintain homeothermy and homeostasis [1]. Heat production in dairy cows mainly comes from basal metabolism, rumen fermentation, nutrient absorption, growth, lactation, gestation, immunisation, and exercise, whereas heat loss is mainly through heat transfer mechanisms: conduction, convection, radiation, and evaporation [2].

Cows can maintain a balance of heat production and heat dissipation in the thermoneutral zone, in which they perform the lowest physiological and immune costs, and the highest productivity [1]. When the thermal environment exceeds the thermoneutral zone and suppresses the efficiency of non-evaporative heat loss, evaporative heat loss, or both [3,4], various mechanisms are activated in dairy cows to dissipate excess heat and maintain homeothermy [5]. This demand made by the environment on animals for heat dissipation is defined as heat stress [6]. **Figure 1-1** shows some main indicators representing environmental stimuli and their induced animal responses.

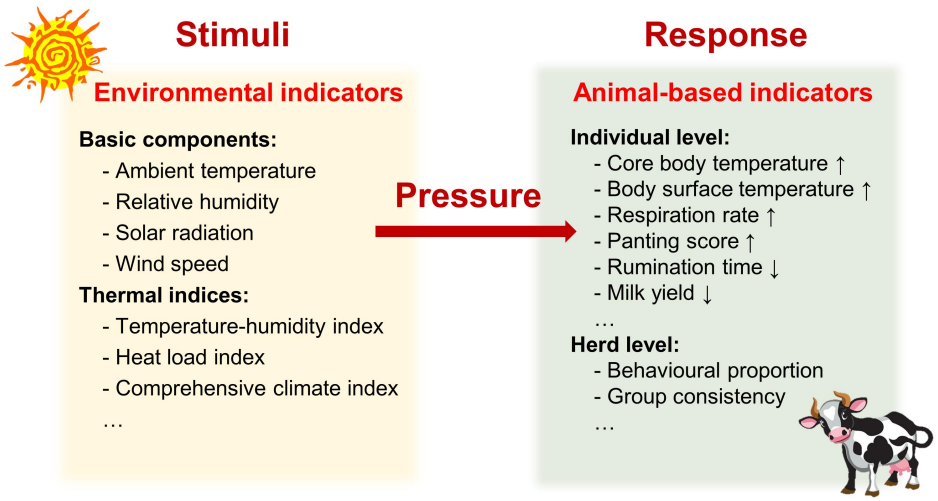


Figure 1- 1: Relationship between environmental stimuli and induced animal response.

1.2. State-of-the-art solutions for heat stress mitigation

Present-day management of heat stress in dairy farms leans on several complementary strategies. Environmental regulation, nutritional intervention, and genetic selection for thermotolerant breeds are the three pillars on which the prevention and the mitigation of heat stress for cows in dairy farms rely [7,8].

1.2.1. Environmental regulation

Regulation of the barn environment aims to improve cow comfort by changing perceived heat stress and by supporting more efficient heat dissipation. These strategies include providing shade and cooling measures, as well as optimising barn structures.

Providing shade involves creating shelters or shaded areas where cows can retreat from direct sunlight which contains direct solar radiation (SR) to increase the heat balance of cows. These shaded areas have significantly lower ambient temperature (Ta) compared to areas that are exposed to direct sunlight. Relevant structures can range from natural trees or artificial shades on the pastures to well-constructed barns in housing systems. A roof with good insulation and proper slope can not only reduce the radiation of sunlight on the roof but also promote natural ventilation which is an effective, least-cost option to refresh indoor air conditions and reduce the difference between the inside and outside temperature [9].

Additional cooling measures are always required during heat stress seasons when indoor temperatures surpass the upper critical threshold of the cows' thermoneutral zone. Among all cooling measures, active ventilation systems are the most basic one. These systems promote airflow inside the barn thus affecting both convective and evaporative heat losses. They can be very efficient when cows are sweating as the wind can remove body heat while sweeping away the just-evaporated water vapour. Commonly used ventilation systems range from more affordable panel fans, basket fans, and ceiling fans to more expensive tunnel ventilation [10]. Techniques such as low-profile cross ventilation [11] and numerical simulation using computational fluid dynamics and particle image velocimetry [12] are increasingly applied for ventilation optimisation.

Evaporative cooling systems combine ventilation systems with droplet-delivering systems (e.g., sprinklers, soakers, misters) [10] or water-flowed cooling pads [11] to more effectively reduce indoor temperature. The idea is to lower the temperature surrounding or on the surface of the cows by delivering water drops into the air or onto their surface while increasing wind speed (WS) to evaporate the water. Since the latent heat of water evaporation is provided by the air, this approach decreases Ta, boosting the thermal gradient between cows and their environment or cows' core and their surface, facilitating more efficient heat dissipation. However, the use of droplet-delivering systems in humid climates (>75% relative humidity (RH)) is not recommended, as they do not provide any physiological relief [10]. This is because any potential drop in Ta is offset by the increase in RH. Moreover, excessive humidity may wet bedding materials thus affecting cow comfort.

For intensive housing systems, proper barn design should be considered to improve the indoor thermal environment. For example, the layout of the barn should be designed in consideration of diurnal and seasonal variations in sunlight and wind direction [13]. In addition, barn optimisation also includes modifying bedding materials to promote conduction while cows lying down. This can be welcomed since cows spend most of their daily time lying and lying time can reflect their

comfort and welfare. Some bedding materials have been studied and sand was found to have higher heat flux for cows compared with straw and mattresses filled with rubber granules [14] and dried manure [15]. In addition, water-cooled heat exchangers which are buried below the bedding surface seem to be an alternative way to increase the heat flow from cows to the environment and are better used in conjunction with sand as the bedding material [15].

Additionally, reducing stocking densities is promising to mitigate heat stress [16,17]. This is because heat can be transferred between animals by radiation and increases with animal density [18,19]. Evidence has shown that early lactating cows with a stocking density of 100% had higher oxidative stress status compared with those with a stocking density of 75% [20]. The provision of additional space for sheep at high Ta mitigated their physiological and behavioural responses induced by heat stress [21].

1.2.2. Nutrition management

The effective management of nutrition for heat-stressed dairy cows is a multifaceted approach crucial for their well-being and productivity. Under heat stress, cows reduce their feed intake, necessitating an increased nutrient concentration to maintain an adequate intake of all essential nutrients, which becomes critical in high-producing cows. Optimal diets should balance high-quality fibres, proteins, and fats, ensuring adequate rumen functioning and sufficient nutrient absorption. High-quality forages are more welcomed in this case since they are digested faster and thus produce less heat compared with low-quality stemmy forages. Increasing the energy density may require the use of larger amounts of concentrates. Collectively, increasing the energy density of the diet, using high-quality forages, and feeding more concentrates can help animals maintain their energy needs, even if they consume less dry matter. However, care should be taken to properly balance the diet to avoid digestive system disorders such as rumen acidosis [22]. Heat stress has been found to increase acidosis prevalence from 0.3 to 4.1% [23]. Moreover, the timing and frequency of feeding play a significant role. Adjusting feeding schedules to cooler periods of the day and increasing feeding frequency can help mitigate the reduced appetite associated with heat stress. This also involves ensuring feed quality, as heat can rapidly degrade nutrients in stored feed, especially wet ones, like silage.

Furthermore, maintaining adequate fresh water is required during heat stress as it can help cows dissipate body heat by sweating and breathing. Cold water at 10 °C has been found to be more effective in reducing cow's body temperature (0.75 °C) compared to water at 28 °C (0.46 °C) [24]. Additionally, electrolyte balance is essential for cellular functions and maintaining body fluid balance. Heat stress can disrupt electrolyte balance, leading to issues such as dehydration or heat exhaustion. Supplementing diets with electrolytes, particularly sodium, potassium, and chloride, can help maintain electrolyte equilibrium, supporting better hydration and heat stress tolerance. This supplementation can be achieved through mineral licks, water additives, or directly through the diet. The precise electrolyte requirements may vary

based on factors such as lactation stage and parity, necessitating a tailored approach for each herd.

The use of feed additives is increasingly recognised for its potential benefits in relieving heat stress. These additives include yeast cultures, enzymes, various plant extracts, neuromodulators, and others. Yeast cultures are known to improve rumen function and overall gut health and metabolism [25,26], which can be compromised during heat stress. Enzymes aid in the efficient breakdown of feed components, ensuring optimal nutrient absorption even when feed intake is reduced [27]. Plant extracts from bupleurum and honeysuckle or other herbs containing polyphenols and flavonoids can alleviate heat stress due to their potential functions in lowering cortisol and modulating relevant signalling pathways such as HSP70/NF-κB and AMPK/mTOR [28]. Among neuromodulators, γ-aminobutyric acid has been used in the diet of heat-stressed cows in recent years, achieving good results in many aspects, such as improving production performance, immune status, and antioxidant levels [29,30]. Other additives, such as N-carbamylglutamate [31] and L-theanine [32], have the potential to relieve heat stress to some extent. These additives not only enhance digestive efficiency but also potentially bolster the immune system, helping the animals to better withstand the stressors associated with high temperatures. Their inclusion in the diet, however, must be carefully calibrated to suit the specific needs of the herd, considering factors like age, lactation stage, and existing dietary composition.

1.2.3. Genetic selection for thermotolerance

The abovementioned environmental and nutritional interventions are essential but only short-term solutions. They may be unsustainable due to potential economic losses for farmers caused by the costs. Also, the needs for changing environments and feeds may also bring environmental pollution. The dairy industry is turning towards a more sustainable and long-term solution: breeding cows that are genetically predisposed to withstand heat stress [33].

Heat tolerance in cattle, defined as the ability to maintain thermal balance under extreme climatic conditions, is a complex and heritable parameter. Traits for selection include physiological responses, milk production and composition, reproduction, and other biomarkers related to heat stress. For example, about 17% of the variation in rectal temperature (RT) among cows during heat stress can be attributed to the variation in genetics [34]. In vivo and in vitro heat challenges indicate that individuals with high heat shock protein 70 concentration may be more resilient to heat stress [35,36]. In addition, many key genes and pathways associated with thermal stress have been identified as potential indicators for marker-assisted selection of both high-yielding and thermotolerant Holstein dairy cattle [37,38]. These findings are pivotal as they open the door to selective breeding programs aimed at enhancing thermotolerance. By identifying and selectively breeding individuals with naturally higher heat tolerance, herds that can better maintain productivity and health in warmer climates can be gradually developed. However, one drawback of selecting heat-tolerant animals is the antagonism between heat

tolerance and productivity [39]. Thus, a better understanding of the underlying genetic mechanisms of heat tolerance is required to develop selection programs which maintain productivity to some extent.

Additionally, biomarkers for heat tolerance can be detected by using metabolomics. The idea is to identify which metabolites are produced when animals are heat-stressed. For instance, studies have identified dozens of specific metabolites affected by heat stress [40-42], including glucose, lactate, and pyruvate, which aids in understanding its impact on energy and nucleotide metabolism in dairy cattle. Combining these findings with physiological parameters can enhance the selection of thermotolerant genotypes [43]. Therefore, it is of great importance to collect both physiological and molecular biomarkers properly and accurately.

Moreover, the introduction of heat-tolerance genes from other cattle breeds offers an intriguing possibility. For instance, the incorporation of the SLICK gene from the Senepol cattle into Holsteins has shown promising results, with offspring exhibiting better ability to regulate body temperature [44]. Such crossbreeding strategies not only enhance thermotolerance but also contribute to genetic diversity within the breed. Notably, systems with ample resources will benefit more from this approach, while systems with limited resources may gain more advantages from hybridisation or breeding with local populations [39].

To sum up, genetic selection for thermotolerance is a forward-thinking strategy that combines scientific research with practical breeding techniques. While it presents an opportunity to mitigate the adverse effects of global warming on dairy farming, it also necessitates careful consideration of potential drawbacks, such as negative impacts on other traits (e.g., cold stress resistance and productivity) and adaptation to different environmental conditions.

2. Detection of heat stress using animal-based indicators

As discussed above, interventions could be taken from the environmental, nutritional, and genetic perspectives to alleviate the impact of heat stress on animals. No matter what interventions would be adopted, the detection of heat stress is necessary to make an evidence-based decision (**Figure 1-2**). Short-term interventions (i.e., environmental regulation and nutrition management) should be adopted when weather changes and animals show resulting responses. Genetic selection, as a long-term solution, requires a proper, accurate, comprehensive collection of phenotypic information to identify heat-tolerant animals.

To understand how animals cope with heat stress, the data acquisition of animal-based indicators is of particular importance [45,46]. Direct measurement and indirect prediction are two major routes to acquire data from the animal side. Unfortunately, there has long been a lack of labour- and time-efficient methods to measure or predict animal responses [47]. As an alternative, thermal indices have been developed by reducing the dimensionality of thermal stressors (i.e., Ta, RH, WS, and SR). An example of such indices is the family of temperature-humidity

indices (THIs), which is a series of highly correlated indices that were developed by combining temperature and humidity [48]. These environmental indicators have been linked to animal-based indicators to assess heat stress indirectly with easily accessible weather data. Notably, environmental indicators have difficulties dealing with individual or herd differences in the onset of heat stress and showing whether animals are being adequately cooled [45,49].

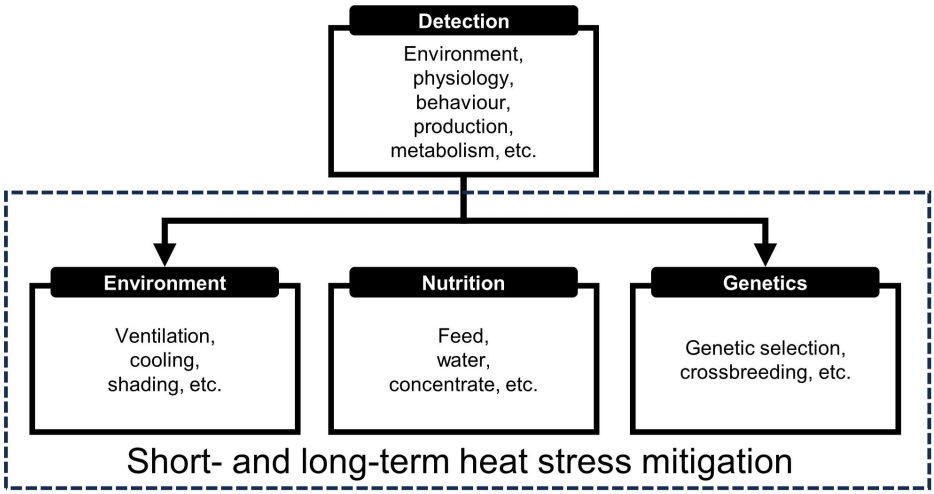


Figure 1- 2: Detection of heat stress and mitigations that could be taken.

For decades, studies focusing on the onset of heat stress have primarily emphasised production traits [50]. This is understandable since milk quantity and quality are what dairy practitioners most care about. It is well-known that heat stress can reduce milk yield and various components such as fat and protein as a result of reduced dry matter intake and impaired mammary synthesis, especially in high-producing cows, meanwhile increasing somatic cell count due to impaired immunity [51]. However, changes in production usually lag the changing environment by about one to three days [52-54]. Thus, production traits are more like results of thermoregulation rather than early indicators of heat stress [45].

In fact, heat stress induces a lot of responses at different stages, ranging from instant vasodilation to genetic adaptations that happen over generations [55]. Recently, many efforts have been made to identify more sensitive animal-based indicators to detect heat stress in the early stage [45,56]. Candidate indicators include physiological and behavioural traits (e.g., respiration rate (RR), body temperature, standing and lying behaviours) [5], as well as other molecular biomarkers such as milk metabolites, oxidative biomarkers, and specific fatty acids especially C18:1 [57-59]. Milk traits can indicate the effect of heat stress on milk mammary synthesis [51], while oxidative biomarkers can indicate the faster formation of reactive oxygen species against antioxidants during heat stress [60]. Although advances have been made in identifying heat stress-related molecules

which can help understand the underlying mechanism in metabolic pathways of carbohydrate, amino acid, lipid, and gut microbiome-derived metabolism [40,41,61], it is not doubtful that physiological and behavioural indicators are easier to achieve real-time, continuous, and timely heat stress detection.

Recently, dairy herds in many regions have become more susceptible to heat stress due to the intentional selection of high-producing dairy cows and the adoption of animals from the Holstein ascend [45]. Indeed, productivity is inversely proportional to heat tolerance, as high-producing cows produce more heat while having poor heat dissipation capacities at the same time [39]. More importantly, global warming is projected to cause heat wave events to occur more frequently, leading to a significant increase in economic losses from heat stress in dairy cows [62,63]. All this information emphasises the importance of improving heat stress detection in dairy cows by using earlier animal-based indicators to avoid the subsequent reduction in production and welfare, as well as the increase in disorders and mortality.

Precision livestock farming (PLF) aims to provide real-time monitoring for significant events in the animal husbandry sector so that interventions can be implemented immediately [64]. Recently, many methods and technologies have been developed to measure or predict heat stress response [46,65]. After data acquisition, critical thresholds are naturally developed to quantify and assess heat stress. This process mainly relies on statistical analysis and data mining. As mentioned earlier, the fact that high-producing dairy cows are raised more highlights the urgent need to validate and update the current thresholds.

Thus, the present review focuses on the recent efforts in developing detection methods of heat stress in dairy cattle, with a special emphasis on those based on body temperatures, respiratory dynamics, and behavioural changes. The primary objective was to summarise the progress on data acquisition techniques, including both direct measurement and predictive modelling. The secondary objective was to summarise the existing thresholds from both animal and environmental viewpoints.

3. Measurement of heat stress in dairy cows

Direct measurements typically provide the most accurate results and are always regarded as the gold standard or ground truth. This section introduces the recent progress in measuring animal-based indicators, with a special emphasis on body temperatures, respiratory dynamics, and behavioural changes.

3.1. Core body temperatures

Core body temperatures (CBT) are the most commonly used animal-based indicator of heat stress in dairy cows [56]. Normally, the range of CBT of homeotherms is very narrow with a circadian rhythm due to the effort of thermoregulation [66]. When cows are exposed to uncomfortably hot environments, their CBT will increase abnormally if they are unable to dissipate excess body heat. The increase in CBT can therefore be used as an indicator of heat stress. The most recent measurement methods of CBT are listed in **Table 1-1**.

Table 1- 1: Summary of the state-of-the-art measurement methods for dairy cattle body temperatures.

Indicator	Technology	Automatic	Continuous	Real-time [a]	Accuracy	Reference
Rectal temperature	Digital thermometer	No	No	No	±0.1 °C	[67]
	A thermistor probe attached to a recorder	No	No	No	Unknown	[68]
	Radiofrequency-based digital thermometer	No	No	No	±0.5 °C	[69]
	An indwelling temperature data logger supported by a customised tail harness or piping	Yes	Yes	No	±0.2 to 0.5 °C	[70,71]
Vaginal temperature	Temperature probe with a finger anchor	Yes	Yes	No	±0.2 °C	[72]
	Temperature data loggers and modified vaginal controlled internal drug release	Yes	Yes	No	±0.1 to 1 °C	[73]
	A wireless vaginal temperature device	Yes	Yes	Yes	±0.1 °C	[74]
	An indwelling device with temperature sensors, a data collector, and a computer system	Yes	Yes	Yes	Mean difference of 0.02 °C with a 95% confidence interval of -0.23 to 0.26 °C	[75]
Reticulorumen temperature	Ruminal and reticular bolus	Yes	Yes	Yes	±0.25 °C	[76]
Tympanic temperature	Infrared ear thermometer	No	No	No	±0.2 °C	[77]
	Ear-canal sensor	Yes	Yes	Yes	±0.1% °F	[78]
Subcutaneous temperature	Implanted temperature data loggers	Yes	Yes	No	±0.1 to 1 °C	[79]
	Percutaneous microchips with scanner	No	No	No	±0.1 °C	[77]
	Implanted thermistors with Bluetooth	Yes	Yes	Yes	±0.05 °C	[80]
	Percutaneous microchips combined with wearable scanners, Long-Range, and Wi-Fi techniques	Yes	Yes	Yes	±0.1 °C	[81]

Milk temperature	Robotic milking system	Yes	Yes	Yes	Unknown	[82]
Body surface temperature	Temperature data loggers	Yes	Yes	No	±0.1 to 1 °C	[83]
	Thermistor sensor	Yes	Yes	No	±0.07 °C	[84]
	Handheld infrared camera	No	No	No	±1.5 °C	[85]
	Handheld infrared gun	No	No	No	±1.5 °C	[86]
	Fixed infrared camera	Yes	Yes	No	±2%	[87]

[a] Ability to transmit in real time.

3.1.1. Rectal temperature

Rectal temperature (RT) is the most popular proxy of CBT. Traditionally, RT is measured with a digital thermometer inserted about 9 and 15 cm into the rectum for calves and adult cattle, respectively [88]. Spiers et al. [68] measured RT with a stainless-steel thermistor probe attached to a recorder (Cole Parmer Instruments, Chicago, IL). Debnath et al. [69] achieved a fast RT measurement in 3 s using a radiofrequency-based digital thermometer. However, despite their usefulness, these methods still rely heavily on human operations. As a result, they can be time-consuming and labour-intensive for herd-level measurement, while also lacking the capability for continuous measurements [89].

To achieve a continuous measurement, some researchers designed indwelling devices which typically consisted of a temperature data logger and a carrier for fixing. For example, Reuter et al. [70] fixed a data logger to a custom-fabricated aluminium tail harness and measured RT at 1-min intervals for 10 h. However, this device has been questioned on the potential damage to the tail and the influence of faecal temperature [46,90]. Lees et al. [71] measured RT at an interval of 20 s for 18.5 h in grazing heifers using a device consisting of a data logger (iButton DS1922L; Maxim Integrated, San Jose, CA, USA) and a soft polyethylene piping.

As summarised by Burfeind et al. [91], the RT readings could be biased by the type of thermometer, the depth into the rectum, and the operation itself. Indeed, the invasive operation puts cattle under additional stress, potentially biasing the results.

3.1.2. Vaginal temperature

Vaginal temperature (VT) is another representative CBT. The measurement of VT is very similar to that of RT. Various temperature data loggers used in conjunction with support devices have been fixed into cows' vaginas. A plastic anchor was first introduced as a support device by Hillman et al. [72] and further validated by Lee et al. [92]. It is demonstrated that long-finger plastic anchors were superior due to better stability when inserted in the vagina [92]. However, the application of these anchors may be difficult due to inaccessibility [90].

The most popular way nowadays is to use a modified vaginal controlled internal drug release insert which contains a temperature data logger [90,93,94]. The price of data loggers varies distinctly, where inexpensive and low-accuracy loggers have been reported to under- and overestimate VT compared with expensive and high-accuracy loggers [73]. Notably, none of these methods can monitor VT in real time due to the lack of wireless transmission. Temperature readings are only accessible after the sensors have been retrieved.

Sakatani et al. [74] achieved wireless transmission of VT at a measurement interval of 5 min by using a temperature sensor (Gyuonkei; Remote, Oita, Japan) in conjunction with Wi-Fi technology. More recently, Wang et al. [75] designed a wireless measuring system consisting of an indwelling device equipped with temperature sensors (ADT7320; Analog Devices, Norwood, MA, USA), a data collector, and a computer system. VT was measured every 30 min, wirelessly transmitted to the data collector every 2 h, and then uploaded to the computer system.

All known indwelling devices for measuring VT were placed intravaginally for a maximum of 24 days (around a week in most applications) [73], possibly due to the increased concern of irritation and infection [90]. That said, this method may be only applicable for short-term research purposes. Other than the type of sensors, the temperatures recorded vaginally can be affected by logger displacement and lying bouts [94].

3.1.3. Reticulorumen temperature

Rumen and reticulum are also locations of interest to measure CBT. Many boluses have been designed and applied to measure ruminal and reticular parameters (e.g., temperature, pH) [95-99]. A rumen or reticulum bolus is a configured pill placed in the reticulum or the junction between the rumen and reticulum, consisting of a temperature sensor, a telemetry system, and a battery [46]. Hicks et al. [96] first measured the temperature ruminally in three cows using CorTemp sensors (HTI Technologies, Palmetto, FL, USA). Initially, this method required a fistula or an implant surgery, which compromises animal welfare. At present, commercially available boluses can be administered orally via a customised bolus applicator or a balling gun [100,101]. With the development of wireless and battery technology, the most recent boluses are administered once and can work for at most 6 years with adjustable measurement intervals (10-min as default) [102]. SmaXtec bolus (Animal care GmbH, Graz, Austria), one of the popular products on the market, has shown its ability in detecting heat stress in grazing dairy cattle [103]. Rumen boluses have recently been customised for dairy calves to monitor infections, but whether these devices can be used for heat stress detection in dairy calves remains unknown [104].

The temperatures recorded in the rumen and reticulum are greater than RT by approximately 0.5 °C due to the heat produced by rumen microorganisms [98]. A drinking bout can decrease reticulorumen temperature (RET) by 9.2 °C and it would take up to 3.5 h to return to the original temperature before drinking [98,105]. Thus, RET can be used to monitor drinking behaviour [106]. Moreover, given that the reestablishing time depends on both the temperature and amount of water drunk, it can be further used to predict water intake, which also plays a part in heat stress in cattle [107]. To mitigate the impact of water intake on RET, Ammer et al. [108] calculated median and mean RET for 2 h preceding RT and VT measurement and demonstrated that the median RET was more highly correlated to RT and VT (r = 0.48, 0.53) compared with mean RET (r = 0.43, 0.46) and single RET (r = 0.40, 0.48). Since calves drink less cool water than adult cattle, and their milk or milk supplements are preheated, the intake of water by calves may not bias RET as much as it does for adults [104].

3.1.4. Tympanic temperature

Tympanic temperature (TT) was initially measured in beef cattle through data loggers placed in the ear canal [109-111]. This method has also been used in dairy cows to measure heat stress [112]. However, the use of this method in adult cattle has been questioned due to the discomfort caused by foreign objects and the potential for infections over long-term use [113]. In addition, the probes need to be

properly located near the tympanic membrane, and any dislocation could result in inaccurate readings [46,97,114].

To date, wireless temperature sensors that can be fixed in the ear canal have already been commercialised for dairy calves [114,115]. The latest model of these temperature-sensing ear tags (TempVerified FeverTag; Amarillo, Texas, USA) has an accuracy of ±0.1% in °F [78]. Notably, TT is the only validated indicator of CBT in calves [77]. In addition, for instant use, an infrared ear thermometer provides a reading of TT in 1 s (Vet-Temp Instant Ear Thermometer, Advanced Monitors, San Diego, CA, USA) [77].

3.1.5. Subcutaneous temperature

Subcutaneous temperatures (ST) are also candidates to represent CBT since heat-stressed cows would drive more blood from the core to the peripheral [81]. ST can vary distinctly among different implantation locations and many efforts have been made to determine the best locations as the proxy for bovine CBT. Lee et al. [79] measured ST in seven Holstein steers using data loggers (iButton DS1922L) in three sites (lateral neck, upper, and lower scapula). Due to the lack of wireless transmission, temperatures recorded were not available until recovered from an explant surgery. This method is definitely only suitable for research purposes.

In consideration of animal welfare, wireless transmission is required for ST measurements to avoid explant or retrieval operations. Commercially available microchips can now be used to measure bovine ST. For example, Reid et al. [116] used implantable radiofrequency identification (RFID) microchips (Bio-Thermo; Digital Angel, Saint Paul, MN, USA) to measure Holstein steers' ST in three sites, including the base of the ear, posterior to the poll, and the area beneath the umbilical fold. In a recent study of dairy calves [77], similar microchips (Bio-Thermo; Allflex, USA) were implanted to measure ST in three sites (the ear scutulum, the upper scapula, and intramuscularly in the trapezius muscle of the neck). Moreover, the nasal submucosal temperature of 20 beef heifers was measured using LifeChip microchips with Bio-Thermo technology (Destron Fearin, TX, USA) implanted in the left and right nasal submucosa about 10 cm caudal to the alar cartilages [117]. It should be noted that the products on market must be used with a reader to collect temperature readings, meaning that manual operation is still required.

In order to reduce human efforts regarding temperature scanning on the field, Chung et al. [81] achieved a long-distance transmission by combining the LifeChip microchips with RFID scanning, long-range wireless communication, and Wi-Fi technologies. Ear base temperature was measured and transmitted every 30 s. In addition, Iwasaki et al. [80] implanted Bluetooth-based wireless thermistor sensors into a Japanese hybrid cow to measure hourly ST in 10 different anatomical locations. Five receivers were placed in the barn to collect data from the sensors and further transferred them to a personal computer.

3.1.6. Milk temperature

Milk temperature (MT) has been proposed as an indicator for heat stress in dairy cows due to the high degree of vascularity in the mammary gland [52]. Strong

correlations have been reported between MT with other gold-standard CBT [118]. Nowadays, MT can be collected and stored automatically per milking per cow by the robotic milking system, e.g., LELY T4C (Lely Industries NV, Maassluis, the Netherlands) [82]. Basically, temperature sensors inside the robot's arm measure the MT of each quarter individually and the highest reading measured during the whole milking event is often recorded in the database.

3.2. Body surface temperatures

When dairy cows first feel discomfort during heat exposure, vasodilation allows more blood to divert from the core to the skin, which increases BST and allows for increased heat exchange with the environment [81]. BST have been increasingly mentioned in recent years as their non-invasive measurement is more in line with animal welfare and PLF [119-121]. The most recent measurement methods of BST are listed in **Table 1-1**.

Body surface temperatures (BST) can be measured in many ways among which infrared thermography (IRT) is the most popular means due to its non-contact nature and quick readings [122]. IRT can be used in both portable and fixed ways. For portable use, an infrared camera or infrared gun can be held to measure data according to users' needs [85,86]. However, manual operation limits the frequency of data collection. For fixed use, an infrared camera can be fixed at a specific location (e.g., milking parlour, calf feeder, and water trough) [87,123,124]. The cameras can be integrated with existing systems to achieve automated collection of infrared data. An example is Schaefer et al. [123] where a FLIR S60 broadband camera (FLIR, Boston, MA) was interfaced to a control system including RFID sensors such that the camera shot every time a calf attended the water station.

No matter where and how the infrared camera is used, regions of interest (RoI) have to be defined to collect specific BST. The common RoI used by dairy studies include eyes, nostrils, foreheads, ears, muzzles, and udders. The definition of RoI is often done after data collection by manual localisation using certain shapes on the processing software. This operation is time- and labour-consuming. Several tricks have been used to increase the efficiency of RoI definition. In the study by Hoffmann et al. [124], the RoI (i.e., body and head) were manually defined at the beginning of videoing, and no replicated operation was needed due to the fixed position of the calves and the camera. In the study of Jorquera-Chavez et al. [87], the RoI was also manually defined first by the users and the Kanade-Lucas-Tomasi algorithm was used to track the selected RoI among sequential frames. In addition, object detectors based on computer vision and deep learning techniques have been introduced to achieve automated detection of RoI [125,126].

Generally, top-tier infrared cameras with high accuracy are quite expensive, whereas affordable infrared guns may produce less reliable results. In addition, the results of IRT are affected by the dirt on the surface, the skin fold, the direct sun exposure, the distance and angle to the object animal, etc. [45], indicating that additional image processing is needed. However, data processing techniques may

work well for some cameras but not for others due to the different frame rate, image resolution, and accuracy of the cameras [127].

In addition to IRT, BST can also be measured through wearable devices. Data loggers (iButton DS1922L) were placed on the skin above the tail vein opposing the rectum using flexible tape [83]. Thermocouples and thermistors were also fixed to the skin to measure BST [84,128]. The results of the thermistors (EU-UU-VL5–0, Grant Instruments, Cambridge, UK) and data loggers (iButton DS1922L) were demonstrated to have a very good agreement when measuring BST [129]. However, wearable devices are difficult to fix on the skin surface and prone to displacement, and the results recorded were easily affected by hair and air movement [80]. To minimise the influence of wind and hair, Kou et al. [84] designed an automated measurement for BST by using a shell shape to securely position the thermistor on the metatarsus of the hind leg, where there is less hair.

3.3. Respiratory dynamics

Respiration-based indicators are relevant indicators of heat stress in dairy cows. More importantly, the respiratory system is activated before CBT, as measured in rectal, vaginal, and tympanic areas, begin to rise [130,131]. Respiration-based indicators may be the most appropriate indicators to early monitor heat stress in dairy cows due to their high sensitivity in response to heat stressors and cost-effective measurement [132]. The most recent measurement methods of respiratory dynamics are listed in **Table 1-2**.

3.3.1. Respiration rate

The traditional method to measure RR of cattle is to manually count the movement of the flank and convert it into breaths per minute (bpm). Such a visual observation is both time- and labour-consuming. In addition, the interaction between observers and cows may bias the result [133]. Many efforts have been made recently to design automated measurements for RR. Generally, these new methods can be categorised into contact and non-contact.

Contact measurements rely on wearable devices which are mounted to specific locations of cows (e.g., face, flank, chest, neck). These devices collect respiration-related electrical signals based on different methodologies and further convert them into RR. Eigenberg et al. [134] developed a thoracic belt for cows based on a device originally designed for humans. The device consists of a force transducer deriving electrical signals from rib cage movement and a small battery-powered micro-computer further converting the signal into RR. Its memory could collect RR data for four consecutive days under the condition of collecting 1 min every 15 min. Atkins et al. [135] designed similar devices and equipped them on eight lactating Holstein cows to detect breathing-related abdominal expansion. Two methods (peak-to-peak and fast Fourier transform) were used to derive RR from the raw signal.

Table 1- 2: Summary of the state-of-the-art measurement methods for dairy cattle respiration-related data.

Indicator	Methodology	Technology	Automatic	Continuous	Real-time [a]	Accuracy	Reference
Respiration rate	Temperature changes around nostrils	Infrared thermography	Yes	Yes	No	8.4 ± 3.4 (mean ± SD) bpm lower	[87]
		Thermistor sensor	Yes	Yes	No	±2 bpm for 80% of the time	[133]
	Body movements	Pressure sensor	Yes	Yes	No	±2-3 bpm	[134]
						±0.47 bpm	[135]
		Accelerometer-based collar	Yes	Yes	No	Unknown	[136]
		Laser distance sensor	Yes	Yes	No	Mean difference of 6 bpm	[137]
		Ultra-wideband radar	Yes	Yes	No	Unknown	[138]
		Frequency-modulated continuous wave radar	Yes	Yes	Yes	Root mean square error of 1.582 bpm	[139]
		RGB camera	Yes	Yes	No	Mean accuracy of 98.58%	[140]
						Mean accuracy of 93.04%	[141]
	Pressure changes around nostrils	Pressure sensor	No	Yes	Yes	Mean difference of -0.2, 0.2, and1.4 bpm when dozing, lying, and standing, respectively	[142]
	Breathing sounds	MP3 recorder	Yes	Yes	No	Mean bias of 2.75 bpm	[143]
	Changes in respiration volume	Spirometer	Yes	Yes	Yes	Unknown	[144]

	Colour changes in facial part	RGB camera	Yes	Yes	No	Unknown	[145]
Panting	Panting scoring system	Visual scoring	No	No	No	Not applicable	[146-148]
	Body movements	Accelerometer-based collar	Yes	Yes	Yes	Unknown	[130]
		Accelerometer-based ear tag	Yes	Yes	Yes	Positive predictive value of 79%	[149]

SD = standard deviation; bpm = breaths per minute.

[a] Ability to transmit in real time.

Milan et al. [133] developed a thermistor-based halter to measure RR according to the temperature changes around nostrils during breathing and found that the results of the device were slightly lower than that of manual counting in four out of five cows. However, they did not validate its reliability when cows were heat-stressed, nor did they achieve a real-time transmission. It can be inferred that this method may be of limited use when Ta is close to exhaled air temperature since no pulses in temperature would then be detected. Strutzke et al. [142] also designed a halter that can detect respiration-related signals based on the pressure difference between inhalation and exhalation of air. As a preliminary investigation, they did not yet develop automated processing software or algorithm, thus still requiring manual counting of the peaks on the signal plot. A wireless local area network was applied to achieve wireless transmission. The results showed no significant difference ($P > 0.05$) between manual counting on the video and that on the sensor signal when cows were dozing or lying, but a potential overestimation by a mean difference of 1.4 bpm when the cows were standing ($P < 0.05$). The authors explained that this could be due to incomplete manual counting. Their device also had many limitations, e.g., short battery life, loss of devices, and the unknown reliability of the device in hot conditions as well. Moreover, although there was no adverse event observed during the study, the need to insert a flexible tube 10 cm deep into the nasal cavity may cause some problems since cattle like to lick their noses [150]. The authors expected to improve this method by developing processing software and including information about breath depth. In addition, a facial mask equipped with an indirect calorimetry system is very important for research purposes due to the ability to measure comprehensive respiratory characteristics, including RR, tidal volume, and the composition of the exhaled breath [144].

Recently, an accelerometer-based collar was further developed to measure RR [136]. Using Fourier transform, respiration-related signals could be distinguished from those of rumination, and RR could be obtained directly from the frequency domain. In addition, the amplitude may be indicative of breath depth. More recently, an acoustic method was developed by using an MP3 audio recorder mounted to the cow's halter [143]. However, the RR was retrospectively analysed after data collection and the battery could only last up to two days. Wearable devices are expected to continuously measure RR in real-time if battery life, wearing stability, and remote wireless transmission can be further improved.

The temperature changes around the nostrils during breathing can also be captured in a non-contact way by using hand-held IRT cameras [151,152]. The mean difference between IRT and manual counting was 0.83 bpm in adult cattle and 0.02 bpm in calves. More recently, Jorquera-Chavez et al. [87] used a fixed thermal camera and computer vision algorithms to calculate RR based on the changes in pixel intensity within the nose area. Again, the systemic defect that the Ta and exhaled air temperature are difficult to distinguish in a hot environment makes this method only suitable for specific cool indoor areas. Pastell et al. [137] designed a novel method based on a laser distance sensor. The laser rangefinder (L-Gage LT3; Banner, Minneapolis, MN, USA) was fixed on the side of the cow and the RR was

calculated from the change in the distance to the flank when the cow was breathing. The mean difference was 6 bpm when the detection frequency was set to 100 Hz. The signal of displacements from cows' flank movements can also be detected using radar sensors. For example, ultra-wideband radar [138] and frequency-modulated continuous wave radar [139] have been used to extract RR signals from handled or milking cows. To sum up, non-contact RR measurement methods are generally costly and not mobile, and therefore would be better to use in specific scenarios where cows are handled separately (e.g., milking parlours, calf feeders).

In addition, vision-based methods are potentially low-cost solutions which rely solely on RGB cameras. These methods typically rely on periodic body movements or changes in facial blood flow during breathing. On the one hand, periodic body movements can be detected using optical flow methods which estimate the motion of sparse or dense pixels by measuring the relative speed changes between consecutive frames. For example, Song et al. [140] developed a RR detection method for lying cows based on Lucas-Kanade sparse optical flow algorithm and yielded an average accuracy of 98.58%. The respiration-related body movements can be weaker while cows are standing. To solve this problem, Wu et al. [141] combined a phase-based video magnification algorithm with Lucas-Kanade sparse optical flow and achieved a mean accuracy of 93.04% for 70 video clips (each about 20 s in length) of individual standing cows. In their most recent work, an instance segmentation method was incorporated to achieve a multi-target measurement for two cows [153]. On the other hand, methods based on the colour or brightness changes in facial blood flow typically follow the remote photoplethysmography principle. For example, Fuentes et al. [145] used a computer vision algorithm to analyse RR for individually handled cows based on the luminosity changes on the green to red colour channel from CIELab scale. To sum up, these methods are more affordable but were only designed for measuring single or very few targets. Further improvements are needed to increase their applicability in more diverse scenarios.

3.3.2. Panting score

Panting, as an important physiological pathway of evaporative heat loss, has long been proposed to describe heat stress in dairy cows [6]. Panting is considered a behavioural change strategy but is introduced here due to its relevance with RR.

With the accumulation of heat in dairy cattle, cows drool first, then open their mouth and protrude their tongues [154]. Panting score (PS) has been developed in numerical scales to rate these signs [146-149]. A recent study proposed a new scoring system to assess heat stress in grazing dairy cows [155]. This score is very similar to PS, also relying on various behavioural signs to assess heat stress on a scale of 0 to 4. To measure PS or related scores, trained observers rate a cow's respiratory dynamics according to the numerical scale, and the process would take about 10 s per observation per cow. Although measuring PS is relatively easier than RR, the inter-observer difference can be an important source of error.

The ability of a commercially available collar to accurately reflect heat stress in dairy cows has been validated [130]. 'Heavy breathing' detected by this tag is

defined by characteristic movements (e.g., forward-backward heaving and increased chest movement). An accelerometer-based ear tag from the same manufacturer was also validated in steers [149]. The sensor was designed to classify each minute segment into one of five behaviours, probably leading to a high false-negative rate by rating breathing heavily cows as other more prominent behaviours. As proposed by both Bar et al. [130] and Islam et al. [149], these accelerometer-based sensors have difficulties in detecting slight panting and only give a binary result (panting or not). Although the detection of panting or not is enough for the start of heat abatement protocol, a more sophisticated classification would be more informative for customised interventions. Further studies can therefore focus on improving the sensitivity and classification of the sensors. In fact, the determination of 'heavy breathing' is based on a default threshold of the frequency of respiration-related movements. Therefore, 'heavy breathing' can be seen as an extension of RR.

In addition, computer vision is feasible to measure panting but requires a direct view of facial areas to detect drooling, mouth opening, and tongue protruding. Otherwise, measurements relying on other motion characteristics, i.e., forward-backward heaving and increased chest movement, are equivalent to measuring changes in RR.

3.4. Behavioural changes

Except for panting, cows take many other behavioural change strategies to cope with changing thermal environment. For example, they increase water intake and standing bouts, decrease feed intake, and seek cooling and shade [5,156-158]. Accordingly, agonistic social behaviours, such as replacing others at the water trough, would increase [159]. The most recent measurement methods of behavioural indicators are listed in **Table 1-3**.

The traditional way to measure cow behaviours is through manual observation either on the field or on video recordings, both of which require massive time and human effort. Thanks to wearable technology, basic behaviours including lying, standing, rumination, walking, drinking, and feeding can be accurately measured by using accelerometer-based sensors. Many commercially available products (such as collars, ear tags, leg sensors, and boluses) have been validated in dairy research. For example, Heinicke et al. [160] measured lying behaviour in 196 cows by using IceTag3D leg sensors (ceRobotics, Edinburgh, UK). In another study, smart collars (SCR Engineers, Netanya, Israel) were equipped on 864 cows to measure rumination [161]. More importantly, these devices can provide the daily behavioural duration of individual cows as an output, allowing for the analysis of individual differences in dealing with heat stress.

In addition, basic behaviours such as standing, lying, feeding, and drinking can be easily detected through non-contact vision-based methods due to their representative postures and relative position with facilities (e.g., drinking at the water trough, feeding at the feeding trough). For example, Viola-Jones algorithm, a classic object detector, has been trained for detecting lying, standing, and feeding behaviours from video recordings [162,163]. Furthermore, the number of detections of the specific

Table 1- 3: Summary of the state-of-the-art measurement methods for dairy cattle behavioural indicators.

Indicator	Data acquisition	Cow information	Algorithm	Sample size	Accuracy	Reference
Lying and standing	Accelerometer-based leg sensor	Not applicable	Not applicable	Not applicable	Unknown	[160]
Daily duration of lying, standing, rumination, walking, feeding, drinking	Accelerometer-based collar, ear tag, and bolus	Not applicable	Not applicable	Not applicable	Unknown	[161,164]
Percentage of cows lying, standing, and feeding	Top-view video recording	15 Holstein dairy cows	Violae-Jones algorithm	1145 positive image samples	Sensitivity of 0.92, 0.87, 0.86 for lying, feeding, standing	[162,163]
	Video recording	51 cows	YOLOv3	316 images	Coefficient of determination of 0.67 for lying and standing	[165]
A total of 15 individual and social behaviours	Side-view video recording	A Hanwoo herd	YOLOv3 with spatio-temporal information	350 videos with an average duration of 12 min	Mean average precision of 0.856	[166]
Drinking	Top-view cameras	25 Holstein dairy cows	Tiny YOLOv3	1000 images	F1 score of 0.987	[167]
Drinking, rumination, walking, standing, and lying	Side-view video recording	Individual Holstein cows	VGG16 feature extractor and Bi-LSTM classifier	4566 videos with duration of 10 to 55 s	Average recognition accuracy of 0.976	[168]
Drinking and its competing	Electronic water bins	42 Holstein cows	Not applicable	268 events	Specificity and sensitivity of 1 for drinking	[169]
		20 lactating Holstein cows	Thresholding	669 events	Recall and precision above 0.8 for competing	[170]

behaviour was divided by the total number of cows in the barn to produce herd-level behavioural indicators, including cow standing index, cow lying index, and cow feeding index. The measurement of these indicators allows for further analysis on a herd basis.

Recently, deep learning-based algorithms have significantly increased the accuracy of cow behaviour recognition. Plus, more detailed and dynamic behavioural classification can be fulfiled by utilising temporal information [166,168]. The deep learning-based approach reportedly has already achieved an average accuracy of up to 0.976 in detecting basic cow behaviours (i.e., drinking, ruminating, walking, standing, and lying) [168]. However, it should be noted that algorithms are mostly developed for specific purposes on specific farms, often using small datasets [171]. The lack of external validation raises concerns about the generalisability of the developed methods in broader real-world scenarios. Important issues, such as the position and angle of camera mounting, the ability to cover the entire herd, and the ability to work during day and night, should be further investigated before such vision-based methods can be extensively used in commercial environments.

Behaviours of accessing environmental resources can also be obtained by checking the visiting record of electronic devices. For example, drinking can be measured automatically through Insentec water bins (Marknesse, the Netherlands) [169]. This process requires a good connection with on-cow fitted identification devices (e.g., electronic ear tags and neck collars). Agonistic social behaviours, competing for water bins in this case, can be further obtained by using certain processing algorithms embedded into the hardware. An event of replacement is often defined as a brief interval between two consecutive visits at the same waterer. The optimal threshold was reported at ≤29 s by McDonald et al. [170] and between 22 to 27 s by Foris et al. [172].

4. Prediction of heat stress in dairy cows

Predictive models aim to predict difficult-to-access parameters using easily accessible ones. They can be simply classified into empirical and mechanistic methods. The former is driven by data while the latter is driven by theory.

4.1. Empirical models

Typically, empirical modelling is used either to infer rules and relationships within real-world data or to create a model to predict future observations. That is to say, the modelling begins from data acquisition and is highly dependent on the quantity and quality of the datasets. The recent efforts in empirical models are listed in **Table 1-4**.

The classic and traditional way to construct an empirical model is to use statistical models, such as linear regression models. Technically, statistical models are more commonly used to infer relationships between variables. Although these methods can make predictions, the accuracy of the predictions is not their strength. A recent statistical model for predicting heat stress in dairy cows is presented by Li et al. [173], in which herd average RR was predicted from environmental parameters, previous milk yield, and time block. Although a very good coefficient of

determination was collected, the use of herd means neglects individual differences and cannot make a prediction at the individual level.

In recent decades, machine learning (ML) techniques are gaining popularity in data science. ML techniques can be simply divided into two categories: supervised and unsupervised learning. On the one hand, supervised learning involves training a model with known predictors (also called features) and outcomes (also called labels) to predict future observations. Unsupervised learning (e.g., clustering), on the other hand, is often used for explorative data mining of the hidden and intrinsic patterns within new datasets. In the case of heat stress prediction, nearly all relevant works rely on supervised learning.

Supervised learning is a technique that first trains a model using a subset of data and then tests the model on a different subset of data to collect its true performance. Notably, supervised learning can still perform statistical models. For example, linear regression, a statistical method, can be trained as a linear regressor and produce the same results as a statistical regression model which aims to minimise the squared error between observations. In fact, these linear regression models are frequently used as baselines or controls for other non-linear ML models.

Supervised ML algorithms can be used for regression and classification tasks. As shown in **Table 1-4**, the majority of works employed regression models for predicting physiological responses, whereas only two works used classification models to predict a four-level heat stress score and panting score, respectively [155,174]. Overall, the applied algorithms include linear regression, logistic regression, naïve Bayes, gradient boosted machines, random forests, artificial neural networks, recurrent neural networks, and neuro-fuzzy networks.

Since different studies used different datasets, the horizontal comparison of their results is unfair. However, artificial neural networks and random forests performed always the best when multiple algorithms were compared. This fact at least demonstrates the advantage of these two algorithms in fitting heat stress responses and can be used as a reference for future studies to choose appropriate algorithms. In addition, the number of data points used for training varies dramatically among studies from 150 to more than 6000. The relationship between sample size and predictability can be complex and depends on various factors, such as the variability in the data and the complexity of the underlying patterns. A larger sample size, however, generally avoids overfitting and increases the chances of obtaining accurate and reliable predictions. The rule of thumb is that the data points should be at least ten times as many as the predictors [175,176].

Some studies regard BST as an outcome, such as Fuentes et al. [177] and Gorczyca and Gebremedhin [178], whereas some other studies treat BST as a predictor of other gold-standard indicators such as Pacheco et al. [179] and Sousa et al. [180]. These two routines reflect two viewpoints on the roles that these variables play in thermoregulation. Theoretically, predictors should at least change before the outcome to make a temporally logical prediction. To some extent, eye temperature,

Table 1- 4: Summary of the recent empirical models for predicting dairy cattle body temperatures and respiratory dynamics

Reference	Methodology	Cow information	Thermal condition	Sample size	Predictor	Response	Accuracy [a]
[173]	Linear regression	45 high-producing Holstein cows	Ta ranged from 6.9 to 33.3 °C and RH ranged from 31.8% to 99.7%	253 records of herd average	Ta, RH, wind speed, milk yield, and time block	Respiration rate	R^2 of 0.836
[181]	Artificial neural networks, neuro-fuzzy networks	Holstein cows	Unknown	6676 records	Ta and RH	Respiration rate	RMSE of 6.67 bpm
[179]	Artificial neural networks	21 and 14 Holstein primiparous and multiparous cows for summer and winter sessions, respectively	Summer: Ta and RH averaged 29.2 ± 6.20 °C and 62.6% ± 20.51%, respectively. Winter: Ta and RH averaged 19.2 ± 4.97 °C and 70.8% ± 18.73%, respectively	About 1850 records	Ta, RH, dew point temperature, and maximum ocular temperature	Rectal temperature	RMSE of 0.40 °C
					Ta, RH, and mean front temperature	Respiration rate	RMSE of 10.01 bpm
[178]	Penalised linear regression, random forests, gradient boosted machines, and artificial neural networks	19 lactating Holstein cows with an average daily milk yield of 39 ± 4 kg	Ta and RH averaged 29.2 ± 6.20 °C and 62.6% ± 20.51%, respectively	486 records	Ta, RH, wind speed, and solar radiation	Vaginal temperature	RMSE of 0.434 °C
						Body surface temperature	RMSE of 0.334 °C
						Respiration rate	RMSE of 9.695 bpm

[177]	Artificial neural networks	102 Holstein cows	Ta from 11 to 13 °C and RH from 74% to 82%	150 records	Respiration rate, heart rate, abrupt movements, and nine weather parameters	Body surface temperature: eyes	Unknown
[155]	Random forests, logistic regression, and Gaussian naïve Bayes	18 lactating Holstein cows and 9 lactating Jersey cows	Temperature-humidity index averaged 76.81 ± 4.99	856 records	18 features including weather parameters, cow-related factors, production parameters, and behavioural parameters	Heat stress score based on respiration rate and its related behaviours	Accuracy of 0.888
[80]	Artificial neural networks	A Japanese hybrid cow	Ta averaged 4.7 ± 3.5 °C	342 records	7 subcutaneous temperatures	Inner flank wall temperature	Correlation coefficient of 0.93
[81]	Long short-term memory networks	3 Holstein cows	Unknown	Unknown	Temperature-humidity index and ear base subcutaneous temperature	Vaginal temperature	RMSE of 0.081 °C
[174]	Logistic regression and long short-term memory networks	200 and 96 mixed-breed cattle for training and validation sets, respectively	Ta averaged 23.8 ± 5.56 °C and 25.84 ± 4.24 °C, respectively	Unknown	Historic and forecast weather data and the herd's heat vulnerability dilation factor	Heavy breathing (panting)	Accuracy of 0.802

Ta = ambient temperature; RH = relative humidity; R^2 = coefficient of determination; RMSE = root mean square error; bpm = breaths per minute.

[a] When multiple algorithms were compared, only the best one was shown.

despite being classified as a kind of BST, would be better treated as an outcome given its close relationship with CBT in both values and temporal sequence [182].

Although ML models typically have good predictive ability, their interpretability is always sacrificed. Fortunately, some newly developed model-agnostic approaches are able to extract post-hoc explanations, such as feature importance, partial dependence plots, Local Interpretive Model-Agnostic Explanations, among others. These techniques have already been used for interpreting ML models in predicting milk production [183] and land use [184]. Thus, it is interesting to see if these techniques can help interpret the nonlinear relationship between the thermal environment and animal responses.

Another concern of ML models is their external validity and generalisability. Proper splitting between training and testing sets is crucial for model generalisation [185]. Traditional random splitting can lead to an over-optimistic estimation of ML regression model performance by up to 48%, as multiple samples from the same subject might be present in both sets, leading to data leakage [186,187]. Subject-wise splitting should provide better generalisation because it can ensure complete independence between sets and thus better represent real-world applications. However, all studies adopted random splitting (**Table 1-4**) and it remains unknown to what extent the potential overestimation is. Also, appropriate modelling techniques (e.g., regularisation) can be used to avoid overfitting in the training set [188]. However, even if all these points have been properly addressed, generalising trained models to datasets collected elsewhere remains challenging due to significant differences in environments, animals, and management. To some extent, this problem can only be solved by training with combined datasets collected from different settings. Undoubtedly, the more exposure the ML models have to the overall distribution of data, the better they can adapt to new data samples.

4.2. Mechanistic models

Mechanistic models aim to mimic and predict what will happen in a real-world system in a deterministic way through the fundamental laws of natural sciences, such as physical, chemical, and biological principles, [65]. Ideally, mechanistic models are capable of simulating situations that have not been seen, while empirical models cannot because their predictions are based on previously trained samples [189]. In addition, mechanistic models need less experimental data to calibrate the model and determine unknown model parameters. Another advantage of mechanistic models is that the model parameters have an actual physical meaning, which allows for easier scientific interpretation of the results.

Unlike empirical models which start with the data, mechanistic models start by assuming that a particular set of processes are affecting the outcome of interest. In the case of modelling humans or animals' thermal balance, the strategy is to start with treating humans or animals as a control system, then simulating the heat transfer process between the body and environment using mathematical equations, and finally relying on differential equations to solve the unknowns. The recent works in this respect are listed in **Table 1-5**.

Table 1- 5: Summary of the recent mechanistic models for predicting dairy cattle body temperatures and respiratory dynamics.

Reference	Structure	Input	Output	Accuracy [a]
[190]	Three-layer model: core, skin, and coat with heat production, solar radiation, air temperature, and wind speed sub-models	Environmental inputs: air temperature, solar radiation, wind speed, and vapor pressure. Animal inputs: body-core weight, genotype-specific parameters, dry matter intake rates, and diet composition	Core body temperature	Unknown
[191]	One dimensional model	Environmental inputs: air temperature and wind speed. Animal inputs: tissue density and specific heat, metabolic heat production, etc	Body surface temperature	Unknown
[192]	One-dimensional model	Environmental inputs: air temperature, relative humidity, and wind speed. Animal inputs: tissue resistance, fur thermal conductivity, coat thickness, hair density, hair diameter, etc	Body surface temperature	Unknown
[193]	Five-layer model: body core, top and bottom skin, and top and bottom coat	Environmental inputs: Julian day number, air temperature, relative humidity, wind speed, annual average air temperature, location. Animal inputs: body mass, milk yield, feed intake, and feed ingredients	Core, skin, and coat temperatures	RMSE of 0.4 and 1.16 °C for 2 datasets
[194]	A total of 10 two-layer and three-layer models based on existing models and equations	Environmental inputs: air temperature, relative humidity, and wind speed. Animal inputs: body mass, milk yield, coat thickness, tissue resistance, hair diameter, etc	Body surface temperature	RMSE of 0.649 °C
[195]	Three-layer model: core, skin, and coat with physiological regulation sub-models	Environmental inputs: air temperature, relative humidity, and wind speed. Animal inputs: body mass, body core diameter, milk yield, coat thickness, tissue resistance, hair coat thermal resistance, etc	Core body temperature	RMSE of 0.30 °C
			Body surface temperature	RMSE of 1.20 °C
			Respiration rate	11 breaths per minute

RMSE = root mean square error.

[a] When multiple algorithms were compared, only the best one was shown.

Following the development of mechanistic models in human biometeorology, mechanistic models were introduced into cattle biometeorology around the 1980s. McArthur [196] developed a heat-balance model for ruminants that could predict CBT under static environments. However, the majority of subsequent efforts were made in estimating heat transfer mechanisms such as evaporative and convective heat losses [3,197,198] while neglecting physiological responses. In the recent decade, global attention in mechanistic modelling has again focused on the prediction of physiological responses, possibly driven by increased interest in precise heat stress detection.

Specifically, the main assumption of relevant works involves treating the animal body or part of the body as a number of concentric cylinders with closed ends, e.g., the entire cow body can be structured into three layers: core, skin, and coat as per a classic model developed by McGovern and Bruce [199]. Also, a significant amount of information is required, such as tissue heat transfer properties, metabolic heat production, and thermoregulatory mechanisms [65]. However, certain input variables, including various biophysical parameters that are required for calibrating and deploying such models, are rather difficult to collect in practical farms.

Many studies [190,199] took advantage of the model structure of McArthur [196] since it contains physiological responses (i.e., body temperature, sweating, panting, and vasodilation). Notably, the model of McArthur [196] can only predict body temperature under a static environment where microenvironmental conditions do not change. Taking a step further, Thompson et al. [190] developed a dynamic heat exchange model by connecting the heat flows with body temperature. This model successfully predicts dynamic CBT in response to constantly changing environmental factors such as Ta, WS, and SR.

It is crucial to properly incorporate heat transfer and thermoregulation principles as well as animal-related factors to accurately predict animal responses to heat stress [200,201]. Mechanistic models have become increasingly complex over time as researchers continue to incorporate new elements and refine the models to better capture the complex interactions between environmental factors and animal physiology. This increased complexity is somehow necessary to ensure that the models accurately represent the underlying biological and physiological processes that govern animal health and productivity under various environmental conditions. For example, Li et al. [193] improved the model of Thompson et al. [190] by introducing a sub-model for describing conduction between the lying animal and the ground surface. Also, more layers (i.e., core, muscle, fat, and skin) can be considered due to different heat transfer mechanisms (e.g., heat generated by external work occurs only in the muscle layer) as per human studies [65,202]. Moreover, radiation, as another relevant heat transfer mechanism that can happen between cows and the environment as well as between individual cows, requires more investigation [65].

Another research hotspot is to validate previously developed model structures with regard to the interpretation of physiological responses using nowadays high-producing dairy cows. This is driven by the fact that modern cows are increasingly

sensitive to heat stress due to increased milk production. For example, Zhou et al. [195] developed a three-node mechanistic model to predict body core, skin, and coat temperatures and obtained satisfactory accuracy on recent data from high-producing Holstein dairy cows. Another study by Yan et al. [194] compared ten thermal models with different combinations of previous equations and models to predict the BST of high-producing dairy cows. The two-layer mechanistic model was identified as the best, however, requires the input of measured CBT.

To sum up, the validation of these models, especially those developed earlier, in nowadays real-world data is typically insufficient. The complexity of the models makes this process even harder since more parameters have to be refined [203]. In addition, the complex equations and parameters may be too technical for frontline workers who are unfamiliar with the heat transfer process [189]. These reasons may contribute to mechanistic modelling being less popular in the research community than empirical modelling [204].

5. Assessment of heat stress in dairy cows

Once relevant data are available, the next step is to assess and quantify heat stress. Unlike humans, animals are unable to express their subjective feelings of comfort. Therefore, the assessment of animal thermal comfort relies heavily on physiological and behavioural indicators. Many studies compared the physiological or behavioural responses of animals under different THI classes [67,164,205]. Some studies separated the records according to different climate or management conditions (e.g., summer versus spring, cooling versus no cooling) [89,206]. These methods can identify a strong association between animal- and environment-based variables but are unable to obtain critical thresholds.

5.1. Development of critical thresholds

The most common methods for describing the relationship between environment- and animal-based indicators are piecewise and polynomial regression models. Although polynomial or other nonparametric models (e.g., generalised additive model) may bring better fitting [207,208], the fact that all data conform to the global equations makes them easily subject to changes in data distribution. Also, their results are insufficiently informative. In contrast, piecewise regression models, also known as segmented or broken-line models, rely on multiple linear regression models that are fit independently to each segment, allowing for a more flexible fit compared to a single linear regression model. These models offer helpful information, including the breakpoint as well as the slope before and after the breakpoint, which can inform the development of management strategies to better regulate animal responses to environmental stressors.

Spline regression models combine the advantages of piecewise and polynomial regression models. They use piecewise polynomial functions, known as splines, to model the relationship between dependent and independent variables. Splines are designed to be continuous and have continuous derivatives up to a specified order at the segment boundaries (known as knots). This allows spline regression models to

capture smooth changes in the relationship between the variables and provides a more flexible fit compared with piecewise regression models. The more knots, the more flexible the model, and the more critical thresholds. However, too many knots can also lead to overfitting problems. In summary, the key difference between piecewise and spline regressions lies in the continuity and smoothness of the fitted function at segment boundaries: piecewise regression models allow for abrupt changes, while spline regression models enforce smooth transitions.

Despite the abovementioned advantages, spline regression models are rarely used in dairy cattle heat stress studies. An example of using spline models to fit the effect of THI on milk yield is Ekine-Dzivenu et al. [205]. The cubic spline model had the best predictive performance however no thresholds were given. The unpopularity of spline models is possibly due to the assumed thermoregulation principles, which can be simply explained by a threshold effect. Below the upper critical threshold, animals remain comfortably within the thermoneutral zone, while above it, animals become exposed to a hot zone [1]. This upper critical threshold can be easily modelled by a simple piecewise regression model, which is particularly suited for capturing sudden changes in the response variable. Consequently, these models may offer a more intuitive representation of the underlying biological processes associated with the thermal comfort of dairy cattle. The recent critical thresholds on both environmental and animal sides are listed in **Table 1-6**.

The consideration of the plateau period when using piecewise functions is quite different among relevant studies, which could have a great impact on the results. For example, Li et al. [173] and Peng et al. [85] collected data from the same dairy farm and both used segmented regression to obtain heat stress thresholds for RT in dairy cows. The former assumed the existence of a plateau while the latter did not. The THI thresholds were 70 and 74.1, respectively. As the environmental condition and animal status were almost identical, it can be inferred that the introduction of the plateau period could lower the THI threshold to some extent. In fact, the decision should be based on the response pattern of specific physiological indicators. For example, it is sensible to use a plateau to fit CBT as the normal CBT of homeotherms fluctuates within a narrow range. In contrast, the plateau may be not applicable to fit other indicators, such as BST and RR, which consistently change with the thermal environment. An example can be found in a recent study by Dado-Senn et al. [209], where RT remained nearly stable with a flat slope before the breakpoint, while RR increased at a slight slope before the breakpoint.

In addition, the range of thermal conditions during the testing days is important to the results. For example, Dado-Senn et al. [210] and Kovács et al. [211] both used piecewise regression models to determine the THI thresholds for RT, RR, and BST in dairy calves. However, the THI thresholds of the former were between 65 and 69, while the THI thresholds of the latter were between 82.4 and 88.1. It is mostly because the former was conducted under a much broader range of thermal exposures (THI of 60 to 85) than the latter which had an average THI greater than 80 on three out of four experimental days. Therefore, the thresholds obtained by the latter may be the point in which heat stress was aggravated rather than triggered.

Table 1- 6: Summary of heat stress thresholds in dairy cattle based on body temperatures, respiratory dynamics, and behavioural changes.

Indicator	Reference	Thermal condition	Cow information	Threshold for animal-based indicator	Threshold for environmental indicator
Rectal temperature	[212]	A continental climate with a THI range of 17.8 to 85.1	High-producing Holstein dairy cows	38.4 °C	THI [a] 70
	[173]	A hot climate with a Ta range of 9.5 to 30.8 °C	High-producing Holstein dairy cows	38.6 °C	THI [a] 70 Ta 20.4 °C
	[85]	A temperate continental monsoon climate with a THI range of 58 to 84	High-producing Holstein dairy cows	38.55 °C	THI [a] 74.1
	[213]	BGHI higher than 72	½ Holstein dairy cows	Unknown	BGHI 76.44
			¾ Holstein dairy cows	Unknown	BGHI 73.51
	[210]	A subtropical climate with an averaged THI of 78 (shade only)	24 preweaning dairy calves	Unknown	THI [a] 67
	[209]	A continental climate with a THI range of 60.8 to 77.3	63 Holstein dairy calves	38.5 °C	THI [a] 69 Ta 21.5 °C
	[211]	A hot climate with a THI range of 70.3 to 94	Preweaning Holstein bull calves	Unknown	THI [b] 88.1
	[214]	A subtropical climate with an averaged THI of 76.45	111 dry cows	Unknown	THI [a] 77
Vaginal temperature	[215]	A temperate continental monsoon climate with a daily mean Ta range of 14.7 to 25.8 °C	20 high-producing Holstein dairy cows	38.7 °C	Ta 25.3 °C
	[216]	THI ranged from 55.8 to 79.9	Multiparous nonpregnant Holstein cows	Unknown	THI [c] 69
	[217]	A hot-humid climate with an average THI of 82.4	Dairy cows	38.9 °C	Unknown
	[135]	Unknown	Lactating Holstein cows	Unknown	THI [a] 70

Milk temperature	[82]	A subtropical climate with a Ta range of 2.0 to 38.0 °C	Lactating Holstein cows	Unknown	Dynamic thresholds
Body surface temperatures: forehead, eye, and muzzle (mean, maximum)	[215]	A temperate continental monsoon climate with a daily mean Ta range of 14.7 to 25.8 °C	20 high-producing Holstein dairy cows	30.1 °C (mean forehead temperature) to 37.5 °C (maximum eye temperature)	Ta from 24.1 °C (maximum forehead temperature) to 25.4 °C (mean and maximum muzzle temperature)
Body surface temperatures: head, eye, cheek, ear, neck, trunk, udder, foreleg, and hindleg (mean, maximum)	[218]	A temperate climate	233 high-producing Holstein cows	Unknown	THI [a] from 69.1 (mean cheek temperature) to 77.9 (maximum udder temperature)
Body surface temperatures: forehead (mean, maximum)	[85]	A warm temperate semi-humid continental monsoon climate with a THI range of 58 to 84	488 high-producing Holstein cows	30.05 °C, 30.34 °C	THI [a] 71.4, 66.8
Body surface temperatures: eye, hindquarter, nose, horn, and ear	[219]	THI ranged from 75.1 to 84.7	Hanwoo heifers	Unknown	THI [a] 65 (eye) to 70 (hindquarter)
Body surface temperature: ear	[211]	A hot climate with a THI ranging from 70.3 to 94	Preweaning Holstein bull calves	Unknown	THI [b] 83.0
Respiration rate	[173]	A hot climate with a Ta range of 9.5 to 30.8 °C	High-producing Holstein cows	48 bpm	THI [a] 70 Ta 20.4 °C
	[215]	A temperate continental monsoon climate with a daily mean Ta range of 14.7 to 25.8 °C	20 high-producing Holstein dairy cows	53.8 bpm	Ta 24.4 °C

	[212]	A continental climate with a THI range of 17.8 to 85.1	Standing and lying high-producing Holstein cows	37, 39 bpm	THI [a] 70, 65
	[213]	BGHI higher than 72	½ Holstein dairy cows	30 bpm	BGHI 73.61
			¾ Holstein dairy cows	45 bpm	BGHI 72.29
	[135]	Unknown	Lactating Holstein cows	Unknown	THI [a] 70
	[220]	Unknown	Dry cows	61 bpm	Unknown
	[210]	A subtropical climate with an averaged THI of 78 (shade only)	24 preweaning dairy calves	Unknown	THI [a] 65
		A subtropical climate with an averaged THI of 78.25 (shade plus fans)	24 preweaning dairy calves	Unknown	THI [a] 69
	[209]	A continental climate with a THI range of 60.8 to 77.3	63 Holstein dairy calves	40 bpm	THI [a] 69
					Ta 21 °C
	[211]	A hot climate with a THI range of 70.3 to 94	Preweaning Holstein bull calves	Unknown	THI [b] 82.4
	[214]	A subtropical climate with an averaged THI of 76.45	111 dry cows	Unknown	THI [a] 75 to 77
Average drinking length	[167]	A THI range from 68 to 93	25 lactating Holstein cows	Unknown	THI [a] 84
Daily lying time	[160]	A moderate climate with a THI range of 42 to 77	196 lactating high-producing Holstein cows	660 min/d	Daily average THI [a] 67
Daily rumination time	[161]	A temperate climate with a THI range of 25.6 to 84.8	864 lactating Holstein cows	554 to 542 min/d	Daily average THI [c] 63
Daily rumination time	[221]	A moderate climate with a THI range of 20.4 to 85.9	183 lactating high-producing Holstein cows	535 min/d	Daily average THI [a] 52
Milk replacer intake	[210]	A subtropical climate with an averaged THI of 78 (shade only)	24 preweaning dairy calves	Unknown	THI [a] 82

THI = temperature-humidity index; Ta = ambient temperature; BGHI = black globe-humidity index; bpm = breaths per minute.
[a] THI equation from National Research Council [222]. [b] THI equation from Bianca [223]. [c] THI equation from Mader et al. [146].

Many studies only measured animal-based data once or twice a day. This sampling frequency, however, may affect the accuracy of the developed threshold. If the indicators of interest change more frequently than the frequency of measurements, it is difficult to observe the real effect [70]. For example, VT should be measured every 120 min or more to reflect daily average CBT [73]. In addition, RR should be measured every 90 min to provide an accurate reflection of heat stress [154].

Various factors have an impact on the sensitivity of dairy cows to their thermal environments, including internal cow-related factors (e.g., age, breed, production level, lactation stage, parity, body posture, previous thermal exposure, acclimation) [1,68,212] and external management factors (e.g., milking, cooling, nutrition, shading) [93,210,224]. Therefore, it is insensible to assess the heat stress state of cows based on a generic threshold.

To deal with these dilemmas, efforts have been made to adjust or customise the thresholds according to various animal and management factors. The heat load index was developed for feedlot beef cattle by Gaughan et al. [147]. Genotype, coat colour, health status, acclimatisation, shade, days on feed, manure management, and drinking water temperature were considered for adjustment when determining the threshold for reference animals. Other studies tried to customise thresholds according to different levels of productivity [82], posture [212], cooling strategy [210], and lactation number and stage [225]. Further studies are required to continuously customise heat stress thresholds for internal and external factors that have an impact on the sensitivity to heat stress.

Notably, the difference in the sensitivity of heat stress manifests both on environmental and physiological sides. For instance, high-producing cows may enter heat stress at a lower THI and a higher RT [82,226]. However, many studies did not report the threshold and its confidence interval of animal-based indicators, which might have been discouraged by the lack of automated measurement for animal-related data in actual farms.

Most of the present critical thresholds assume a simultaneous change of physiological indicators with the thermal environment. For the detection of early signs of heat stress, the high sensitivity of the indicators to environments may render the time effect negligible. However, when predicting long-term effects, such as milk production, the duration of heat exposure becomes a strong influencing factor [53]. In addition, when dairy cows have been exposed to long-term heat stress, some of them may enter a new physiological state and can better dissipate excess body heat, manifesting as increased RR and BST as well as restored milk yield [227]. Therefore, existing RR and BST thresholds might be biased upward to some extent if too much thermotolerant data were used for threshold development.

5.2. Sensitivity of animal-based indicators in response to heat stress

As discussed above, many physiological and behavioural indicators can reflect heat stress in dairy cows, but they function in different stages. Indeed, it has long been recognised that some indicators like RR represent the effort made by

thermoregulation while an increased CBT is more like the result of thermal equilibrium [228].

From a thermodynamics point of view, homeotherms take two ways to dissipate body heat: sensible heat loss (non-evaporation) and latent heat loss (evaporation). The former consists of conduction, convection, and radiation while the latter consists of respiratory and cutaneous evaporation. Cutaneous evaporation can account for up to 80% of latent heat loss in a tropical environment [4], but it is rather difficult to measure. When Ta increases beyond the upper critical temperature of the thermoneutral zone, the non-evaporative ways of heat loss appear to be far less efficient [195]. When Ta is close to 32 °C, Holstein cows even start to gain heat from the environment through sensible heat transfer [3]. As the thermal environment continues to deteriorate, cows can succumb to hyperthermia with an increased CBT at some point where they fail to maintain thermoneutrality [1].

From a physiological point of view, when Ta rises in the first place, homeotherm will take three major physiological procedures to increase heat dissipation: vasodilatation, sweating, and panting [5]. Vasodilatation is to drive more blood from the core to the skin and peripheral to increase BST so that promotes sensible heat loss. The remaining two procedures are to promote evaporation. At the same time, cows will adopt a series of behavioural change strategies in response to heat stress [5]. Strategies for promoting heat loss include increasing standing time to promote convection through increased surface area, as well as increasing water intake to promote evaporation. Strategies for reducing heat production include decreasing physical activity, feed intake, rumination time, etc. Only when these preliminary efforts fail to dissipate excess body heat would CBT increase abnormally [196,220]. When the increase in BST and RR is effective to keep thermoneutrality, RT can remain unchanged [229].

As summarised in **Table 1-6**, the temporal sequence of animal responses to heat stress is supported by many studies, with BST and RR increasing at lower THI or Ta thresholds relative to CBT, and BST responding even before RR due to direct contact with the ambient environment [85,209,210,212,213,215]. RR was reported to increase approximately an hour earlier than RT [131], and panting was found to respond about 15 min earlier than VT both when increasing and decreasing [130]. This evidence supports the increase in CBT as a consequence of insufficient thermoregulation.

Although MT is a time- and labour-saving way to detect heat stress in dairy cows, it is only accessible to lactating cows [79]. Also, MT may be less effective to inform early cooling decisions due to infrequent measurements. At the same time, MT cannot provide suggestions for an intensive cooling session in the waiting room prior to milking. Instead, MT is more like a quality controller to assure that the intensive cooling protocol implemented before milking truly works. It should also be mentioned that bringing cows to the milking parlour may cause additional stress and thus biasing the use of MT in indicating heat stress [230].

In addition, behavioural indicators often detect the onset of heat stress at lower THI thresholds compared with physiological indicators, yet they are less frequently used for this purpose (**Table 1-6**). Of note, THI thresholds for basic behaviours are often summarised by daily average since the categorical outcome (presence or absence of certain behaviour) cannot be linked to environmental indicators directly with a piecewise model. One possible solution is to summarise these behaviours in herd proportions, thereby converting categorical data into numeric. This may allow a fairer comparison of behavioural and physiological indicators based on their associated THI thresholds. Anyway, automated behavioural recognition, especially non-contact methods based on deep learning, should be better leveraged to improve decision-making in managing heat stress.

Greater energy demanded by preliminary thermoregulatory efforts (e.g., panting and sweating) combined with lower energy intake (decreased feed intake) results in a decrease in the energy used for production [231,232]. Cooling measures are used to reduce the extra energy used by cows to maintain thermoneutrality, thereby consistently maximising their production performance. Cooling support may be more efficient when cows are still able to dissipate heat through their thermoregulation system. It has been reported that providing convective cooling could better reduce heat stress response and yield higher milk production when the temperature gradient from the skin to the surrounding environment was larger [233].

Recently, there has been a trend to use more sensitive animal-based indicators to inform heat stress management [210,212], probably driven by the increased interest in PLF. Since BST are highly linearly related to Ta, it may be difficult to find a valuable inflection point. Linking BST to other indicators like RR may help determine the best time at which interventions should be implemented. In addition, herd-level behavioural indicators may provide new solutions for heat stress detection and should thus be considered in future studies.

5.3. Selection of thermal indices for specific environments

As pointed out by other reviews, there is no recognised best thermal index at present [234,235]. This means that environmental indicators should be selected to find the one that best represents the specific thermal environment. Thermal indices mentioned in this review are summarised in **Table 1-7**.

The selection of environmental indicators often relies on the relationships with animal-based indicators [236,237]. In a recent study by Yan et al. [238], nine typical cattle thermal indices were compared using 273 lactating Holstein-Friesian dairy cows raised in a naturally ventilated commercial barn. Their results revealed that comprehensive climate index correlated most with BST ($r = 0.849$) while some indices even showed no significant correlation at all. However, in another recent study where the correlations of daily milk yield and MT with ten thermal indices were compared, THI or Ta alone showed the best results [239]. Interestingly, Ta showed the best correlation with animal-based indicators than other more complicated thermal indices in many studies [48,136,209,239,240].

Table 1- 7: Summary of the thermal indices mentioned in this review.

Thermal index	Formula	Resource
Temperature-humidity indices (THIs)	$\mathrm{THI} = (1.8 \times \mathrm{Ta} + 32) - (0.55 - 0.0055 \times \mathrm{RH}) \times (1.8 \times \mathrm{Ta} - 26)$	[222]
	$\mathrm{THI} = (0.35 \times \mathrm{Ta} + 0.65 \times \mathrm{Tw}) \times 1.8 + 32$	[223]
	$\mathrm{THI} = 0.8 \times \mathrm{Ta} + (\mathrm{RH}/100) \times (\mathrm{Ta} - 14.4) + 46.4$	[146]
Black globe-humidity index (BGHI)	$\mathrm{BGHI} = \mathrm{Tbg} + 0.36 \times \mathrm{Tdp} + 41.5$	[241]
Comprehensive climate index (CCI)	$\mathrm{CCI} = \mathrm{Ta} + \mathrm{Eq.}\,[1] + \mathrm{Eq.}\,[2] + \mathrm{Eq.}\,[3]$ $\mathrm{Eq.}\,[1] = \mathrm{e}^{(0.00182 \times \mathrm{RH} + 1.8 \times 10^{-5} \times \mathrm{Ta} \times \mathrm{RH})} \times (0.000054 \times \mathrm{Ta}^2 + 0.00192 \times \mathrm{Ta} - 0.0246) \times (\mathrm{RH} - 30)$ $\mathrm{Eq.}\,[2] = \frac{-6.65}{\mathrm{e}^{\left(\frac{1}{(2.26 \times \mathrm{WS} + 0.23)^{0.45 \times \left(2.9 + 1.14 \times 10^{-6} \times \mathrm{WS}^{2.5} - \log_{0.3}(2.26 \times \mathrm{WS} + 0.33)^{-2}\right)}}\right)}$ $- 0.00566 \times \mathrm{WS}^2 + 3.33$ $\mathrm{Eq.}\,[3] = 0.0076 \times \mathrm{SR} - 0.00002 \times \mathrm{SR} \times Ta + 0.00005 \times \mathrm{Ta}^2 \times \sqrt{\mathrm{SR}} + 0.1 \times \mathrm{Ta} - 2$	[242]
Heat load index (HLI)	$\mathrm{HLI}(\mathrm{Tbg} < 25) = 10.66 + 0.28 \times \mathrm{RH} + 1.9 \times \mathrm{Tbg–WS}$ $\mathrm{HLI}(\mathrm{Tbg} > 25) = 8.62 + 0.38 \times \mathrm{RH} + 1.55 \times \mathrm{Tbg} - 0.5 \times \mathrm{WS} + \mathrm{e}^{2.4 - \mathrm{WS}}$	[147]

Ta = ambient temperature (°C); RH = relative humidity; Tw = wet bulb temperature (°C); Tbg = black globe temperature (°C); Tdp = dewpoint temperature (°C); WS = wind speed (m/s); SR = solar radiation (W/m^2).

The different performances of thermal indices in different studies could be attributed to the different conditions under which they were originally modelled. For example, THIs with larger weights on humidity are more suitable for use in areas with higher humidity, and vice versa [243]. In addition, the source of weather data varies among studies. Most studies use on-farm measurements or data from a nearby weather station. In the study of Ji et al. [239], the thermal indices calculated using data from on-farm measurements and the local weather station had similar correlations with MT. However, the THI calculated from on-farm data was significantly higher than those calculated from the closest weather station in the study of Shock et al. [244]. Ideally, environmental measurements should be in close proximity to animals to better represent their microenvironments.

To summarise, all of these inconsistencies highlight the need for future studies to compare and select thermal indices for specific farm locations, or at least to use thermal indices that were developed in a similar environment. Plus, on-farm measurements should be used to provide precise representations of

microenvironments. After all, inappropriate use of thermal indices may obscure true animal responses and bias latter-developed thresholds.

6. Future development of heat stress detection methods

This section discusses the extent to which a comprehensive data acquisition strategy, including both direct measurement and predictive modelling, as well as internet of things (IoT) technologies, can facilitate precise heat stress assessment.

6.1. Better acquisition of animal-based indicators by PLF techniques

In pursuit of PLF and animal welfare, an ideal measurement method for heat stress response should be automatic, accurate, continuous, non-invasive, low-cost, and real-time [49,120,121].

Indwelling devices measuring CBT rectally and vaginally cannot work for a long time due to interference with normal physiological activities. These devices are better used for short-term research purposes. Ear-canal sensors and rumen boluses may be more appropriate as they are less invasive. Implantable devices can measure CBT subcutaneously, but their cost and long-term safety should be further examined. In addition, wearable devices can accurately measure basic behaviours with little to no invasion and have shown a possibility to measure BST and RR, and further estimate CBT. Especially, wearable devices equipped with wireless transmission and positioning technologies are of particular importance in the setting of grazing pastures due to their ability to remotely monitor the thermal comfort of individual animals. Battery life, biocompatibility, stability, remote transmission, and accuracy should be improved for these devices targeting grazing situations [245].

Non-contact measurements based on IRT are non-invasive to dairy cows but are too expensive and difficult to achieve real-time measurement at the herd level. They are more suitable for quick monitoring at specific locations such as calf feeders and pathways to milking. RGB camera-based methods provide a low-cost solution for measuring heat stress response. With the help of advanced computer vision and deep learning algorithms, both physiological and behavioural data can be detected from RGB images and video streams. Further work in this regard should increase the number of animals covered by the camera's view to measure, evaluate, and relieve heat stress at the herd level.

Identification is necessary when individual-level analysis and management are of interest. Wearable sensors are easier to obtain individual-level measurements due to their inherent one-to-one relationship with animals. In contrast, non-contact methods need to be integrated with identification algorithms or companion wearables in order to map individual information. For example, cow drinking behaviour can be recognised with non-contact antennas at water bins and contact identifiers on cows [169]. In fact, non-contact methods typically hold the advantage of measuring social behaviour since their broader measurement scale can cover more animals. An

example of combining cow identities and their social behaviour is the study of Ren et al. [246], in which ultra-wideband technology was used for cow identification and location, alongside computer vision for social behaviour recognition.

In addition, behavioural dynamics can be measured and analysed not only at the individual level but also at the herd level. An interesting demonstration is the work of Laurindo et al. [165] where cow lying and standing behaviours were recognised and analysed by herd percentages. The herd-level seasonal pattern was thus obtained. Another example is the study of Xu et al. [247] where group behaviours were connected with herd thermal comfort. The behaviour of pigs was classified through the clustering and deep learning method and their aggregation degree showed a positive association with indoor temperature.

Predictive models are always expected as a non-invasive way to obtain heat stress response. In fact, predictive models have a two-way relationship with measurement and can do more than just acquire data (**Figure 1-3**). The accuracy of a predictive model is always highly dependent on the quality and quantity of the measurements used to train or tune it, and the model can in turn guide the measurement. For example, the interpretation of a model can help infer which variables are of greater importance for making an accurate prediction. Understanding the complex interactions between these factors allows practitioners to develop more effective and tailored interventions against heat stress and maximise cow health and productivity.

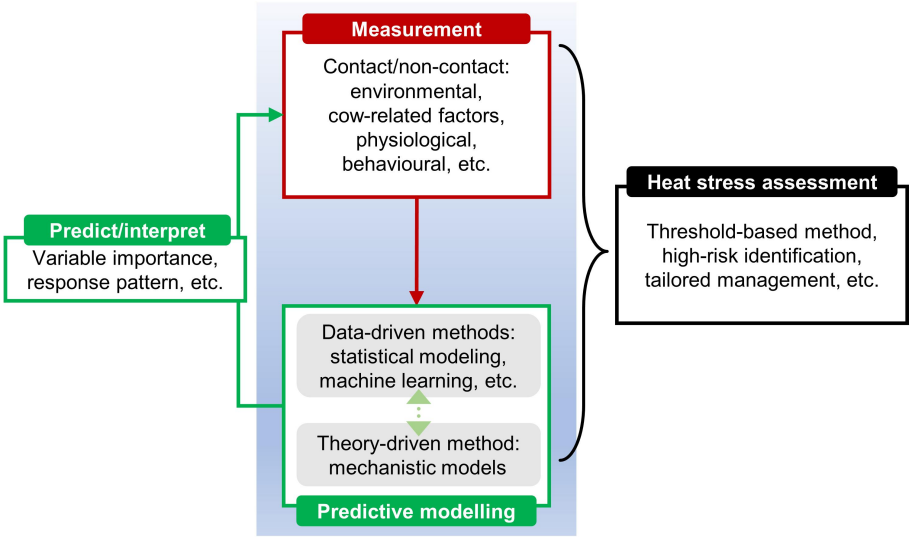


Figure 1-3: Measurement, prediction, and assessment of heat stress in dairy cows.

Some ideas about connecting ML and mechanistic models have been proposed [189]. For example, mechanistic models are able to create data to feed ML models [65], and ML outputs may also be used as inputs to mechanistic models [248]. However, these ideas remain in the data acquisition stage and do not alter the nature of modelling techniques. The more prospective way is to combine the advantages of

both methods to create a so-called grey-box model. This hybrid model is supposed to train the black-box model (i.e., ML model) while taking into consideration specific prior knowledge from the white-box model (i.e., mechanistic model). For example, some advanced grey-box models make use of Bayesian approaches [249], which rely more on the data when data is reliable and fall back to a theory-based method when the data is unreliable. This strategy should improve the accuracy and interpretability, as well as reduce the amount of data required for training. However, it requires developers to have a higher level of mathematical expertise.

To summarise, wearable devices were developed earlier and thus are more readily available in today's market. With the advancement of deep learning algorithms, vision-based non-contact methods are gaining popularity in the research community and have particular advantages in measuring interactive behaviours. The development of contact and non-contact measurement methods can accelerate the development of data-driven predictive models by allowing access to more relevant parameters. Plus, mechanistic models can also benefit from high-quality measurements when refining parameters or validating theories.

6.2. IoT technologies help with precision management

Once the thresholds are determined, it is crucial to assess how to use them in practice. Currently, the prevalent method is to indirectly evaluate the heat stress level of the entire herd through the thermal indices calculated from easily accessible meteorological data, such as data from nearby weather stations, or average values from temperature and humidity sensors placed in the barn. Considering the significant variation in microclimates to which individual dairy cows are exposed [250], a more precise approach would involve deploying more sensors to divide the barn into smaller measurement and control units. By leveraging IoT technology, the cooling facilities within specific units can be continuously monitored, enabling more effective and efficient management of heat stress.

It is always important to remember that environmental indicators neither truly reflect how animals respond to the changes in the thermal environment nor show whether animals are being adequately cooled [49]. Stress responses were reported to remain at a high level when THI was low at night [135], which provides a good temperature gradient supporting heat dissipation from the animals to the environment. Indeed, cooling sessions at night have been demonstrated to be effective in relieving cows' accumulated heat during the day [251], but they can be ignored by farms relying solely on the thresholds of environmental indicators. In addition, individual differences in the sensitivity of heat stress response exist even within the same subgroup of cows [68,149], meaning that cooling measures may be wasted on low-risk animals. Therefore, direct monitoring of animal-based indicators at the individual level should help better manage heat stress.

At present, many large-scale commercial dairy farms have used smart ear tags or collars to monitor several daily behaviours of individual cows. Combined with positioning or detection technologies, the nearby cooling facilities can be turned on automatically when a large proportion of cows breathe heavily [252]. Such an IoT

system is very promising not only because it can detect heat stress response more accurately and earlier based on animals' feelings, but also because it can save energy. Another location to integrate such technologies is on the way to the parlour, where heat-stressed cows can be automatically detected and directed to a separate room for heat abatement [253].

Integrating both environment- and animal-based indicators for a comprehensive assessment of heat stress also has a good prospect. Due to different response dynamics, it is possible to distinguish heat stress from other events that may lead to increased body temperatures and accelerated breathing, such as exercise [69], diseases (e.g., infections [116]), physiological processes (e.g., oestrous [254]), and management factors (e.g., bringing cows to the milking parlour [230]). Plus, considering that thermal comfort is essentially a highly ambiguous concept, some efforts have been made to build a comprehensive index by using fuzzy mathematical methods [255,256]. These works fused parameters monitored by both on-farm and on-animal sensors to obtain a comprehensive assessment of thermal comfort. However, the determination of weights, fuzzy sets, and membership functions in fuzzy mathematics often rely on expert experience and researchers' subjective judgment, which may lead to uncertainty and instability of the results.

7. Conclusion

Some key conclusions can be drawn from this review:

- BST, RR, and relevant behaviours are the most appropriate animal-based indicators for the detection of heat stress in dairy cows due to their ease of acquisition and sensitivity to heat stressors.
- Many studies that relied on piecewise regression models did not report thresholds for animal-based indicators probably due to the lack of automated measurement for animal-related data in actual farms.
- The existing environmental thresholds should be used carefully due to the differences in environmental conditions and animal information. Future studies should evaluate thermal indices and customise thresholds for specific locations.
- Commercially available collars and boluses can measure basic behaviours and CBT in real time while vision-based non-contact methods are gaining popularity in the research community due to their potential for non-invasively measuring various heat stress responses.
- The development of measurement methods can accelerate the development of data-driven predictive models, and these models can in turn guide the measurement with their interpretations.
- Combined with IoT technologies, a comprehensive strategy based on both animal- and environment-based indicators is expected to increase the precision of heat stress detection in dairy cows.

Chapter 2

Problematic, research aim, thesis outline, and experimental design

1. Problematic

Heat stress detection in dairy cows has long relied on the relationship between environmental and production indicators. However, the reduction in milk yield lags behind the exposure to heat stress events for about two days. Plus, environmental indicators can neither capture the true need for a response by the animals nor deal with differences between individuals. Other stress responses, such as physiological and behavioural changes, are well documented to be activated by dairy cows in the earlier stage dealing with heat stress compared with production loss. Among all candidate indicators, body surface temperatures (BST), respiration rate (RR), and relevant behaviours have been concluded to be the most appropriate indicators due to their ease of acquisition and high sensitivity to heat stressors. Vision-based methods are promising for accurate measurements while adhering to animal welfare principles. Meanwhile, predictive models show a non-invasive alternative to obtain these data and can provide useful insights with their interpretations. Thus, this project was conducted to deal with the following research question: how can vision-based methods and predictive models be effectively utilised to improve heat stress detection in dairy cows?

2. Research aim

This project aimed to leverage artificial intelligence techniques, specifically computer vision and machine learning, to develop non-invasive solutions for the detection of heat stress in dairy cows. To accomplish this aim, the thesis was fractionated into several objectives from various research scopes:

- developing non-contact measurement methods for the selected animal-based indicators of heat stress (i.e., BST, RR, and relevant behaviours);
- developing critical thresholds for the onset of heat stress based on automated measurements from the developed measurement methods;
- predicting and interpreting physiological responses of dairy cows by using machine learning methods.

3. Thesis outline

This thesis was organised into seven chapters. **The first chapter** provided a comprehensive review of the progress in the measurement, prediction, and assessment of heat stress in dairy cattle. In light of the insights obtained**, the second chapter** defined the problem, aim, outline, and experimental design of this thesis.

The following experimental chapters were designed in order to fill the identified gaps in methodological approaches for the automated detection of heat stress. **The third chapter** was a semantic segmentation mission aiming to collect facial temperatures from cattle infrared images. **The fourth chapter** involved an instance segmentation mission aiming to detect RR for multiple cows in video clips. **The fifth chapter** focused on an object detection mission aiming to recognise cow behaviours and determine the onset of heat stress at the herd level. **The sixth chapter** was a predictive modelling study aiming to predict the physiological

responses of dairy cows using easily acquired data and help identify cows at the highest risk through model interpretation.

Finally, **the seventh chapter** was devoted to general discussion and conclusions of all the notable findings and contributed to future research on heat stress detection in dairy cows through some recommendations and perspectives.

4. Experimental design

As shown in **Figure 2-1**, field experiments were first conducted to collect various animal and environmental data. Next, raw data were processed manually to obtain the ground truth on which the subsequent works were done. Then, on the one hand, vision-based measurement methods were developed for the selected animal-based indicators (i.e., BST, RR, and relevant behaviours) following different machine vision methodologies. On the other hand, machine learning-based predictive modelling was performed, with animal and environmental information as predictors and ground-truth physiological indicators as outcomes. After that, the automated measurements, along with the predictions from the predictive models, were examined for their agreement with the ground truth. Finally, critical thresholds were determined using piecewise regression models with automated measured animal-based indicators as outcomes and environmental indicators as predictors. High-risk animals were identified through the interpretation of the predictive models. Together, these two elements constituted a comprehensive assessment of heat stress in dairy cows.

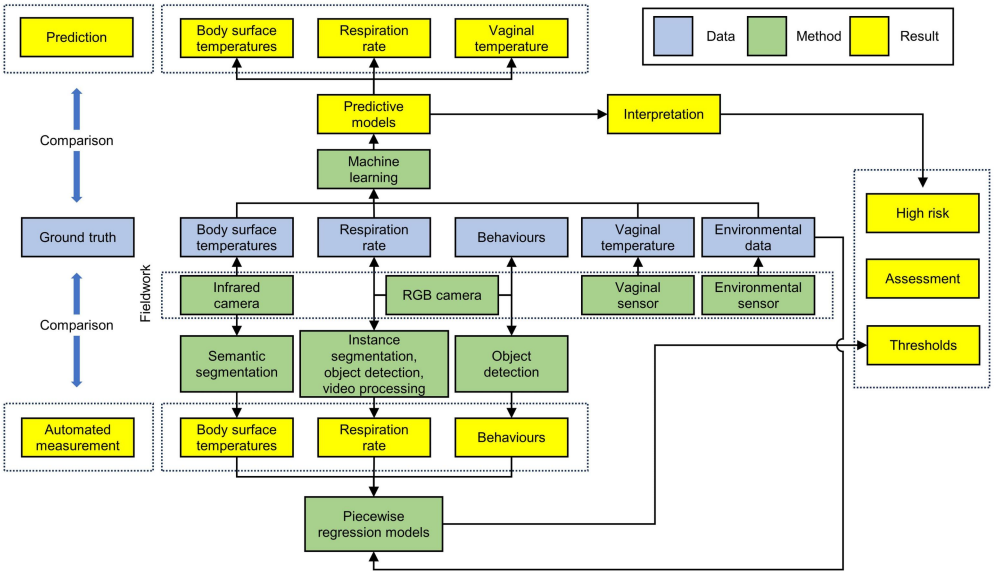


Figure 2- 1: Technical route of the project 'Advancing heat stress detection in dairy cows through machine learning and computer vision'.

Context – Chapter 3

According to the findings reported in the first chapter, body surface temperatures are promising indicators of heat stress in dairy cows due to their ease of acquisition and sensitivity to heat stressors. The infrared thermography-based non-contact measurement method is becoming increasingly popular because of its welfare-friendly nature. However, traditional temperature collection requires massive manual operations on professional software. Therefore, the third chapter aimed to propose a tool for automated temperature collection from cattle facial landmarks (i.e., eyes, muzzle, nostrils, ears, and horns) and to determine critical thresholds for the onset of heat stress based on automated measurements.

Chapter 3

Improved UNet for facial temperature analysis in dairy cow infrared images

Adapted from:

Shu, H., Wang, K., Guo, L., Bindelle, J., Wang, W. *Automated collection of facial temperatures in dairy cows via improved UNet*. Computers and Electronics in Agriculture, 2024. https://doi.org/10.1016/j.compag.2024.108614

Abstract

In cattle, facial temperatures (FT) captured by infrared thermography provide useful information from physiological aspects for researchers and local practitioners. Traditional temperature collection requires massive manual operations on relevant software. Therefore, this paper aimed to propose a tool for automated temperature collection from cattle facial landmarks (i.e., eyes, muzzle, nostrils, ears, and horns) and to determine critical thresholds for the onset of heat stress based on automated measurements. An improved UNet was designed by replacing the traditional convolutional layers in the decoder with Ghost modules and adding Efficient Channel Attention (ECA) modules. The improved model was trained on our open-source cattle infrared image dataset. The results show that Ghost modules reduced computational complexity and ECA modules further improved segmentation performance. The improved UNet outperformed other comparable models on the testing set, with the highest mean Intersection of Union of 80.76% and a slightly slower but still good inference speed of 32.7 frames per second. Further agreement analysis reveals small to negligible differences between the temperatures obtained automatically in the areas of eyes and ears and the ground truth. The piecewise regression models based on the automated FT measurements determined an ambient temperature threshold at 26.1 °C when the mean eye temperature reached 35.3 °C. Collectively, this study demonstrates the potential of the proposed method for automated FT collection and heat stress detection through infrared thermography. Further modelling and correction with data collected in more complex conditions are required before it can be integrated into on-farm monitoring of animal health and welfare.

Keywords: precision livestock farming, animal welfare, deep learning, body surface temperatures, heat stress

1. Introduction

Infrared thermography (IRT) is the technique of detecting infrared radiation from an object, converting it to temperature, and visualising the temperature distribution with an image [257]. Due to its non-contact advantage, IRT has been widely used in human fever detection and health evaluation. In husbandry, IRT can contribute to precision livestock farming which aims to provide an automated protocol for monitoring animal health and welfare parameters [258].

In cattle, temperatures obtained from specific facial landmarks (e.g., eyes, forehead, nostril, ears, horns, cheek) have been widely used as indicators or predictors of health conditions such as physiological state [259], bovine respiratory disease [123], and foot-and-mouth disease [260]; animal welfare issues such as temperament [261], emotions [262], and heat stress [85]; and productivity issues such as feed efficiency [263] and meat quality [264].

In order to collect the temperature of the abovementioned facial landmarks, facial regions of interest (RoI) must first be defined. In most literature, RoI are defined manually in infrared images using relevant processing software due to the lack of reliable detection tools for cattle facial landmarks [127,265]. Thus, there has been a growing interest among researchers to develop such a tool to increase the efficiency of dealing with cattle infrared images. In previous studies, RoI such as eyes, ear base, cheek, and nose have been localised in infrared images for specific purposes [87,125,266,267]. However, very limited effort has been contributed yet to a comprehensive method for separating multi-class facial RoI in cattle infrared images. This method should be robust against usual interfering factors such as camera angle and changing microenvironment.

Of note, most previous works use traditional image processing techniques for detecting facial landmarks in cattle infrared images, such as Haar cascade classifiers [125] and thresholding [268]. The recent development of deep learning provides alternative solutions. For example, recent studies have achieved automatic ocular temperature collection using an object detection method based on improved YOLOv4 [126], YOLOv5 [269], and YOLOv7 [270]. In addition, semantic segmentation, as another central computer vision task that separates each pixel into pre-defined classes [271]. Its usefulness is particularly evident in the field of face parsing, which aims to create pixel-wise segmentation maps for facial parts in human RGB images [272]. It can be therefore imagined that this pixel-wise outlining can lead to more accurate and comprehensive temperature assessments from infrared images. However, no relevant work or attempts have been found yet.

Semantic segmentation in infrared images has to deal with some challenges [273]. One of the main challenges is the lower resolution of infrared images compared with RGB images, which can result in less detailed information and difficulty in accurately delineating object boundaries. Another challenge is the phenomenon of thermal crossover, where objects at similar temperatures blend into the background, making it harder to distinguish them. Advancements of more sophisticated algorithms and deep learning models, such as convolutional neural networks (CNN),

Vision Transformer, and other attention mechanisms, have markedly improved accuracy in recognition across various data, such as speech [274] and image [275]. Therefore, it is of great interest to explore to what extent these techniques can help in segmenting facial landmarks in cattle infrared images.

Thus, the primary objective of this study was to propose a semantic segmentation-based tool for automated facial temperature collection from cattle infrared images. Specifically, a baseline semantic segmentation network, namely UNet, was modified, trained, and compared its performance with other state-of-the-art models in segmenting cattle facial landmarks. Then, the temperatures obtained from the predictions of the improved UNet were compared with those obtained from the ground-truth annotations. Additionally, the secondary objective was to determine critical heat stress thresholds based on automated measurements from appropriate facial landmarks.

2. Materials and methods

Since there is no public infrared image dataset appropriate for the semantic segmentation of cattle facial landmarks, a field experiment was conducted for data acquisition. The experimental protocol was approved by the Experimental Animal Care and Use Committee of Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (approval number IAS2021-220).

2.1. Data acquisition

The experimental farm is located in Shandong, China (34◦50'37"N and 115◦26'11"E), and belongs to a temperate continental monsoon climate with hot and humid summers. It is worth noting that the temperature difference between the background and animals would change dramatically from non-heat-stressed months to heat-stressed months. Ignoring this fact would definitely affect the robustness of the trained network in practice. Thus, the experiment was conducted from May to August 2021 to cover a wide range of thermal environments from warm to hot. The free-stall pen was covered by a double-pitched roof, and therefore, most of the solar radiation was prevented from reaching the cows inside the barn. Electronic fans (1.1 m in diameter; capacity: 25000 m^3/h each) and sprinklers (flow rate: 1.5 L/min each; 1 min on and 4 min off) were automatically turned on when the indoor temperature reached 20 °C and 25 °C, respectively. Indoor microenvironmental parameters including ambient temperature (Ta) and relative humidity (RH) were measured by using six Kestrel 5000 and 5400 environment meters that were equally spaced in the barn (measurement interval: 10 min, accuracy: ± 0.4 °C Ta and ± 1% RH; Nielsen-Kellerman, Boothwyn, PA, USA). The temperature-humidity index (THI) was calculated according to Eq. (1) [222].

$$THI = (1.8 \times Ta + 32) - (0.55 - 0.0055 \times RH) \times (1.8 \times Ta - 26) \quad (1)$$

A total of 59 primiparous and multiparous Holstein dairy cows were selected for infrared thermal imaging. The infrared images were taken with a portable infrared camera (VarioCAM HR, InfraTec, Dresden, Germany) which has a spectral range from 7.5 to 14 μm, a temperature measuring range from -40 to 2000 °C, an accuracy

of ± 1.5 °C, and a resolution of 640 × 480 pixels. All images were taken at a distance of approximately 1 to 1.5 m from the cow. To increase the robustness of the proposed method in actual farms, cows were not restrained during photography and a wide range of situations including heterogeneous postures were covered. Thermal imaging was carried out between 08:00 and 17:00 h. All cows were healthy during the entire experiment.

2.2. Data pre-processing

Infrared images were initiated with IRBIS 3 Standard software (YSHY, Beijing, China). Before formal processing, images with low quality and multiple faces were manually eliminated, contributing to a dataset with 1000 images. All images were calibrated by setting the emissivity to 0.98 [276], and by inputting the averaged Ta record from the sensors corresponding to the time when they were taken. The images were outputted into greyscale JPEG format (640 × 480 pixels) with the temperature scale set to 295 to 315 degrees Kelvin. The temperature matrices were also outputted into CSV format for further temperature collection from segmentations. For a given greyscale image, facial landmarks that have been frequently used in dairy research (i.e., eyes, muzzle, nostrils, ears, horns) were annotated with polygons using Labelme (https://github.com/wkentaro/labelme.git) (**Figure 3-1**).

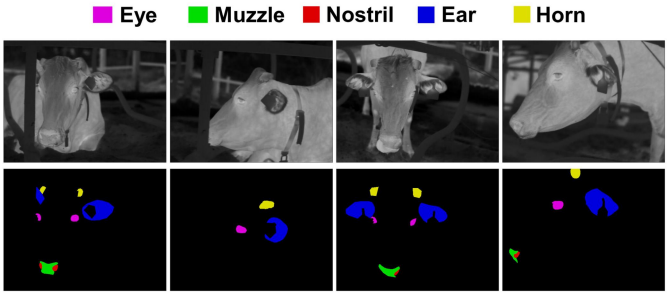


Figure 3- 1: Examples of cattle infrared images and their ground-truth annotations for facial landmarks.

Two splitting strategies were used to prepare the training, validation, and testing sets: (1) image-level splitting, where different images of the same cow were pooled together and then randomly split at a ratio of 8:1:1, and (2) cow-level splitting, where all cows were randomly allocated at a rough ratio of 8:1:1 to make sure that images from the same cow were always in the same set, thus making the three sets completely independent (**Table 3-1**). These two splitting strategies were compared to evaluate the magnitude of the overestimation induced by a simple random splitting since evidence has shown that it can lead to an overestimation of the model's performance when applied to unseen subjects [277].

Table 3- 1: Overview of the datasets split using two strategies.

Strategy	Training		Validation		Testing	
	No. images	No. cows	No. images	No. cows	No. images	No. cows
Image-level	800	59	100	59	100	59
Cow-level	782	46	125	6	93	7

The training sets were used to train the networks, the validation sets to tune the hyperparameters and obtain an initial assessment of accuracy, and the testing sets to collect the final performance. Data augmentation methods, such as flipping, rotation, brightness changing, contrast changing, sharpening, Gaussian noise adding, and elastic deformation, were performed in the training sets to improve the accuracy and generalisation capacity of the trained network (see **Appendix 1. Supplementary material Figure A3-1**). Thus, the training sets were seven-fold augmented. The images were resized to 512 × 512 pixels before they were fed to the networks.

2.3. Segmentation network architectures

2.3.1. UNet model

UNet, as a popular semantic segmentation network, has a symmetric U-shaped architecture of a contracting path for capturing global context and an expansive path for precise localisation, and uses skip connections between two paths to transfer context information to higher resolution layers [278]. In this study, the UNet with a VGG16 [279] encoder was used as the baseline (**Figure 3-2**). The downsampling block is repeated by two or three 3 × 3 convolutions (activated by ReLU functions) and one 2 × 2 max-pooling operation. Thus, the image is halved in size after each block, compensated with a doubled number of feature map channels. In the decoder part, repeated blocks including an upsampling (bilinear method), a concatenation with the corresponding feature map from the encoder, and two repeated 3 × 3 convolutions (each followed by a ReLU function) are used to fuse and reconstruct feature maps from both local details and global context. Finally, a 1 × 1 convolution with the number of channels set to the number of classes is used to generate class-wise classification results for each pixel.

2.3.2. Improved UNet model

The idea for improving the baseline UNet architecture was to lighten the network by reducing the number of parameters brought by 3 × 3 convolution while introducing an attention mechanism to suppress the transmission of irrelevant features. Specifically, as shown in **Figure 3-2**, the modification on the baseline UNet model happened to the decoder where two consecutive 3 × 3 convolutional layers are replaced by a combination of a Ghost with Efficient Channel Attention (GhostECA) module and a Ghost module. The detailed structure of the improved UNet is shown in **Table 3-2**.

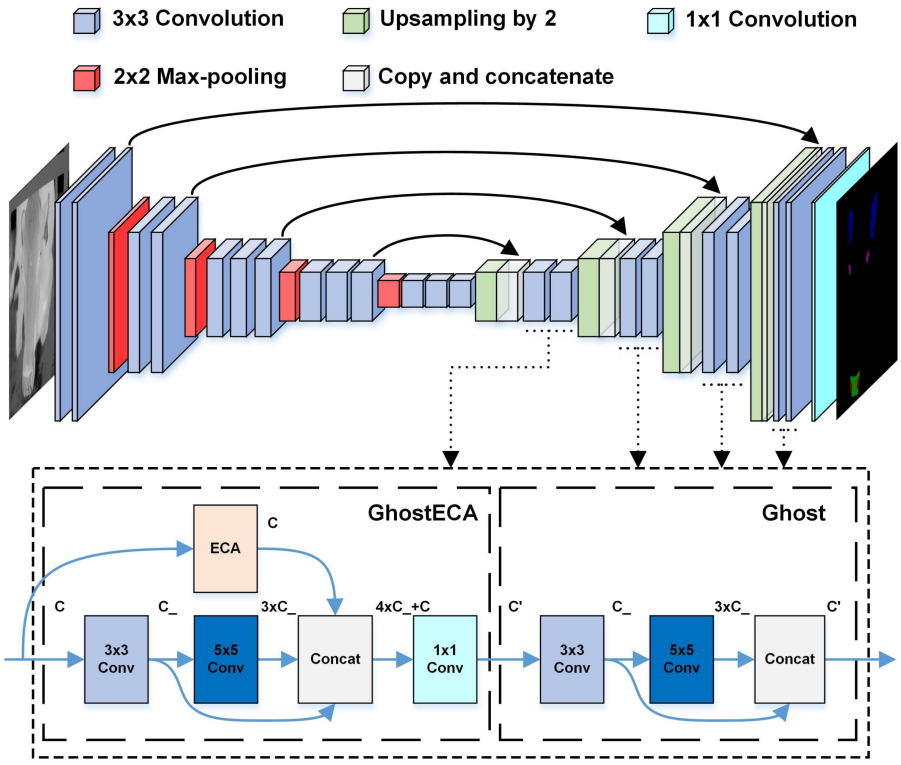


Figure 3- 2: Architecture of the baseline UNet model and the improved UNet model by replacing convolutions in the decoder with Ghost and GhostECA modules (shown in dashed boxes, where C and C' represent the input and output channel numbers, respectively, and C_=C'/4).

The Ghost module is a plug-and-play component that can be used to replace common convolutional operations in any classical CNN [280]. The idea of Ghost modules came from the observation of the intermediate feature maps calculated by mainstream CNN. The authors found redundancy in feature maps, in which some feature maps are very similar in pairs as if one of the pair is a 'ghost' of the other. This means that one feature map of the pair can be obtained by transforming the other feature map with cheap operations. Therefore, Ghost modules aim to generate more feature maps with fewer parameters and cheaper operations.

The applied Ghost module, as shown in **Figure 3-2**, consists of two consecutive convolutions. In the first convolution, the number of input channels is reduced to one-quarter, using a kernel size of 3 and a stride of 1. This reduction in channel dimensions helps reduce model complexity and computational cost. The second convolution expands the reduced channels by three times, using a larger kernel size of 5 and a stride of 1. Finally, the output of the second convolutional operation is concatenated with that of the first convolution to ensure that the number of output channels equals the number of original input channels.

Table 3-2: Structure of the improved UNet.

Layer	Kernel size & stride	Output shape	Connect to
Input	-	512 × 512 × 3	Convolution1
Convolution1	3 × 3, 1	512 × 512 × 64	Convolution2
Convolution2	3 × 3, 1	512 × 512 × 64	Max-pooling1 & Concatenate4
Max-pooling1	2 × 2, 2	256 × 256 × 64	Convolution3
Convolution3	3 × 3, 1	256 × 256 × 128	Convolution4
Convolution4	3 × 3, 1	256 × 256 × 128	Max-pooling2 & Concatenate3
Max-pooling2	2 × 2, 2	128 × 128 × 128	Convolution5
Convolution5	3 × 3, 1	128 × 128 × 256	Convolution6
Convolution6	3 × 3, 1	128 × 128 × 256	Convolution7
Convolution7	3 × 3, 1	128 × 128 × 256	Max-pooling3 & Concatenate2
Max-pooling3	2 × 2, 2	64 × 64 × 256	Convolution7
Convolution8	3 × 3, 1	64 × 64 × 512	Convolution8
Convolution9	3 × 3, 1	64 × 64 × 512	Convolution9
Convolution10	3 × 3, 1	64 × 64 × 512	Max-pooling4 & Concatenate1
Max-pooling4	2 × 2, 2	32 × 32 × 512	Convolution9
Convolution11	3 × 3, 1	32 × 32 × 512	Convolution10
Convolution12	3 × 3, 1	32 × 32 × 512	Convolution13
Convolution13	2 × 2, 2	32 × 32 × 512	Upsampling1
Upsampling1	-	64 × 64 × 512	Concatenate1
Concatenate1	-	64 × 64 × 1024	GhostECA1
GhostECA1	-	64 × 64 × 512	Ghost1
Ghost1	-	64 × 64 × 512	Upsampling2
Upsampling2	-	128 × 128 × 512	Concatenate2
Concatenate2	-	128 × 128 × 768	GhostECA2
GhostECA2	-	128 × 128 × 256	Ghost2

Ghost2	-	128 × 128 × 256	Upsampling3
Upsampling3	-	256 × 256 × 256	Concatenate3
Concatenate3	-	256 × 256 × 384	GhostECA3
GhostECA3	-	256 × 256 × 128	Ghost3
Ghost3	-	256 × 256 × 128	Upsampling4
Upsampling4	-	512 × 512 × 128	Concatenate4
Concatenate4	-	512 × 512 × 192	GhostECA4
GhostECA4	-	512 × 512 × 64	Ghost4
Ghost4	-	512 × 512 × 64	Convolution14
Convolution14	1 × 1, 1	512 × 512 × 6	-

However, most of the redundant feature maps generated by the Ghost module are not related to the useful features. Therefore, an attention mechanism was further integrated to make the network pay more attention to the most relevant features. It is well known that adding attention mechanisms to CNN can improve their performance. The attention mechanism in deep learning works similarly to human selective visual attention in that both aim to identify and emphasise the most important information from large amounts of data. Efficient channel attention (ECA) is an extremely efficient and lightweight channel attention mechanism proposed by Wang et al. [281]. The applied ECA module consists of three main steps (**Figure 3-3**). Firstly, a global average pooling operation is applied to the input feature maps, squeezing the spatial dimensions W × H to 1 × 1 while retaining channel-wise information. Next, a one-dimensional convolution with a kernel size of 3 is performed to achieve local cross-channel interaction and capture channel-wise dependencies. A sigmoid activation function is then used to compute channel-wise attention weights. Finally, the attention weights are multiplied element-wise with the input feature maps, allowing the network to selectively emphasise relevant information.

In this study, an ECA module is integrated into the first Ghost module of each upsampling block in the decoder to enhance its performance (**Figure 3-2**). The integrated GhostECA module has one more convolutional operation after concatenating the outputs of the first two convolutions and the ECA module in order to adjust the number of output channels.

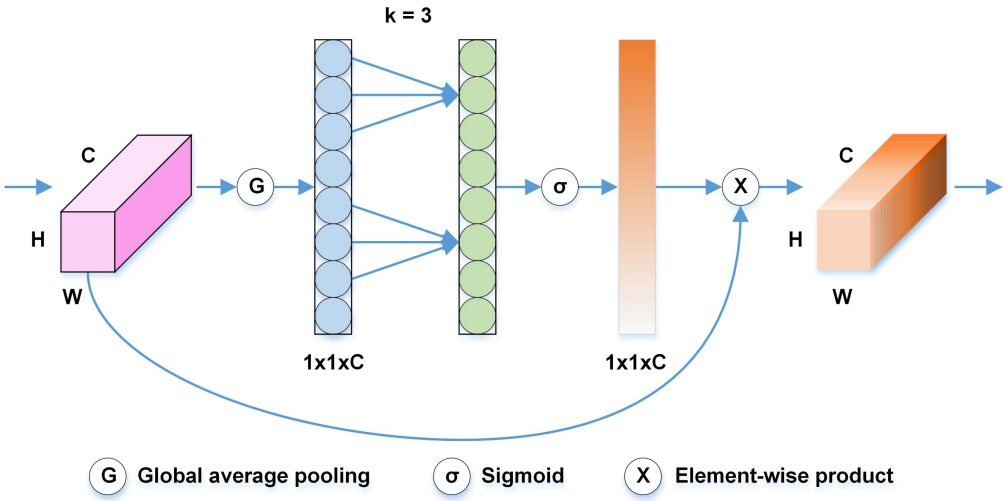


Figure 3- 3: Architecture of the efficient channel attention (ECA) module.

2.4. Segmentation network training

The training was performed in Python 3.7 language with Pytorch 1.13.0 on a 64-bit Windows 11 computer with NVIDIA GeForce RTX 3090 GPU. Transfer learning can significantly reduce the number of required images and increase training efficiency compared with training from scratch with randomly initialised weights. In this study, the initialised weights of all encoders were transferred from the networks pre-trained on the ImageNet dataset [282]. The epoch was set to 300, the batch size to 16, and the learning rate to 0.0001 with an Adam optimiser. A combination of cross-entropy and dice coefficient was used as the loss function (L_{CE} + L_{Dice}), as defined in Eqs. (2 and 3). Dice loss was used because it can effectively handle the pixel imbalance between foreground and background.

$$L_{CE} = -\sum_{i=1}^{C} t_i \log p_i \quad (2)$$

$$L_{Dice} = 1 - \frac{2TP}{2TP+FP+FN} \quad (3)$$

The C takes 5, indicating five classes of interest (i.e., 'eye', 'muzzle', 'nostril', 'ear', 'horn'), t_i and p_i are the ground truth and the Softmax probability of each pixel for each class i, respectively, TP denotes true positive (pixels correctly classified as a class of interest), FP denotes false positive (pixels incorrectly classified as a class of interest), TN denotes true negative (pixels correctly classified as the background), and FN denotes false negative (pixels incorrectly classified as the background or a wrong class). The model from the epoch with the lowest validation loss was used for testing.

2.5. Ablation and comparison studies

Ablation tests were conducted: (1) UNet with VGG16 as the backbone was used as the baseline model; (2) based on the baseline model, the convolutional layers in the

decoder were replaced by Ghost modules. This model is referred to UNet+Ghost; and (3) based on UNet+Ghost, ECA was integrated into the first Ghost module of each decoder. This is the proposed model to be compared, which is referred to UNet+GhostECA.

To show the competitiveness of the improved UNet model, it was compared with other popular semantic segmentation models in the field, including FCN [283] with VGG16 as the backbone (FCN-VGG16), PSPNet [284] with MobileNetV2 [285] and ResNet50 [286] as the backbone, respectively (PSPNet-MobileNetV2 and PSPNet-ResNet50), DeepLabV3+ [287] with MobileNetV2 [285] and Xception [288] as the backbone, respectively (DeepLabV3+-MobileNetV2 and DeepLabV3+-Xception), UNet with ResNet50 as the backbone (UNet-ResNet50), as well as SegFormer [289] with B5 as the backbone (SegFormer-B5).

2.6. External validation

Since the testing sets of the two splitting strategies are different, an external validation based on the data collected from another farm in 2017 [85] was performed to fairly compare the generalisability between the two splitting strategies. A total of 100 infrared images from 95 cows, taken with the same camera and annotated following the same procedure, were used for comparing only the proposed architecture.

2.7. Performance evaluation

The per-class segmentation results were shown using the Intersection over Union (IoU), Recall, and Precision, as expressed in Eqs. (4-6). The IoU is the intersection of the prediction and ground truth divided by their union. The Recall indicates the proportion of all positive labels that are classified correctly. The Precision indicates the proportion of all positive predictions that are classified correctly.

$$IoU = \frac{TP}{TP+FP+FN} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

The overall segmentation performance was evaluated by the mean Intersection over Union (mIoU) and mean pixel accuracy (mPA) where mIoU was the primary metric. The mIoU is the mean IoU of the background and five classes of interest, whereas the mPA is the average of their pixel accuracy, as expressed in Eqs. (7) and (8), respectively:

$$mIoU = \frac{1}{C+1}\sum_{i=0}^{C} IoU \times 100\% \quad (7)$$

$$mPA = \frac{1}{C+1}\sum_{i=0}^{C} \frac{p_{ii}}{\sum_{j=0}^{C} p_{ij}} \times 100\% \quad (8)$$

where C + 1 equals 6 indicating the background and five classes of interest. P_{ii} and P_{ij} are the total numbers of pixels belonging to class i that are predicted to belong to i and j, respectively.

In addition, secondary metrics included number of parameters, model size, and floating-point operations (FLOPs) to show model complexity and computational requirements, and frames per second (FPS) to indicate the inference speed (Eq. (9)).

$$FPS = \frac{N}{t_N} \quad (9)$$

where t_N is the total inference time (s) on N images.

2.8. Data analysis

2.8.1. Agreement with ground truth

Agreement analysis was done using the images from the testing set. The predicted segmentations by the improved UNet as well as the ground-truth annotations were used for generating temperature parameters (i.e., mean and maximum) of the RoI. This was done by using a self-written program in Python that maps the coordinate matrices of the segmentations and annotations with the original temperature matrices. The results of the ground-truth annotations were regarded as ground truth. Additionally, the baseline UNet model with VGG16 encoder was used as a control for temperature collection to investigate how much the proposed model has improved the results. Moreover, the temperature parameters were also generated by the traditional method with RoI manually defined on the software using appropriate shapes such as ellipses and rectangles (see **Appendix 1. Supplementary material Figure A3-2**). This was to represent the common practice in relevant studies. This method was denoted as manual collection.

Finally, the results obtained by the three methods, namely the proposed improved UNet model, the baseline UNet model, and the traditional method based on manual collection, were examined for their agreement with the ground truth using Bland-Altman plots [290].

2.8.2. Threshold development

In order to explore to what extent cattle facial temperatures (FT) can indicate the onset of heat stress, piecewise regression models were used to fit their response to Ta and THI, and locate the breakpoint at which this response changed in trend. This was done by using the 'segmented' package which determines the breakpoint based on the Davies test [291]. Note that the analysis relied on the whole dataset (n = 1000). Specifically, a simple linear regression was first modelled to fit the response of certain FT to an environmental variable (i.e., Ta or THI) using the lm function. The piecewise model was then built to update the simple linear regression model, written as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X_{bp}) X_k + \varepsilon_i, X_k = \begin{cases} 0 \; if \; X \le X_{bp} \\ 1 \; if \; X > X_{bp} \end{cases} \quad (10)$$

where Y is certain FT, β_0 is the population intercept, X is the environmental variables (i.e., Ta and THI), X_{bp} is the breakpoint, X_k is the dummy variable, β_1 is the left slope, β_2 is the difference between right slope and left slope, and ε_i is the random residual for the i-th observation.

3. Results and discussion

3.1. Overview of the datasets

During the experimental period, Ta averaged 30.1 °C (range from 22.4 to 37.6 °C), RH averaged 61.1% (range from 19.5% to 94%), and THI averaged 79.8 (range from 70.3 to 85.9). The standard deviation of daily mean Ta, RH, and THI were 2.7 °C, 16.3%, and 3.2, respectively. The THI distribution shown in **Figure 3-4** indicates a good consistency between training and testing sets as well as wide coverage of thermal environments in both image- and cow-level datasets. According to the THI threshold customised for high-producing dairy cows, heat stress occurs at a THI of 68, mild-moderate at 72, moderate-severe at 80, and severe at 90 [50]. Thus, our test cows experienced no to moderate-severe heat stress during observations.

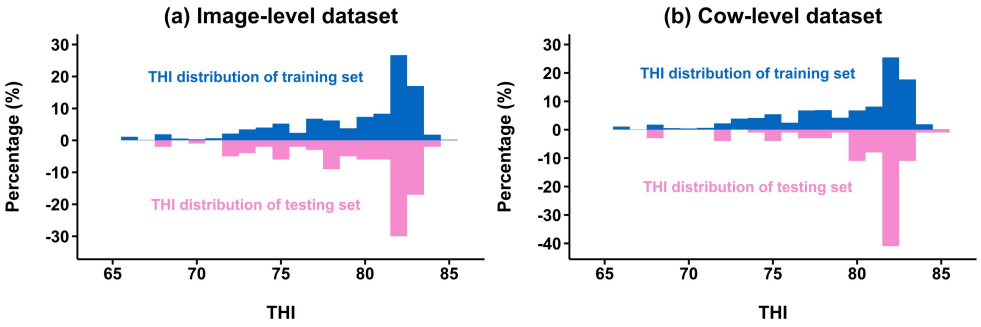


Figure 3- 4: Distribution of temperature-humidity index (THI) during photography in (a) image-and (b) cow-level datasets, summarised by training (including validation and before augmentation) and testing sets, respectively.

3.2. Results of training and ablation study

To the best of our knowledge, this is the first attempt at pixel-level facial landmark segmentation in cattle infrared images. As shown in **Figure 3-5**, there was a rapid increase in validation mIoU as loss decreased at the early stage of the training of the proposed architecture on the cow-level dataset, and the model converged at the middle stage. The lowest loss (0.137) on the validation set was obtained at epoch 153. The training curve of the proposed architecture on the image-level dataset can be found in **Appendix 1. Supplementary material Figure A3-3**.

The ablation study on the cow-level dataset showed a performance gain by introducing Ghost and GhostECA modules (**Table 3-3**). By replacing convolutions in the decoder with Ghost modules, the UNet+Ghost model had an increased mIoU and mPA by 0.91% and 0.46%, respectively, compared with the baseline UNet model. This can be explained by the enlarged receptive field as a result of the addition of larger convolutional kernels in the decoder. Plus, its number of parameters, model size, and FLOPs decreased by 30.49%, 31.25%, and 47.49%, respectively, resulting in a slight increase in FPS by 3.89%. These results are consistent with previous reports that Ghost modules can reduce computational

complexity by exploiting redundancy in intermediate feature maps calculated by mainstream CNN [126,292].

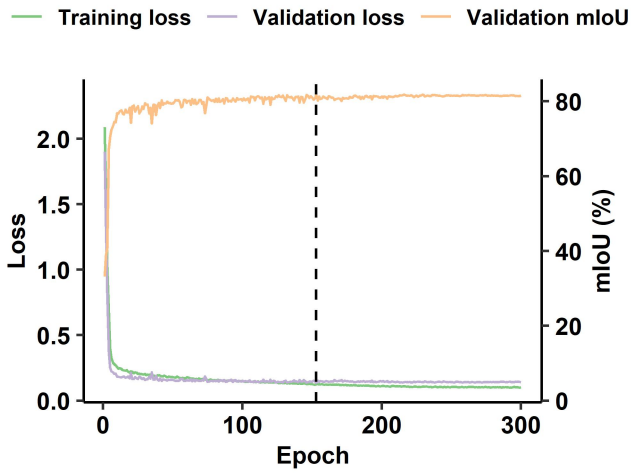


Figure 3- 5: Loss and mean Intersection over Union (mIoU) curve of the proposed architecture trained using the cow-level dataset. The dashed lines show the epoch (153) with the lowest validation loss (0.137).

Table 3- 3: Performance of ablation study trained using the cow-level dataset (n = 93).

Model	Backbone	mIoU (%)	mPA (%)	Number of parameters (M)	Model size (MB)	FLOPs (G)	FPS
UNet	VGG16	79.17	87.82	24.89	96	451.81	36
UNet+Ghost	VGG16	80.08	88.28	17.3	66	237.24	37.4
UNet+GhostECA (proposed)	VGG16	80.76	88.92	18.43	70.3	269.63	32.7

mIoU = mean Intersection over Union; mPA = mean pixel accuracy; FLOPs = floating-point operations; FPS = frames per second.

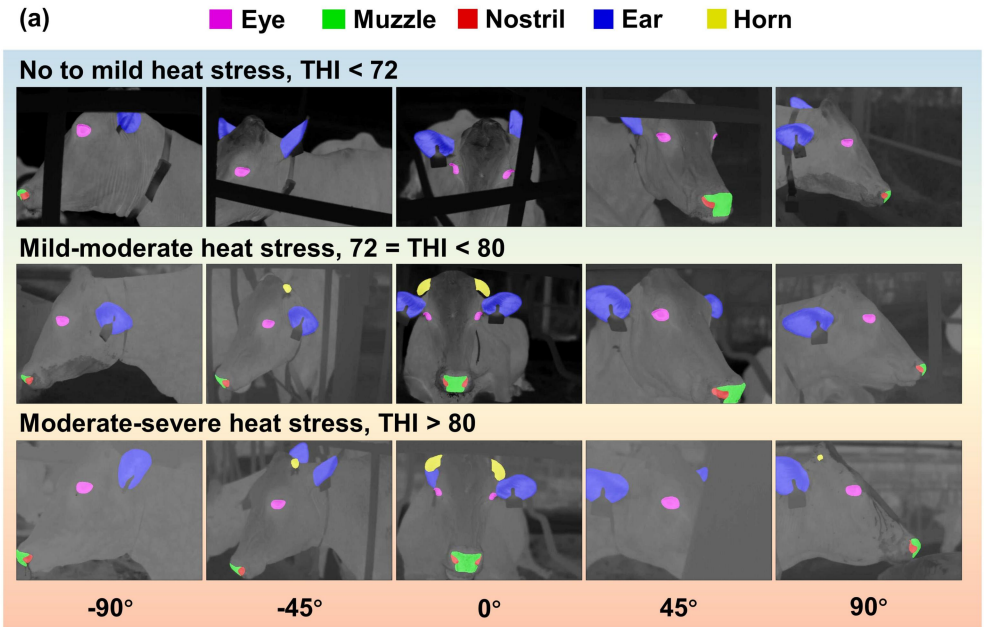
By further integrating the ECA module into Ghost modules, the UNet+GhostECA model further increased mIoU by 0.68% at the basis of the UNet+Ghost model, compensated by an increased number of parameters, model size, and FLOPs by 6.53%, 6.52%, and 13.65%, respectively. Still, the proposed UNet+GhostECA model outperformed the baseline model in all metrics except for FPS, which decreased by 9.17%. These results are as expected, as the integration of attention mechanisms increases performance by suppressing the gradient transmission of irrelevant information, but often requires more computational resources due to the more complex structure. Collectively, the UNet+GhostECA model should be

considered a successful improvement due to its leading mIoU, smaller computational requirements, and good inference speed.

The results of the ablation study, which was trained using the image-level dataset, are listed in **Appendix 1. Supplementary material Table A3-1.** As expected, these models shared the same trend but had generally better performance since image-level splitting mixed all cows' images in training and testing sets, which could have led to data leakage between both sets.

3.3. Segmentation results of the improved UNet

The detailed performance of the improved UNet model illustrated in **Figure 3-6** shows robust segmentations against usual interfering factors including camera angle and extreme ambient environment. More importantly, all RoI yielded an IoU higher than 50% which is a commonly used threshold above which a result is considered to be accurate. The IoU, Recall, and Precision shared a similar trend, with the best performance obtained by 'ear' and 'eye', while the worst by 'muzzle', 'horn', and 'nostril'. The misclassification of eyes was primarily due to partially open or closed eyes. Unfortunately, this is hard to solve due to relatively limited negative samples in the current datasets. The relatively poor segmentation in the nose areas was most likely due to the misclassification of pixels between nostrils and muzzles. Since the nose area of cattle is often covered by foreign matter such as mud, water, and saliva, especially during hot seasons [293], the detection of nostrils and muzzle is more difficult than other RoI. However, a better segmentation can be speculated by combining 'muzzle' and 'nostril' as one unified label class of nose areas.



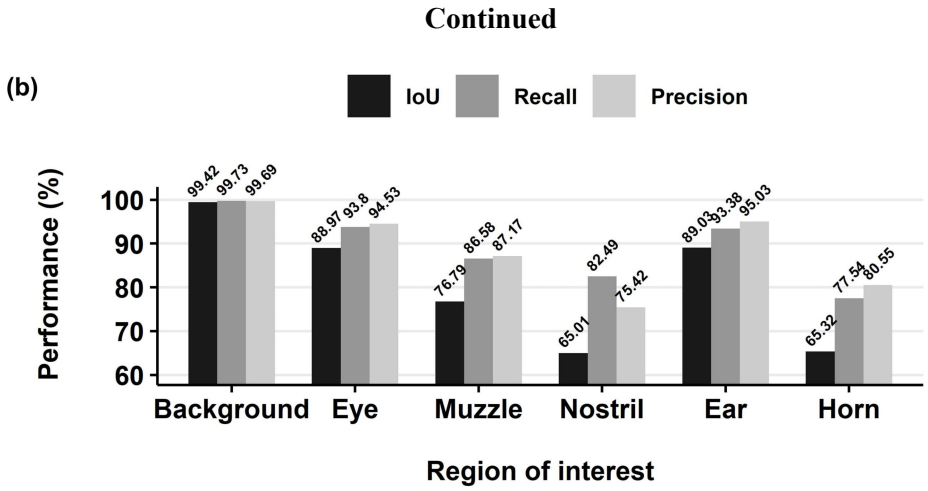


Figure 3- 6: Detailed segmentation results on the cow-level testing set (n = 93). (a) Predictions on some example images against varying thermal conditions classified by temperature-humidity index (THI), shown at camera angles ranging from -90° to 90°. (b) Per-class Intersection over Union (IoU), Recall, and Precision.

The worse results of horns can be explained by relatively fewer training instances since not all cows had horns. Also, their misclassification is partially due to incomplete horn removal. Calves were disbud using hot iron on the experimental dairy farm at around 40 days of age. If the horn bud remained subdermal, such skin surface would show a blurred region on the infrared image and could be misclassified as a horn, especially from certain side views. This finding suggests a novel method for veterinarians to confirm the effectiveness of horn removal and determine whether a second operation is required.

3.4. Comparison with other semantic segmentation models

The results of the comparison study based on the cow-level dataset demonstrate the highest mIoU (80.76%) by the proposed UNet+GhostECA model (**Table 3-4**). The trade-off between segmentation performance and inference speed is obvious. Overall, more complex and deep networks (such as UNet and DeepLabV3+) and backbones (such as Xception and ResNet50) had better performance metrics but lower inference speed compared with lightweight networks (such as PSPNet) and backbones (such as MobileNetV2). The only exception was SegFormer-B5 which had the largest model size and complexity but performed almost the worst results. It should be noted that SegFormer, as a transformer-based framework, was pre-trained on a dataset with cityscapes as classes of interest. On the contrary, UNet was primarily designed for segmenting medical images which are similar to our grey-scale images. This may explain why SegFormer performed worse than UNet on our infrared dataset. Moreover, DeeplabV3+, as a recent network, typically has better segmentation results when dealing with challenging tasks but performed poorly than

UNet on our dataset. The poor performance of more recent and complex models (such as DeeplabV3+ and SegFormer) could be attributed to our relatively few and simple images. Indeed, UNet architecture has been reported to be more appropriate for training with limited training images and fewer deep-level features [294].

Table 3- 4: Performance of comparison study trained using the cow-level dataset (n = 93).

Model	Backbone	mIoU (%)	mPA (%)	Number of parameters (M)	Model size (MB)	FLOPs (G)	FPS
FCN	VGG16	76.64	84.38	19.17	73.1	204.34	45.2
PSPNet	MobileNetV2	73.73	83.38	2.38	9.3	6.03	141.4
PSPNet	ResNet50	78.82	87.4	46.7	178	118.43	85.1
DeepLabV3+	MobileNetV2	77.84	88.76	5.81	22.4	52.9	88.1
DeepLabV3+	Xception	79.14	90.33	54.71	209	166.88	28.8
UNet	VGG16	79.17	87.82	24.89	94.9	451.81	36
UNet	ResNet50	78.85	90.94	43.93	167	184.23	46.8
SegFormer	B5	70.93	81.49	84.6	969	986.48	21.8
Proposed	VGG16	80.76	88.92	18.43	70.3	269.63	32.7

mIoU = mean Intersection over Union; mPA = mean pixel accuracy; FLOPs = floating-point operations; FPS = frames per second.

The results of the comparison study, which was trained with the image-level dataset, are listed in **Appendix 1. Supplementary material Table A3-2.** Similarly, they shared the same trend but outperformed cow-level models due to data leakage.

The segmentation examples shown in **Figure 3-7** confirm the better performance of the proposed UNet+GhostECA model. It can be seen that all models have good segmentation ability when the Ta was much lower than cattle FT and the contour of the RoI was obvious. However, when the Ta increased closely to cattle FT and the boundary between cattle and their environments became blurry, some comparable models became less effective in segmenting the pixels at the edge while the UNet+GhostECA model still had smooth and precise segmentation.

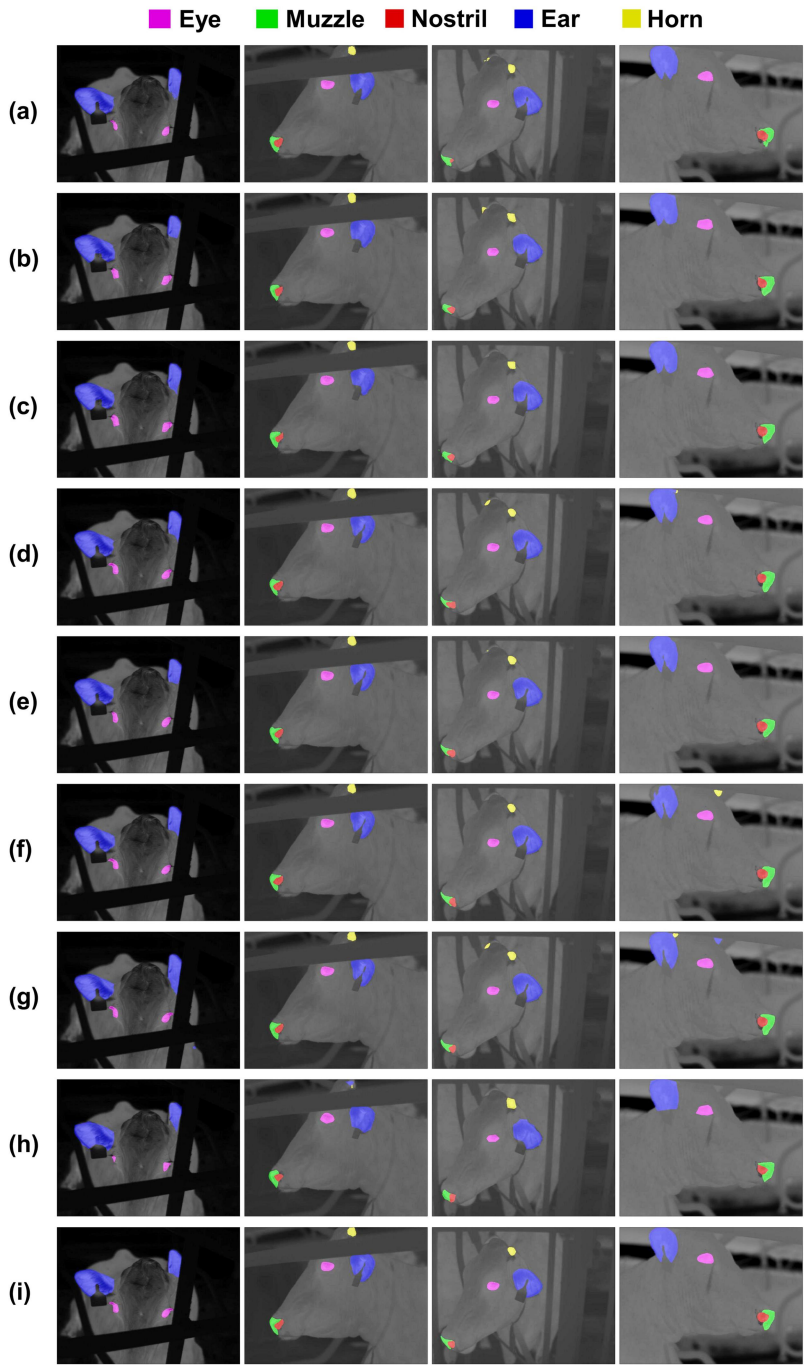


Figure 3- 7: Segmentation results of different semantic segmentation models on the cow-level testing set (n = 93). (a) FCN-VGG16; (b) PSPNet-MobileNetV2; (c) PSPNet-ResNet50; (d) DeepLabV3+-MobileNetV2; (e) DeepLabV3+-Xception; (f) UNet-VGG16; (g) UNet-ResNet50; (h) SegFormer-B5; (i) Proposed.

3.5. Comparison of the generalisability of the two splitting strategies

When comparing the performance of the improved UNet models by the two splitting strategies on their testing sets, image-level mIoU is 4.31% higher than cow-level mIoU (**Table 3-5**).

Table 3- 5: mean Intersection over Union of the improved UNet models, trained separately using the (a) image- and (b) cow-level datasets, on the testing and external testing sets, respectively.

Splitting strategy	Testing set	External testing set
Image-level	85.07 (n = 100)	73.74 (n = 100)
Cow-level	80.76 (n = 93)	73.97 (n = 100)

mIoU = mean Intersection over Union; mPA = mean pixel accuracy; FLOPs = floating-point operations; FPS = frames per second.

Detailed IoU shown in **Figure 3-8(a)** indicates comparable results in most RoI except for 'horn' where cow-level IoU is lower by 21.74%. However, the external validation shows comparable mIoU (**Table 3-5**) and detailed IoU in all RoI between the models trained with two splitting strategies **Figure 3-8(b)**. These results demonstrate that image-level splitting had an overestimated result for segmenting horns on the testing set. This is because the test cows had different horn shapes and image-level splitting contained images from all cows in its training and testing sets, while cow-level splitting had independent cows in each set. This advantage, however, disappeared when applied to unseen cows with unseen horn shapes in the external testing set.

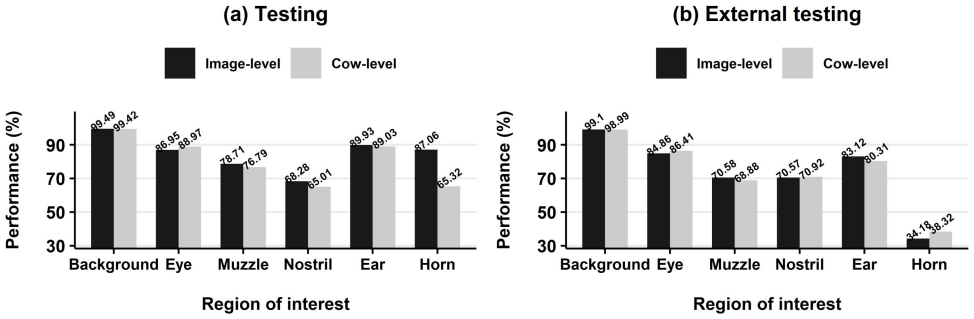


Figure 3- 8: Per-class Intersection over Union of the improved models, trained separately using the image- and cow-level datasets, on the (a) testing and (b) external testing sets, respectively.

The performance reduction when applying to new images from different cows, farms, and years is consistent with the reality that cows from different farms have

greater variation. The detailed performance shows that the IoU of 'eye', 'muzzle', and 'nostril' are comparable between testing and external testing sets, and the performance reduction of both models is mainly due to worse results of 'ear' and 'horn'. Ears were segmented worse in the external testing set because these new cows wore two ear tags (**Figure 3-9**), but old cows only one. In addition, horns were segmented much worse in the new cows by both models due to much fewer training instances. To reiterate, cows can have different shapes of horns, thus having enough number of cows during training is necessary for a good generalisability of horns.

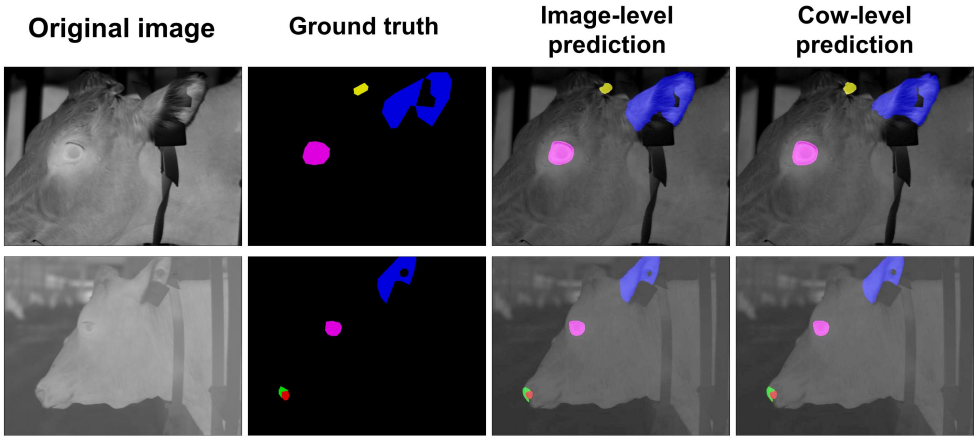


Figure 3-9: Predictions of two example images from the external testing set with cows wearing two ear tags.

Interestingly, other facial landmarks were not overestimated by image-level splitting, with comparable results between both splitting strategies in both internal and external testing. This could be explained by the fact that these facial landmarks (i.e., eyes, muzzles, and nostrils) have relatively identical shapes among cows of the same breed. Consequently, their features have been learned sufficiently, making them easier to generalise to new cows. To some extent, the similarity in features in these facial landmarks across different cows allows them to be regarded as one, thereby ensuring good generalisability regardless of which data-splitting strategy was employed.

3.6. Agreement of automated and manual methods with ground truth

The UNet+GhostECA model trained with the cow-level dataset was used for further automated temperature collection due to its outperforming segmentation performance as well as realistic and unbiased generalisation. The temperature results of the proposed automated method show that eye temperature (ET) always had the highest values (**Figure 3-10**). This is consistent with previous knowledge that ET is the closest proxy of core body temperatures among other candidate body surface temperatures [260,295].

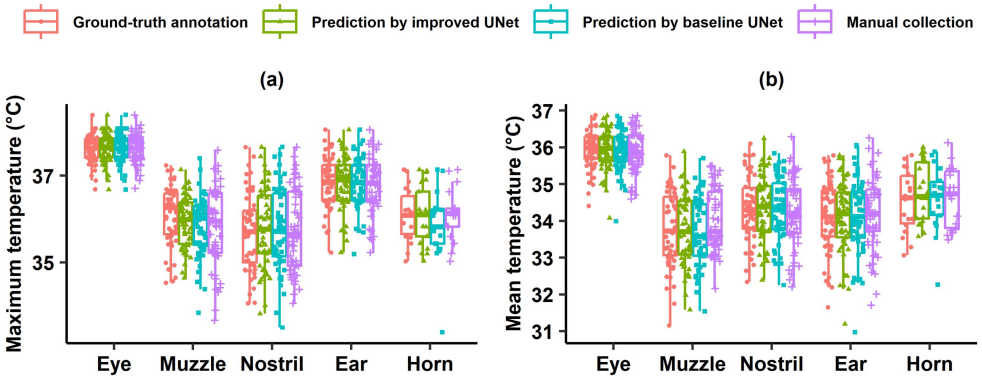


Figure 3- 10: Overview of the (a) maximum and (b) mean temperatures of ground-truth annotation, predicted segmentation by the proposed model, and manual collection on the cow-level testing set (n = 93).

As shown in **Figure 3-11**, the mean differences between the temperatures obtained automatically by the proposed model and the ground truth are small, particularly in eyes and ears. More importantly, the differences between the temperatures obtained by the proposed model and the ground truth (**Figure 3-11**) had generally narrower limits of agreement compared with those between the temperatures obtained by the baseline UNet model and the ground truth (**Figure 3-12**), indicating a general better agreement.

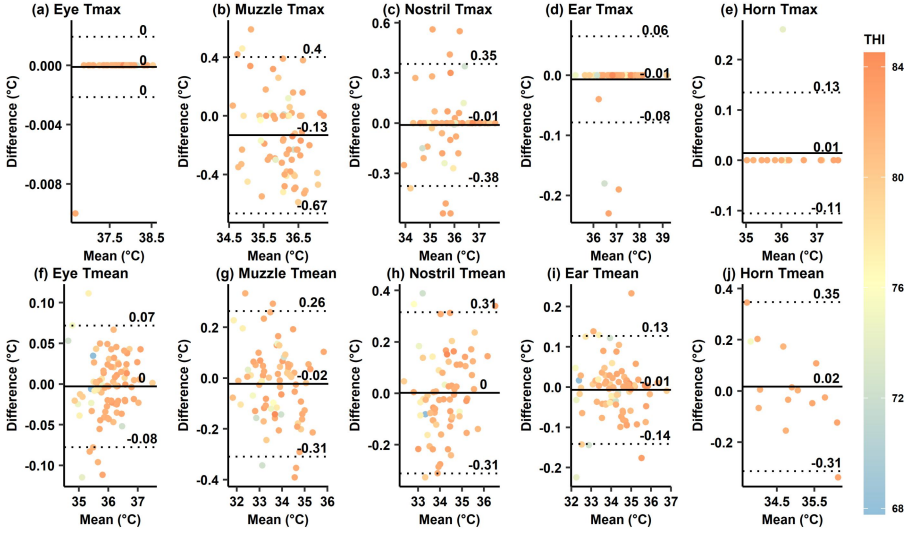


Figure 3- 11: Bland-Altman plots showing the agreement between the predicted segmentation of the improved UNet model and ground-truth annotation on the cow-level testing set (n = 93) in terms of the maximum and mean temperatures (Tmax and Tmean) of five facial landmarks. The solid and dashed lines represent mean difference and 95% limits of agreement, respectively. Datapoints are coloured by temperature-humidity index (THI).

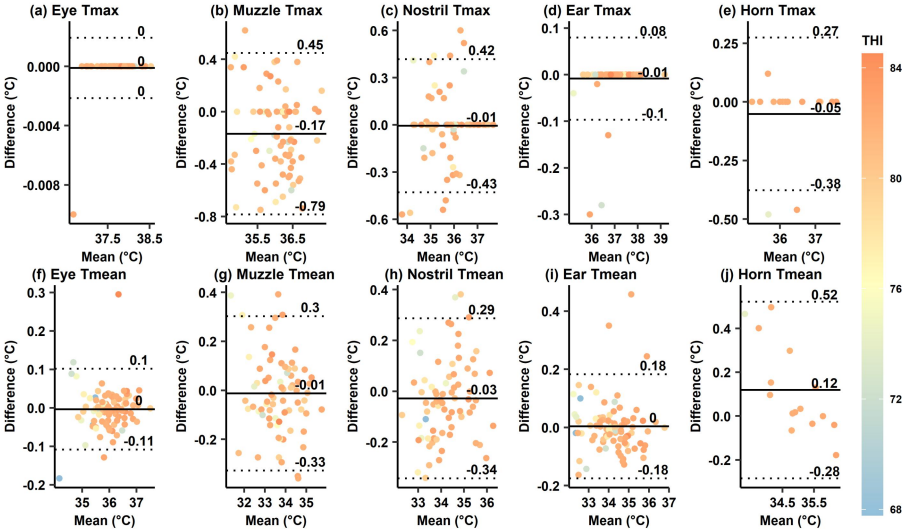


Figure 3- 12: Bland-Altman plots showing the agreement between the predicted segmentation of the baseline UNet model and ground-truth annotation on the cow-level testing set (n = 93) in terms of the maximum and mean temperatures (Tmax and Tmean) of five facial landmarks. The solid and dashed lines represent mean difference and 95% limits of agreement, respectively. Datapoints are coloured by temperature-humidity index (THI).

In addition, the differences between the temperatures obtained by the proposed model and the ground truth (**Figure 3-11**) had narrower limits of agreement in most cases compared with those between the temperatures obtained manually and the ground truth (**Figure 3-13**), indicating a general better agreement. This is reasonable since manual collection using professional software can only achieve rough coverage of the RoI rather than pixel-wise segmentation. This common practice of collecting temperatures manually may work for maximum temperatures due to being less impacted by the overall pixels, as well as for landmarks with fewer obstacles and typical outlines (e.g., eyes). However, it can suffer when collecting mean temperatures at other irregularly shaped landmarks (e.g., ears being seriously influenced by ear tags). Thus, the common practice of using manual collection in relevant studies may obtain under- or over-estimated temperatures.

Besides, the differences between the proposed model and the ground truth (**Figure 3-11**) show a homogeneous distribution around their mean differences in most cases, indicating no visible proportional error of one method versus the other. A THI-related colour code was added to confirm whether ambient environments had affected the agreement between the results and their ground truth. It is obvious that the temperature difference stayed homogeneous over the THI range we observed, demonstrating the good robustness of the proposed method against extreme thermal conditions.

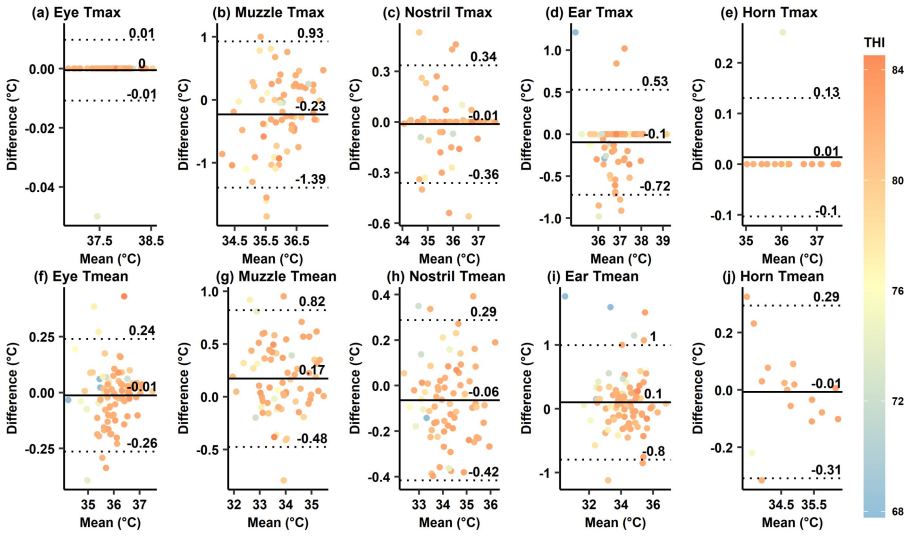


Figure 3- 13: Bland-Altman plots showing the agreement between manual collection and ground-truth annotation on the cow-level testing set (n = 93) in terms of the maximum and mean temperatures (Tmax and Tmean) of five facial landmarks. The solid and dashed lines represent mean difference and 95% limits of agreement, respectively. Datapoints are coloured by temperature-humidity index (THI).

Other studies compared automated measured and manually collected temperatures. For example, Lowe et al. [125] reported an average difference between automated measured and manually collected maximum ET of 0 ± 0.001 °C. Wang et al. [126] obtained an average difference of 0.051 °C and 0.042 °C between the automated measured and manually collected temperatures of the left and right maximum ET, respectively. We did not compare automated measurements with manual collections, and thus cannot compare our results with theirs. As discussed before, manually collected maximum eye temperature can be a good proxy of ground truth. Therefore, it is sensible to admit that the present study, as well as the abovementioned studies, all work for automated maximum ET collection. However, our proposed semantic segmentation model can provide more valuable information about mean temperatures which have been determined to more appropriately reflect core body temperatures [218].

3.7. Response of FT to heat stress

Since the temperatures measured at eyes and ears showed negligible differences with the ground truth, piecewise regression models were used to fit their response to Ta and THI. Piecewise regression models were converged in all cases (**Figure 3-14**).

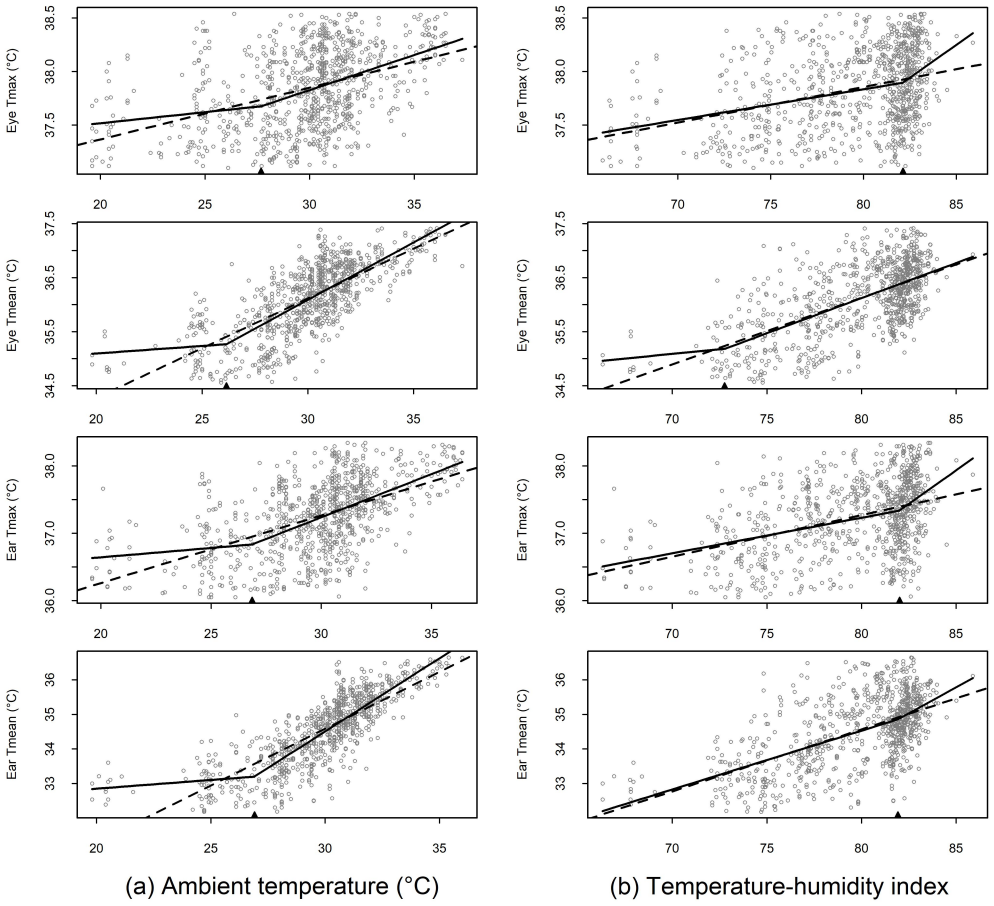


Figure 3- 14: Automated measurements of the maximum temperature (Tmax) and mean temperature (Tmean) of eyes and ears (n = 1000) and their fitted profiles from linear regression (dashed line) and piecewise regression (solid line) models with (a) ambient temperature and (b) temperature-humidity index as the predictor, respectively. The breakpoints are marked as a black triangle above the x-axis.

As detailed in **Table 3-6**, the earliest Ta threshold (26.1 °C) was obtained by mean ET. The corresponding mean ET threshold was 35.3 °C. In the case of THI, all breakpoints unsurprisingly occurred where THI datapoints were most intensively gathered. It seems that THI is less representative of the current environment than Ta due to the inability to show environmental differences. This finding was also identified by our previous pilot study in the same environment where Ta showed a higher correlation with animal-based indicators than THI [215]. Generally, mean temperatures had lower thresholds than maximum temperatures which is consistent with the study of Yan et al. [218]. However, their determined THI thresholds were 77.3 and 77.2 for maximum and mean ET, and 75.6 and 70.2 for maximum and mean ear temperature, much lower than ours. This is potentially because they turned

off sprinklers during the experimental period making their subject cows more easily stressed. The presence of cooling measures in the current study appeared to effectively cool the animals and delayed their onset of heat stress.

Table 3- 6: Parameter estimates (mean ± standard error) of the piecewise regression models predicting eye and ear temperatures with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively (n = 1000).

Predictor	Outcome	Intercept	Breakpoint	Left slope	ΔSlope	AIC	
						Linear	Piecewise
Ta	Eye Tmax	37.1 ± 0.3	27.7 ± 1.1	0.02 ± 0.01	0.04 ± 0.01	566.8	555.7
	Eye Tmean	34.5 ± 0.8	26.1 ± 0.4	0.03 ± 0.03	0.18 ± 0.03	1044.9	1003.8
	Ear Tmax	36.1 ± 0.5	26.9 ± 0.7	0.03 ± 0.02	0.10 ± 0.02	1109.7	1088.9
	Ear Tmean	31.8 ± 0.8	26.9 ± 0.3	0.05 ± 0.03	0.37 ± 0.03	1543.4	1398.0
THI	Eye Tmax	35.5 ± 0.3	82.1 ± 0.4	0.03 ± 0.004	0.09 ± 0.04	623.6	619.3
	Eye Tmean	32.6 ± 3.0	72.8 ± 1.4	0.03 ± 0.04	0.10 ± 0.04	1298.8	1295.6
	Ear Tmax	33.0 ± 0.5	82.0 ± 0.4	0.05 ± 0.006	0.15 ± 0.05	1248.0	1241.2
	Ear Tmean	20.9 ± 0.8	81.9 ± 0.7	0.17 ± 0.01	0.13 ± 0.09	2073.5	2072.9

ΔSlope = difference between right and left slopes; AIC = Akaike information criterion; Tmax = maximum temperature; Tmean = mean temperature.

3.8. Limitations and perspectives

Infrared images taken under a thermoneutral environment were limited since the data was originally collected for a heat stress study. Thus, our method should be prioritised for studies in heat stress evaluation. For example, automated heat stress recognition can be achieved by inputting our segmentation results into a deep learning-based thermal level classification [296]. For applications under a thermally comfortable situation, we speculate that our model would remain robust since the larger temperature difference between the animals and their environments in such a situation should increase the separability of RoI in the infrared images. However, further studies should be conducted in which more data from thermoneutral environments are collected to validate and improve the generalisability of the proposed network.

The facial IRT measurements were used as a proxy of FT. However, the measured facial infrared radiation is a function of actual FT and certain measurement setups, e.g., camera performance, distance to camera, angle of view, as well as ambient wind speed (WS), temperature, and humidity. Although Ta was calibrated, other undealt factors might have affected the representativeness of IRT measurements to

actual FT. The images used for modelling were taken at a fixed distance (i.e., 1 to 1.5 m) from the cows to produce a consistent size of cattle faces. This distance is aligned with the recommendation of previous studies which validated IRT as a means to measure bovine temperatures [85,276]. However, it may be difficult to reach such a close distance to the cows in practice, except for specific locations like feeding stations [125] and the entry to the milking parlour [266]. Direct application of the proposed network to a real-world situation with a greater distance between the cows and the camera should result in a lower segmentation accuracy as well as lower temperature results since facial landmarks in the image taken from a greater distance will be represented by a lower number of pixels [297].

Plus, the angle of view between the camera and the object is known to influence the infrared emissivity and thus affect the IRT temperature [298]. Although we have successfully segmented specific facial landmarks in infrared images taken from different camera angles to the cows, the variation in the angle of view would definitely lower the measured temperature in general [299]. This effect should be homogeneous among cases thus the determined environmental thresholds should be robust with the FT thresholds being underestimated. A slightly lower threshold is good for increasing Recall (i.e., the ability to detect a heat-stressed cow). Temperature correction is not the focus of this study but should be investigated in future studies in order to obtain reliable temperature readings in challenging practical cases. A recent study is a promising step forward, in which a response surface method was developed for correcting the effect of distance and the angle of view on the IRT temperature of pigs [299].

In addition, many other parameters would affect the results of IRT, such as sunlight exposure, WS, coat colour, hair thickness, dirt, and the resolution and accuracy of the camera [300]. For example, direct sunlight can increase ET by 0.56 ± 0.36 °C [297]. This would lead to a false positive alert of heat stress, but should not be a problem in an indoor scenario with adequate shading from direct sunlight. WS over 3.3 m/s can decrease ET by 0.78 ± 0.33 °C [297]. The average WS was 1.3 ± 0.9 m/s during this study. Thus, the developed FT thresholds should be used carefully in environments with different levels of WS. Wang et al. [301] proposed a signal processing flow to reduce the effect of WS and other ambient factors.

The emissivity does not need to be changed based on an individual's skin tone according to a human study [302]. This is because the infrared spectrum emissivity depends on surface structure and composition, unlike the visible spectrum which depends on colour (wavelength) [297]. Still, darker-coated cows should have slightly higher BST and CBT because they naturally absorb more heat [303]. Interestingly, no differences were found between black and white cows in BST measurements across multiple RoI in the study of Anzures-Olvera et al. [304]. A pig study found hairy skin areas or skin with no blood perfusion had slightly lower emissivity by 0.018 ± 0.010 and 0.012 ± 0.006, respectively [305]. Collectively, the proposed tool and thresholds may not be directly applicable to different parameter settings. These gaps highlight the need for further studies based on data collected under more complex and practical conditions.

4. Conclusions

Collectively, our work provides relevant studies with an automated tool for collecting FT from cattle infrared images and determining their heat stress states. This method is robust against usual interfering factors including camera angle and extreme ambient environment. However, additional training with data supplemented on a variety of influencing factors (e.g., the distance between the camera and the cows, coat colour) and temperature correction against these factors are required before it can be integrated into on-farm automated monitoring of animal health and welfare.

Context – Chapter 4

According to the findings reported in the first chapter, respiration rate (RR) is a promising indicator of heat stress in dairy cows due to its ease of acquisition and sensitivity to heat stressors. Traditional visual observation requires massive labour and is impractical for large-scale commercial farms. With the help of advanced computer vision and deep learning algorithms, RR can be detected from RGB video streams. However, recently developed vision-based RR measurement methods are highly limited to measuring a small number of animals in a controlled environment. Therefore, the fourth chapter aimed to propose a vision-based multi-object RR measurement method for dairy cows lying on free stalls and to determine critical thresholds for the onset of heat stress based on automated measurements.

Chapter 4

Non-contact respiration rate measurement of multiple cows using computer vision

Adapted from:

Shu, H., Bindelle, J., Gu, X. *Non-contact respiration rate measurement of multiple cows in a free-stall barn using computer vision methods*. Computers and Electronics in Agriculture, 2024. https://doi.org/10.1016/j.compag.2024.108678

Abstract

In cattle, respiration rate (RR) provides researchers and practitioners with valuable physiological information. However, traditional visual observation requires massive labour and is impractical for large-scale commercial farms. Recently developed vision-based RR measurement methods are highly limited to measuring a small number of animals in a controlled environment. Therefore, this paper aimed to propose a vision-based multi-object RR measurement method for dairy cows lying in free stalls. An RGB camera was aimed at the lying zone and was able to cover about 16 stalls. The proposed framework first utilised two YOLOv5-based networks to segment cow instances and detect cow flank objects. Next, an object tracker was used to link the predictions of each cow throughout the video clip. The Lucas-Kanade optical flow was then calculated specifically on the overlapped area of the cow mask and the flank bounding box. Finally, RR was extracted using Fast Fourier Transform. The results show that the proposed method had a precise RR measurement with a correlation coefficient of 0.945, a root mean square error of 5.24 breaths per minute, and an intraclass correlation coefficient of 0.98 when compared to visual observation. The piecewise regression models identified a change in RR when the ambient temperature reached 23.6 °C or the temperature-humidity index reached 72. The corresponding RR thresholds were 61.1 and 60.4 breaths per minute, respectively. Collectively, these results can be used to inform an automated local cooling system, e.g., fans in the lying area. However, more experiments and calibration with data collected using more cost-effective video recording systems are required before this technology can be applied on farms.

Keywords: precision livestock farming, animal welfare, deep learning, multi-object measurement, heat stress

1. Introduction

Respiration rate (RR) is a basic vital sign of animals that indicates their physiological state and overall health condition. For example, an abnormal RR in cattle can be used as an alert for the onset of common events of interest, such as respiratory disease [306], transport stress [307], and heat stress [131], among others. However, traditional RR measurement by manually counting flank movements over a set period of time and converting them to breaths per minute (bpm) can be tedious and time-consuming in frontline practice and is also prone to biased results due to the lack of professionals [308].

In order to solve the abovementioned problems, many precision livestock farming techniques have recently been put forward to achieve an automated RR measurement. These techniques rely on different methodologies to collect respiration-related signals, which are then processed to determine RR using proper signal processing methods or more advanced deep learning methods. Depending on whether physical contact with animals is required, these methods can be divided into two categories: contact and non-contact methods.

On the one hand, contact methods typically involve placing devices on the cows' bodies to collect respiration-related signals. For example, an MP3 recorder on a halter can collect the acoustic signal of breathing [143], while pressure sensors mounted to a belt [134] or a harness [135] can measure RR relying on chest and abdomen movements. Pressure sensors [142] or thermistors [133] can also be mounted to halters making use of the pressure and temperature difference between inhaled and exhaled air, respectively. More recently, accelerometers contained in a collar [130,136] or an ear tag [149] have been validated to be able to detect small movements during heavy breathing.

On the other hand, non-contact methods for measuring RR are drawing increased interest since they better meet animal welfare requirements by eliminating the stress and discomfort of cows. In cattle studies, various techniques and devices have been applied to measure RR, such as radar [139], laser [137], infrared cameras [265], and RGB cameras [177]. Radar and laser methods rely on respiration-related movements, while the infrared method relies on the changes in skin or air temperature around nose areas caused by breathing, all of which require expensive devices and thus may be unaffordable for most farms.

Also, respiration-related body movements can be detected more cost-effectively through an RGB camera and processed further by cattle segmentation, optical flow calculation, and statistical analysis to generate RR estimation [141,153]. However, the reported method was only designed for a controlled scenario in which a single or two targeted cows were held in a fixed sideway pose in the field of view. Moreover, respiration-related body movements also result in periodic fluctuations in the brightness of colour channels [309]. Consequently, RR can be directly estimated from the variations in pixel intensity of the RGB channels of certain regions of interest (e.g., chest and abdomen). This method has shown its effectiveness in a multi-animal environment for measuring up to five pigs [310] and four cows [311],

but has yet to be tested in cows raised in a larger-scale free-stall pen. In addition, the colour or brightness changes in the skin due to changes in blood volume during respiration cycles can be measured in proper colour channels using the remote photoplethysmography (rPPG) principle. This method, however, requires a closer distance between the camera and the animal in order to capture the subtle respiration-related blood flow changes in veins. rPPG-based method has been used to measure RR for hairless animals like pigs [312]. However, in the case of dairy cows, which typically have longer hair, the requirement of a direct view of body areas with less hair (e.g., facial area) by rPPG [145], renders the large-scale measurement in free stalls impossible.

In summary, the existing RR measurement methods for dairy cows require improvement as they are either too expensive, require physical contact with the animal, or are limited to measuring very few animals in a controlled scenario. For a larger-scale application in a practical farm, it is necessary to further integrate certain automated technologies such as cow and regions of interest recognition, multi-object tracking, and RR calculation. Thus, this paper aimed to (1) propose a vision-based multi-object RR measurement pipeline for dairy cows housed in an uncontrolled scenario, i.e., a commercial free-stall barn, and (2) further investigate its feasibility in indicating heat stress. By leveraging deep learning and traditional computer vision techniques, the proposed method was expected to provide a more cost-effective and non-invasive solution for farmers and veterinarians to detect significant events, such as heat stress, in an early stage.

2. Materials and methods

All protocols involving animals were approved by the Experimental Animal Care and Use Committee of Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (approval number 2016IAS018).

2.1. Experimental design

The field experiment was carried out on a free-stall dairy farm in Beijing, China for 35 days from late August to early October, 2017. The location is characterised by a temperate continental monsoon climate with hot and humid summers.

The experimental barn (oriented along the north–south axis) had four pens with a typical 4-row head-to-head design and was covered by a roof. All four pens were used in this experiment, housing 79, 70, 80, and 60 lactating Holstein-Friesian dairy cows, respectively. For each pen, ambient temperature (Ta, ℃) and relative humidity (RH, %) were measured at an interval of 5 min using three Kestrel 5000 environment meters (accuracy: ± 0.4 °C Ta, ± 1% RH; Nielsen-Kellerman, Boothwyn, PA, USA). The sensors were fixed at a height of 2 m and were evenly placed at one-fourth, one-half, and three-fourths of the total length of the pen, respectively. The temperature-humidity index (THI) was calculated as per Eq. (1) [222].

$$THI = (1.8 \times Ta + 32) - (0.55 - 0.0055 \times RH) \times (1.8 \times Ta - 26) \quad (1)$$

A Sony HDR-CX405 camera (Sony, Tokyo, Japan) was fixed on a telescopic camera stick with an angle of approximately 45° downward. The frame rate of the recordings was set to 30 frames per second (FPS). In order to cover as many cows as possible, the camera was placed at a height of 4 m along the feeding alley. The camera was aimed at the lying zone and was able to cover approximately 16 stalls. The reason for focusing on lying cows is that the RR of standing cows is rather difficult to detect due to their extremely subtle breathing movements and frequent accompanying interfering movements. Lying cows, on the contrary, reflect heat stress response to the ambient environment earlier [212], and thus provide a good representative of the herd. The video clips of 10-25 s were captured seven times daily, from 07:30 to 17:30 h, for the four pens sequentially.

2.2. Data preparation

A video frame was randomly extracted from each video clip by using a self-written Python program. This was done to avoid duplicating similar frames since cows had relatively small and slow displacements in the video clips. The extracted frames were in JPG format with a resolution of 1920 by 1080 pixels. Low-quality frames, such as those blurred or covered by flies, were discarded. Finally, a total of 750 images were chosen.

All images were annotated using the popular annotation tool Labelme (https://github.com/wkentaro/labelme.git). On the one hand, the pixel-level cow masks were annotated with polygons, after which they were labelled either 'Standing' or 'Lying' according to their body postures. Cow instances were discarded when their posture was unable to recognise, e.g., being at the edges of the image or being too obscured (more than 75% pixels) by other cows or facilities. On the other hand, the flanks of lying cows were annotated as a new flank object detection dataset by using rectangles. When lying down, cows position their hind legs to one side, either stretched or not [313]. Thus, the flank on the side where the hind legs were placed was labelled as 'Flank_up' and the flank on the other side as 'Flank_down'. Flanks with strong occlusion (more than 75% pixels) were discarded. Some examples of annotation in both datasets are visualised in **Figure 4-1**.

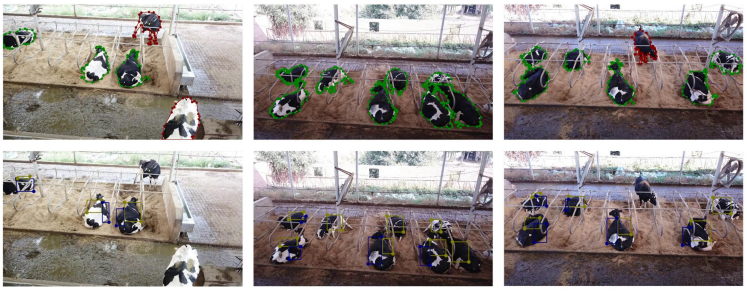


Figure 4- 1: Example annotations for cow instances (top) and flank objects (bottom). Polygons in red and green represent standing and lying cow instances, respectively. Rectangles in yellow and blue represent the flank objects on the hind leg side and the opposite side, respectively.

Two splitting strategies were used to prepare the training, validation, and testing sets for both datasets: (1) image-level splitting, where different images of the same pen were pooled together and then randomly split at a ratio of 6:2:2, and (2) pen-level splitting, where four pens were allocated at a ratio of 2:1:1 to make sure that images from the same pen were always in the same set, thus making the three sets completely independent in cows. These two splitting strategies were compared since evidence has shown that a simple random splitting can lead to an overestimation of the model's performance [277]. Training and validation sets were used to train the networks while the testing sets were used to evaluate the final performance of the trained models. A detailed description of the datasets with regard to the number of instances or objects per class is given in **Table 4-1** and **Table 4-2**.

Table 4-1: Overview of the cow segmentation datasets split using two strategies.

Dataset	Number of images	Number of pens	Number of cow instances		
			Standing	Lying	Total
Image-level					
Training	450	4	1187	2476	3663
Validation	150	4	411	808	1219
Testing	150	4	407	788	1195
Pen-level					
Training	451	2	1252	2144	3396
Validation	148	1	387	964	1351
Testing	151	1	366	964	1330
Total	750	4	2005	4072	6077

Table 4-2: Overview of the flank detection datasets split using two strategies.

Dataset	Number of images	Number of pens	Number of flank objects		
			Flank-up	Flank-down	Total
Image-level					
Training	432	4	1989	1417	3406
Validation	145	4	693	501	1194
Testing	145	4	734	508	1242
Pen-level					
Training	429	2	1885	1285	3170
Validation	144	1	657	474	1131
Testing	149	1	874	667	1541
Total	722	4	3416	2426	5842

2.3. Muti-target RR measurement for lying cows based on deep learning and optical flow method

The overall working flow of the proposed computer vision-based RR measurement for multiple lying cows is shown in **Figure 4-2**. All operations to be presented in

this section were done in Python. Two programs were written, and ran in parallel, with the first one running two trained YOLO (You Only Look Once) models and the second one tracking predictions and calculating RR. Multithreading was used to speed up. The YOLO models were set to make inferences once every second to fulfil real-time measurements. All source code is available at https://github.com/Kaiwen-Robotics/CattleRR.git.

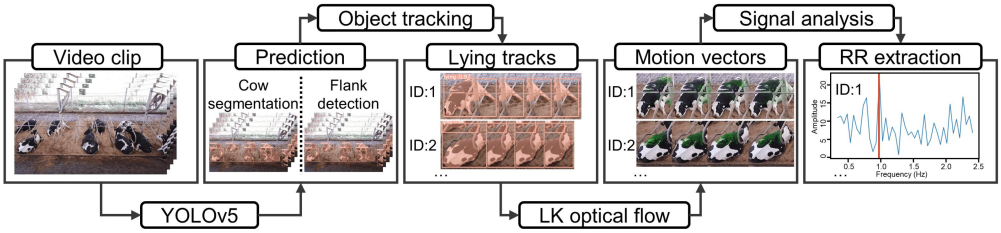


Figure 4- 2: Pipeline for muti-target respiration rate (RR) measurement for lying cows based on deep learning and optical flow method. Lucas-Kanade = LK.

2.3.1. Architecture, training, and evaluation of deep learning networks

The YOLO series of algorithms is a popular framework for object detection that is well known for excellent performance and inference speed. YOLOv5, as a recent release, has gained significant attention in the computer vision community. Notably, its latest 7.0 version has been supported with instance segmentation and has been reported as a new state-of-the-art benchmark [314]. Instance segmentation, as an extension of object detection, focuses not only on localising an object but also on generating a pixel-level mask for each object detected. As emphasised by the authors, these off-the-shelf models are roughly easy to train and deploy. Thus, YOLOv5 was used for object detection and instance segmentation tasks in this study. For clarity, instance segmentation networks were noted as YOLOv5-seg.

The training was performed in Python 3.7 language with Pytorch 1.7.1 on a computer with a 64-bit Windows 11 system and an NVIDIA GeForce RTX 3090 GPU. Transfer learning was used to accelerate the training process. The initialised weights used in this study were transferred from the network pre-trained on the Microsoft COCO benchmark [315]. During the training process of the instance segmentation networks, a batch size of 4 was used, which was the maximum allowed by the hardware. In addition, a batch size of 16 was used for the object detection networks. In both cases, three sizes of the architecture were trained (i.e., YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5s-seg, YOLOv5m-seg, and YOLOv5l-seg) and the epoch was always set to 100. Other hyperparameters were set as default.

The mean average precision (mAP), as the most commonly used performance metric for object detection and instance segmentation tasks, was used in this study. It builds on top of some basic concepts, including the Intersection of Union (IoU), precision (Pr), and recall (R). The IoU measures the intersection area between two bounding boxes or, two masks in the case of instance segmentation, as expressed in Eq. (2):

$$IoU = \frac{Prediction \cap Ground\ truth}{Prediction \cup Ground\ truth} \quad (2)$$

As a convention, an IoU threshold of 0.5 was used to qualify the prediction. For example, if the IoU between the ground-truth bounding box (or mask) and the predicted bounding box (or mask) was greater than 0.5, the predicted bounding box (or mask) was determined to be true positive (TP), otherwise it was false positive (FP). Additionally, a ground-truth bounding box (or mask) presented but not predicted was determined to be false negative (FN). Following that, Pr and R were determined, which represent how many positive predictions are actual positives and how many actual positives are correctly predicted as positive, respectively. Their equations are shown in Eqs. (3 and 4):

$$Pr = \frac{TP}{TP+FP} \quad (3)$$

$$R = \frac{TP}{TP+FN} \quad (4)$$

Subsequently, average precision (AP) was used to summarise Pr and R by computing the area under the Pr-R curve, and mean average precision (mAP) was used to summarise the AP over all classes. Their equations are shown in Eqs. (5 and 6):

$$AP = \int_0^1 Pr_{(R)}\, dR \quad (5)$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \quad (6)$$

where C always takes 2, indicating 2 classes in both cases, i.e., 'Lying' and 'Standing' in the case of cow segmentation and 'Flank_up' and 'Flank_down' in the case of flank detection. The higher the mAP, the better the object detection or instance segmentation network. In addition, FPS was used to indicate the inference speed, as expressed in Eq. (7):

$$FPS = \frac{N}{t_N} \quad (7)$$

where t_N is the total inference time (s) on N images.

2.3.2. Optical flow

In this study, the optical flow method was used to track YOLO predictions, deal with camera shake, and extract respiration-related signals. Optical flow refers to the motion pattern of pixels in an image sequence resulting from either object or camera movement. It is represented as a two-dimensional vector field, with each vector representing the displacement of pixels from one frame to the next. Optical flow can be calculated using either sparse or dense methods. Sparse optical flow focuses on tracking a chosen set of pixels across frames, whereas dense optical flow aims to compute motion vectors for all pixels in the frame. The sparse optical flow was adopted in this study in consideration of efficiency.

Optical flow relies on several assumptions: (i) the pixel intensities of an object do not change between consecutive frames; (ii) the movement of the pixels over time is small.

Assumption (i) can be written as:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (8)$$

where I(x,y,t) represents the pixel intensity of a pixel with a coordinate of (x, y) at the frame t and it moves by (dx, dy) in the next frame after dt time.

According to assumption (ii), the image constraint at I(x,y,t) to the right side of Eq. (8) can be expanded by the Taylor series as:

$$I(x+dx, y+dy, t+dt) = I(x,y,t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt \quad (9)$$

that is to say, it follows:

$$\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt = 0 \quad (10)$$

divided by *dt* and results in:

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0 \quad (11)$$

where V_x and V_y represent the x and y components of the optical flow of I(x,y,t), respectively, and $\partial I/\partial x$, $\partial I/\partial y$, and $\partial I/\partial t$ are the derivatives in the corresponding directions. Let $I_x=\partial I/\partial x$, $I_y=\partial I/\partial y$, and $I_t=\partial I/\partial t$, Eq. (11) can be further written as:

$$I_x V_x + I_Y V_y + I_t = 0 \quad (12)$$

To solve the two unknowns (i.e., V_x and V_y), the Lucas-Kanade (LK) sparse optical flow introduces another assumption that neighbouring pixels have similar motions. A search window is used to find neighbouring pixels around the target point, from which equations are then created, as shown in Eq. (13). Finally, V_x and V_y are solved by using the least squares principle. The LK optical flow was calculated by using the OpenCV function calcOpticalFlowPyrLK().

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t1} \\ -I_{t2} \\ \vdots \\ -I_{tn} \end{bmatrix} \quad (13)$$

2.3.3. Object tracking and stabilisation

Object tracking is necessary to link the predictions of each cow throughout the video clip. Tracking was done in a series of steps. First, the predictions of cow masks and their bounding boxes as well as those of the flank bounding boxes were obtained in each frame. To ensure reliable tracking, a minimum bounding box size threshold of 100 pixels was set to discard small objects. Next, the LK Optical Flow was used to track these detected objects across consecutive frames, mainly focusing on specific feature points of the objects identified by using the Shi-Tomasi corner detection method. At the same time, the flank mask was extracted by cropping the flank bounding box with the cow mask if the inclusion rate of the flank bounding box in the cow bounding box reached more than 90%. The inclusion rate was calculated with Eq. (14):

$$Inclusion\ rate\ (a\ in\ b) = \frac{a \cap b}{a} \quad (14)$$

where a and b represent flank and cow bounding boxes, respectively. Tracks that were shorter than 10 frames were regarded as false positives and eliminated. Tracks with a missing flank mask for at least 10 frames were considered occluded and thus

dropped. Tracks were split when labels were switched for at least 10 frames. Only tracks of 'Lying' labels were entered into the following analysis.

Due to camera motion, there is some unwanted video shake existing in the current clips. Thus, an optical flow-based stabilisation was performed along with object tracking. First, the frames were initialised by converting to greyscale images, which simplified subsequent processing. Next, the Shi-Tomasi corner detection method was used to select feature points on the background area which was extracted by removing the cow mask from the bounding box, and the LK optical flows were calculated for the selected feature points. A transformation matrix was then computed based on the mean optical flows. Finally, the object was subjected to an affine transformation based on the translation matrix to obtain a stabilised alignment map as the output. In summary, the tracker tracked the motion of an object in a sequence of frames while the stabiliser aligned the object.

2.3.4. Optical flow calculation at abdominal areas

Previous studies have demonstrated that removing noise from the background and uninterested body areas increases the efficiency and accuracy of subsequent RR calculations [316]. Wang et al. [310] used oriented bounding boxes to localise animals in order to include animal pixels more accurately while excluding irrelevant background pixels. Our study adopted a different strategy to completely eliminate background noise by taking advantage of the instance segmentation technique. Specifically, the abdomen areas were obtained by intersecting the flank detections (from a YOLOv5 model) with the cow segmentations (from a YOLOv5-seg model) (**Figure 4-3**).

Figure 4-3: Three regions of interest to compute sparse optical flows: (a) cow bounding box, (b) cow mask, and (c) cow mask overlapped by flank bounding boxes. The schematic of the regions of interest and the computed optical flows are shown in the top and bottom rows, respectively.

Preliminary investigations show that the optical flow calculated by using the LK sparse optical flow method specifically at the abdomen areas could better reflect respiration-related pixel movements (**Figure 4-3**). LK sparse optical flow was used because it has shown better efficiency and accuracy in extracting respiration-related motions compared with dense methods [140]. Indeed, sparse optical flow is more computationally efficient, less susceptible to noise and occlusions, and requires less memory than dense optical flow.

Similarly to the processing of the stabiliser, the video frames were initialised by converting them to greyscale images. The difference is the location for calculating optical flows. Specifically, the gradient map of the greyscale images was first obtained and the background of the regions of interest (i.e., the intersections of cow masks and flanks) were excluded from the gradient map by setting them to zero. Next, grid points were obtained at a specified interval. LK sparse optical flow analysis was then performed between consecutive greyscale images using only grid points with a non-zero gradient.

2.3.5. Signal processing for extracting RR

The sinusoidal values of the resulting optical flow vectors were derived from the Vx and Vy. These sinusoidal values were then appended to a list that contained the sinusoidal values of all grid points across multiple frames. Subsequently, a Fast Fourier Transform was performed on each grid point's 120-frame sinusoidal values. A band-pass filtering of 0.4 to 2.4Hz (i.e., 24 to 144 bpm) was applied to focus specifically on the respiration-related frequency range and remove unwanted noisy signals (e.g., occasional body wiggle and high-frequency abdominal twitching).

The frequency spectrums obtained from all grid points were then summed together. The instantaneous RR was determined by using peak detection on the summed spectrum, i.e., identifying the frequency with the maximum magnitude. This frequency was constantly updated and appended to a list of RR while processing the clip. The final RR across all frames was calculated by averaging the list of RR values at the end of the processing.

2.4. Performance analysis

A total of 55 video clips were randomly selected to measure RR by visual observation (**Figure 4-4**). These clips had a mean ± standard deviation duration of 16 ± 4 s and contained a total of 241 cow instances. First, lying cows were selected, that is 220 instances. Then, two trained observers with an intraclass correlation coefficient of 0.91 measured RR by manually counting the flank movement of the tracks over the clips and converting it to bpm. The results from the two observers were averaged. Cases with frequent and large unwanted movements, as well as those with visual occlusion of the abdomen, were discarded due to the inability to obtain ground-truth RR. As a result, ground-truth RR was obtained for a total of 180 tracks of lying cows. The mean ± standard deviation number of available tracks per video clip was found to be 3.3 ± 1.7, with a range between 1 and 8.

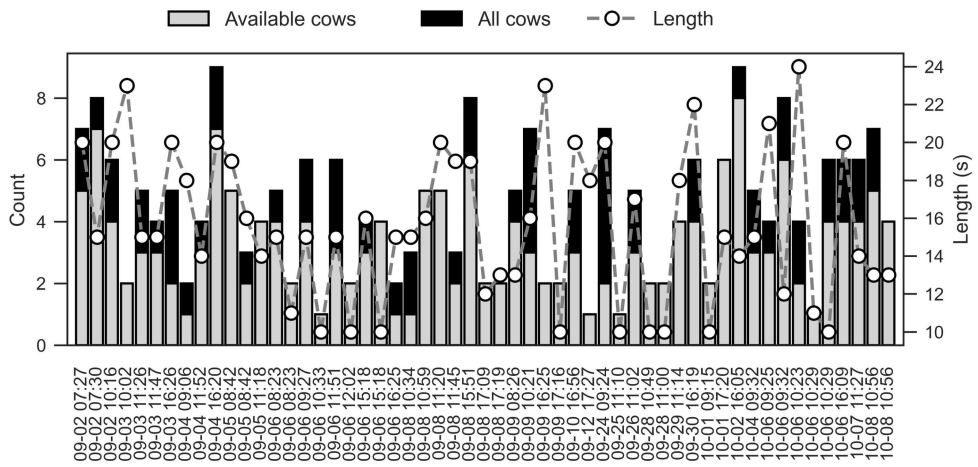


Figure 4- 4: Description of the dataset for the development of respiration rate measurement. The 55 video clips are sorted by datetime (mm-dd hh:mm).

The statistical relationship between automated and manual RR measurements was evaluated through linear regression, with the Pearson correlation coefficient calculated using the cor function (R version 3.4.4; https://R-project.org). The agreement between automated and manual RR measurements was explored using a Bland-Altman plot [290] and further quantified using the intraclass correlation coefficient calculated with the icc function from the 'irr' package. Additionally, the overall inference speed was evaluated by using the average processing time and FPS of RR calculation on the 55 clips. Plus, an investigation was conducted to explore how the number of objects to be processed affected processing time and FPS.

The potential effects of flank occlusion and flank colour on the calculated RR were examined using one-way analyses of variance with the aov function. To do so, all cases were classified by flank occlusion into two classes: flank partially occluded by cubicles or pillars (at least one pixel of such obstacles existed), and no occlusion; and by flank colour into three classes: no texture but with clear sand or ribs, no texture and no clear reference, with textures. Furthermore, the potential effect of location on the RR accuracy was examined by plotting the coordinates of all cow bounding boxes with their measurement error.

In addition, a colour-based method was applied to extract respiration-related signals, as a comparison to the state of the art. The detailed methodology was adapted from Mantovani et al. [311] which focused on non-contact RR measurement of group-housed lying cows. More specifically, the same object tracker and video stabiliser were used. The extraction of respiration-related signals was based on the averaged pixel intensity of the colour channels (R, G, B) of all pixels on the flank areas (i.e., intersection areas of flank detections and cow segmentations). RR was calculated using the same signal processing method.

2.5. Threshold development for heat stress

Piecewise regression models were used to fit the response of RR to heat stress (i.e., Ta and THI) and locate the breakpoint at which this response changed in trend. This was done by using the 'segmented' package which determines the breakpoint based on the Davies test [291]. Note that the analysis was based on the RR measured by the proposed framework on 55 video clips (n = 180). Specifically, simple linear regression models were first modelled to fit the response of RR to environmental variables (i.e., Ta and THI) using the lm function. Piecewise models were then built to update the simple models, written as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X_{bp}) X_k + \varepsilon_i, X_k = \begin{cases} 0 \; if \; X \le X_{bp} \\ 1 \; if \; X > X_{bp} \end{cases} \quad (15)$$

where Y is the RR, β_0 is the population intercept, X is the environmental variables (i.e., Ta and THI), X_{bp} is the breakpoint, X_k is the dummy variable, β_1 is the left slope, β_2 is the difference between right slope and left slope, and ε_i is the random residual for the i-th observation.

3. Results and discussion

3.1. Performance of cow segmentation and flank detection

The results on the testing sets demonstrate extremely good performance in segmenting cow instances with all three architectures trained with both image- and pen-level datasets achieving a bounding box mAP and a mask mAP exceeding 0.98 (**Table 4-3**).

Table 4-3: Performance of the three YOLOv5-seg architectures for segmenting cow instances, each trained separately using the image- and pen-level datasets.

Model	AP (bounding box)		mAP	AP (mask)		mAP	FPS	Model size (MB)
	Standing	Lying		Standing	Lying			
Image-level								
YOLOv5s-seg	0.993	0.994	0.993	0.993	0.988	0.991	53	15.3
YOLOv5m-seg	0.993	0.995	0.994	0.993	0.995	0.994	36	42.6
YOLOv5l-seg	0.993	0.995	0.994	0.993	0.995	0.994	24	364
Pen-level								
YOLOv5s-seg	0.981	0.99	0.985	0.985	0.984	0.985	49	15.3
YOLOv5m-seg	0.984	0.991	0.988	0.984	0.987	0.986	35	42.6
YOLOv5l-seg	0.991	0.993	0.992	0.989	0.993	0.991	24	364

AP = average precision; mAP = mean average precision; FPS = frames per second.

Additionally, all six object detection models trained with image- and pen-level datasets, respectively, had an mAP greater than 0.92 in detecting flank objects (**Table 4-4**). Notably, models trained on the image-level dataset always had a slightly higher mAP (at most 0.03) on the testing set compared with those trained on the pen-level dataset.

Table 4- 4: Performance of the three YOLOv5 architectures for detecting flank objects, each trained separately using the image- and pen-level datasets.

Model	AP (bounding box)		mAP	FPS	Model size (MB)
	Flank-up	Flank-down			
Image-level					
YOLOv5s	0.967	0.924	0.945	108	13.6
YOLOv5m	0.971	0.942	0.956	85	40.1
YOLOv5l	0.978	0.936	0.957	77	88.4
Pen-level					
YOLOv5s	0.957	0.885	0.921	103	13.6
YOLOv5m	0.961	0.891	0.926	83	40.1
YOLOv5l	0.975	0.904	0.94	75	88.4

AP = average precision; mAP = mean average precision; FPS = frames per second.

The similar performance between image- and pen-level splitting may suggest that data leakage of cow information is not a critical issue when multiple cows are present in the view and a sufficiently large sampling interval is maintained. Even if the images containing the same cows showed up across training, validation, and testing sets, they can be regarded as independent of each other as the same cows can have completely different postures and locations. The slightly higher mAP of the models trained with image-level datasets is mostly likely due to prior knowledge about pen facilities since they have already seen images from all four pens during training.

As expected, larger models had higher mAP but lower inference speed (**Table 4-3** and **Table 4-4**). In the case of cow instance segmentation by pen-level splitting, the YOLOv5s-seg model had a 0.7% lower box mAP and 0.6% lower mask mAP, but 104.2% faster speed compared with the YOLOv5l-seg model. In the case of flank object detection by pen-level splitting, the YOLOv5s model had a 1.9% lower mAP and 37.3% faster speed compared with the YOLOv5l model. The cases with image-level splitting shared the same trend. Therefore, the smallest models in both cases with cow-level splitting (i.e., the YOLOv5s-seg and YOLOv5-s models) were selected for further RR measurement due to their sufficiently good performance and fastest speed, as well as realistic and unbiased performance when generalising to new pens. Some example predictions of the trained YOLOv5s-seg and YOLOv5s models on the pen-level testing set are shown in **Figure 4-5**.

Figure 4- 5: Example predictions of cow segmentation (top) and flank detection (bottom) on the pen-level testing set. Predictions in red and green represent standing and lying cow instances, respectively. Predictions in yellow and blue represent the flank objects on the hind leg side and the opposite side, respectively.

Many previous works have segmented pixel-level masks for cattle. For example, Salau and Krieter [317] trained a Mask R-CNN network to segment Holstein-Friesian dairy cows housed in a similar free-stall barn. Their results showed a bounding box AP of 0.91 and a mask AP of 0.85. In addition, Xu et al. [318] trained several Mask R-CNN networks to segment cattle in different outdoor scenarios. They reported a bounding box AP of 0.95 and a mask AP of 0.94 for cattle on pastures, and a bounding box AP of 0.91 and a mask AP of 0.90 for cattle in feedlots. It should be noted that the instance segmentation performance obtained by YOLOv5 in this study is much better than that obtained by Mask R-CNN in the abovementioned studies, suggesting YOLOv5 or its more recent release as a new benchmark for instance segmentation in cattle research.

3.2. Performance of RR measurement based on optical flow method

This study aimed to develop a computer vision-based framework for measuring RR in dairy cows lying in free stalls. As shown in **Figure 4-6(a)**, the RR measured by the proposed optical flow-based method had a higher Pearson correlation coefficient (r = 0.945) and a lower root mean square error (5.24 bpm; RMSE) with the ground truth, outperforming the colour-based method. In addition, the proposed method had a narrower 95% limits of agreement (-8.97 to 11.13 bpm) on the Bland-Altman plot with the differences distributed homogeneously along the mean difference-mean axis (**Figure 4-6(b)**), suggesting the absence of proportional error. The mean difference of 1.08 bpm means that the proposed method generally counted 1.08 more bpm than visual observation. Additionally, the proposed method's intraclass correlation coefficient with the visual observation (0.98) outperformed that of the colour-based method (0.867) by 11.3%, confirming its superior agreement with the ground truth.

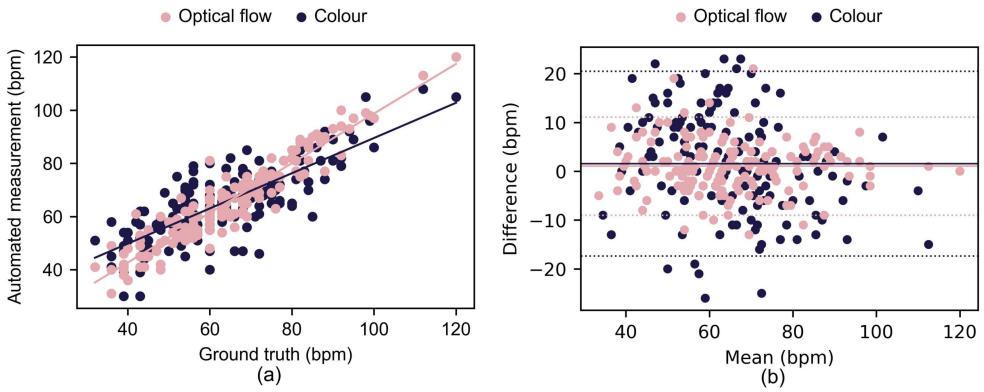
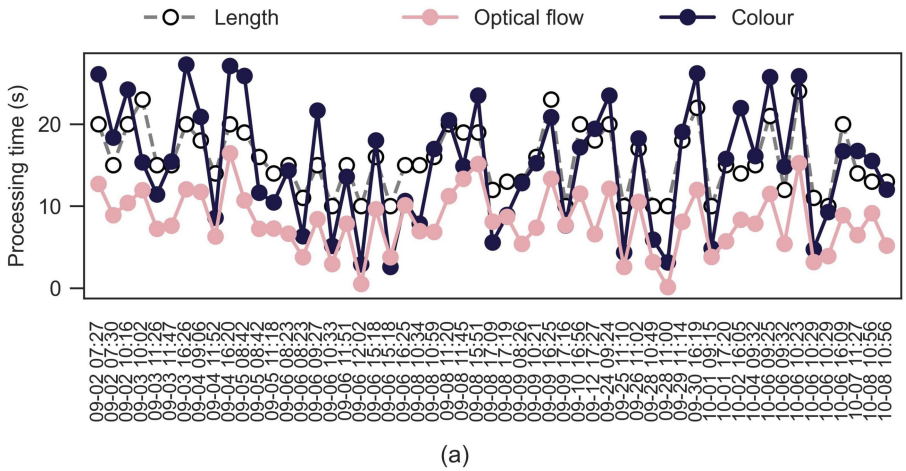


Figure 4- 6: Statistical analysis between automated and manual respiration rate measurements (breaths per minute (bpm), n = 180). Automated measurements relied on optical flow- and colour-based methods. (a) Linear regression plot, where the Pearson correlation coefficient of 0.945 and 0.787, respectively, and the root mean square error of 5.24 and 9.79 bpm, respectively. (b) Bland-Altman plot, where the solid lines represent the mean difference: 1.08 and 1.58 bpm, respectively, and the dashed lines represent 95% limits of agreement: -8.97 to 11.13 bpm and -17.35 to 20.51 bpm, respectively.

As for the overall inference speed, the optical flow-based method had an average processing time of 8.2 s and an average FPS of 64 per test video clip, which is 6.9 s and 26 FPS faster than the colour-based method, respectively (**Figure 4-7(a)**). Both methods had a processing time positively related to the number of objects to be processed, with the proposed method being affected less, as indicated by the lower slope in **Figure 4-7(b)**. Additionally, both methods had much lower processing FPS when handling a greater number of animals (**Figure 4-7(c)**). Anyway, all clips processed with the optical flow-based method had an average FPS over 30, demonstrating the achievement of real-time measurement. As a control, 23 clips processed with the colour-based method had an average FPS below 30.



Continued

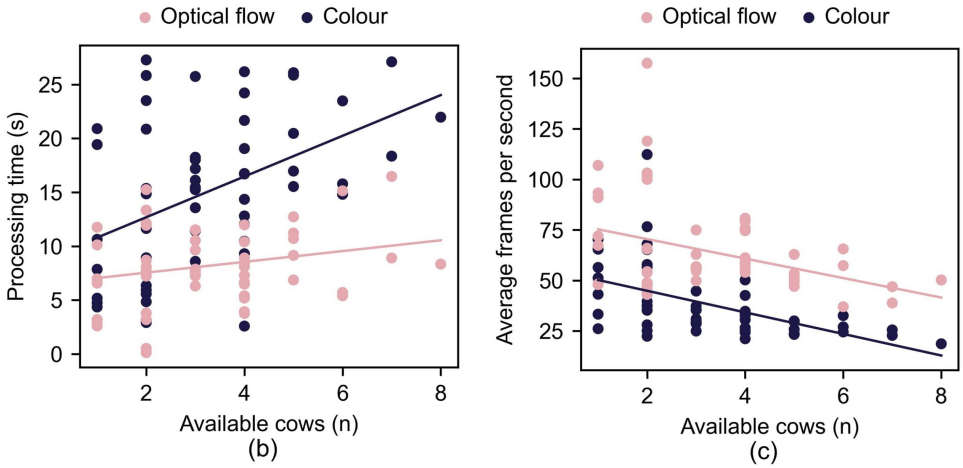


Figure 4- 7: Inference speed of the automated respiration rate measurements relied on optical flow- and colour-based methods. (a) Processing time of the 55 clips sorted by datetime (mm-dd hh:mm). (b) Relationship between the number of available cows to be processed and processing time. (c) Relationship between the number of available cows to be processed and average frames per second.

It should be acknowledged that previous studies have achieved optical flow-based RR measurement for an individual lying [140] or standing [141] cow. However, these methods require the field of view to be filled with the region of interest (i.e., abdomen) which is less efficient for a large-scale dairy farm. Wu et al. [153] achieved a multi-target RR measurement for cows in the exercise area by incorporating an instance segmentation network. However, their dataset only contained videos of two resting cows and the robustness of their method in free stalls and heat stress conditions remains unknown. Although our RMSE (5.24 bpm) is slightly higher than that of Wu et al. [153] (3.74 bpm), our study covered much broader thermal conditions and achieved multi-object RR measurement for cows lying in an area of approximately 16 free stalls. It should also be noted that the abovementioned studies did not apply object tracking techniques since they used a pre-selected dataset with only fixed and limited objects. For such a real-world scenario in our study, we included an object tracker into the proposed framework to deal with the identity disorder due to changing positions of interfering objects, as well as a video stabiliser to deal with unwanted video shakes. In addition, the control method using the averaged pixel intensity of R, G, and B channels had a slightly higher RMSE of 9.79 bpm than 8.3 bpm in the original study of Mantovani et al. [311], which may suggest that the current dataset is more challenging for colour-based methods, possibly due to a larger distance between animals and the camera.

Theoretically, both optical flow- and colour-based methods may have issues where the estimated signal may not accurately represent body motions when being overexposed or having less body texture. However, the colour-based method was

witnessed to suffer more since its calculation was based on the entire abdomen area which contained a bunch of pixels that did not change periodically with respiration. In contrast, optical flow-based methods pre-selected pixels in the abdomen area where the gradient was not zero. These pixels often belonged to the edges of patterns or dirt on the body, which had a higher amplitude of periodic changes, resulting in more stable RR estimation. In addition, the colour-based method's higher computational costs due to using all pixels for signal processing always led to lower processing speed. On the contrary, optical flow-based methods were faster since they only performed calculations on pixels with a non-zero gradient. It should be noted that our results do not necessarily indicate the superiority of optical flow methods over colour-based methods in all cases. It can only be concluded that colour-based methods require sophisticated filtering to select periodically changed pixels which can represent respiration. A possible solution, as suggested by Wang et al. [310], is to filter the pixels by using standard deviation. However, it was not adopted as it failed to filter non-periodic noise in our case.

3.3. Factors that might have affected the results

As shown in **Figure 4-8**, no clear evidence was found that flank colour has decreased the accuracy of RR calculation. Nine cases are outside the 95% limits of agreement, among which three (33.3%) were found to suffer from no texture. With good illumination and camera pixels, the flank areas which had no texture were still able to be clearly seen by sand or ribs. These provided enough pixels for reliable optical flow-based RR calculation. However, if being less illuminated or overexposed, flank pixels can be totally black and white and no details can be seen (that is no texture and no clear reference). Seven cases had this issue, and most pixels of the flank areas were excluded due to having a zero gradient while only pixels on the edge of the flank were kept due to having a non-zero gradient. Luckily, six out of the seven cases lay in the 95% limits of agreement since the pixels on the edge of the flank were effectively selected to support a precise RR calculation.

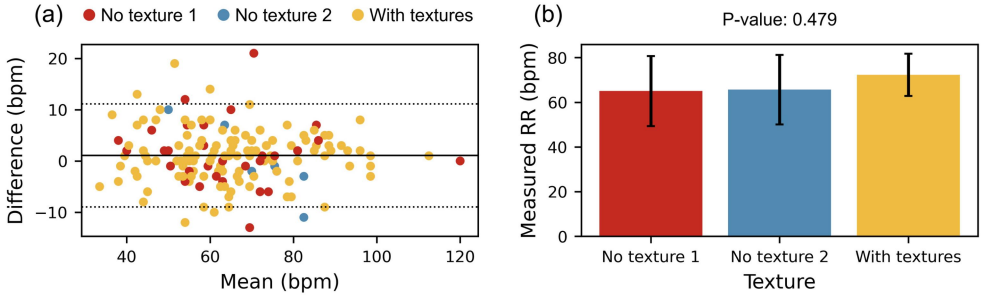


Figure 4-8: The effect of flank colour on measured respiration rate (RR, breaths per minute (bpm)). (a) Bland-Altman plot between automated measurements and manual observations. (b) Bar plot of one-way analysis of variance. No texture 1 = No texture but with clear sand or ribs (n = 34), No texture 2 = No texture and no clear reference (n = 7), With textures = with clear textures (n = 139).

In addition, three of the nine cases (33.3%) outside the 95% limits of agreement were found to suffer from partial occlusion (**Figure 4-9(a)**). No significant difference was found in measured RR between cases with partial occlusion and those without occlusion (P = 0.756; **Figure 4-9(b)**). Only partial occlusion was analysed since strong occlusion was not annotated at the beginning. It should be noted that this manual determination and removal of strong occlusions can be arbitrary, since there may be still some breathing-related pixels left. A more precise annotation and segmentation of cows which excludes all occlusion is expected to achieve RR measurement in more difficult cases.

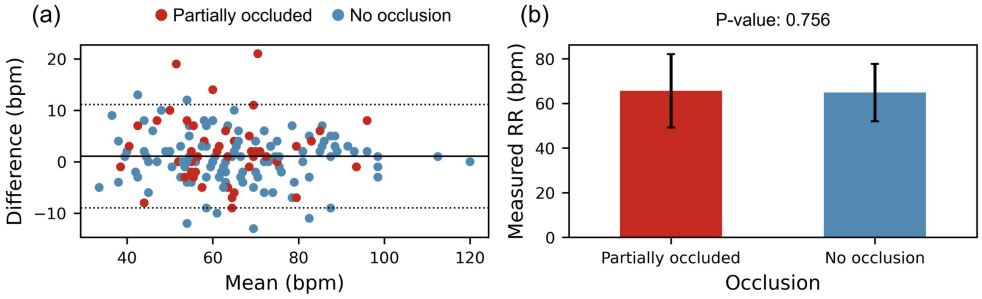


Figure 4-9: The effect of flank occlusion on measured respiration rate (RR, breaths per minute (bpm)). (a) Bland-Altman plot between automated measurements and manual observations. (b) Bar plot of one-way analysis of variance. Partially occluded = Flank partially occluded by cubicles or pillars (n = 52), No occlusion = Flank not occluded (n = 128).

Similarly, no evidence shows that location has decreased the accuracy of RR calculation (**Figure 4-10**). This is expected since cows showing up to the edge of the view with incomplete and strange postures were not annotated on their flank areas at first and thus were not taken into account for RR calculation. Therefore, no measurement was actually affected by the location.

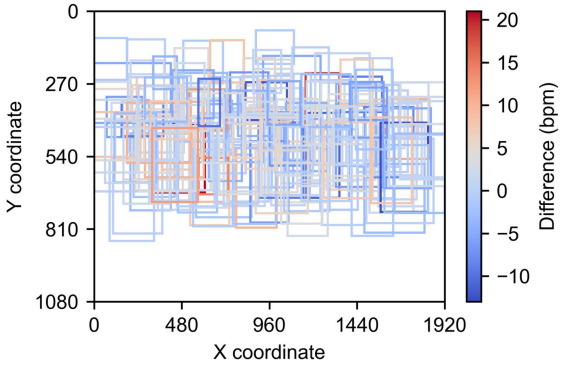


Figure 4-10: Visualisation of all cow bounding boxes in the last video frame, coloured by the difference between automated measured and manual observed respiration rate (breaths per minute, bpm).

Additionally, four of the five cases with the highest measurement error were observed to experience a sudden cessation of breathing-related movements in the abdomen for about 2 to 3 s. Due to the use of FFT, the current results reflected cyclical changes over a period of time, instead of counting the exact number of flank movements, thus overestimating the ground-truth manual observation. However, to some extent, the result from FFT can better represent actual RR because abdominal stillness due to the onset of rumination does not affect the cow's normal breathing. In other words, manual observation of RR by counting flank movements in this case may have underestimated the actual RR.

3.4. Response of RR to heat stress

The linear and piecewise regression models with Ta and THI as predictors are illustrated in **Figure 4-11** and detailed in **Table 4-5**. The slope of the model indicates a slow increase in RR of 0.4 bpm per unit increase in Ta below 23.6 °C and a much steeper increase in RR by 1.82 bpm per unit increase in Ta above 23.6 °C. This Ta threshold is slightly lower than the Ta threshold of 24.4 °C determined by Shu et al. [215]. It is probably because the current study included only lying cows which are more sensitive to heat stress [212].

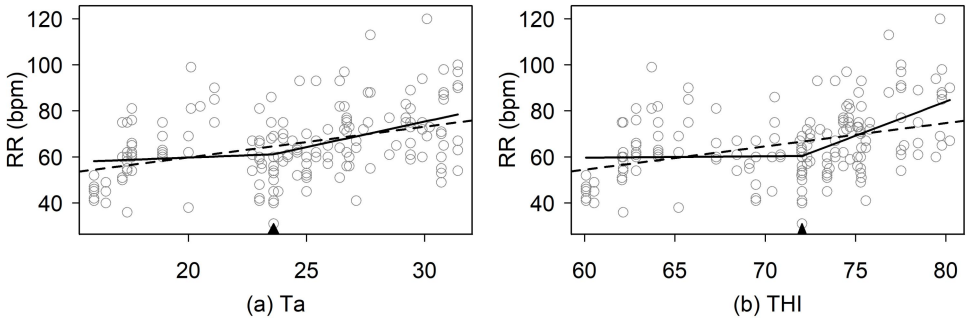


Figure 4- 11: Automated measurements of respiration rate (RR; breaths per minute (bpm), n = 180) and their fitted profiles from linear regression (dashed line) and piecewise regression (solid line) models with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictors. The breakpoints are marked as black triangles above the x-axis.

Table 4- 5: Parameter estimates (mean ± standard error) of the piecewise regression models predicting respiration rate with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively (n = 180).

Predictor	Intercept	Breakpoint	Left slope	ΔSlope	AIC	
					Linear	Piecewise
Ta	51.8 ± 11.2	23.6 ± 2.1	0.40 ± 0.55	1.82 ± 0.81	1453.8	1452.7
THI	56.0 ± 27.5	72.0 ± 1.4	0.06 ± 0.42	2.91 ± 0.71	1457.1	1444.7

ΔSlope = difference between right and left slopes; AIC = Akaike information criterion.

Additionally, RR increased slowly by 0.06 bpm per degree below 72 THI, and more steeply by 2.91 bpm per degree above 72 THI. This result is much higher than that of Pinto et al. [212], who determined a THI threshold of 65 for RR in lying cows. We speculate that the difference in the determined THI thresholds among studies can be attributed to the difference in production level since high productivity is well-known as a risk factor for heat stress [55]. The cows in the four experimental pens had an average milk yield of 32.3 kg/d during the study, which is much lower than the herd in the study of Pinto et al. [212] which had an average milk yield of 41.1 kg/d.

In addition, the determined THI threshold is lower than the threshold of 77 determined for dry cows [214]. This is sensible since the lactating state of our experimental cows made them more sensitive to heat stress than dry cows. Other cow-related factors, such as body condition score, coat colour, and breed, can all have an effect on animals' heat tolerance [319]. However, body condition and coat colour were not considered as grouping factors in this study, and no THI thresholds for RR for other breeds of dairy cows were found in previous studies. Thus, it is impossible to make a comparison with the current results.

The corresponding RR thresholds determined by the piecewise regression models were 61.1 and 60.4 bpm by Ta and THI, respectively (**Table 4-5**), which are consistent with Collier et al. [50] who determined the onset of heat stress when RR exceeded 60 bpm. In addition, the determined RR threshold is higher than those determined by Dalcin et al. [213] for ½ Holstein × Girolando dairy cows (30) and ¾ Holstein × Girolando dairy cows (45). This can be explained by the higher physiological states of pure-bred Holstein cows compared with mixed cows.

3.5. Limitations and perspectives

No sensitivity or comparison study was conducted to explore the impact of camera resolution and frame rate on the calculated RR. It is also notable that high-end cameras are too expensive to provide a complete view of the barn. Network recording cameras, although more economically accessible in commercial farms, do suffer from practical issues such as more severe camera shake and lens distortion. In the next step, we will test and improve our method on a video recording system.

It should be noted that the proposed method was only trained and tested on the data collected from one free-stall barn with one specific herd and breed. Although pen-level splitting was used to improve generalisability, further targeted training and testing are required before applying it to other facilities, herds, and breeds given the potentially significant environmental and animal differences. Anyway, this can be accelerated by using transfer learning techniques.

Additionally, night data was not collected in this study. Note that cows would take advantage of the greater temperature gradients from their bodies to the environment at night to relieve their accumulated heat load during the daytime [320]. It is, therefore, necessary to complete 24-h measurements for continuous monitoring of cows' stress state. Although lights are turned on at night in the barn, the change in

environmental illumination should have an impact on the robustness of the proposed method, which should be refined by further studies.

4. Conclusions

Collectively, our work represents an important step towards a non-contact RR measurement for dairy cows lying in free stalls. The proposed method can cover a lying area of approximately 16 free stalls and provide a precise and real-time RR measurement with a correlation coefficient of 0.945, an RMSE of 5.24 bpm, and an intraclass correlation coefficient of 0.98 when compared to visual observation, as well as an FPS of 64. Automated local cooling is promising by combining RR measurements and the critical thresholds to inform possible cooling systems at different places in the barn. However, further experiments and calibration with data collected by using low-cost video recording systems, recording during day- and night-time, are required before this technology can be applied on farms.

Context – Chapter 5

According to the findings reported in the first chapter, some behavioural changes are promising indicators of heat stress in dairy cows due to their ease of acquisition and sensitivity to heat stressors. RGB camera-based methods provide a low-cost solution for measuring behavioural dynamics. Especially, these behavioural changes can be measured and analysed not only at the individual level but also at the herd level, for instance by herd percentage. Therefore, the fifth chapter aimed to propose a deep learning-based model for recognising cow behaviours and to determine critical thresholds for the onset of heat stress at the herd level.

Chapter 5

Determining heat stress in a dairy herd via automated behaviour recognition

Adapted from:

Shu, H., Bindelle, J., Guo, L., Gu, X. *Determining the onset of heat stress in a dairy herd based on automated behaviour recognition*. Biosystems Engineering, 2023. https://doi.org/10.1016/j.biosystemseng.2023.01.009

Abstract

Dairy cows have various strategies for dealing with heat stress, including a change in behaviour. The aim of this study was to propose a deep learning-based model for recognising cow behaviours and to determine critical thresholds for the onset of heat stress at the herd level. A total of 1000 herd behaviour images taken in a free-stall pen were allocated with labels of five behaviours that are known to be influenced by the thermal environment. Three YOLOv5 architectures were trained by the transfer learning method. The results show the superiority of YOLOv5s with a mean average precision of 0.985 and an inference speed of 73 frames per second on the testing set. Further validation demonstrates excellent agreement in herd-level behavioural parameters between automated measurement and manual observation (intraclass correlation coefficient = 0.97). The analysis of automated behavioural measurements during a 10-day experiment with no to moderate heat stress reveals that lying and standing indices were most responding to heat stress and the test dairy herd began to change their behaviour at the earliest ambient temperature of 23.8 ℃ or temperature-humidity index of 68.5. Time effects were observed to alter the behavioural indicators values rather than their corresponding environmental thresholds. The proposed method enables a low-cost herd-level heat stress alert without imposing any burden on dairy cows.

Keywords: smart livestock farming, animal welfare, thermal comfort, group measurement, behavioural index

1. Introduction

Homeotherms, including dairy cows, constantly maintain thermal equilibrium with their environments through thermoregulation [1]. Heat stress is defined as the demand made by the environment for heat dissipation [6]. It is triggered when the thermal environment exceeds the upper critical threshold of the thermoneutral zone, inducing a variety of physiological and behavioural responses to reduce heat production and increase heat dissipation [5]. Due to the lack of real-time, large-scale, and automated measurement of animal-based indicators, heat mitigation in practice has long depended on environmental indicators and their critical thresholds [308]. Environmental indicators, however, do not reflect the actual response of the animal, making it difficult to evaluate the effectiveness of cooling measures [321].

Dairy cows take a series of behavioural changes to cope with heat stress. As environmental temperature increases, cows will spend more time standing to increase surface area for better heat dissipation [322]. However, this may result in a significant reduction in sleeping time, posing a potential risk to cow welfare [5]. Cows will drink more frequently under heat stress but with less water each time [56]. Cows will also reduce feed intake and subsequent rumination to reduce metabolic heat production [323]. Therefore, recognising changes in behavioural patterns ascribable to heat stress can help quantify the true response of cows.

Recently, deep learning-based methods have allowed the automated recognition of basic cow behaviours such as lying, standing, and drinking with an accuracy of up to 0.976 [166,168]. Further quantification of the results with association to animal growth, health, and welfare, and the extent to which they can be used to improve decision making have been highlighted for future work [324]. Tsai et al. [167] analysed how drinking time and frequency were affected by heat stress after detecting drinking behaviour with a convolutional neural network. Still, further application of deep learning techniques is required so that the detection of multiple behaviours enables a more comprehensive analysis of when heat stress is triggered.

Although progress has been made in computer vision-based individual identification in free-stall barns [325], issues such as lack of colour pattern and occlusion still lead to poor identification. In addition, detections from deep learning methods provide an opportunity to calculate herd-level behavioural indices which have been commonly used for evaluating cow comfort. For example, cow lying index, which is defined as the number of cows lying in the stall divided by the total number of cows, has been calculated automatically with a computer vision-based system [162]. Other indices related to free-stall usage and cow comfort would require knowing whether the cow is standing on the stall bed and whether the cow is eating or drinking. Therefore, a detailed behavioural recognition method that addresses the above questions is still required to compute these indices in an automated way.

Scan sampling, as a common method in animal research, is often used to record herd-level behavioural indices at predetermined intervals. Traditionally, scan sampling requires manual checks through direct observations or video recording,

both of which are time- and labour-consuming. With the help of automated behaviour recognition, video frames can be processed in real time. However, continuous processing and storing of data would be a waste of time and memory. Scan sampling is still of great value, especially for behaviours that basically follow continuous and diurnal patterns, such as lying and standing. The sampling interval should always be determined in accordance with specific purposes. For example, studies or regular checks aimed at ascertaining standing and lying patterns may use sampling intervals of 30 or even 60 min [326].

By combining deep learning and scan sampling methods, the onset of heat stress is promising to be determined at the herd level with the memory and power of local devices greatly saved. Therefore, the aim of this study was (1) to train and validate a deep learning-based model to recognise cow behaviours and further calculate herd-level behavioural indicators, and (2) to develop critical thresholds of heat stress at herd level based on behavioural indicators.

2. Materials and methods

All protocols involving animals were approved by the Experimental Animal Care and Use Committee of Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (approval number IAS2021-220).

2.1. Experimental design

The experiment was conducted on a free-stall dairy farm in May 2021 in Shandong, China, which has a temperate continental monsoon climate with hot and humid summers. The experiment consisted of two periods, in which the first three days in early May were designed to collect data for training behaviour recognition models while the other ten days during mid-May when the environment got warmer were designed to explore how the onset of heat stress can be determined through automated behaviour recognition.

2.1.1. Housing, animals, and management

The barn had four pens with a 4-row head-to-head design and was covered by a double-pitched roof (gradient 15%). The experimental pen (11 m × 96 m, oriented along the north–south axis) housed 79 lactating Holstein-Friesian dairy cows with 128 stalls and 128 headlocks. The pen could be evenly divided into eight areas, with each having a feeding zone of 16 headlocks and a parallel resting zone of 10 to 20 stalls with or without a water trough (see **Appendix 1. Supplementary material Figure A5-1**). At the beginning of the experiment, the cows had a mean ± SD milk yield of 30.4 ± 11.8 kg/day, parity of 2.8 ± 1.4, and days in milk of 273.1 ± 117.3. The cows were milked three times daily at 08:30, 16:30, and 00:00 h in a parlour that was about 20 m away from the barn. All cows were observed to return to the pen within 1 h of departure. A total mixed ration was delivered three times daily after milking. Clean drinking water was delivered in five troughs. Both feed and water were provided ad libitum to all cows. Electronic fans (1.1 m in diameter; capacity: 25000 m^3/h each; see **Figure A5-1**) were turned on when the indoor temperature reached 20 °C whereas sprinklers remained closed during the

experiment. Stalls were sand-bedded to a depth of about 150 mm and raked once daily while the cows were away for morning milking.

2.1.2. Behavioural and environmental measurement

Cow behaviour was recorded using eight closed-circuit video cameras (DS-IPC-K14L-WT; Hikvision, Hangzhou, China) which were evenly spaced and placed opposite the pen's longitudinal axis at a height of about 6 m and an angle of about 45° downward (see **Figure A5-1**). Each camera was able to capture a feeding zone with 16 headlocks and a parallel resting zone, allowing the complete side view of the pen to be captured. The eight cameras were linked, synchronised, and controlled using a Wi-Fi router (DS-3WR23-E; Hikvision, Hangzhou, China) and an eight-channel video recording system (DS-7808NB-K1/W; Hikvision, Hangzhou, China). It is recommended that video recordings taken between 08:00 and 15:00 h is best for representing daily behavioural pattern in summer [327]. Besides, behavioural assessment should be performed at least 1 h after cows return from morning milking to avoid being affected by intensive feeding [328]. Accordingly, video recording was performed from 10:30 to 15:00 h on each test day as adapting the previous recommendations to the actual schedule.

Environmental parameters including ambient temperature (Ta, ℃) and relative humidity (RH, %) were measured at an interval of 10 min using a total of six Kestrel 5000 environment meters and Kestrel 5400 heat stress trackers (accuracy: ± 0.4 °C Ta, ± 1% RH; Nielsen-Kellerman, Boothwyn, PA, USA; see **Figure A5-1**). These sensors were evenly distributed in the pen and were fixed at a height of 2.2 m. The measurements from all sensors were averaged for representing the global environment inside the pen. The temperature-humidity index (THI) was calculated according to Eq. (1) [222].

$$THI = (1.8 \times Ta + 32) - (0.55 - 0.0055 \times RH) \times (1.8 \times Ta - 26) \quad (1)$$

2.2. Development of behaviour recognition model

2.2.1. Data preparation

Video frames from the first three-day experiment were extracted at an interval of 6 min by using a self-written program in Python. This interval was set for increasing the heterogeneity of the training data since the cows changed their behaviour less frequently. The extracted frames were in JPG format and were further corrected for distortion and cropped using OpenCV. The final images for training and evaluation had a resolution of 1920 by 1080 pixels. Finally, a total of 1000 images were chosen after eliminating low-quality frames (e.g., lens covered by flies).

To maintain the good quality of training data, all cows presented in all images were annotated as per the definition presented in **Table 5-1**. Five target behaviours known to be influenced by the thermal environment were carefully defined to be exclusive, meaning that the cows were not able to perform two behaviours simultaneously. 'Standing-in' was previously subdivided into 'perching' and 'standing' based on whether all four feet or just two feet touched a stall [329]. However, this was abandoned due to insufficient data in each subgroup. The annotation tool LabelImg (https://github.com/tzutalin/labelImg) was used to allocate

the appropriate class to each cow per image with a bounding box. A total of 1000 annotated images were randomly split into a training set (60%), a validation set (20%), and a testing set (20%). A detailed description of the behaviour recognition dataset with regards to the number of labels per class is given in **Table 5-2**.

Table 5- 1: Definition of the target cow behaviours.

Behaviour	Definition
Drinking	Standing by a water trough with mouth in the trough
Eating	Standing with neck in a feeding rack
Lying	Lying in total lateral or sternal recumbency within a stall
Standing-in	Standing with two or more feet touching a stall bed
Standing-out	Standing or walking outside the stall but not eating

Table 5- 2: Overview of the behaviour recognition dataset. Images were extracted using 6-min scan sampling from 10:30 to 15:00 h during the three-day experiment.

Dataset	Number of cow labels					
	Drinking	Eating	Lying	Standing-in	Standing-out	Total
Training (600 images)	164	269	3326	509	2682	6950
Validation (200 images)	49	73	1237	170	836	2365
Testing (200 images)	40	103	1152	186	769	2250
Total	253	445	5715	865	4287	11565

2.2.2. Deep learning algorithm and transfer learning

YOLO (You Only Look Once) is a popular one-stage framework for object detection. Unlike two-stage methods (e.g., Faster R-CNN), one-stage methods skip the region proposal stage, and regard object detection as a regression task with class probability and coordinates of the bounding box as the outcome. As a result, one-stage methods have much higher inference speeds and relatively lower accuracy.

YOLOv5 is a recent popular version of the YOLO-series of algorithms. Adapting from YOLOv4, YOLOv5 uses Focus structure with CSPdarknet53 as the backbone and introduces Spatial Pyramid Pooling method, mosaic training, self-adversary training, and multi-channel feature. It has been reported that YOLOv5 has a much higher inference speed and smaller size compared with its previous versions without sacrificing accuracy [330]. Thus, YOLOv5 was chosen to recognise cow behaviour in the present study. Specifically, three architectures were trained, with size increasing from YOLOv5s, YOLOv5m, to YOLOv5l.

Deep learning methods require a lot of time and memory to train on a large number of images. Transfer learning, which involves transferring previously learned knowledge from a related task, is expected to accelerate the training process and usually produce better results than training from scratch. The pre-trained weights

used in this study were provided by the authors of YOLOv5 based on the COCO dataset which is a benchmark object detection dataset published by Microsoft [315].

The training of the different YOLOv5 architectures was performed in Python 3.8 language with Pytorch 1.8.0. on a 64-bit version Windows 11 laptop with NVIDIA GeForce RTX 3060 GPU and 6 GB video memory. The batch size was set to 8 and the epoch was set to 100. Hyperparameters were set as default.

2.2.3. Performance evaluation

In object detection tasks, an Intersection over Union (IoU) threshold has to be given first to determine whether a predicted bounding box should be classified as positive or negative. The IoU is the overlap of the predicted and ground-truth bounding boxes divided by their union, as expressed in Eq. (2):

$$IoU = \frac{Prediction \cap Ground\ truth}{Prediction \cup Ground\ truth} \quad (2)$$

In this study, an IoU threshold of 0.5 was adopted by convention. Thus, a detection with an IoU ≥0.5 was classified as true positive (TP), a detection with an IoU <0.5 was classified as false positive (FP), and a ground truth presented but failed to be detected is classified as false negative (FN).

Afterward, the precision (Pr), recall (R), and average precision (AP) were calculated according to Eqs. (3 - 5). The Pr indicates the proportion of the predicted bounding boxes being correctly detected whereas the R indicates the proportion of the ground-truth bounding boxes being correctly detected. The ideal object detector should have high Pr and R at the same time. However, there is a trade-off between the two metrics depending on the confidence threshold. Confidence represents the probability (0 to 1) of a bounding box containing an object and the predictions with class probabilities lower than a given confidence threshold will be removed. A very high confidence threshold will discourage the model from making positive predictions, thus increasing Pr and decreasing R, and vice versa. The AP is a commonly recommended metric since it summarises the Pr along with the R at all possible confidence thresholds. This was done by default setting the confidence threshold to 0.001.

$$Pr = \frac{TP}{TP+FP} \quad (3)$$

$$R = \frac{TP}{TP+FN} \quad (4)$$

$$AP = \int_0^1 Pr_{(R)}\, dR \quad (5)$$

In such a multi-class task, the mean average precision (mAP) was used to evaluate the overall performance, as expressed in Eq. (6):

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \quad (6)$$

where C takes 5, indicating 5 classes (i.e., 'Drinking', 'Eating', 'Lying', 'Standing-in', 'Standing-out'). In addition, frames per second (FPS) was used to indicate the inference speed, as expressed in Eq. (7):

$$FPS = \frac{N}{t_N} \quad (7)$$

where t_N is the total inference time (s) on N images.

Unlike AP calculations, where all potential confidence thresholds are required, the confidence threshold for actual inference must be tuned and specified for better detection. The confidence threshold for inference on the testing set was determined by checking the global maximum on the F1 confidence curve. The F1 score is the harmonic mean of Pr and R, as expressed in Eq. (8):

$$F1\ score = 2 \times \frac{Pr \times R}{Pr + R} \quad (8)$$

2.3. Behavioural indicator calculation and heat stress determination

2.3.1. Behaviour recognition

The video frames were further extracted using 30-min scan sampling from 10:30 to 15:00 h since this method has been validated to be effective and efficient for analysing cow behaviour [327,331]. Consequently, 10 scan samples were obtained for each of the 13 test days, each containing eight images from eight cameras. The proposed behaviour recognition model was applied to all images per scan sample.

2.3.2. Detection filtering

The eight cameras captured all 128 headlocks in the feeding zone, but partially overlapped at the far end of the view (i.e., resting zone) due to the fact that faraway objects naturally appear smaller than closer ones. However, it is important to count each cow only once when calculating herd-level behavioural indicators. Thus, processing had to be done to filter each detection per image per camera to ensure that only the detections located within the area of interest were counted. As shown in **Figure 5-1**, the area of interest was predetermined for each camera by using a polygon that covered the entire floor of the feeding zone and its parallel resting zone. The filtering was based on Jordan Curve Theorem [332]. For each predicted bounding box, a ray was first drawn horizontally to the right of the centre of its lower boundary line. This was to ensure that the cows stood or lied within the area. If the number of intersections was odd then the centre was inside the polygon and the predicted bounding box was kept. This filtering was based on the coordinates of the polygon and those of the lower centres of the bounding boxes using a self-written program in Python. The filtered number of detections per class per image was therefore obtained.

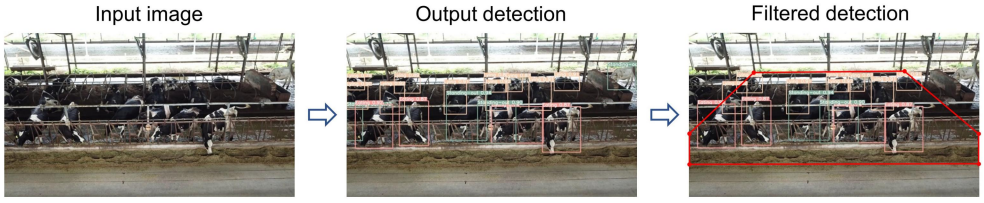


Figure 5- 1: Schematic of detection filtering taking one camera for example. The predefined red polygon represents the area of interest and only the bounding boxes with their lower centres in the area are kept.

2.3.3. Behavioural parameters

The number of detections for each behavioural class per scan was determined after merging the results into each scan sample. The herd-level behaviour distribution of each scan was naturally calculated by dividing the number of detections for each behavioural class by the total number of detections. Except for the percentage of cows lying, several herd-level behavioural indices have often been used for characterising lying and standing behaviours. With the automated measurements, comfort index (CI), stall-use index (SUI), and cow stress index (CSI) were calculated according to Eqs. (9 - 11) [328,331]:

$$CI = lying/(lying + standing\text{–}in) \quad (9)$$

$$SUI = lying/(lying + standing\text{–}in + standing\text{–}out) \quad (10)$$

$$CSI = (standing\text{–}in + standing\text{–}out)/n \quad (11)$$

where *n* denotes the total number of cows.

To further evaluate the proposed behaviour recognition model, the scan samples from the first three-day recording were manually observed to count the frequency of each behaviour. The manual results of behavioural parameters including the percentage of all five target behaviours as well as advanced lying and standing indices (i.e., CI, SUI, and CSI) were then calculated. The intraclass correlation coefficient was computed to assess the agreement between manual and automated methods using the icc function from the 'irr' package (R version 3.4.4; https://R-project.org).

2.3.4. Threshold development

Since the proposed model worked well in recognising herd-level behavioural parameters, the remaining 10-day data were further used to explore the herd-level behavioural pattern with respect to the onset of heat stress. A total of six herd-level behavioural parameters, including the percentages of cows drinking (drinking%), eating (eating%), and lying (lying%), as well as CI, SUI, and CSI, were used as animal-based indicators, whereas Ta and THI were used as environmental indicators. Data from 11:00 to 14:30 h were used for further analyses since their behavioural results show a clear association with environmental indicators.

All statistical analyses in this section were performed using R software. Spearman's rank correlation analysis was performed using the cor function to explore how these indicators were associated with each other. Piecewise regression models were used to fit the response of animal-based indicators to environmental indicators and locate the breakpoint at which this response changed in trend. The piecewise models were built with the 'segmented' package which works by updating the existing models with two or more breakpoints based on Davies test [291]. Therefore, basic piecewise models were built by updating simple linear regression models fitted with the lm function. The models are written as Eq. (12):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X_{bp}) X_k + \varepsilon_i, X_k = \begin{cases} 0 \; if \; X \le X_{bp} \\ 1 \; if \; X > X_{bp} \end{cases} \quad (12)$$

where Y_i is the animal-based indicators, β_0 is the population intercept, β_1 is the left slope, X_i is the environmental indicators, β_2 is the difference between the right slope

and left slope, X_{bp} is the breakpoint, X_k is the dummy variable, and ε_i is the random residual for the i-th observation.

If the basic piecewise models converged, advanced piecewise models with the random effect of time of day were built to separate the profiles of different hours. Otherwise, the results would be shown with simple linear regression models. Advanced piecewise models were built by updating linear mixed models fitted with the lme function included in the 'nlme' package. The random effect of time of day was included for every model parameter, including intercept, slope difference, and breakpoint. For j-th time of day, the model can be written as Eq. (13):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \beta_{2j}(X_{ij} - X_{bpj})X_k + \varepsilon_{ij}, X_k = \begin{cases} 0 \; if \; X \le X_{bpj} \\ 1 \; if \; X > X_{bpj} \end{cases} \quad (13)$$

where each parameter is given by the sum of fixed and random effects (e.g., β_{0j} is the sum of fixed term β_0 and random term d_j).

3. Results and discussion

3.1. Performance of behaviour recognition

Figure 5-2 showing the training process of three YOLOv5 architectures indicates a faster convergence always in the validation mAP than the training loss. As expected, the training converged earlier as the network deepens in size. As shown in **Table 5-3**, the mAP on the testing set was roughly close among YOLOv5 models, in which 'Drinking' was consistently the most difficult to detect, probably due to the most limited data to train. This is somehow inevitable when using such an imbalanced dataset with 'Standing-out' being almost 16 times more represented than 'Drinking'. In the case of FPS, models with smaller sizes show much higher inference speed, with YOLOv5s increasing by 32.7% compared with YOLOv5m, and by 65.9% compared with YOLOv5l. Other algorithms commonly used for object detection were not compared since the YOLOv5 algorithms have already shown extremely good performance. In fact, YOLOv5 has been reported to have a dramatic increase in inference speed compared with YOLOv3 and YOLOv4 [333]. Our results show that all three YOLOv5 architectures should be effective and capable of being deployed on mobile terminals. Anyway, we simply chose the YOLOv5s model for further application due to its ability to further compress the weight size while maintaining accuracy.

The F1 confidence curve indicates that the F1 score of all behavioural classes peaked at a confidence threshold of 0.696 (**Figure 5-3(a)**). Thus, further evaluation and application were performed with confidence threshold set to 0.696. The confusion matrix shown in **Figure 5-3(b)** indicates that the major misclassification was marked between drinking and standing-out behaviours and between standing-in and lying behaviours, with 1 out of 40 'Drinking' labels being misclassified as 'Standing-out' and 4 out of 186 'Standing-in' labels being misclassified as 'Lying', respectively. The example results shown in **Figure 5-3(c)** demonstrate a good ability in dealing with occlusion caused by facilities or other cows.

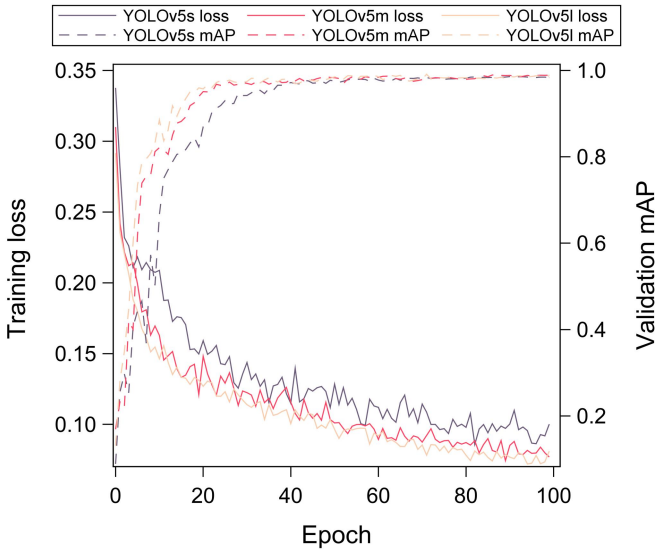


Figure 5- 2: Comparison of training loss and validation mean average precision (mAP) for three YOLOv5 architectures.

Table 5- 3: Performance comparison of three YOLOv5 architectures.

Model	Average precision					mAP	FPS	Size (M)
	Drinking	Eating	Lying	Standing-in	Standing-out			
YOLOv5s	0.944	0.995	0.995	0.995	0.994	0.985	73	13.7
YOLOv5m	0.963	0.995	0.995	0.995	0.995	0.988	55	40.2
YOLOv5l	0.956	0.995	0.995	0.994	0.995	0.987	44	88.5

mAP = mean average precision; FPS = frames per second.

When calculating herd-level behavioural parameters for the 30 scan samples from the first three-day recording, the intraclass correlation coefficient between manual and automated methods was 0.97. Moreover, the overall linear relationship shows that only 11 out of 240 observations were classified to be outliers by the 95% prediction limits (**Figure 5-4**). Collectively, these results demonstrate an excellent agreement between manual and automated methods in obtaining herd-level behavioural parameters. Of note, given that the view provided by our cameras might be difficult to show the relationship between a cow's head and the trough when there was a strong occlusion, both the ground truth and detections could have underestimated the incidence of drinking, thereby affecting the calculated behavioural parameters. For example, the drinking% and the SUI would be underestimated whereas the CSI would be overestimated. This bias should cause negligible effects on the current results since such occlusion happened in only five scan samples (a total of eight occluded cows) during the entire 13 test days and

drinking played a limited role in the equation compared with standing and lying. However, it may have a stronger impact on behavioural parameter calculation when facing a higher farming density and more occlusions. A top-view camera, as presented by Tsai et al. [167], could be a solution to eliminating this measurement bias. Another possible way is to introduce new labels describing cow behaviour in the drinking area. This is of interest since it can be very crowded in drinking areas during hot seasons and cows would even compete for troughs [159].

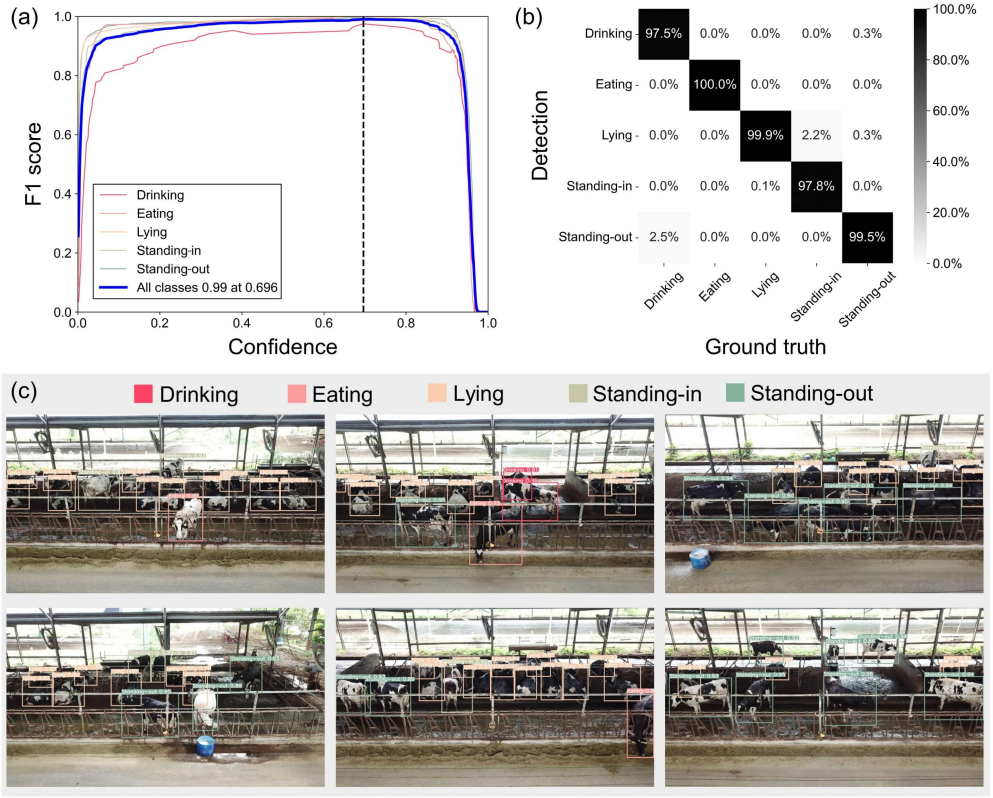


Figure 5-3: Detailed performance of the YOLOv5s model. (a) F1 confidence curve with the black vertical dashed line indicating the best F1 score at a confidence threshold of 0.696. (b) Confusion matrix (normalised by column) with confidence threshold set to 0.696. (c) Example results of behavioural recognition with confidence threshold set to 0.696.

3.2. Behavioural pattern under heat stress conditions

As shown in **Figure 5-5**, the 10-day experiment conducted in mid-May well captured the beginning of heat stress with daily mean Ta rising from 14.7 to 25.8 °C and daily mean THI rising from 58.4 to 74.4. The heat stress threshold for high-producing dairy cows (≥35 kg/day) has been updated to a daily mean THI of 68 by Collier et al. [50]. According to their revised THI thresholds, the study herd first

stayed within the thermoneutral zone and then experienced mild to moderate heat stress during the 10-day experiment.

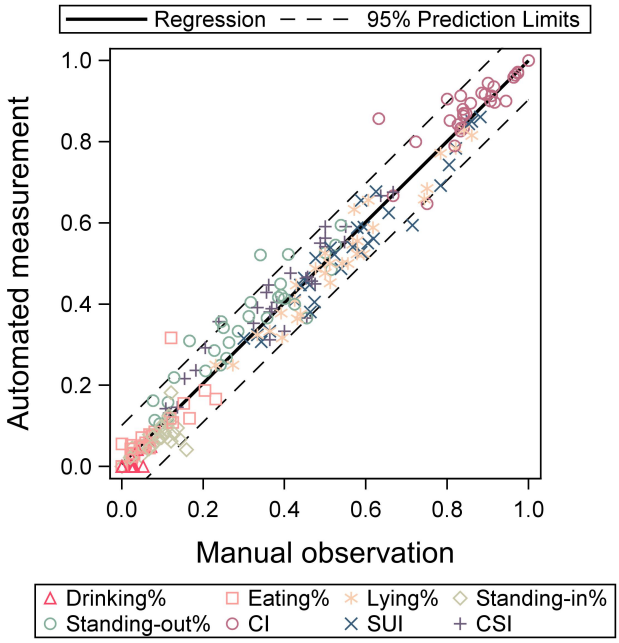


Figure 5- 4: Comparison of herd-level behavioural parameters measured half-hourly by manual and automated methods from 10:30 to 15:00 h during the three-day experiment. Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; Standing-in% = percentage of cows standing-in; Standing-out% = percentage of cows standing-out; CI = comfort index; SUI = stall-use index; CSI = cow stress index.

The 100% bar chart presents the herd-level behavioural pattern from 10:30 to 15:00 h (**Figure 5-6**). The complete statistics used for plotting can be found in **Appendix 1. Supplementary material Table A5-1**. As expected, lying consistently occupied the largest proportion. Lying has long been used for indicating cow comfort and welfare, and cows should spend most idle time lying [334]. Besides, cows had the lowest lying% and a relatively high eating% at 10:30 h compared with other scan samples, indicating that the effect of intensive feeding was still lasting 2 h after leaving for the morning milking which was scheduled at 08:30 h. Afterward, the lying% raised dramatically until it peaked at 11:00 h when the majority of cows were resting on their stall beds. The lying% then followed a decreasing trend from 11:00 to 14:30 h, which was right opposite to the rising Ta and THI. These patterns are consistent with Overton et al. [328] and Mattachini et al. [331], who found that a herd needs 2 to 3 h after leaving for milking to finish feeding and return to rest, and increasing environmental temperature will decrease the lying% during idle time. Therefore, the eight scan samples from 11:00 to 14:30 h were used for further

analyses exploring the effect of heat stress on cow behaviour due to a clear relationship between cow behaviour and thermal environment during this period.

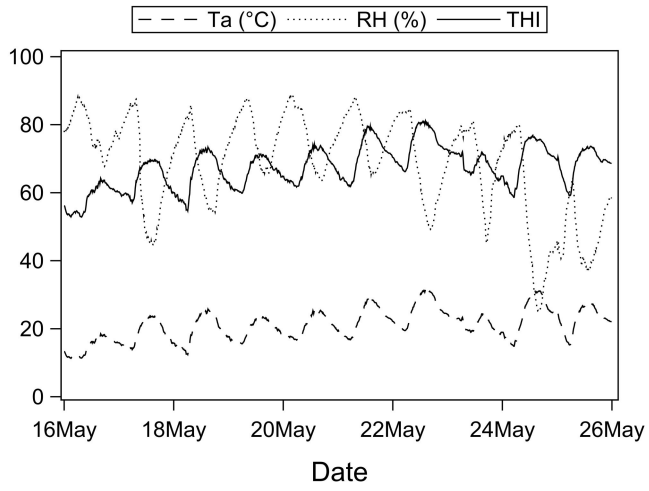


Figure 5- 5: Overall variation of indoor ambient temperature (Ta, °C), relative humidity (RH, %), and temperature-humidity index (THI) during the 10-day experiment with a measurement interval of 10 min.

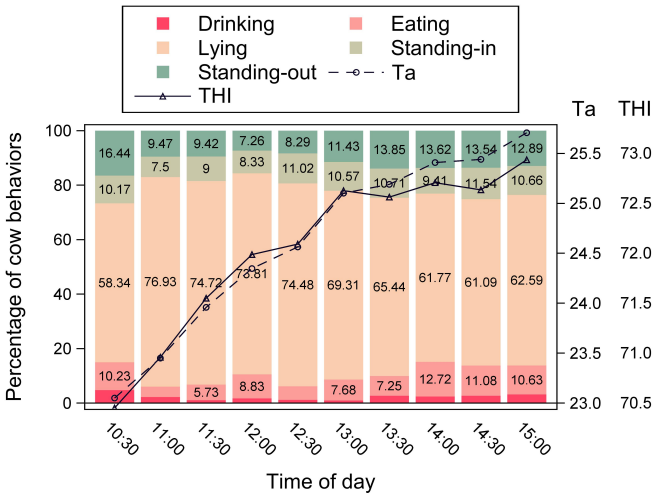


Figure 5- 6: Herd-level behaviour distribution measured half-hourly by the proposed automated method, as well as the corresponding ambient temperature (Ta, °C) and temperature-humidity index (THI), averaged during the 10-day experiment.

The spaghetti plot shows the temporal pattern of six herd-level behavioural indicators with regard to two environmental indicators (**Figure 5-7**). The complete statistics used for plotting can be found in **Appendix 1. Supplementary material**

Table A5-1. Consistent with lying%, indices describing lying behaviour (i.e., CI and SUI) increased to a peak at 11:00 h and followed an overall decreasing trend until 14:30 h. The only difference between lying%, CI, and SUI was the calculation of the denominator. Neither CI nor SUI takes eating into account for calculation. An increasing eating% would therefore lower the denominator and finally increase their results. This leads to fluctuations in CI and SUI from 11:30 to 12:00 h and from 13:30 to 14:00 h. Moreover, CSI, as an index describing idle standing behaviour, showed an almost horizontally symmetrical trajectory to SUI. These trends of lying%, SUI, and CSI are comparable with Mattachini et al. [331] who manually observed cow behaviour through video recording with a 60-min scan sampling. Once again, this demonstrates the effectiveness of our proposed automated method for behaviour recognition.

Generally, lying and standing indices (i.e., lying%, CI, SUI, and CSI) and drinking% had strong correlations with Ta and THI (all $P < 0.01$), whereas eating% did not appear to have any correlation (both $P > 0.05$) (**Figure 5-8**). Of note, Ta had a stronger correlation with cow behaviour than THI with exception of drinking% and eating%. A better correlation of Ta with animal-based indicators was also observed in our previous study when physiological indicators were used [215]. To some extent, Ta seems to better describe environmental stress than THI in this specific environmental condition, probably because the effect of RH not yet playing an important role at the very beginning of summer.

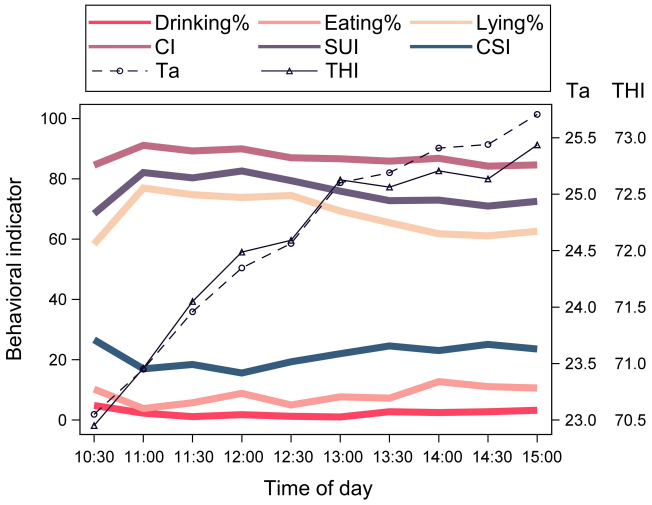


Figure 5- 7: Herd-level behavioural indicators measured half-hourly by the proposed automated method, as well as the corresponding ambient temperature (Ta, °C) and temperature-humidity index (THI), averaged during the 10-day experiment. Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index.

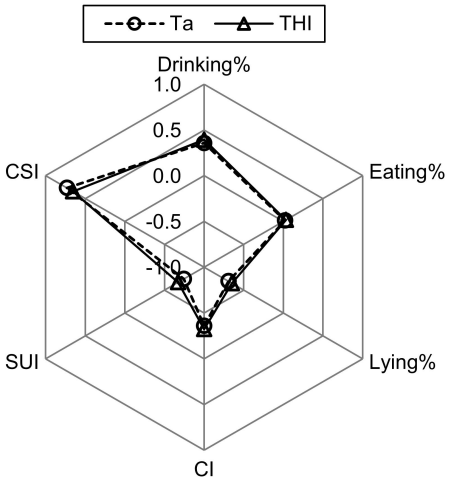


Figure 5- 8: Spearman's rank correlation coefficients between herd-level behavioural indicators measured half-hourly by the proposed automated method from 10:30 to 15:00 h during the 10-day experiment and ambient temperature (Ta, °C) and temperature-humidity index (THI). Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index.

Among lying and standing indices, CI correlated the weakest with Ta ($r = -0.360$, $P < 0.001$) and THI ($r = -0.322$, $P = 0.002$), whereas SUI correlated the strongest with Ta ($r = -0.744$, $P < 0.001$) and THI ($r = -0.673$, $P < 0.001$). Similarly, CI was found less susceptible to environmental temperature compared with lying% and SUI by Overton et al. [335], and the strongest correlation coefficient was found between SUI and Ta (-0.762) by Mattachini et al. [331]. It is well known that cows will change their drinking and feeding patterns to better coping with heat stress [1]. The positive correlation of drinking% with Ta ($r = 0.357$, $P < 0.001$) and THI ($r = 0.393$, $P < 0.001$) observed in this study might be attributed to increased visit to the trough and increased time per visit, as previously identified by Tsai et al. [167] and McDonald et al. [159]. Moreover, eating patterns can change through different strategies by cows with different production stages [336] or social ranks [337]. Thus, the small to no effect of heat stress on eating% is probably due to heterogeneous strategies taken among the herd. Indeed, some cows may reduce their feed intake with a longer eating time [158,336]. Anyway, eating% alone appears to be insufficient to quantify eating behaviour, and more data from longer duration or smaller sampling intervals are required to evaluate its ability as a herd-level heat stress indicator.

3.3. Critical threshold for determining the onset of heat stress

The basic linear and piecewise models visualised in **Figure 5-9** show the benefit of using lying%, SUI, and CSI for indicating the onset of heat stress. Their statistical results presented in **Table 5-4** indicate that the lowest upper critical Ta was

associated with SUI and CSI (both 23.8 °C), and the lowest upper critical THI was associated with lying% (68.5). The slope differences were always higher with Ta than with THI, once again suggesting the superiority of Ta in representing the stress imposed by the current environment. As for the behavioural indicators that were not converged in the piecewise regression, drinking% and CI were positively and negatively related to environmental indicators, respectively, whereas eating% appeared to be independent of Ta and THI. These results are consistent with those from the correlation analysis.

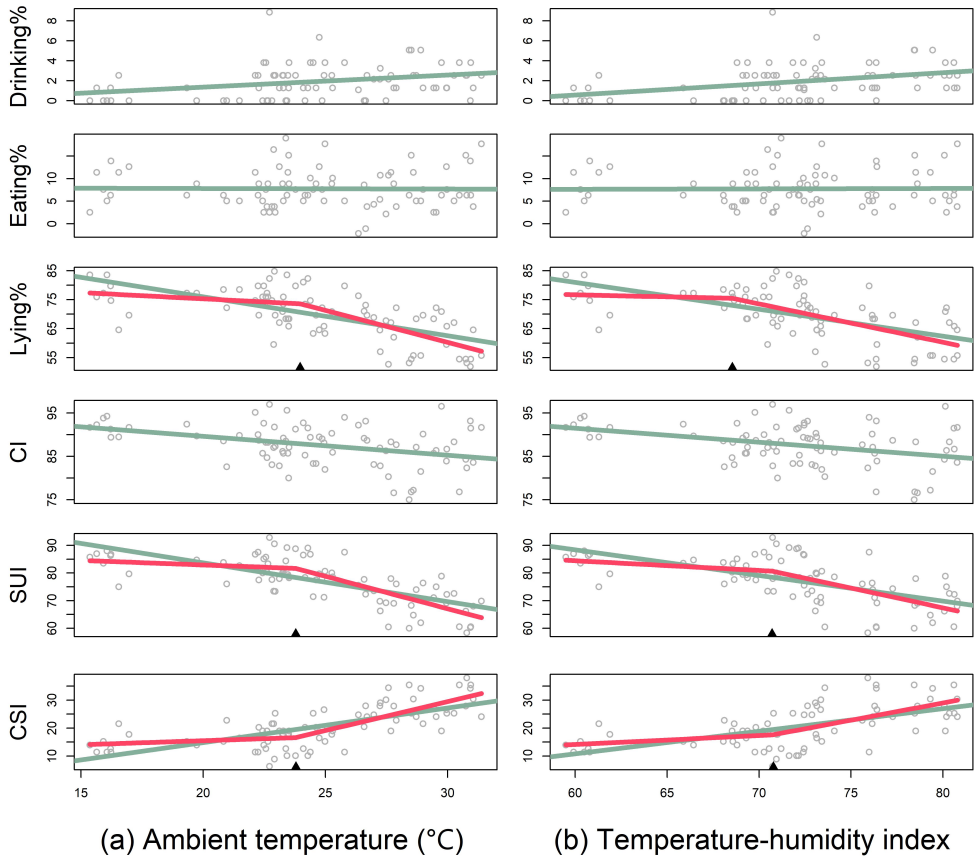


Figure 5- 9: Automated measurements of herd-level behavioural indicators and their fitted profiles from linear regression (green) and piecewise regression (red) models with (a) ambient temperature and (b) temperature-humidity index as the predictor, respectively. Breakpoints are marked as a black triangle above the x-axis. Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index.

Several studies have compared cow behaviour in different environmental classes (e.g., THI classes), and the upper limit of the class at which significance occurred

was stated as a critical point. For example, a THI of 68 was determined since the THI class <68 had a significantly lower percentage of cows standing compared with other predetermined THI classes [320]. However, this method may be arbitrary and can lose information. Moreover, piecewise or segmented models, are used for developing critical thresholds due to their indicative parameters (i.e., slope and breakpoint). For example, Heinicke et al. [160] determined a THI threshold (67) for total lying/standing time, number of lying/standing bouts, and lying bout duration. Our results, however, cannot be compared directly with theirs since THI and behavioural indicators were summarised as daily averages in their study.

Table 5- 4: Parameter estimates (mean ± standard error) of basic piecewise regression models with ambient temperature (Ta, °C) and temperature-humidity index (THI) as the predictor, respectively. Behavioural indicators were measured half-hourly by the proposed automated method from 11:00 to 14:30 h during the 10-day experiment. [a]

Predictor	Outcome	Intercept	Breakpoint	Left slope	ΔSlope	AIC	
						Linear	Piecewise
Ta	Drinking%	-1.1 ± 1.1	N/A	0.12 ± 0.04	N/A	309.6	N/A
	Eating%	8.1 ± 2.9	N/A	-0.01 ± 0.12	N/A	464.9	N/A
	Lying%	84.0 ± 7.9	24.0 ± 1.3	-0.43 ± 0.37	-2.21 ± 0.42	534.0	527.8
	CI	98.4 ± 3.0	N/A	-0.44 ± 0.12	N/A	471.3	N/A
	SUI	89.5 ± 6.7	23.8 ± 1.0	-0.33 ± 0.32	-2.35 ± 0.36	516.0	502.2
	CSI	9.7 ± 6.2	23.8 ± 1.0	0.29 ± 0.29	2.07 ± 0.31	499.3	486.4
THI	Drinking%	-6.1 ± 2.4	N/A	0.11 ± 0.03	N/A	306.1	N/A
	Eating%	7.1 ± 6.5	N/A	0.01 ± 0.09	N/A	464.9	N/A
	Lying%	85.2 ± 35.6	68.5 ± 3.3	-0.14 ± 0.56	-1.32 ± 0.23	544.4	542.4
	CI	110.8 ± 6.8	N/A	-0.32 ± 0.09	N/A	472.5	N/A
	SUI	105.5 ± 20.8	70.7 ± 2.1	-0.35 ± 0.31	-1.43 ± 0.27	534.0	530.9
	CSI	-5.1 ± 18.9	70.7 ± 2.3	0.32 ± 0.28	1.23 ± 0.24	517.9	515.8

ΔSlope = difference between right and left slopes; AIC = Akaike information criterion; Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; CI = comfort index; SUI = stall-use index; CSI = cow stress index.

[a] N/A indicates that piecewise regression failed to converge. Intercept and left slope, in this case, represent the parameters of simple linear regressions.

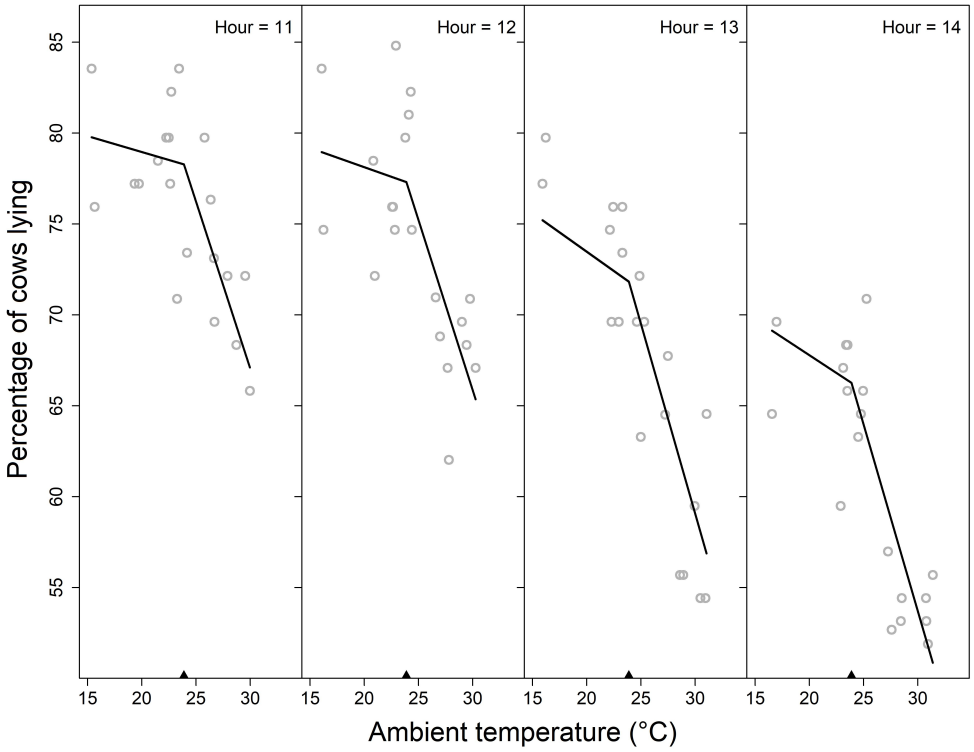


Figure 5- 10: Automated measurements of the percentage of cows lying and their fitted profile from advanced piecewise regression with ambient temperature as the predictor. Random effects of time of day (h) were introduced for intercept, slope difference, and breakpoint. Breakpoints are marked as a black triangle above the x-axis.

The advanced piecewise models with THI as the predictor all failed to converge. The profiles of the advanced piecewise models with Ta as the predictor and lying%, SUI, and CSI as the outcomes are shown in **Figure 5-10**, **Figure 5-11**, and **Figure 5-12**. For each behavioural indicator, Ta breakpoints were roughly the same at different times of day, with differences only being observed since the ten thousandth place. The rounded Ta breakpoints among times of day were 23.88, 23.70, and 23.65 °C, for lying%, SUI, and CSI, respectively. The behavioural pattern can be found on the Y-axis with standing increasing and lying decreasing from 11:00 to 14:30 h. According to these findings, accumulated heat load over the observed period of time did not make cows more sensitive to heat stress at a particular time point. In a recent chamber study, the critical Ta threshold of respiration rate was found to be lower in the afternoon than in the morning, indicating that cows were more sensitive after longer exposure to heat stress [338]. Anyway, more data is required to confirm our results since the subgroup sample size used in the analysis might not support a precise localisation of breakpoints. To sum up, any attempt to

integrate behavioural indicators for heat stress evaluation should carefully consider the temporal pattern.

3.4. Strength and limitations

To the best of our knowledge, this is the first study to determine heat stress based on herd-level behavioural recognition. Although individual measurements can help identify animals with the greatest risks and thereby customise heat abatement, this can be costly in free-stall barns since individual measurement and abatement require a lot of improvement on the existing facilities. In many cases, even if individual measurements have been done, the data have to be summarised to reflect the herd mean [76]. Indeed, as long as heat abatement is implemented at the herd level, information on the individual level is not necessary for decision making [339].

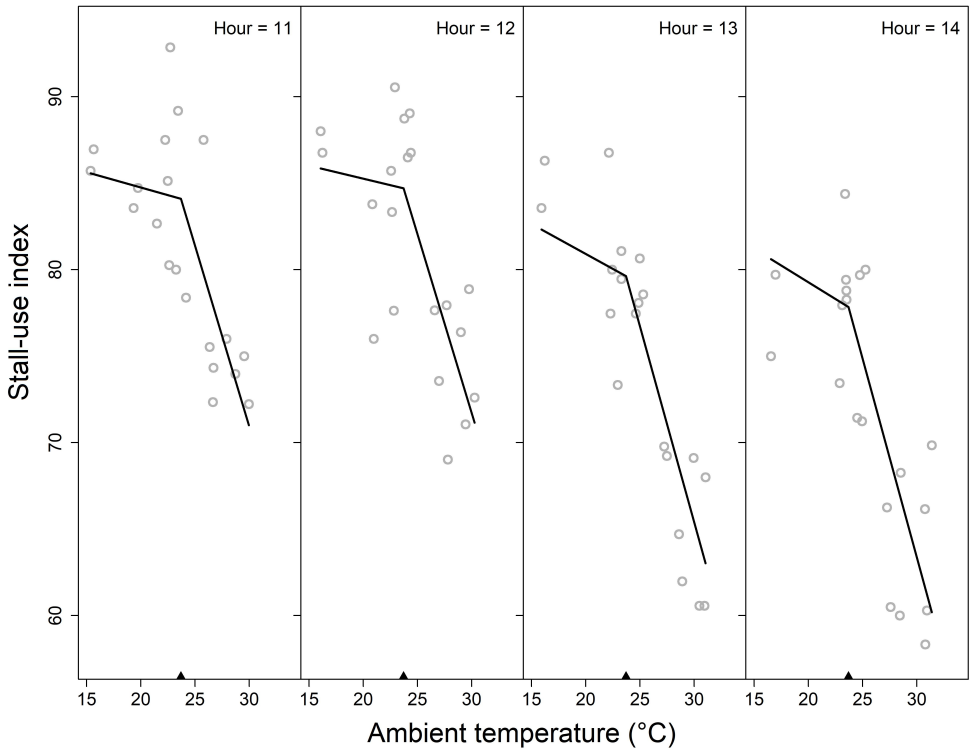


Figure 5- 11: Automated measurements of stall-use index and their fitted profile from advanced piecewise regression with ambient temperature as the predictor. Random effects of time of day (h) were introduced for intercept, slope difference, and breakpoint. Breakpoints are marked as a black triangle above the x-axis.

With the help of computer vision and scan sampling, our work offers a low-cost herd-level heat stress alert without imposing any burden on dairy cows. Besides, the effectiveness of heat abatement can be tracked by introducing more flexible scan sampling or even continuous measurement when necessary. It should be noted that

cows would take advantage of nights when the temperature is thermally comfortable to relieve their accumulated heat load throughout the daytime [320]. However, such nighttime behavioural data is missing in this study. Future works with reliable nighttime recording and transfer learning techniques are required to develop a behavioural recognition model that can work day and night as well as to customise heat stress thresholds for nighttime hours. By doing so, it is promising to develop a comprehensive protocol for heat stress detection, mitigation, and evaluation.

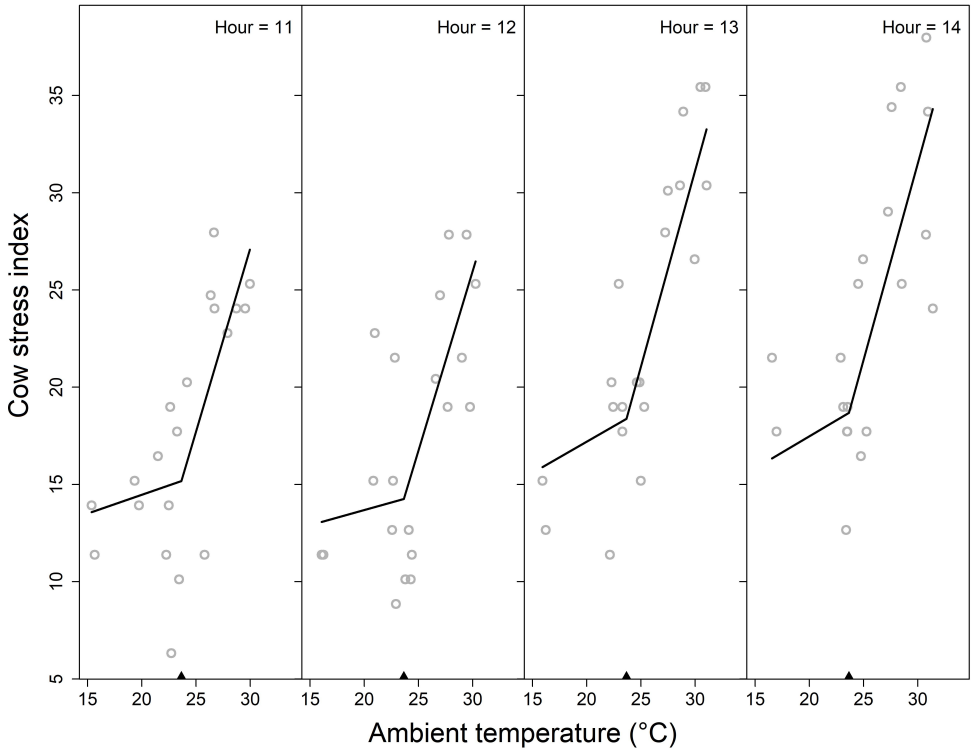


Figure 5- 12: Automated measurements of cow stress index and their fitted profile from advanced piecewise regression with ambient temperature as the predictor. Random effects of time of day (h) were introduced for intercept, slope difference, and breakpoint. Breakpoints are marked as a black triangle above the x-axis.

Our method is best suited for intensive farms where facilities and animals are highly standardised. From the facility perspective, our method for camera mounting can be directly applied in similar settings, but further evaluation in other designs (e.g., 4-row tail-to-tail) is required. In addition, our solution requires fewer cameras than the top-view method [163], because it includes more cattle in each view. Top-view cameras, as stated before, can also be supplemented in specific areas (e.g., troughs) to provide more reliable recordings. From the animal perspective, cows raised on intensive farms are typically grouped by common influencing factors (e.g.,

productivity, lactation stage, and parity) and thus behave relatively homogeneously against heat stress. This will help to shrink within-herd variation, allowing abatement measures based on herd means to be useful for the majority of cows. This also allows precision management by customising herd-level heat stress thresholds with respect to different levels of heat sensitivity. Of note, any extrapolation and interpretation of the determined critical thresholds should carefully consider the impact of different management or facility conditions on cow behaviour, such as overstocking, milking frequency, and bedding materials [340,341].

4. Conclusions

This study has proposed a YOLOv5-based method for recognising cow behaviour with excellent mAP and inference speed. The ability of the proposed model in measuring herd-level behavioural indicators has been validated in comparison to manual observation. The automated measurements taken during the 10-day experiment reveal that lying and standing indices (i.e., lying%, SUI, and CSI) were most responding to heat stress and the test dairy herd began to change their behaviour at the earliest Ta of 23.8 ℃ or THI of 68.5. Collectively, the model and results presented in this paper have achieved a low-cost heat stress alert for the study herd without imposing any burden on dairy cows. Further study using multiple herds with varying characteristics is promising to customise herd-level heat stress thresholds.

Context – Chapter 6

According to the findings reported in the first chapter, data-driven predictive models can not only offer a non-invasive way to obtain heat stress-related data but also provide insights for decision-making through their interpretations. Previous studies used limited variables as predictors of physiological responses, and the developed models poorly predict animal responses in evaporatively cooled environments. Therefore, the sixth chapter aimed to build machine learning models using comprehensive variables to predict the physiological responses of dairy cows raised on an actual dairy farm equipped with sprinklers.

Chapter 6

Predicting physiological responses of dairy cows using machine learning

Adapted from:

Shu, H., Li, Y., Bindelle, J., Jin, Z., Fang, T., Xing, M., Guo, L., Wang, W. *Predicting physiological responses of dairy cows using comprehensive variables*. Computers and Electronics in Agriculture, 2023. https://doi.org/10.1016/j.compag.2023.107752

Abstract

Heat stress is increasingly affecting the production, health, and reproduction of dairy cows. Previous studies used limited variables as predictors of physiological responses, and the developed models poorly predict animal responses in evaporatively cooled environments. The aim of this study was to build machine learning models using comprehensive variables to predict the physiological responses of dairy cows raised on an actual dairy farm equipped with sprinklers. Four algorithms including random forests, gradient boosting machines, artificial neural networks (ANN), and regularised linear regression were used to predict respiration rate (RR), vaginal temperature (VT), and eye temperature (ET) with 13 predictor variables from three dimensions: production, cow-related, and environmental factors. The classification performance of the predicted values in recognising individual heat stress states was compared with commonly used thermal indices. The performance on the testing sets shows that the ANN models yielded the lowest root mean square error for predicting RR (13.24 breaths per min), VT (0.30 ℃), and ET (0.29 ℃). The results interpreted with partial dependence plots and Local Interpretable Model-agnostic Explanations show that P.M. measurements and winter calving contributed most to high RR and VT predictions, whereas lying posture, high ambient temperature, and low wind speed contributed most to high ET predictions. When determining the ground-truth heat stress state by the actual RR, the best classification performance was yielded by the predicted RR with an accuracy of 77.7%; when determining the ground-truth heat stress state by the actual VT, the best classification performance was yielded by the predicted VT with an accuracy of 75.3%. This study demonstrates the ability of ANN in predicting the physiological responses of dairy cows raised on actual farms with access to sprinklers. Adding more predictors other than meteorological parameters into training could increase predictive performance. Recognising the heat stress state of individual animals, especially those at the highest risk, based on the predicted physiological responses and their interpretations can inform better heat abatement decisions.

Keywords: precision livestock farming, animal welfare, predictive modelling, thermal comfort, interpretability

1. Introduction

In the dairy industry, heat stress is increasingly affecting the production, health, and reproduction of dairy cows [342]. Thermal indices have long been developed and applied to describe the amplitude and duration of heat stress [235]. However, environmental indicators can neither reflect the true response of animals nor address individual variation within a herd [46,130].

Heat stress induces acute responses which are driven by the autonomic nervous system to maintain homeostasis, as well as chronic responses which are driven by the endocrine system to achieve a new physiological state [55]. Physiological responses, such as respiration rate (RR), core body temperatures (CBT), and body surface temperatures (BST), can be obtained by direct measurement or predictive modelling, and further used for determining the heat stress state of animals. Although the direct measurement of such responses can provide the most accurate results, it is difficult to achieve continuous alarms since traditional manual measurements are invasive, tedious, and time-consuming [308]. Predictive modelling offers a non-invasive alternative to predict physiological responses from more easily accessible data such as meteorological parameters [48,210].

Machine learning (ML) models have gained much interest in animal science research due to their advantage in predicting nonlinear relationships and being less subject to assumptions about data distribution [178]. Tree-based algorithms and neural networks are two typical methods that have been extensively used in regression (e.g., prediction of productivity, energy consumption, physiological state) and classification (e.g., behaviour recognition, disease detection, body condition scoring) tasks [343,344]. For such a regression task predicting physiological responses, random forests (RF), gradient boosting machines (GBM), and artificial neural networks (ANN) have shown much better predictive ability than traditional linear models in broilers [345], dairy cows [179,181], beef cattle [180,346], sheep [307], and pigs [248].

The choice of predictors is of particular importance for the predictive ability of the model. Many studies relied solely on meteorological parameters to predict physiological responses [178,181,347]. Some studies used subcutaneous temperatures [80,81] or BST [179,346] as predictors since vasodilation during heat stress drives more blood from the core to the periphery. Li et al. [173] incorporated previous milk yield and time block into their predictive model, reflecting production level and diurnal changes in cow physiology, respectively. Moreover, lots of cow-related factors are well documented to have an impact on cows' susceptibility to heat stress, including age, breed, lactation stage, parity, and body posture [5,68,105]. Accordingly, inputting cow-related factors is supposed to better deal with individual variation in heat stress predictions [348]. Days in milk (DIM) and parity have been incorporated into ML models for predicting milk productivity and quality [253,349]. However, few attempts have been made to incorporate cow-related factors into ML models for predicting physiological responses of dairy cows exposed to heat stress.

The other concern about using ML methods is that the models can only be applied and interpreted in the environment similar to where they were originally developed due to their data-driven nature. Dairy farms employ a variety of cooling strategies to alleviate heat stress, the most efficient and widely used of which is electric fans coupled with evaporative cooling (e.g., misters and sprinklers) [350]. Fans plus misters systems produce small droplets and cool the air through evaporation as they disperse, whereas fans plus sprinklers systems produce much larger droplets to wet the skin surface of cows and cool the surface directly through evaporation of the water [351,352]. All these microenvironmental changes can be captured by ambient sensors and their effect on cow thermal comfort can still be explained. However, the fact that the large droplets delivered by sprinklers wet the surface of cows making them more efficiently cooled by fans would be neglected by previous models which were developed in the absence of sprinklers. The lack of these data makes it unknown whether ML methods would remain useful in such a complex environment in order to guide decision making with regards to cow management.

To explore the abovementioned questions, this study aimed to build and compare ML models for predicting physiological responses (RR, CBT, and BST) of dairy cows from previous milk yield, cow-related factors, microenvironmental parameters, and time block on an actual farm which was equipped with sprinklers. We hypothesised that ML models developed using comprehensive variables would perform well in this real-world environment.

2. Materials and methods

The experimental protocols were approved by the Experimental Animal Care and Use Committee of Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (approval number IAS2021-220).

2.1. Location, facilities, and animals

The experiment was conducted from May to August 2021 at an intensive dairy farm, located in Shandong, China (coordinates: 34◦50’37”N, 115◦26’11”E; altitude: 52 m), characterised by a temperate continental monsoon climate with hot and humid summers. The experiment was conducted over three different phases, firstly during 20 days in late spring and early summer (May to June), secondly during 10 days in mid-summer (July), and thirdly during 10 days in late summer (August). These phases were expected to cover a wide range of thermal environments, thus facilitating the training of ML models.

For each experimental phase, a new group of 19 to 20 primi- and multiparous Holstein dairy cows reared in a free-stall pen (11 m × 96 m) were selected based on similar parity, lactation stage, and body condition score [353], so that cows were comparable at the beginning of each phase (**Table 6-1**). The barn was covered by a double-pitched roof, and therefore, most of the solar radiation was prevented from reaching the cows inside the barn. The pen was equipped with a total of 22 fans (diameter: 1.1 m; capacity: 25000 m^3/h each; installation height: 2.8 m) and 46 sprinklers (flow rate: 1.5 L/min each; installation height: 2 m; 1 min on and 4 min

off). Fans and sprinklers were automatically turned on when the indoor temperature reached 20 °C and 25 °C, respectively. Cows were milked three times per day at 08:30, 16:30, and 00:00 h, and were fed a total mixed ration three times per day after milked. Cows had free access to drinking water but no access to outdoor pasture. One cow in phase 1 and another in phase 3 were withdrawn from the experiment due to health issues, namely high somatic cell count and gastroenteritis, respectively.

Table 6- 1: Summary of the cows at the beginning of each phase.

Variable	First phase (n = 20)		Second phase (n = 20)		Third phase (n = 19)	
	Mean ± SD	Min, Max	Mean ± SD	Min, Max	Mean ± SD	Min, Max
Parity	2.7 ± 0.9	1, 5	2.5 ± 1.1	1, 5	2.5 ± 1.2	1, 5
Days in milk	150.2 ± 21.1	109, 179	150.9 ± 16.9	113, 178	150.0 ± 15.9	119, 178
Body condition score[a]	3.0 ± 0.2	2.8, 3.5	3.1 ± 0.3	2.8, 3.5	3.1 ± 0.2	2.8, 3.5
Daily milk yield (kg/day)	43.0 ± 5.3	30.9, 52.9	39.3 ± 5.0	30.4, 50.4	37.4 ± 5.5	28.8, 50.4

SD = standard deviation; Min = minimum; Max= maximum.

[a] Body condition score was measured using a 1 (severe undercondition) to 5 (severe overcondition) scale as per Wildman et al. [353].

2.2. Variables

Respiration rate (RR), vaginal temperature (VT), and eye temperature (ET) were the three response variables for predictive modelling, while candidate predictor variables were determined along three dimensions: production factors, cow-related factors, and environmental factors (**Figure 6-1**).

Production factors included the single daily milk yield (DMY) of three days before the test day. The use of DMY was initially motivated by the fact that high-producing cows typically suffer more from heat stress [50]. The milk yield in the last few days could represent the mean production level of individual cows. Besides, the changing dynamics or accumulated response [53] of DMY over previous days was intended to show how the animals were coordinated by acclimatisation. This information is important since exposure to heat stress would induce stress responses in either acute (minutes to days) or chronic (days to weeks) ways, manifesting into different physiological states [55]. For example, cows showed a much steeper increase in RR and VT during acute stress than during chronic stress [354]. Given that acute stress takes at least three days to achieve thermal balance [355], the previous three days' DMY was finally nominated.

Cow-related factors including birth season, calving season, DIM, parity, age in months (AIM), and posture were nominated since they have long been identified as influencing factors of individual sensitivity to heat stress [5,68,105]. Environmental factors included ambient temperature (Ta), relative humidity (RH), and wind speed

(WS) which represented indoor microenvironments, as well as time block which represented time effect.

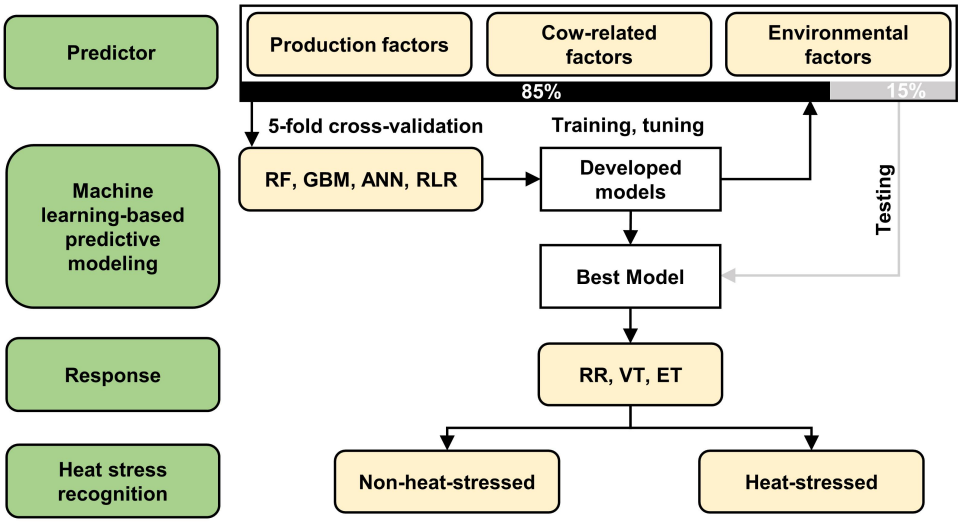


Figure 6- 1: Flow chart showing the strategic plan of the present study. RF = random forests; GBM = gradient boosting machines; ANN = artificial neural networks; RLR = regularised linear regression; RR = respiration rate; VT = vaginal temperature; ET = eye temperature.

2.3. Data collection

Vaginal temperature (VT) was recorded automatically at an interval of 5 min by using data loggers (DS1922L, accuracy: ± 0.5 °C, resolution: ± 0.0625 °C; Maxim Integrated, San Jose, CA, USA) attached to modified vaginal controlled internal drug releases (Pfizer Animal Health, New York, NY, USA). The devices were removed after a week in vivo for each cow per phase to avoid interfering with artificial insemination and risking harming the foetuses. Meteorological parameters were measured automatically at an interval of 10 min by using six Kestrel environmental data loggers (model: 5000 and 5400; accuracy: ± 0.4 °C Ta, ± 1% RH, ± 1.66% WS; Nielsen-Kellerman, Boothwyn, PA, USA) which were evenly placed in the pen at a height of 2.2 m. These readings were used for describing the overall thermal condition throughout the entire experiment.

The manual field measurement of physiological (RR and ET) and microenvironmental variables (Ta, RH, and WS) was conducted twice on each test day, once during A.M. (08:00-11:30 h) and once during P.M. (13:30-16:30 h). For each measurement, each cow was expected to be measured twice, once while lying and once while standing, with at least a 30-min gap between the two observations. However, due to their unconstrained nature, the number of times the cows were measured varied for each measurement, with mean ± standard deviation observations of 2.0 ± 0.9, 2.4 ± 0.9, and 2.3 ± 0.8 per cow per measurement for three

periods, respectively. For each observation, RR was recorded by two trained observers (intra-class correlation coefficient: 0.91) by timing 15 flank movements (and converting to breaths per min); ET was measured from the cows' side with an angle of approximately 90° and a distance of approximately 1.5 m by a photographer using a portable infrared camera (VarioCAM HR, accuracy: ± 1.5 °C, resolution: 640 × 480 pixels; InfraTec, Dresden, Germany) which was fully warmed up as per Howell et al. [356]; and microenvironmental parameters (i.e., Ta, RH, and WS) were manually collected from the closest Kestrel data logger.

Previous DMY (kg/day) and cow-related factors including birth season, calving season, DIM, parity, and AIM were acquired from the automatic milking system (Afimilk, Kibbutz Afikim, Israel). Birth season and calving season were coded to spring (March to May), summer (June to August), autumn (September to November), and winter (December to February). Body posture (lying or standing) was recorded manually for every observation. Time block (A.M. or P.M.) was recorded for two separate field measurements on each test day.

2.4. Data processing

The infrared images were interpreted using IRBIS 3 Standard software (YSHY, Beijing, China). Low-quality images, specifically those that were blurred or had closed eyes, were manually removed. ET was determined using the maximum temperature of the medial canthus area, as per Shu et al. [357]. The data of the two sick cows on the day they were withdrawn from the experiment were removed from the dataset for data quality control.

Further data processing was done using R software (version 4.1.0; https://www.R-project.org/). To ensure a high-quality prediction, it is important that the observations used for training and testing are included in the same distribution and are not subject to outliers. This can be done by calculating Mahalanobis distance which is an effective distance metric that measures the distance between an observation and the barycenter defined in the multi-dimensional space [358]. Thus, a principal component analysis was first performed using the *PCA* function from the *FactoMineR* package to reduce the dimensionality of the predictor matrix. The Mahalanobis distances between observations were then calculated, referring to the method described in Soyeurt et al. [359]. The presence of outliers among observations was detected using a Mahalanobis distance threshold of 5 as convention. Moreover, although multicollinearity does not typically affect accuracy, it can be a problem when interpreting the results, and thus should be carefully dealt with. Multicollinearity was detected using variance inflation factors with a threshold of 5 [360]. This was done by first building a regression model that included all predictor variables using the *lm* function, and then applying the *vif* function of the *car* package to the regression model. Besides, correlation matrices were built to visualise the correlations among 13 candidate predictor variables using the *cor* function.

2.5. Predictive modelling

Predictive modelling was performed using the *h2o* package. For each response variable, the *h2o.splitFrame* function was used to randomly divide 85% of the data as the training set and 15% as the testing set (**Figure 6-1**). In addition, a cow-level splitting was performed as a control where cows were randomly split at a rough ratio of 85:15 and their observations were allocated to training and testing sets, respectively.

The training set was used to fit the model and the testing set was used to collect the final performance. Moreover, 5-fold cross-validation was performed to enhance the model reliability and avoid issues with 'lucky' data split. Four ML algorithms, including RF, GBM, ANN, and linear regression with elastic net regularisation, were used for modelling (**Figure 6-1**). The reason for choosing these algorithms was that they are typical methods in such a regression task and are easily accessible from popular software. The grid search was performed to identify the best combination of hyperparameters using the *h2o.grid* function. For RF, GBM, and ANN, a random grid search was performed with the parameter *max_models* setting to 2000. For regularised linear regression, a cartesian grid search was performed due to much fewer options of hyperparameters. For each algorithm, the model with the lowest cross-validation root mean square error (RMSE) was selected as the best performing model. These selected models were further evaluated for their performance on the testing set, and the one with the lowest RMSE and the highest coefficient of determination (R^2) was selected as the overall best model.

To interpret and visualise the results, partial dependence plots which are available in the *h2o* package were built to understand how the response variables changed with the predictor variables. Understanding which variables are most influential and how different levels of variables affect animal response to heat stress is important to make an accurate prediction at the individual level. The state-of-the-art post-hoc local interpretability technique, Local Interpretable Model-agnostic Explanations (LIME), was performed using the *lime* package to gain further insight into individual predictions. The top five influential predictor variables that best explained the linear model were used for plotting LIME heatmaps.

2.5.1. Random forests

Random forests (RF) is a powerful tree-based algorithm that is commonly used in both classification and regression tasks [361]. RF is a kind of ensemble algorithm, adopting the method of bagging in which decision trees are trained with replacement sampling and the mean prediction of all trees is the output. The hyperparameters randomly searched included the number of trees (10 to 250, increased by 10), the maximum tree depth (10 to 100, increased by 10), the number of variables to consider at each split (3 to 13, increased by 1), and the minimum number of observations for a leaf (1, 2, 10, 20, and 30).

2.5.2. Gradient boosting machines

Gradient Boosting Machines (GBM) is another tree-based ensemble method [362]. Unlike RF, GBM uses the method of boosting in which the weights of all training

samples are adjusted according to the residual gradient so that the next base learner pays more attention to the wrongly classified samples. The hyperparameters randomly searched included the number of trees (10 to 250, increased by 50), the maximum tree depth (10 to 100, increased by 10), the minimum number of observations for a leaf (1, 2, 10, 20, and 30), and the learning rate (0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, and 0.5).

2.5.3. Artificial neural networks

Artificial neural networks (ANN) can fit arbitrary nonlinear functions through reasonable network architecture configuration [363]. In the present study, the feedforward ANN model with a multi-layer architecture was trained with stochastic gradient descent using back-propagation. The hyperparameters randomly searched included the activation function (ReLU or Tanh), the number of hidden layers (1, 2, and 3) with the number of neurons (20, 50, 100, and 200) in each hidden layer, the dropout rate (0 to 0.5, increased by 0.05), and the epochs (5 to 500, increased by 5).

2.5.4. Regularised linear regression

Multiple linear regression can suffer from multicollinearity and overfitting, especially on small datasets. Hence, several regularisation methods have been introduced to deal with these problems by shrinking the regression coefficients toward zero. The regularisation method used in this study was elastic net which combines L1 (LASSO) and L2 (RIDGE) regularisation [364]. The hyperparameters were α (0 to 1, increased by 0.01) and λ (searched automatically by setting the parameter *lambda_search* to 'TRUE') in which α controls the weights of L1 and L2 regularisation, and λ controls the strength of regularisation.

2.6. Recognition of heat stress state

The predicted values of RR and VT making use of the revised thresholds for high-producing (> 35kg/day) dairy cows (RR of 60 breaths per min and VT of 38.5 ℃) [50] were further tested for their ability to serve as classifiers for recognising the cows' heat stress state (**Figure 6-1**). The predicted ET was not considered here due to the lack of commonly recognised threshold. These proposed classifiers were compared with the most commonly used temperature-humidity index (THI) classifiers: 68 [50], 70 [365], 72 [366]; adjusted temperature-humidity index (THIadj) classifier: 74 [146]; and more recent equivalent temperature index for cattle (ETIC) classifier: 23 [367]. The THI was calculated according to Eq. (1) as recommended by National Research Council [222]:

$$THI = (1.8 \times Ta + 32) - (0.55 - 0.0055 \times RH) \times (1.8 \times Ta - 26) \quad (1)$$

THIadj and ETIC were calculated according to Eq. (2) and Eq. (3), respectively:

$$THIadj = \left[0.8 \times Ta + \frac{RH}{100} \times (Ta - 14.4) + 46.4\right] + 4.51 - 1.992 \times \mathrm{WS} + 0.0068 \times \mathrm{SR} \quad (2)$$

$$ETIC = Ta - 0.0038 \times Ta \times (100 - \mathrm{RH}) - 0.1173 \times WS^{0.707} \times (39.2 - \mathrm{Ta}) + 1.86 \times 10^{-4} \times Ta \times SR) \quad (3)$$

where SR represents solar radiation and was set to zero as per the original authors' instructions for use in indoor situations.

The ground truth of individual heat stress states (heat-stressed or non-heat-stressed) was determined by using measured values (RR and VT) and the corresponding thresholds (60 breaths per min and 38.5 °C), respectively. The classification performance was evaluated using four metrics: recall, precision, F1-score, and accuracy. Recall measures how many cows that are truly heat-stressed can be correctly classified as being heat-stressed whereas precision measures how many cows that are classified as being heat-stressed are truly heat-stressed. F1-score is a comprehensive measure that strikes a balance between recall and precision, while accuracy indicates the overall rate of correctly classified cows. The equations are as follows:

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$F1-score = \frac{2TP}{2TP+FP+FN} \times 100\% \quad (6)$$

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (7)$$

where TP denotes true positive (heat-stressed cows correctly classified as heat-stressed cows), FP denotes false positive (non-heat-stressed cows incorrectly classified as heat-stressed cows), TN denotes true negative (non-heat-stressed cows correctly classified as non-heat-stressed cows), and FN denotes false negative (heat-stressed cows incorrectly classified as non-heat-stressed cows).

3. Results and discussion

The daily patterns of meteorological variables during the entire experimental period are shown in **Figure 6-2**. Ta and RH had opposite trajectories, while THI began to increase in May and remained stably high from June to August. Additionally, all three variables showed a shrinking diurnal change from May to early August. These facts indicate an increased intensity and duration of daily exposure to heat stress during the experimental phases. In addition to acute responses, heat stress is well documented to induce chronic responses in dairy cows, even during temperature drops in Autumn [227]. Our results indicate that this experiment well covered the onset and development of heat stress. Thus, the induced physiological responses in dairy cows including both acute and chronic responses were expected to be well collected. This heterogeneous data would definitively contribute to better training of ML algorithms in terms of the non-linear response of dairy cows to heat stress.

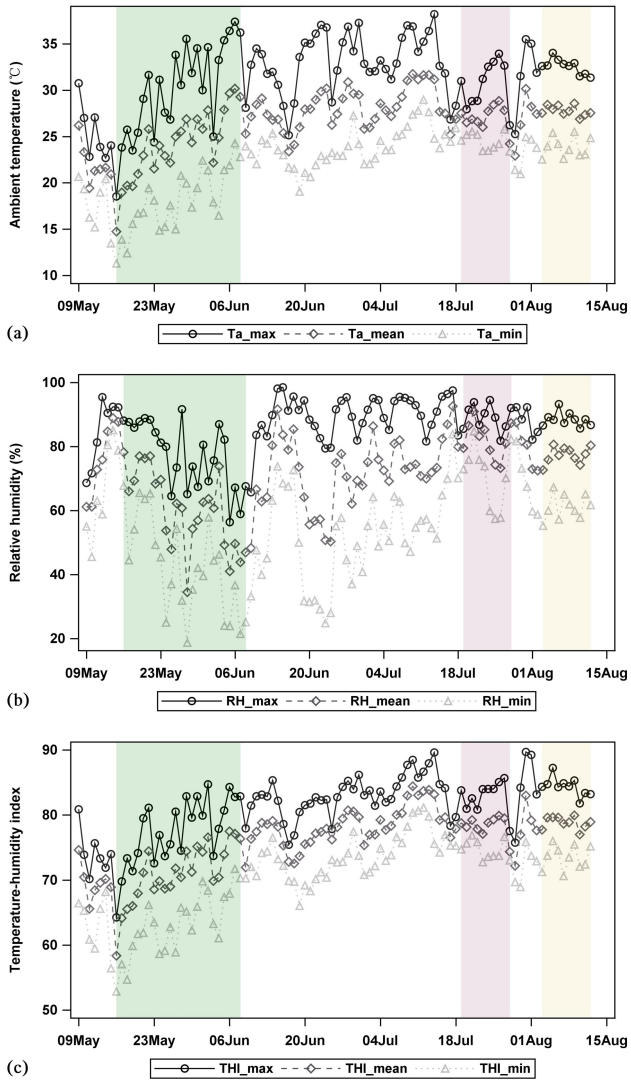


Figure 6- 2: Daily patterns of (a) ambient temperature (Ta), (b) relative humidity (RH), and (c) temperature-humidity index (THI). Zones in green, purple, and yellow represent three experimental phases, respectively. Ta_mean, Ta_max, and Ta_min are the mean, maximum, and minimum of daily Ta, respectively; RH_mean, RH_max, and RH_min are the mean, maximum, and minimum of daily RH, respectively; THI_mean, THI_max, and THI_min are the mean, maximum, and minimum for daily THI, respectively.

3.1. Data cleaning and descriptive statistics

After removing outliers, a total of 2910, 1561, and 1866 observations were obtained for the datasets modelling RR, VT, and ET, respectively. The variance inflation factors among the 13 candidate predictor variables in RR, VT, and ET sets

were all below 5. The correlation matrices (see **Appendix 1. Supplementary material Figure A6-1**, **Figure A6-2**, and **Figure A6-3**) show that none of the correlations between candidate predictor variables was higher than 0.75. These findings support that multicollinearity was not present in the datasets, and therefore, all candidate variables were used for modelling. The descriptive statistics of the three datasets are summarised in **Table 6-2**.

Table 6- 2: Summary of the datasets for predicting respiration rate (RR), vaginal temperature (VT), and eye temperature (ET). Continuous variables are summarised as mean ± standard deviation, categorical variables are summarised as n (%).

Predictor	RR set (n = 2910)	VT set (n = 1561)	ET set (n = 1866)
DMY1D (kg/day)	39.3 ± 7.0	38.4 ± 6.8	39.1 ± 7.1
DMY2D (kg/day)	39.5 ± 7.1	38.3 ± 6.8	39.2 ± 7.2
DMY3D (kg/day)	39.5 ± 7.1	38.5 ± 7.0	39.2 ± 7.1
Birth season			
Spring	1411 (48.5)	883 (56.6)	926 (49.6)
Summer	377 (13.0)	150 (9.6)	213 (11.4)
Autumn	287 (9.9)	130 (8.3)	183 (9.8)
Winter	835 (28.7)	398 (25.5)	544 (29.2)
Calving season			
Spring	273 (9.4)	219 (14.0)	186 (10.0)
Autumn	514 (17.7)	122 (7.8)	293 (15.7)
Winter	2123 (73.0)	1220 (78.2)	1387 (74.3)
Days in milk	168.0 ± 20.1	167.4 ± 18.9	168.4 ± 19.7
Parity			
1	629 (21.6)	349 (22.4)	389 (20.8)
2	960 (33.0)	527 (33.8)	619 (33.2)
3	814 (28.0)	468 (30.0)	525 (28.1)
4	402 (13.8)	193 (12.4)	271 (14.5)
5	105 (3.6)	24 (1.5)	62 (3.3)
Age in months	48.4 ± 12.7	46.7 ± 13.1	48.4 ± 13.1
Posture			
Standing	1503 (51.6)	786 (50.4)	1103 (59.1)
Lying	1407 (48.4)	775 (49.6)	763 (40.9)
Ambient temperature (°C)	28.9 ± 4.0	28.5 ± 3.9	29.6 ± 3.4
Relative humidity (%)	61.4 ± 17.8	68.1 ± 10.1	61.0 ± 17.6
Wind speed (m/s)	1.3 ± 0.9	1.3 ± 0.9	1.3 ± 0.9
Time block			
A.M.	1417 (48.7)	781 (50.0)	840 (45.0)
P.M.	1493 (51.3)	780 (50.0)	1026 (55.0)

DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day).

3.2. Model performance

The predictive performance of the four candidate algorithms on the testing sets is shown in **Table 6-3**. ANN always performed the best on the testing set with the lowest RMSE of 13.24 breaths per min, 0.30 ℃, and 0.29 ℃, and the highest R^2 of 0.56, 0.45, and 0.45 when predicting RR, VT, and ET, respectively. Besides, ANN had similar results between the training and testing sets, indicating no overfitting occurred. Although GBM had roughly good results on the testing sets, an obvious decrease was always observed between the training set and cross-validation, as well as between the training and testing sets, suggesting the occurrence of overfitting. GBM reportedly has a higher potential for overfitting compared with RF, especially on small datasets [368]. Collectively, our results suggest that ANN is more appropriate to predict physiological responses of dairy cows managed with sprinklers. The linear regressions between measured and predicted values are shown in **Figure 6-3**. In all cases, the regression line between measured and predicted values was close to the regressed diagonal line, indicating a good correlation between predictions and actuals. The data points from different experimental phases show no obvious differential distribution around the diagonal. This is confirmed by the partial results for the three different experimental phases (**Table 6-4**), which show similar performance relative to the overall results (**Table 6-3**).

Table 6- 3: Performance of four candidate algorithms in predicting respiration rate (RR, breaths per min), vaginal temperature (VT, °C), and eye temperature (ET, °C) on the training sets, 5-fold cross-validation, and testing sets.

Response	Algorithm	Training		Cross-validation (SD)		Testing	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
RR	RF	14.59	0.43	14.54 (0.58)	0.44 (0.03)	14.36	0.45
	GBM	11.99	0.62	14.40 (0.49)	0.45 (0.02)	13.34	0.55
	ANN	12.86	0.57	13.26 (0.49)	0.55 (0.02)	13.24	0.56
	RLR	15.79	0.35	15.89 (0.57)	0.34 (0.04)	15.42	0.40
VT	RF	0.35	0.26	0.35 (0.03)	0.26 (0.03)	0.31	0.43
	GBM	0.26	0.59	0.35 (0.02)	0.25 (0.06)	0.31	0.44
	ANN	0.31	0.43	0.35 (0.01)	0.42 (0.05)	0.30	0.45
	RLR	0.36	0.21	0.36 (0.01)	0.19 (0.05)	0.36	0.22
ET	RF	0.28	0.46	0.33 (0.02)	0.44 (0.02)	0.34	0.42
	GBM	0.25	0.68	0.33 (0.02)	0.43 (0.03)	0.31	0.44
	ANN	0.29	0.58	0.33 (0.01)	0.44 (0.03)	0.29	0.45
	RLR	0.35	0.38	0.35 (0.02)	0.36 (0.01)	0.31	0.33

RF = random forests; GBM = gradient boosting machines; ANN = artificial neural networks; RLR = regularised linear regression; SD = standard deviation; RMSE = root mean square error; R^2 = coefficient of determination.

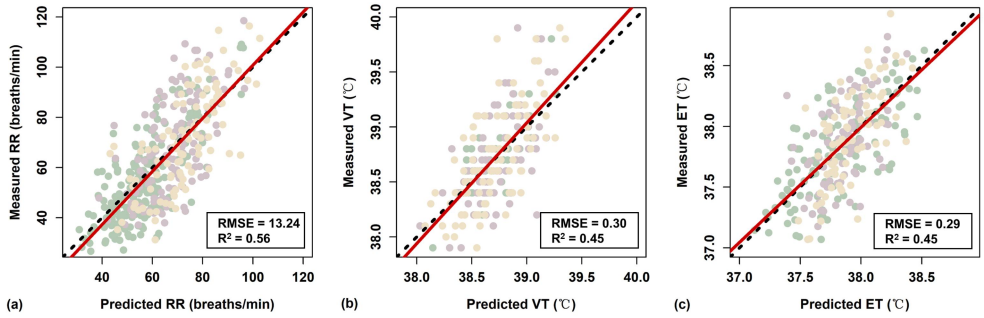


Figure 6- 3: Measured and predicted (a) respiration rate (RR), (b) vaginal temperature (VT), and (c) eye temperature (ET) from the overall best models (artificial neural networks) on the testing sets. The data points in green, purple, and yellow belong to three experimental phases, respectively. The red lines represent the linear regression. The dotted lines represent the diagonal line with a slope of 1. RMSE = root mean square error; R^2 = coefficient of determination.

Table 6- 4: Performance of the overall best models (artificial neural networks) in predicting respiration rate (RR, breaths per min), vaginal temperature (VT, °C), and eye temperature (ET, °C) on the testing sets summarised by three experimental phases.

Response	Phase 1			Phase 2			Phase 3		
	n	RMSE	R^2	n	RMSE	R^2	n	RMSE	R^2
RR	214	12.09	0.58	113	12.82	0.53	99	13.75	0.52
VT	54	0.32	0.53	85	0.33	0.47	96	0.32	0.49
ET	125	0.27	0.51	82	0.30	0.42	81	0.28	0.44

RMSE = root mean square error; R^2 = coefficient of determination.

The predictive performance of the four candidate algorithms on the cow-level testing sets is shown in **Table 6-5**. Similarly, ANN consistently outperformed other algorithms but were systematically worse compared with random splitting. The R^2 values decreased by 19.6%, 57.8%, and 17.8%, for RR, VT, and ET, respectively. The over-optimistic predictive performance of random splitting over subject-wise splitting has been reported in studies using ML for predicting human disease outcomes [369] and cattle milk traits [186]. Interestingly, all relevant studies found to use ML in predicting physiological responses in livestock have previously adopted random splitting [145,177-180,248,345,347]. This means that existing ML models are not ready for generalisation but only for the population included in each specific study. One important reason is that relevant animal studies are rather limited in animal numbers. Our number of cows (n = 59) is in line with previous studies but definitely does not support good representative training samples by using subject-level splitting, therefore affecting the generalisation on testing samples. Consequently, the models trained with random splitting were continuously used for the following analyses despite their predictably poor generalisation.

Table 6- 5: Performance of four candidate algorithms in predicting respiration rate (RR, breaths per min), vaginal temperature (VT, °C), and eye temperature (ET, °C) on the cow-level training sets, 5-fold cross-validation, and testing sets.

Response	Algorithm	Training		Cross-validation (SD)		Testing	
		RMSE	R^2	RMSE	R^2	RMSE	R^2
RR	RF	14.44	0.46	14.13 (0.24)	0.48 (0.02)	16.11	0.29
	GBM	13.17	0.55	14.25 (0.33)	0.48 (0.03)	15.47	0.35
	ANN	14.78	0.45	14.91 (0.56)	0.44 (0.04)	14.16	0.45
	RLR	15.62	0.37	15.73 (0.19)	0.36 (0.02)	16.47	0.26
VT	RF	0.35	0.29	0.34 (0.01)	0.33 (0.07)	0.32	0.10
	GBM	0.32	0.41	0.35 (0.01)	0.28 (0.08)	0.32	0.15
	ANN	0.32	0.40	0.35 (0.01)	0.26 (0.07)	0.31	0.19
	RLR	0.37	0.22	0.37 (0.01)	0.20 (0.05)	0.33	0.04
ET	RF	0.33	0.42	0.33 (0.01)	0.44 (0.04)	0.34	0.31
	GBM	0.20	0.78	0.32 (0.01)	0.45 (0.02)	0.32	0.37
	ANN	0.32	0.48	0.38 (0.01)	0.26 (0.06)	0.32	0.37
	RLR	0.35	0.37	0.35 (0.01)	0.36 (0.04)	0.33	0.33

RF = random forests; GBM = gradient boosting machines; ANN = artificial neural networks; RLR = regularised linear regression; SD = standard deviation; RMSE = root mean square error; R^2 = coefficient of determination.

The advantage of ANN in predicting physiological responses of dairy cows has been reported in previous studies. Under a free-stall barn without evaporative cooling, Pacheco et al. [179] developed ANN models for predicting RR and rectal temperature of 35 Holstein dairy cows. The best model that they selected had an RMSE of 10.01 breaths per min and an R^2 of 0.74 for predicting RR, and an RMSE of 0.40 °C and an R^2 of 0.71 for predicting rectal temperature. Another recent study compared different ML algorithms in predicting physiological responses using historical data collected from 20 Holstein dairy cows restrained in outdoor headlocks and deprived of sprinklers [178]. RF models produced the lowest RMSE for predicting RR (9.70 breaths per min) and BST (0.33 °C), while an ANN model produced the lowest RMSE for predicting VT (0.43 °C). However, efforts have not been done yet to predict physiological responses of dairy cows managed with sprinklers. The RMSE of the overall best models proposed in this study was close to those of the abovementioned studies, particularly VT, which had the lowest RMSE among relevant studies. Our results extend the advantage of non-linear models over linear regression models to situations equipped with sprinklers. More importantly, the gains in performance from non-linear models over linear models are greater than the previous studies. This fact highlights the non-linear effect induced by sprinkler systems and the ability of advanced ML algorithms to fit it.

3.3. Model interpretation

In this study, 13 predictors from both animal and environmental perspectives were used for modelling. This provided a basis for further mining the effects of

comprehensive predictors on physiological responses by applying state-of-the-art post-hoc interpretability methods. The global interpretation of the overall best models shown in **Figure 6-4** helps to understand the relationships between response and predictor variables by visualising the change in predicted values as the specified predictor changes assuming the remaining predictors fixed at their mean value.

The mean response in predicted RR, VT, and ET along with changing DMY of the three days before the test day is shown in **Figure 6-4(a-c)**. The positive association between production level and RR prediction reported by Janni [370] and Li et al. [173] is not clear in our results, probably because the test cows had similar high production levels when entering the study and thus the difference in heat sensitivity between production levels was not discernible. In fact, it is rather difficult to interpret these variables separately because they contained dynamic information as a pattern of response. Acclimatisation manifests in different physiology and production dynamics depending on the intensity and duration of heat stress exposure. Cows typically lose milk production when they enter the acute phase of heat stress, but their productivity can be restored to some extent during chronic stress [55]. Thus, the input of these production variables should contribute to a better prediction because they function like sensors, recording the results of cows' acclimatisation to heat stress and its mitigation.

Cows born in summer had much lower RR and VT compared with those born in other seasons (**Figure 6-4(d)**). This result is consistent with previous knowledge that in-utero heat stress would affect thermoregulation during the entire life of newborn cattle [371]. Lower RR and VT, in this case, could represent a better ability to maintain thermal homeostasis, which benefits from foetal heat acclimatisation [372]. Autumn calving cows had the lowest mean predicted RR and VT compared with cows calving in winter and spring (**Figure 6-4(e)**). The lower RR and VT of autumn calving cows are probably because they were more advance in their lactation thus having less heat load due to lower milk production during heat waves [373]. This is confirmed by a decreased VT since approximately 170 DIM (**Figure 6-4(f)**). Similarly, a decreased VT was reported during late lactation (> 150 DIM) in a recent study using 826 Holstein cows [225]. It is hard to say why RR stayed at a high level during late lactation (**Figure 6-4(f)**), possibly due to its more primary function in thermoregulation or simply because the effect of DIM was masked by that of calving season. Indeed, an increase in RR is used by cows to reduce heat load and hence prevent an increase in CBT [131].

Parity was found to have a positive relationship with predicted RR and VT (**Figure 6-4(g)**), and possibly masked the effects of AIM (**Figure 6-4(h)**). Older cows are known to be more susceptible to heat stress [208]. A trend of RR increasing with parity was also found in the study of Yan et al. [225]. In addition, Choukeir et al. [374] reported that multiparous cows had a higher VT than primiparous cows. Lying cows had higher RR, VT, and ET than standing cows (**Figure 6-4(i)**). Similarly, lying was related to increased RR and CBT in the study of Atkins et al. [135]. Increasing standing time is well known as a behavioural change strategy taken by heat-stressed cows to dissipate excessive heat [375]. The

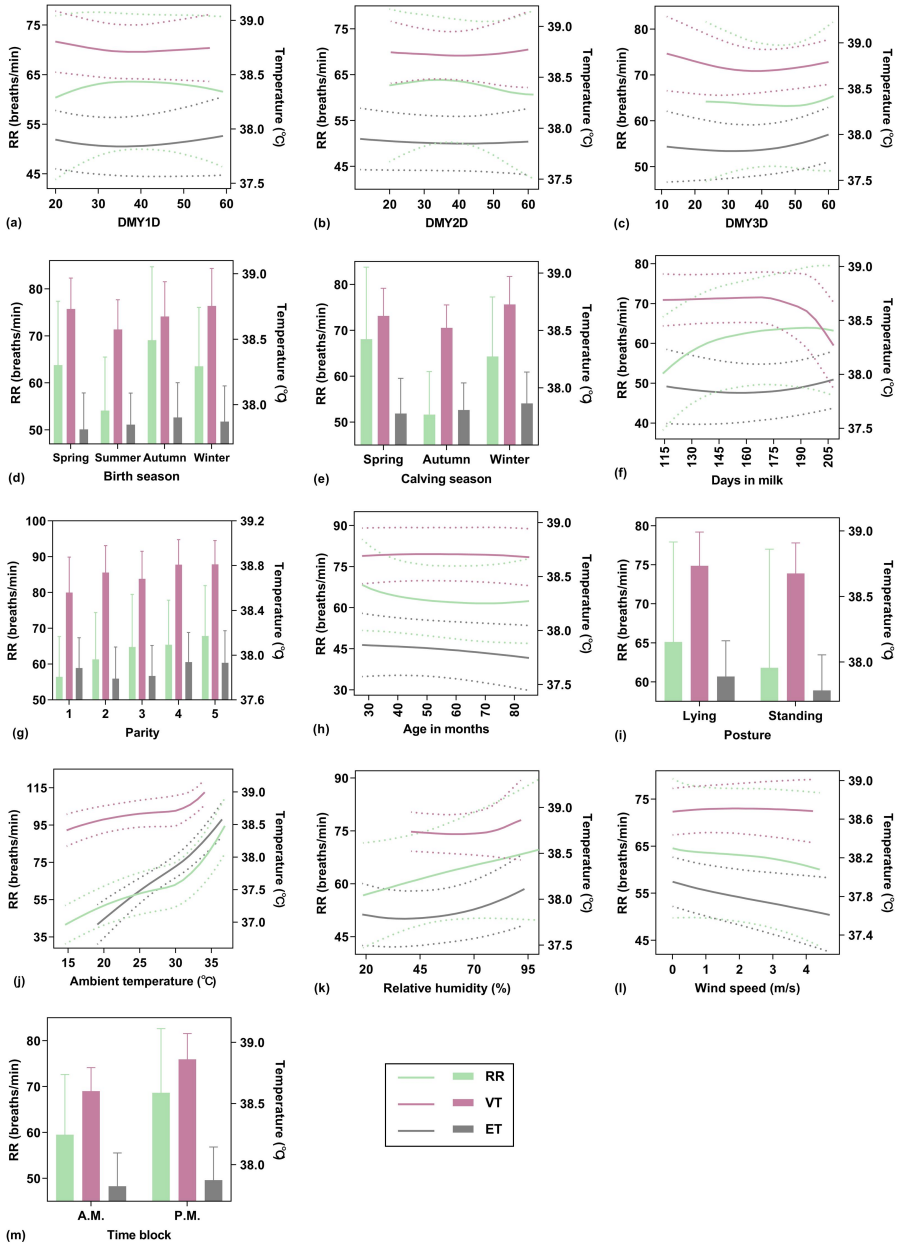


Figure 6- 4: Partial dependence plots of the overall best models (artificial neural networks) on the testing sets showing the effect of production (a-c), cow-related (d-i), and environmental factors (j-m) on respiration rate (RR), vaginal temperature (VT), and eye temperature (ET). The 95% confidence intervals for continuous and categorical variables are shown with dotted lines and error bars, respectively. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day).

THI threshold for RR was found to be 65 in lying cows and 70 in standing cows, suggesting that lying cows are more susceptible to heat stress [212]. The inapparent association of ET with the abovementioned cow-related factors is most likely due to its direct exposure to external environments and thus being less connected with internal animal factors.

The predicted physiological responses had a clear positive relationship with Ta (**Figure 6-4(j)**). The predicted RR and VT increased with a gradually decreasing slope until around 30 °C, after which they increased at a much steeper slope. These results demonstrate the effectiveness of sprinklers in alleviating heat stress in dairy cows by delaying the upper critical temperature of 25 °C [1] to about 30 °C. In addition, ET increased almost linearly with Ta, suggesting that BST is a better representation of microenvironments and a dominant front-line heat dissipator. In the case of RH, ET and VT increased sequentially at around 45% and 70% RH, respectively, whereas RR increased almost linearly with RH within the measured range (**Figure 6-4(k)**). These findings are not surprising since high RH would significantly inhibit latent heat dissipation and result in a high physiological level [4]. In the case of WS, RR and ET had an overall decreasing trend with WS, whereas VT stayed relatively stable (**Figure 6-4(l)**). These results are consistent with previous knowledge that RR and ET respond much more promptly to microenvironmental changes than VT [215]. As expected, observations measured in P.M. had higher RR, VT, and ET than those measured in A.M. (**Figure 6-4(m)**), which is consistent with previous studies on the circadian rhythm of physiological indicators related to heat stress [86]. Again, the above results re-emphasise the non-linear relationship between environmental parameters and physiological responses of dairy cows managed with sprinklers and the ability of ML algorithms to fit it.

The R^2 (mean ± standard deviation) obtained for the local interpretation of the overall best models were 0.81 ± 0.14, 0.78 ± 0.15, and 0.82 ± 0.09 for RR, VT, and ET, respectively, indicating good quality of the interpretations of LIME. As shown in **Figure 6-5**, **Figure 6-6**, and **Figure 6-7**, time block and calving season were common features that strongly influenced RR and VT predictions, whereas posture, Ta, and WS had consistently strong influences on ET predictions. The cases with the highest predicted RR and VT were marked by P.M. measurements plus certain levels of cow-related factors (e.g., births in autumn and spring calving for RR, winter calving for VT, and multiparity for both) (**Figure 6-8**), whereas the cases with the lowest predicted RR and VT were marked by A.M. measurements plus other levels of cow-related factors (e.g., autumn calving, births in summer, and standing posture for RR) (**Figure 6-9**). In the case of ET, Ta was the factor that consistently had a strong positive influence on the predictions, whereas WS and standing posture always had the greatest negative influence (**Figure 6-8** and **Figure 6-9**). These results are consistent with those from the partial dependence plots, suggesting that cow-related factors had a greater impact on RR and VT than on ET, which was more determined by microenvironmental factors. Identifying common contributing features among the most extreme cases can provide useful information for targeted cooling. Collectively, our results suggest herd homogeneity in response

to heat stress as well as a potential for customised heat abatement in different subgroups of cows.

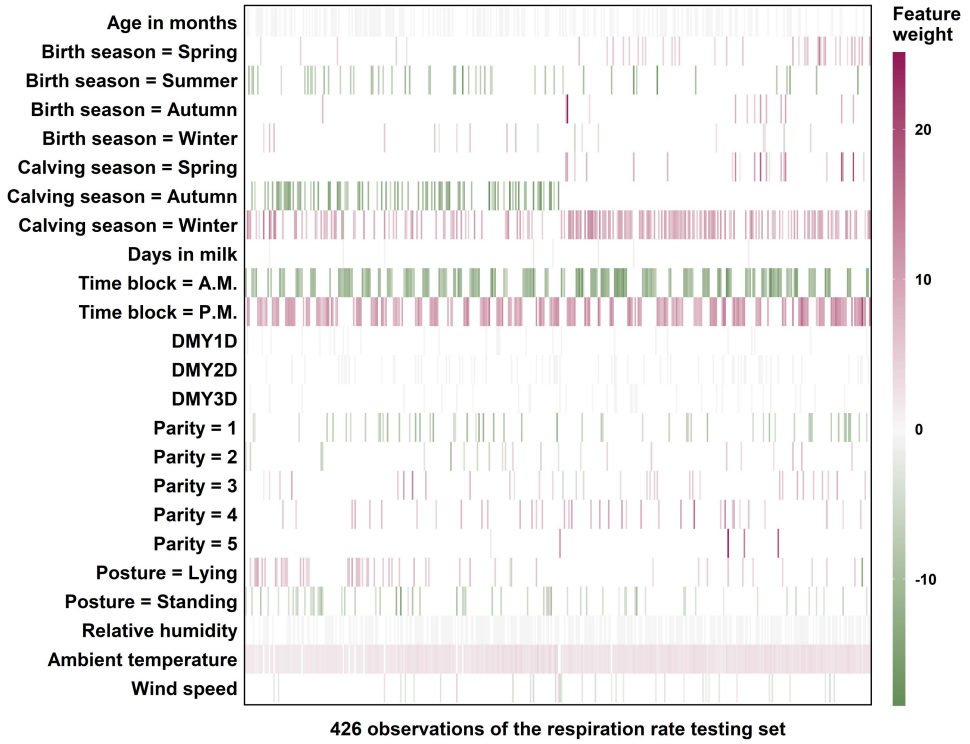


Figure 6-5: Local interpretation heatmap of the overall best respiration rate model (an artificial neural network) showing the influence of different predictor variables on the prediction of 426 observations of the testing set. The top five influential predictor variables that best explained each observation were used for plotting. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day).

3.4. Recognition of heat stress state

The classification performance of the proposed classifiers (i.e., predicted RR and VT), THI classifiers, THIadj classifier, and ETIC classifier in recognising heat stress state is listed in **Table 6-6**. As expected, the F1-score and accuracy of the predicted values were the highest, indicating a good ability in recognising the actual heat stress state. In general, the predicted RR and VT had lower recall than the environmental classifiers, implying a worse ability for detecting heat-stressed animals. However, the better recall of THI and ETIC thresholds was compromised by the lower precision and accuracy, with a huge number of non-heat-stressed cows being misclassified as heat-stressed cows. In fact, THI (68) and ETIC (23) almost classified all the cows as being heat-stressed. Accordingly, environmental classifiers

performed poorly on the RR set where 49.1% of the cases were truly heat stressed, but performed very well on the VT set where most cases were truly heat stressed (69.4%). It can be reasonably speculated that the proposed classifiers would perform much better than environmental classifiers when processing on a more balanced dataset. Similarly, the high recall of THI thresholds was balanced by the low precision in the study of Li et al. [173]. These facts demonstrate the deficiency of environmental thresholds in dealing with individual variation since different animals behave differently under identical thermal environment. Our findings highlight that heat abatement strategies controlled by environmental thresholds can be abused by wasting unnecessary efforts on non-heat-stressed animals. A better way is to make heat abatement decisions according to the predicted heat stress state of animals.

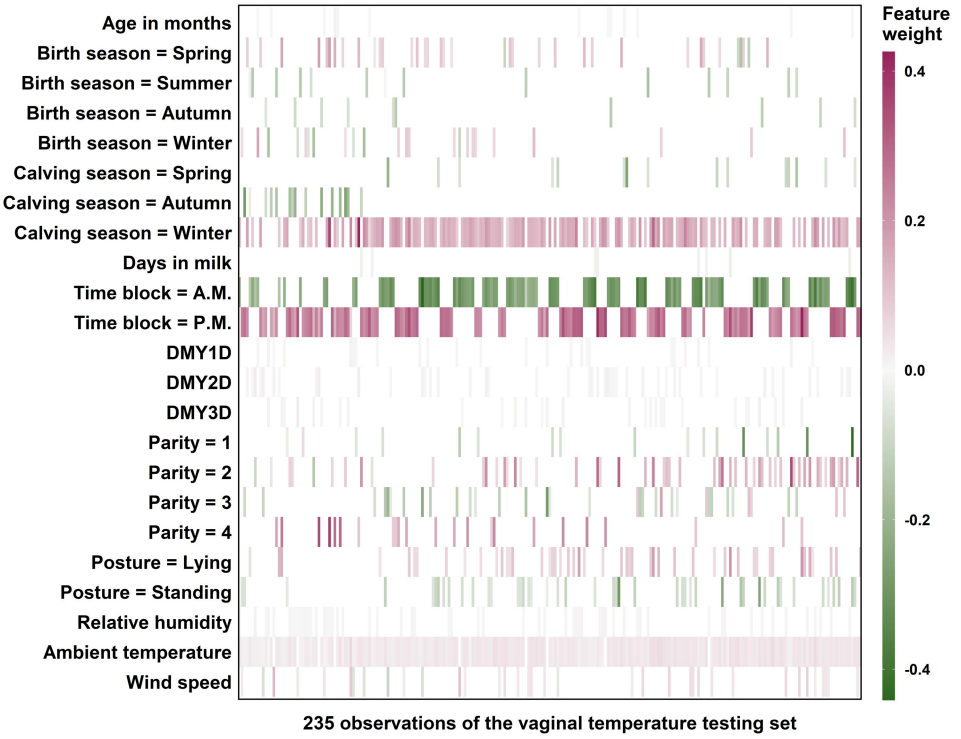


Figure 6- 6: Local interpretation heatmap of the overall best vaginal temperature model (an artificial neural network) showing the influence of different predictor variables on the prediction of 235 observations of the testing set. The top five influential predictor variables that best explained each observation were used for plotting. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day).

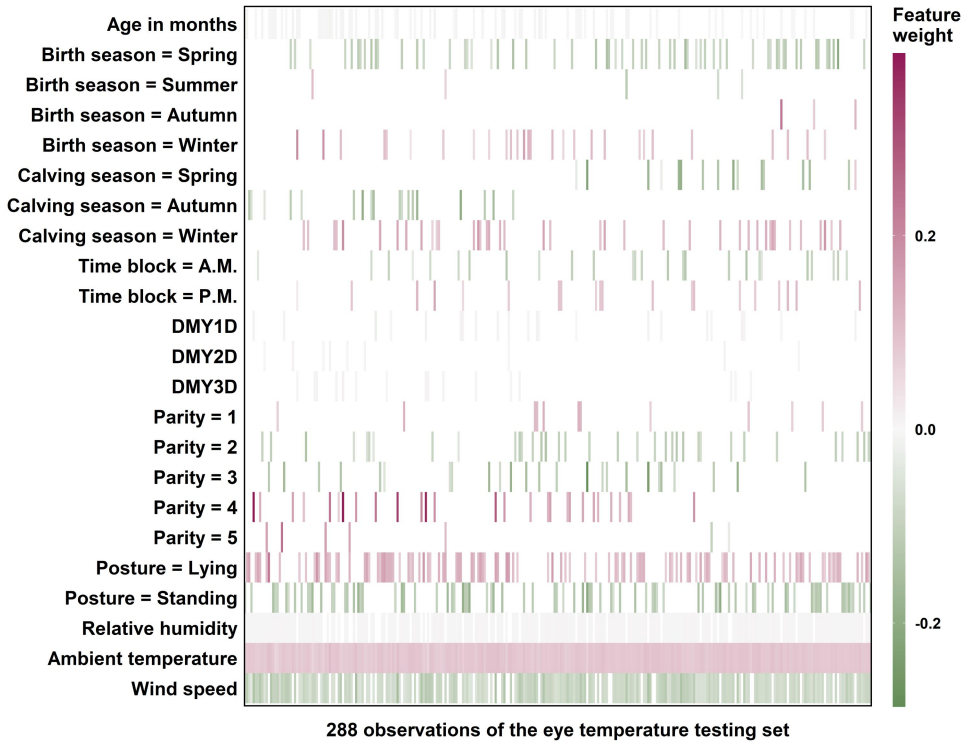


Figure 6- 7: Local interpretation heatmap of the overall best eye temperature model (an artificial neural network) showing the influence of different predictor variables on the prediction of 288 observations of the testing set. The top five influential predictor variables that best explained each observation were used for plotting. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day).

3.5. Limitations and perspectives

One of the biggest limitations of the present study is that data were collected over one single summer. Besides, only mid-lactating cows were included. These limitations resulted in inadequate heterogeneity of DMY and some cow-related factors (i.e., DIM, calving season). Although most interpretations are rational and consistent with previous knowledge, it should be noted that the interpretations could be arbitrary at certain ranges when there was insufficient training data to make a meaningful prediction. Increasing sample size and balancing data distribution, in this respect, is of great importance to improve the interpretations. The dilemma between prediction and interpretation should also be noticed. Although multicollinearity was not detected mathematically, some predictors correlated logistically, e.g., DIM and calving season, AIM and parity. Certain predictors would have to be removed to get more reliable interpretations. However, accuracy would be sacrificed to some extent as a result of information loss.

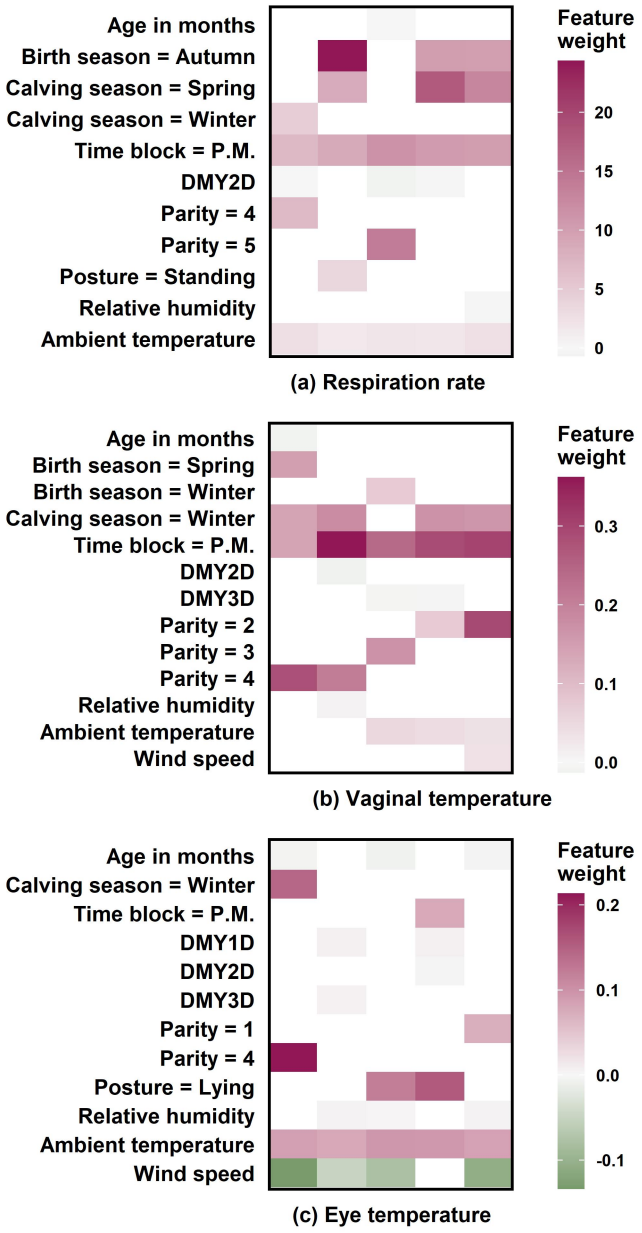


Figure 6- 8: Local interpretation heatmaps of the overall best models (artificial neural networks) plotting the top five influential predictor variables of five observations with the highest prediction selected from the testing set of (a) respiration rate, (b) vaginal temperature, and (c) eye temperature. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day); DMY3D = daily milk yield of the 3rd day before the test day (kg/day).

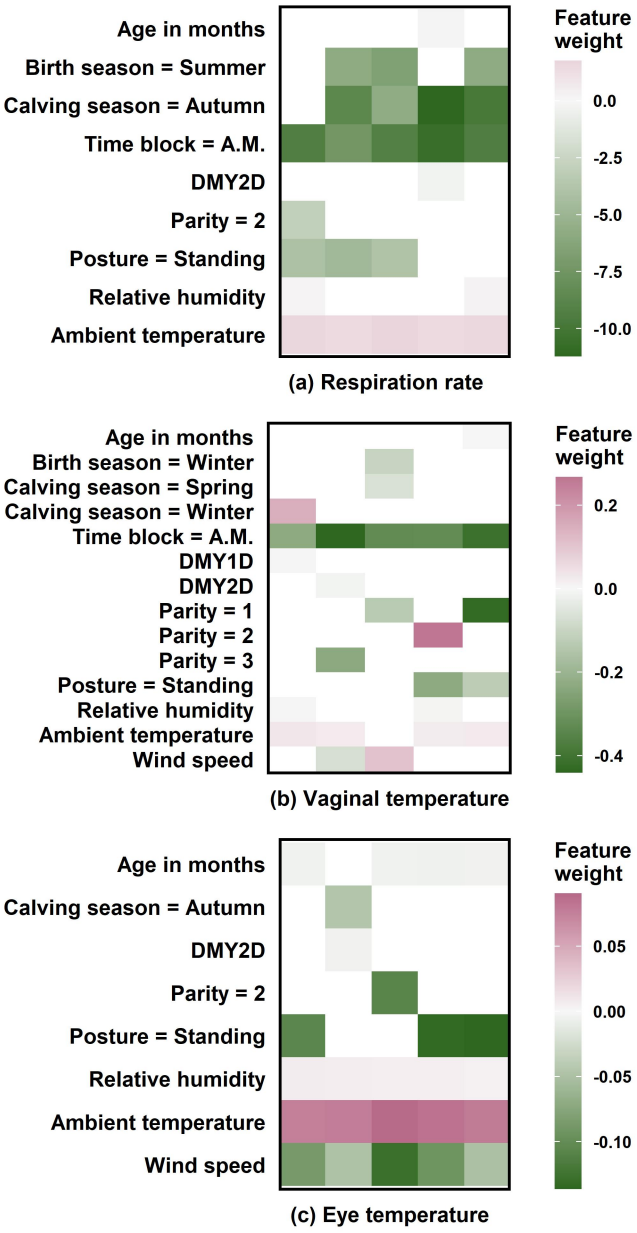


Figure 6-9: Local interpretation heatmaps of the overall best models (artificial neural networks) plotting the top five influential predictor variables of five observations with the lowest prediction selected from the testing set of (a) respiration rate, (b) vaginal temperature, and (c) eye temperature. DMY1D = daily milk yield of the day before the test day (kg/day); DMY2D = daily milk yield of the 2nd day before the test day (kg/day).

Table 6- 6: Performance of the proposed classifiers, temperature-humidity index (THI) classifiers, adjusted temperature-humidity index (THIadj) classifier, and equivalent temperature index for cattle (ETIC) classifier in recognising heat stress state on the testing sets based on measured respiration rate (RR) and vaginal temperature (VT), respectively.

Response	Classifier (threshold)	Recall (%)	Precision (%)	F1-score (%)	Accuracy (%)
RR (n = 426)	Predicted RR (60 breaths per min)	85.6	73.4	79	77.7
	THI (68)	99.5	50.7	67.2	52.3
	THI (70)	98.6	53	68.9	56.3
	THI (72)	97.1	55.9	71	61
	THIadj (74)	93.8	58.9	72.3	64.8
	ETIC (23)	100	49.1	65.8	49.1
VT (n = 235)	Predicted VT (38.5 °C)	89.6	78.1	83.4	75.3
	THI (68)	97.5	69.4	81.1	68.5
	THI (70)	91.4	69.3	78.8	66
	THI (72)	85.9	69.7	76.9	64.3
	THIadj (74)	82.2	70.5	75.9	63.8
	ETIC (23)	100	69.4	81.9	69.4

The moderate R^2 might indicate that there is still room for improvement in the fit of the current dataset. This could be done by introducing more variables that contribute to explaining the variance of cow physiological responses [347]. In this respect, sprinkler-related parameters (e.g., flow rate) and other variables that could reflect the interaction between cows and sprinklers (e.g., how much and how long cows receive watering) should be considered in further studies. These variables may be collected by future sprinklers equipped with cow sensing systems. Another possible way to increase the fit is to increase the data size since ANN would perform better when training with sufficiently large data [155]. Collectively, further studies with data collected over different seasons and years, as well as more useful variables with more reasonable combinations, are required to confirm the results of this study and further improve the model fit and interpretation.

Additionally, although the use of simple random splitting has provided acceptable predictive performance, it has inevitably decreased the probability to generalise the trained models to new unseen cows. Thus, additional work is required to enhance the generalisation and robustness. This includes collecting data from more cows across a variety of application scenarios and then modelling based on cow-level splitting.

4. Conclusions

The proposed models provide acceptable prediction errors and are reliable in a real-world farm equipped with sprinklers. Our work highlights the benefits of inputting more contributing variables in predicting the physiological responses of dairy cows under heat stress. The attempt at global and local interpretation using the state-of-the-art method was basically in line with previous knowledge in this field,

and therefore, will help future studies to deeply explain the non-linear relationships between physiological responses and their influencing factors as well as to identify the most vulnerable animals taking into account individual variations in response to heat stress. Furthermore, recognising the heat stress state of animals based on the predicted physiological responses can inform better heat abatement decisions by saving efforts from non-heat-stressed to heat-stressed. Further studies on larger cow-level datasets with more influencing factors are warranted to improve model fit and generalisation.

Chapter 7

General discussion and conclusion

1. General discussion

The interaction between cows and their thermal environments has been studied for decades. Environment-based indicators and their critical thresholds have long been used to inform cooling decisions in practice. However, environmental indicators cannot reflect true animal responses or show whether a cooling measure has effectively cooled the animals. The better solution is to use the data collected from the animal aspect which directly indicates their true response to heat stress and its alleviation.

The research plan contemplated improving heat stress detection in dairy cows through the development of measurement, prediction, and assessment methods using animal-based indicators and artificial intelligence techniques. **The first chapter**, which is a literature review, determined body surface temperatures (BST), respiration rate (RR), and relevant behaviours as the most appropriate animal-based indicators due to their high feasibility of measurement and high sensitivity to heat stressors.

Next, non-contact measurements were developed for the selected animal-based indicators (i.e., BST, RR, and relevant behaviours) in **the third, fourth, and fifth chapters,** respectively. All the developed methods involved data collection through the use of cameras and data processing using computer vision techniques. The BST on cattle faces were collected using an infrared camera and then processed using semantic segmentation. The RR measurement utilised videos that were collected by an RGB camera, and processed with techniques including instance segmentation, optical flow, and signal processing. The behaviour recognition relied on video collected by a network video recording system, coupled with object detection.

Following that, machine learning-based predictive models were trained in **the sixth chapter**. The outcome variables included BST and RR, as well as the gold standard indicator frequently used by previous studies, i.e., vaginal temperature. The predictors were determined from both animal- and environment- sides. The results were interpreted using state-of-the-art post-hoc interpretability methods.

1.1. The effect of data-splitting strategies on model performance

Machine learning, and especially deep learning tasks, always involve three stages, namely training, validation, and testing. The training stage is the one where the algorithm learns which patterns in the features correlate with the labels, and the data used in this stage is called the training set. The validation stage is the one where the patterns learned during the training stage are tuned to maximise a preset metric (e.g., mean average precision) on a different set of data, which is called the validation set. Finally, the testing stage is where the final performance of the model learned is collected on some new data which is called the testing set. Since models are already tuned towards the validation set, the testing set is important to avoid selection bias and to report an unbiased performance.

When splitting the data into these three sets, random allocation is often the easiest and most commonly used way to avoid bias. However, lots of evidence has shown that simple blind randomisation may lead to an overestimation of the model's performance [186,187,376,377]. This is a kind of data leakage which is defined as information from outside the training set being used to create the model [376]. Therefore, it is always emphasised that the data in different sets should be independent of each other. For example, multiple images from the same subjects should always be in the same set. This awareness is especially emphasised in certain interdisciplinary fields like disease diagnosis [369], medical imaging segmentation [378] and prediction using milk infrared spectral data [185], but very little is mentioned and paid attention to in other popular precision livestock farming (PLF) tasks, such as deep learning-based animal recognition and machine learning-based phenotype prediction [277]. In computer vision tasks, this oversight may be due to the tendency to treat farm animals of the same breed as identical entities, thereby neglecting their independence. However, this may lead to data leakage and overfitting for animals with distinct morphometric and biometric features, such as body condition and colour pattern. In machine learning-based regression modelling, this oversight may be a strategy taken as a last resort to ensure predictive performance due to the limited number of animals and observations.

To explore to what extent this issue can bias the performance and generalisation of the trained deep learning models, simple random splitting has been compared with cow-level splitting and pen-level splitting in **Chapters 3 and 4**, respectively. In addition, to explore to what extent the same issue can bias the predictive performance of the trained machine learning-based regression models, simple random splitting has been compared with cow-level splitting in **Chapter 6**.

In **Chapter 3**, better performance in segmenting horns in cattle infrared images was obtained on the testing set by random splitting than cow-level splitting (by 10.27%), but actually they had the same performance on the external testing set. This demonstrates the occurrence of overestimation by using random splitting. However, this effect was not seen in the other facial landmarks. It is most likely due to the different levels of difficulty in segmenting these landmarks since horns can be in different shapes in different cows but the other facial landmarks like ears and eyes are relatively similar among cows. Moreover, the much smaller sample size of horns increased the difficulty of learning their features [188].

Theoretically, a better cow-based splitting strategy to avoid overfitting and data leakage should lead to a better generalisation in new cows but is unseen in **Chapter 3**. This can happen when the training data is not representative enough for the target application. For example, it is impractical to expect a cow detection model trained exclusively on images of Holstein cows to effectively extrapolate what it has learned to a completely different context, such as Jersey cows. The dataset in **Chapter 3** is too small and can only represent the dairy herds in the study farm. However, there were no images of new cows in the original dataset that could be used to make a fair comparison of the two splitting strategies. Instead, an external testing set which contained data collected from another farm in another year was

used. The performance was reduced significantly for segmenting horns and ears due to different contexts (i.e., different numbers of ear tags and diverse horn shapes). Although eyes and nose areas remain the same accuracy in the external testing, this can be easily inverted when future tested in another new context, e.g., where eye diseases are present or nose rings are worn. Thus, the data for testing generalisability should be chosen in line with the research purposes as different data used during this process can lead to different challenging levels for the model and finally lead to different conclusions [185,379]. Meanwhile, training data should be well collected with enough cow numbers to be a good representative of the study population. In other words, it is critical to collect data with the target application in mind.

In **Chapter 4**, random image-level splitting was found to slightly inflate the model's performance, with a maximum increase of 3% in mean average precision. The trained models should have learned no prior knowledge of the specific cows even if the images casually containing the same cows showed up across training, validation, and testing sets. This is because the same cows had completely different postures and locations in the images to the point where their multiple occurrences can be regarded as independent of each other. The slightly higher performance is most likely due to prior knowledge about pen facilities since the models trained with image-level splitting have already seen images from all four pens during training. Therefore, pen-level splitting more accurately represents realistic and unbiased performance when generalising to new pens although with slightly reduced performance. It is speculated that the models trained with random splitting should have slightly reduced performance but are similar to those with pen-level splitting when applying to a new pen's data due to the challenges of the new environments (e.g., different flooring and bedding materials). However, such external validation like what was done in **Chapter 3** is missing to make a final evaluation of its generalisability.

In **Chapter 5**, images were split randomly since only one pen was used which made it impossible to separate images by pens or cows. However, this should not pose a problem since the results of **Chapter 4** have already demonstrated that images casually containing the same cows can be considered independent of each other, provided they are sampled at a sufficiently large time interval. This independence arises due to the heterogeneity in the compositions of cow identities and behaviours. In addition, the objective was to apply the trained model to the same pen as a demonstration of herd-level heat stress detection. When further applied to other pens, the trained models will undoubtedly experience a performance decline, varying in degree based on their dissimilarity from the study pen. Transfer learning, in this case, could be used to accelerate the retraining process.

In **Chapter 6**, random splitting was used to split the datasets for machine learning-based predictive modelling. This has yielded acceptable predictive performance but has inevitably reduced the generalisability for new unseen cows. Cow-level splitting was also executed as a control but had substantially worse performance. This is understandable since such a small number of cows (n = 59) does not allow a successful cow-level splitting as the training set was not even able to include all

combinations of the different levels of predictors. Unlike image recognition tasks in **Chapters 3, 4, and 5**, which had small numbers of classes and high numbers of instances in single images, and could take advantage of appropriate data augmentation [188], machine learning regression modelling requires a relatively larger sample size depending on the number of predictors. In fact, it is impossible to expect machine learning regression models trained with data of about 40 cows only to make reliable predictions for a few new cows. This can explain the fact that all relevant studies have adopted random splitting instead of cow-level splitting [145,155,177-180,248,345,347]. Anyway, cow-level splitting should still provide strong predictive performance and generalisation and should be adopted when there are enough cow numbers and observations. However, this requires the assistance of advanced measurement methods to collect animal data more massively.

1.2. Choice of individual and herd-level heat stress detection

In this thesis, animal-based heat stress detection has been developed, focusing specifically on the non-contact measurements for individual BST and RR, as well as behaviours at the herd level. Such technologies support cow-centred management by providing helpful information about animal responses to heat stress and its mitigations. One question that can be raised is how to choose between individual and herd-level monitoring.

Precision livestock farming (PLF) ultimately aims to achieve the monitoring of animal key issues at the individual level to fulfil precise management. This goal is in line with the significant individual variabilities despite regular grouping and genetic selection to increase herd uniformity. Therefore, the primary advantage of individual monitoring systems lies in their precision and sensitivity. By closely monitoring variables such as BST and RR, we can gain insights into the health status of each animal or sentinel animals, allowing for early detection of diseases, stress, and other issues. Furthermore, individual data can contribute to a deeper understanding of animal physiology and behaviour under various environmental conditions, aiding in genetic selection and breeding programs aimed at enhancing resilience and productivity. However, individual monitoring systems are not without their drawbacks. The most significant challenge is the logistical and financial burden of implementing and maintaining these systems across large numbers of animals. The complexity of data collection and analysis can also be overwhelming, requiring sophisticated software and skilled personnel. It should be also noted that individual information may still have to be summarised at a herd level to inform decision-making, as environmental regulations like cooling and ventilation in free-stall barns are always operated at the group or herd level.

Herd-level monitoring, on the other hand, offers direct control over the initiation and cessation of environmental regulations based on herd behaviour and welfare. This method is particularly useful in detecting environmental stressors that affect the group as a whole, such as heat stress. Specifically, the proposed method can be used to understand the normal circadian rhythm of each herd and to identify deviations from this norm under different levels of heat stress. Herd monitoring is also more

cost-effective on resource-limited farms compared with individual monitoring. The downside of herd monitoring is its lack of sensitivity. While it excels in identifying general trends and issues affecting the group, it may miss subtler signs of distress or illness in individual animals. This could lead to delayed interventions for some high-risk individuals. Moreover, the interpretation of behavioural data can be subjective and requires careful analysis to avoid misinterpretations.

Looking forward, the integration of individual and herd-level monitoring systems appears to be a promising approach. Combining the BST and RR data from individual monitoring with the broader insights from the herd-level behavioural analysis could provide a more comprehensive assessment of heat stress. This approach not only captures the overall welfare but also ensures attention to high-risk individuals.

1.3. A promising integration of the proposed measurement methods

A detailed discussion from a systematic perspective is provided below, in which a promising system is imagined with all the proposed measurement, prediction, and assessment methods integrated to automatically detect heat stress in dairy cows.

Vision-based measurements have been developed for the determined early indicators (i.e., BST, RR, and relevant behaviours). Note that the scale and method of measurement vary, ranging from individual level measurements for BST and RR to herd level measurements for behaviours, and from handheld infrared cameras for BST to fixed RGB cameras for RR and behaviours. This depends on the technical parameters of the applied cameras. BST relied on an infrared camera with a resolution of 640 × 480 pixels. Thus, a close distance had to be made for an accurate temperature reading. In contrast, RGB cameras were used to record cows for measuring RR and behaviours. Their much higher resolution and wider field of view made it possible to cover more cows in the image. The different measurement scale of these indicators can decide their further application scenarios in commercial farms.

The proposed automated facial BST collection was originally designed to process infrared images already collected for research purposes. It can very much accelerate temperature collection. For an automated measurement of BST, infrared cameras must be fixed somewhere that cows go through on a daily basis and can be captured one by one. One potential location is on the way to the milking parlour or milking robot where cows can be captured good infrared images in order. An example application can be found in a recent study by Wang et al. [126]. This may be the only way to separate cows on their way to milking, as it is too challenging to capture respiration-related signals while cows are walking.

In addition, the proposed measurement methods can be divided into different deep learning tasks. Behavioural recognition relied on object detection, BST on semantic segmentation, and RR on a combination of instance segmentation, object detection, and object tracking. It would be of very interest to integrate some of these tasks together to fulfil a multi-task model. For example, a basic object detection model in

conjunction with an optical flow method is promising to detect RR and behaviour at the same time by using RGB frames as the only input.

Collectively, a systematic application can be imagined as shown in **Figure 7-1**, where RGB cameras are fixed and used to measure RR and behaviours of animals in pens. These measurements can inform the start of local or global cooling measures. At the same time, infrared cameras can be fixed on the way to be milked so that heat stress can be confirmed instantly and stressed cows can be redirected to a holding room for extra cooling (**Figure 7-1**). This strategy ensures that cows with different heat stress states can be milked separately, avoiding mixing milk from stressed cows. One thing that needs to pay attention to is that pushing cows to be milked is a stressful experience and may potentially lead to a higher physiological state [230]. Thus, further studies should investigate the effect of heat stress and moving cows on their physiology and behaviour. In contrast, dairy farms equipped with milking robots might not have this problem due to the achievement of free cow traffic, which has been reported to reduce the stress caused by milking [380].

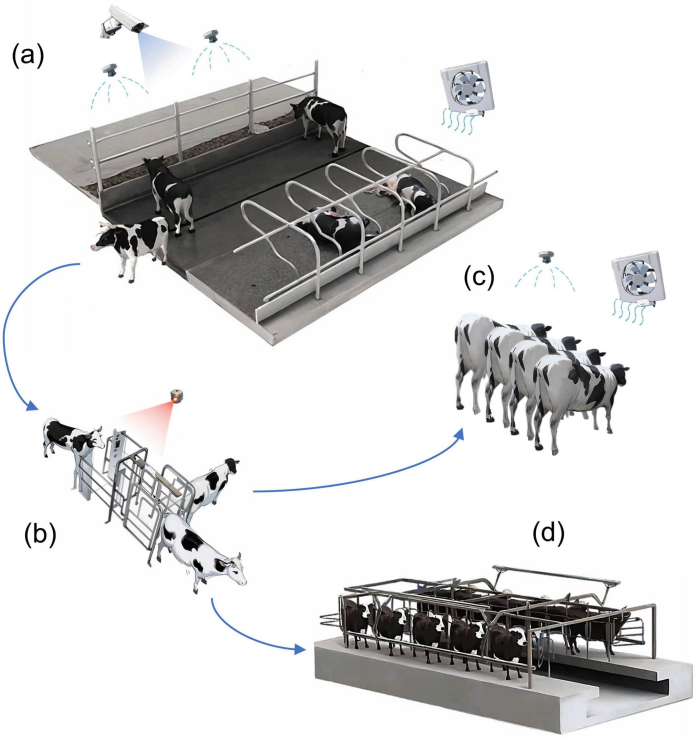


Figure 7- 1: A systematic application of the proposed measurement methods, predictive models, and assessment. (a) Vision-based cow behaviour and respiration rate measurement informing the initiation and cessation of cooling measures; (b) infrared thermography-based facial temperature measurement at the selection gate before milking, allocating heat-stressed cows to (c) an intensive cooling session, and non-heat-stressed cows to (d) the milking parlour.

Moreover, precision tools are required to continue working at night in order to inform night cooling decisions. Nighttime is an important chance for cows to relieve their body heat and night cooling has shown its effectiveness in assisting this process [381]. Meanwhile, night data are also necessary for studying thermodynamic diurnal rhythm and evaluating cooling effects. However, this cannot be fulfiled in this thesis due to the lack of night data collection. It can be speculated that infrared images and further BST can still be collected during the night even without illumination due to the technical advantage of infrared thermography. As for RR and behaviours which rely on RGB cameras, infrared supplement light or illumination during the night, or both, is necessary to have a clear vision of cows. If new data collected at night are ready, transfer learning can be utilised to accelerate the training process. Anyway, future studies should investigate the robustness of the proposed methods at night.

1.3.1. Role of predictive models in this integrative system

The proposed predictive models provide an alternative way to obtain animal-based data. This alternative is totally non-contact and does not require any human work. The results demonstrate comparable root mean square error (RMSE) and coefficient of determination (R^2) compared with relevant studies. However, R^2 is still not good enough, suggesting that although the predictors in the current models were correlated with the outcome variables, they were unable to explain more variability in the outcome variables. Therefore, the model prediction should be further improved.

One possible solution is to increase the training sample size, which can be expected from the proposed measurement methods. By obtaining a larger dataset, the performance of the model can be significantly improved, as it can better capture the underlying patterns and relationships in the data. Another solution can be to include other variables that may provide more informative explanations for the variability. These variables potentially include the duration and amplitude that cows have received sprinkling. It can be expected that these variables can be collected by using a vision-based cow-sensing system. This can be integrated into the abovementioned multi-task model so that more data with more variables will be available to update and improve predictive models. More importantly, including sprinkler-related variables in the model can also make it possible to assess and compare the effectiveness of different sprinkling strategies.

Moreover, the interpretations of the models are relevant with regard to previous knowledge from the field, which demonstrates the feasibility of using machine learning-based predictive models to explore the underlying relationship of heat stress-related variables. Collectively, the proposed models are therefore better used for identifying the most vulnerable animals rather than for precisely obtaining animal-based data. Recognising risk factors that contribute most to heat stress can inform a precise grouping, allowing farmers to tailor their management to address individual needs and mitigate potential risks as early as possible.

1.3.2. Heat stress assessment based on comprehensive information

One might ask which animal-based indicator is the best to determine heat stress. There may not be a clear and common answer to this question. Indeed, thermal comfort is a rather ambiguous concept and animals can take different strategies as initial responses to heat stress. Among the indicators investigated in **the first Chapter**, BST, RR, and relevant behaviours have been concluded to be generally more sensitive to heat stress compared with core body temperatures (CBT). Therefore, we suggest these variations to be considered as much as possible in order to capture the potential activation of intensive heat dissipation in response to heat stress. Ideally, physiological, behavioural, and environmental information should be taken into account together for a better understanding of how animals interact with their thermal environments.

To summarise, the assessment of heat stress in this system will mainly rely on automated measurements and their corresponding critical thresholds. A series of thresholds have been developed based on the data from the proposed measurement methods and the on-farm environmental sensors. These customised thresholds could better represent the relationship between the animals and their environments specifically in the experimental farm. The critical ambient temperature (Ta) or temperature-humidity index (THI) thresholds were determined as follows: 23.8 °C Ta or 68.5 THI for relevant behaviours, 23.6 °C Ta or 72 THI for RR, and 26.1 °C Ta for mean eye temperature. This order somehow reflects the sensitivity of these indicators, with relevant behaviours and RR functioning as frontline indicators and eye temperature as a final examiner. Therefore, comprehensive heat stress detection without blind spots is promising by applying suitable technologies in different scenarios, where much earlier indicators (i.e., relevant behaviours and RR) inform the cooling decision for free-stall barns, and the relatively later indicator (i.e., eye temperature) measured at the selection gate before milking confirms the effectiveness of previous cooling sessions and provides information on any additional cooling needs.

In addition, the assessment of heat stress can also take advantage of the interpretation from the proposed predictive models. For example, animals at the highest risk can be identified based on variable importance. This knowledge allows for targeted interventions to be implemented, with a focus on the specific factors contributing most to heat stress. Through timely and appropriate interventions, such as optimising ventilation and adjusting feed and water availability, the impact of heat stress on the animals can be minimised. Another advantage could be the assessment of the performance and efficiency of the building design and cooling measures by identifying the areas where cows systematically experience more heat stress. This, in turn, allows for improved barn conception.

1.3.3. Pros and cons from a sustainable agriculture perspective

When discussing the application of such proposed PLF technologies in heat stress detection for dairy cows, it is necessary to consider their impacts on sustainable agriculture comprehensively, both from the advantages and potential limitations.

Precision Livestock Farming (PLF) technologies play an important role in promoting agricultural sustainability. Environmentally, the proposed PLF technologies enable more precise detection of dairy cows' heat stress states, guiding more efficient cow-centred management from non-heat-stressed cows to heat-stressed cows and from low-risk cows to high-risk cows. This approach not only reduces resource wastage but also minimises unnecessary ecological impacts from aimless heat abatement strategies. Economically, such PLF technologies can reduce production losses as well as power and energy costs through timely and efficient management, simultaneously promoting health in high-risk animals and thereby cutting costs related to potential disease treatment. With the rising consumer demand for sustainable products, farms utilising PLF technologies are positioned to achieve wider market recognition and enhanced economic benefits. Socially, these non-contact technologies enhance animal welfare and milk quality, aligning with growing public consciousness about sustainable agriculture. Furthermore, the application of PLF technologies can reduce the workload for dairy practitioners, benefiting their welfare and potentially improving job satisfaction.

However, these technologies also face several challenges. Environmentally, while the proposed PLF technologies may help reduce resource consumption and pollution caused by unlimited cooling measures, their operation also relies on energy and resources, potentially impacting the environment. Additionally, the manufacturing and maintenance of vision sensors may involve energy consumption and waste production, necessitating further consideration of their overall environmental impact. Economically, the high initial investment costs of PLF technologies could be burdensome for small-scale or resource-limited farms [382]. The complexity of technologies requires farmers to possess certain technical knowledge and operational skills, potentially leading to additional training and education investments [383]. Furthermore, potential system errors or inaccuracies may lead to production loss [203]. Socially, the implementation of PLF technologies could widen the digital gap in agriculture, potentially sidelining farmers and communities who rely on conventional practices. In addition, over-reliance on technological interventions might raise concerns about the welfare of animals, as the time farmers spend with animals would be reduced and their observational skills would be diminished [384].

In summary, PLF technologies hold significant value in promoting agricultural sustainability, yet they also come with challenges related to economic costs, technical complexity, potential environmental impact, and social and ethical issues. Future developments need to balance these technologies' advantages and limitations, ensuring their application aligns not only with productivity and profit improvements but also with environmental protection, social justice, and ethical principles.

2. Conclusions

This project aimed to leverage artificial intelligence techniques, specifically computer vision and machine learning, to develop non-invasive solutions for the detection of heat stress in dairy cows. By developing edge-cutting models via on-farm collected data and applying them to actual farming, this thesis has shown how

the developed vision-based measurement methods and data-driven predictive models, as well as the further developed critical thresholds, can be effectively utilised to facilitate automated heat stress detection in dairy cows.

The main contributions fitted with the specific objectives of this research are three-fold:

- Three non-contact measurement methods were proposed for BST, RR, and relevant behaviours in dairy cows.
- A series of critical thresholds were determined for the onset of heat stress based on the data measured by the proposed methods.
- Three predictive models were proposed for predicting RR, BST, and vaginal temperature. The interpretation of the models showed the risk factors of heat stress.

Based on these results, an integrative application of all the proposed measurement, prediction, and assessment methods has been suggested, wherein RGB and infrared cameras are used to measure animal-based indicators, and critical thresholds, along with model interpretation, are used to assess the heat stress state of dairy cows. This strategy ensures timely and thorough cooling of cows in all areas of the dairy farm, thereby minimising the negative impact of heat stress to the greatest extent.

It should be acknowledged that certain targeted improvements are required due to the limitations of the current research, such as fusing a multi-task network, addressing the problem of night work, and improving model prediction. Furthermore, additional validation or transfer learning of the proposed methods in a broader real-world context is needed considering the significant differences across dairy farms.

References

'*If I have seen further, it is by standing on the shoulders of giants.*'

Isaac Newton

1. Kadzere, C.T., Murphy, M.R., Silanikove, N., Maltz, E., *Heat stress in lactating dairy cows: a review.* Livestock production science, 2002. **77**(1): p. 59-91.

2. Ouellet, V., Cabrera, V.E., Fadul-Pacheco, L., Charbonneau, É., *The relationship between the number of consecutive days with heat stress and milk production of Holstein dairy cows raised in a humid continental climate.* Journal of dairy science, 2019. **102**(9): p. 8537-8545.

3. Maia, A.S.C., daSilva, R.G., Battiston Loureiro, C.M., *Sensible and latent heat loss from the body surface of Holstein cows in a tropical environment.* International journal of biometeorology, 2005. **50**(1): p. 17-22.

4. Maia, A.S.C., Silva, R.G.d., Loureiro, C.M.B., *Latent heat loss of Holstein cows in a tropical environment: a prediction model.* Revista brasileira de zootecnia, 2008. **37**(10): p. 1837-1843.

5. Becker, C.A., Collier, R.J., Stone, A.E., *Invited review: Physiological and behavioral effects of heat stress in dairy cows.* Journal of dairy science, 2020. **103**(8): p. 6751-6770.

6. Silanikove, N., *Effects of heat stress on the welfare of extensively managed domestic ruminants.* Livestock Production Science, 2000. **67**(1): p. 1-18.

7. Mayorga, E.J., Renaudeau, D., Ramirez, B.C., Ross, J.W., Baumgard, L.H., *Heat stress adaptations in pigs.* Animal Frontiers, 2018. **9**(1): p. 54-61.

8. Krishnan, G., Madiajagan, B., Prathap, P., Vidya, M., Joy, A., Abhijith, A., Sejian, V., *Mitigation of the Heat Stress Impact in Livestock Reproduction*. 2017. p. 64-86.

9. Liberati, P., *Analysis of the effect of the roofing design on heat stress in dairy cow housing.* Journal of Agricultural Engineering, 2012. **39**.

10. Fournel, S., Ouellet, V., Charbonneau, É., *Practices for Alleviating Heat Stress of Dairy Cows in Humid Continental Climates: A Literature Review.* Animals, 2017. **7**(5): p. 37.

11. Garcia, P.R., Silveira, R.M.F., Lensink, J., da Silva, I.J.O., *Thermal performance of a low-profile cross-ventilated freestall dairy barn with evaporative cooling pads in a hot and humid climate.* International Journal of Biometeorology, 2023. **67**(10): p. 1651-1658.

12. Deng, S., Shi, Z., Li, B., *CFD simulation of temperature and humidity distribution in low profile cross ventilated dairy cattle barn.* Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, 2015. **31**: p. 209-214.

13. Shoshani, E., Hetzroni, A., *Optimal barn characteristics for high-yielding Holstein cows as derived by a new heat-stress model.* Animal, 2013. **7**(1): p. 176-182.

14. Radoń, J., Bieda, W., Lendelová, J., Pogran, Š., *Computational model of heat exchange between dairy cow and bedding.* Computers and Electronics in Agriculture, 2014. **107**: p. 29-37.

15. Ortiz, X.A., Smith, J.F., Rojano, F., Choi, C.Y., Bruer, J., Steele, T., Schuring, N., Allen, J., Collier, R.J., *Evaluation of conductive cooling of lactating dairy cows under controlled environmental conditions.* Journal of dairy science, 2015. **98**(3): p. 1759-1771.

16. Zhang, Y.W., McCarl, B.A., Jones, J.P.H., *An Overview of Mitigation and Adaptation Needs and Strategies for the Livestock Sector.* Climate, 2017. **5**(4): p. 95.

17. Ripple, W.J., Smith, P., Haberl, H., Montzka, S.A., McAlpine, C., Boucher, D.H., *Ruminants, climate change and climate policy.* Nature Climate Change, 2014. **4**(1): p. 2-5.

18. Shephard, R., Maloney, S., *A review of thermal stress in cattle.* Australian Veterinary Journal, 2023. **101**(11): p. 417-429.

19. Berman, A., *Inter-animal radiation as potential heat stressor in lying animals.* International Journal of Biometeorology, 2014. **58**(7): p. 1683-1691.

20. Lin, S.-x., Liu, J.-x., Wang, K.-y., Wang, D.-m., *Effects of stocking density on oxidative stress status and mammary gland permeability in early lactating dairy cows.* Animal Science Journal, 2019. **90**(7): p. 894-902.

21. Mayes, B.T., Taylor, P.S., Cowley, F.C., Gaughan, J.B., Morton, J.M., Doyle, B.P., Tait, L.A., *The effects of stocking density on behavior and biological functioning of penned sheep under continuous heat load conditions.* Journal of Animal Science, 2023. **101**.

22. Conte, G., Ciampolini, R., Cassandro, M., Lasagna, E., Calamari, L., Bernabucci, U., Abeni, F., *Feeding and nutrition management of heat-stressed dairy ruminants.* Italian Journal of Animal Science, 2018. **17**(3): p. 604-620.

23. Gantner, V., Kuterovac, K., Potočnik, K., *Effect of Heat Stress on Metabolic Disorders Prevalence Risk and Milk Production in Holstein Cows in Croatia.* Ann. Anim. Sci, 2016. **16**(2): p. 451-461.

24. Stermer, R.A., Brasington, C.F., Coppock, C.E., Lanham, J.K., Milam, K.Z., *Effect of Drinking Water Temperature on Heat Stress of Dairy Cows.* Journal of Dairy Science, 1986. **69**(2): p. 546-551.

25. Ferraretto, L.F., Shaver, R.D., Bertics, S.J., *Effect of dietary supplementation with live-cell yeast at two dosages on lactation performance, ruminal fermentation, and total-tract nutrient digestibility in dairy cows.* Journal of Dairy Science, 2012. **95**(7): p. 4017-4028.

26. Wo, Y., Ma, F., Shan, Q., Gao, D., Jin, Y., Sun, P., *Plasma metabolic profiling reveals that chromium yeast alleviates the negative effects of heat stress in mid-lactation dairy cows.* Animal Nutrition, 2023. **13**: p. 401-410.

27. Shwartz, G., Rhoads, M., VanBaale, M.J., Rhoads, R., Baumgard, L., *Effects of a supplemental yeast culture on heat-stressed lactating Holstein cows.* Journal of dairy science, 2009. **92**: p. 935-42.

28. Guo, Y., Li, L., Yan, S., Shi, B., *Plant Extracts to Alleviating Heat Stress in Dairy Cows.* Animals, 2023. **13**(18): p. 2831.

29. Cheng, J.B., Bu, D.P., Wang, J.Q., Sun, X.Z., Pan, L., Zhou, L.Y., Liu, W., *Effects of rumen-protected γ-aminobutyric acid on performance and nutrient digestibility in heat-stressed dairy cows.* Journal of Dairy Science, 2014. **97**(9): p. 5599-5607.

30. Cheng, J., Zheng, N., Sun, X., Li, S., Wang, J., Zhang, Y., *Feeding rumen-protected gamma-aminobutyric acid enhances the immune response and antioxidant status of heat-stressed lactating dairy cows.* Journal of Thermal Biology, 2016. **60**: p. 103-108.

31. Ma, N., Li, Y., Ren, L., Hu, L., Xu, R., Shen, Y., Cao, Y., Gao, Y., Li, J., *Effects of dietary N-carbamylglutamate supplementation on milk production performance, nutrient digestibility and blood metabolomics of lactating Holstein cows under heat stress.* Animal Feed Science and Technology, 2021. **273**: p. 114797.

32. Yang, L., Zhang, L., Zhang, P., Zhou, Y., Huang, X., Yan, Q., Tan, Z., Tang, S., Wan, F., *Alterations in nutrient digestibility and performance of heat-stressed dairy cows by dietary L-theanine supplementation.* Animal Nutrition, 2022. **11**: p. 350-358.

33. Osei-Amponsah, R., Chauhan, S.S., Leury, B.J., Cheng, L., Cullen, B., Clarke, I.J., Dunshea, F.R., *Genetic Selection for Thermotolerance in Ruminants.* Animals, 2019. **9**(11): p. 948.

34. Dikmen, S., Cole, J.B., Null, D.J., Hansen, P.J., *Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle.* Journal of Dairy Science, 2012. **95**(6): p. 3401-3405.

35. Cartwright, S.L., Schmied, J., Livernois, A., Mallard, B.A., *Effect of In-vivo heat challenge on physiological parameters and function of peripheral blood mononuclear cells in immune phenotyped dairy cattle.* Veterinary Immunology and Immunopathology, 2022. **246**: p. 110405.

36. Cartwright, S.L., McKechnie, M., Schmied, J., Livernois, A.M., Mallard, B.A., *Effect of in-vitro heat stress challenge on the function of blood mononuclear cells from dairy cattle ranked as high, average and low immune responders.* BMC Veterinary Research, 2021. **17**(1): p. 233.

37. Fang, H., Kang, L., Abbas, Z., Hu, L., Wang, Y., Xu, Q., *Identification of key Genes and Pathways Associated With Thermal Stress in Peripheral Blood Mononuclear Cells of Holstein Dairy Cattle.* Frontiers in Genetics, 2021.

38. Li, R., Ahmad, M.J., Hou, M., Wang, X., Liu, S., Li, J., Jiang, Q., Huang, J., Yang, L., *Identification of target genes and pathways related to heat tolerance in Chinese Holstein cows.* Livestock Science, 2023. **271**: p. 105213.

39. Carabaño, M.J., Ramón, M., Menéndez-Buxadera, A., Molina, A., Díaz, C., *Selecting for heat tolerance.* Animal frontiers, 2019. **9**(1): p. 62-68.

40. Fan, C., Su, D., Tian, H., Li, X., Li, Y., Ran, L., Hu, R., Cheng, J., *Liver metabolic perturbations of heat-stressed lactating dairy cows.* Asian-Australasian journal of animal sciences, 2018. **31**(8): p. 1244.

41. Tian, H., Zheng, N., Wang, W., Cheng, J., Li, S., Zhang, Y., Wang, J., *Integrated metabolomics study of the milk of heat-stressed lactating dairy cows.* Scientific reports, 2016. **6**(1): p. 24208.

42. Yue, S., Ding, S., Zhou, J., Yang, C., Hu, X., Zhao, X., Wang, Z., Wang, L., Peng, Q., Xue, B., *Metabolomics Approach Explore Diagnostic Biomarkers and Metabolic Changes in Heat-Stressed Dairy Cows.* Animals, 2020. **10**(10): p. 1741.

43. Fan, C., Di, S., He, T., Hu, R., Lei, R., Ying, Y., Su, Y., Cheng, J., *Milk production and composition and metabolic alterations in the mammary gland of heat-stressed lactating dairy cows.* Journal of Integrative Agriculture, 2019. **18**(12): p. 2844-2853.

44. Dikmen, S., Khan, F.A., Huson, H.J., Sonstegard, T.S., Moss, J.I., Dahl, G.E., Hansen, P.J., *The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows.* Journal of Dairy Science, 2014. **97**(9): p. 5508-5520.

45. Hoffmann, G., Herbut, P., Pinto, S., Heinicke, J., Kuhla, B., Amon, T., *Animal-related, non-invasive indicators for determining heat stress in dairy cows.* Biosystems engineering, 2020. **199**: p. 83-96.

46. Koltes, J.E., Koltes, D.A., Mote, B.E., Tucker, J., Hubbell, D.S., *Automated collection of heat stress data in livestock: new technologies and opportunities.* Translational Animal Science, 2018. **2**(3): p. 319-323.

47. Neethirajan, S., *Transforming the Adaptation Physiology of Farm Animals through Sensors.* Animals, 2020. **10**(9): p. 1512.

48. Dikmen, S., Hansen, P.J., *Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment?* Journal of dairy science, 2009. **92**(1): p. 109-116.

49. Polsky, L., Von Keyserlingk, M.A.G., *Invited review: Effects of heat stress on dairy cattle welfare.* Journal of dairy science, 2017. **100**(11): p. 8645-8657.

50. Collier, R.J., Laun, W.H., Rungruang, S., Zimbleman, R.B. *Quantifying Heat Stress and Its Impact on Metabolism and Performance*. in *Florida Ruminant Nutrition Symposium*. 2012. Gainesville, FL, USA: University of Florida.

51. Prathap, P., Abhijith, A., Joy, A., Sejian, V., Krishnan, G., Madiajagan, B., Ayyasamy, M., Vakayil, B., Kurien, K., Varma, G., *Heat Stress and Dairy Cow: Impact on Both Milk Yield and Composition.* International Journal of Dairy Science, 2016.

52. West, J.W., Mullinix, B.G., Bernard, J.K., *Effects of Hot, Humid Weather on Milk Temperature, Dry Matter Intake, and Milk Yield of Lactating Dairy Cows.* Journal of dairy science, 2003. **86**(1): p. 232-242.

53. Li, G., Chen, J., Peng, D., Gu, X., *Short communication: The lag response of daily milk yield to heat stress in dairy cows.* Journal of Dairy Science, 2021. **104**(1): p. 981-988.

54. Hagiya, K., Bamba, I., Osawa, T., Atagi, Y., Takusari, N., Itoh, F., Yamazaki, T., *Length of lags in responses of milk yield and somatic cell score on test day to heat stress in Holsteins.* Animal Science Journal, 2019. **90**(5): p. 613-618.

55. Collier, R.J., Baumgard, L.H., Zimbelman, R.B., Xiao, Y., *Heat stress: physiology of acclimation and adaptation.* Animal Frontiers, 2019. **9**(1): p. 12-19.

56. Galán, E., Llonch, P., Villagrá, A., Levit, H., Pinto, S., Del Prado, A., *A systematic review of non-productivity-related animal-based indicators of heat stress resilience in dairy cattle.* PloS one, 2018. **13**(11): p. e0206520-e0206520.

57. Lemal, P., May, K., König, S., Schroyen, M., Gengler, N., *Invited review: From heat stress to disease—Immune response and candidate genes involved in cattle thermotolerance.* Journal of Dairy Science, 2023.

58. Dauria, B.D., Sigdel, A., Petrini, J., Bóscollo, P.P., Pilonetto, F., Salvian, M., Rezende, F.M., Pedrosa, V.B., Bittar, C.M.M., Machado, P.F., Coutinho, L.L., Wiggans, G.R., Mourão, G.B., *Genetic effects of heat stress on milk*

fatty acids in Brazilian Holstein cattle. Journal of Dairy Science, 2022. **105**(4): p. 3296-3305.

59. Hammami, H., Vandenplas, J., Vanrobays, M.L., Rekik, B., Bastin, C., Gengler, N., *Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows.* Journal of Dairy Science, 2015. **98**(7): p. 4956-4968.

60. Guo, Z., Gao, S., Ouyang, J., Ma, L., Bu, D., *Impacts of Heat Stress-Induced Oxidative Stress on the Milk Protein Biosynthesis of Dairy Cows.* Animals, 2021. **11**(3): p. 726.

61. Belhadj Slimen, I., Najar, T., Ghram, A., Abdrrabba, M., *Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review.* Journal of Animal Physiology and Animal Nutrition, 2016. **100**(3): p. 401-412.

62. St-Pierre, N.R., Cobanov, B., Schnitkey, G., *Economic Losses from Heat Stress by US Livestock Industries1.* Journal of dairy science, 2003. **86**(S): p. E52-E77.

63. Key, N., Sneeringer, S., *Potential Effects of Climate Change on the Productivity of U.S. Dairies.* American journal of agricultural economics, 2014. **96**(4): p. 1136-1156.

64. Berckmans, D., *General introduction to precision livestock farming.* Animal frontiers, 2017. **7**(1): p. 6-11.

65. Neves, S.F., Silva, M.C., Miranda, J.M., Stilwell, G., Cortez, P.P., *Predictive Models of Dairy Cow Thermal State: A Review from a Technological Perspective.* Veterinary Sciences, 2022. **9**(8): p. 416.

66. Berman, A., Folman, Y., Kaim, M., Mamen, M., Herz, Z., Wolfenson, D., Arieli, A., Graber, Y., *Upper Critical Temperatures and Forced Ventilation Effects for High-Yielding Dairy Cows in a Subtropical Climate.* Journal of dairy science, 1985. **68**(6): p. 1488-1495.

67. Garner, J.B., Douglas, M., Williams, S.R.O., Wales, W.J., Marett, L.C., DiGiacomo, K., Leury, B.J., Hayes, B.J., *Responses of dairy cows to short-term heat stress in controlled-climate chambers.* Animal Production Science, 2017. **57**(7): p. 1233-1241.

68. Spiers, D.E., Spain, J.N., Sampson, J.D., Rhoads, R.P., *Use of physiological parameters to predict milk yield and feed intake in heat-stressed dairy cows.* Journal of thermal biology, 2004. **29**(7-8): p. 759-764.

69. Debnath, T., Bera, S., Pal, P., Debbarma, N., Haldar, A., *Application of radio frequency based digital thermometer for real-time monitoring of dairy cattle rectal temperature.* Veterinary World, 2017. **10**(9): p. 1052-1056.

70. Reuter, R., Carroll, J., Hulbert, L., Dailey, J., Galyean, M., *Technical note: Development of a self-contained, indwelling rectal temperature probe for cattle research.* Journal of Animal Science, 2010. **88**(10): p. 3291-5.

71. Lees, A.M., Lea, J.M., Salvin, H.E., Cafe, L.M., Colditz, I.G., Lee, C., *Relationship between Rectal Temperature and Vaginal Temperature in Grazing Bos taurus Heifers.* Animals, 2018. **8**(9): p. 156.

72. Hillman, P.E., Gebremedhin, K.G., Willard, S.T., Lee, C.N., Kennedy, A.D., *Continuous Measurements of Vaginal Temperature of Female Cattle Using A Data Logger Encased in a Plastic Anchor.* Applied Engineering in Agriculture, 2009. **25**(2): p. 291-296.

73. Tresoldi, G., Schütz, K.E., Tucker, C.B., *Sampling strategy and measurement device affect vaginal temperature outcomes in lactating dairy cattle.* Journal of dairy science, 2020. **103**(6): p. 5414-5421.

74. Sakatani, M., Sugano, T., Higo, A., Naotsuka, K., Hojo, T., Gessei, S., Uehara, H., Takenouchi, N., *Vaginal temperature measurement by a wireless sensor for predicting the onset of calving in Japanese Black cows.* Theriogenology, 2018. **111**: p. 19-24.

75. Wang, S., Zhang, H., Tian, H., Chen, X., Li, S., Lu, Y., Li, L., Wang, D., *Alterations in vaginal temperature during the estrous cycle in dairy cows detected by a new intravaginal device-a pilot study.* Tropical animal health and production, 2020. **52**(5): p. 2265-2271.

76. Levit, H., Pinto, S., Amon, T., Gershon, E., Kleinjan-Elazary, A., Bloch, V., Ben Meir, Y.A., Portnik, Y., Jacoby, S., Arnin, A., Miron, J., Halachmi, I., *Dynamic cooling strategy based on individual animal response mitigated heat stress in dairy cows.* Animal, 2021. **15**(2): p. 100093.

77. Woodrum Setser, M.M., Cantor, M.C., Costa, J.H.C., *A comprehensive evaluation of microchips to measure temperature in dairy calves.* Journal of dairy science, 2020. **103**(10): p. 9290-9300.

78. Mahendran, S.A., Booth, R., Burge, M., Bell, N.J., *Randomised positive control trial of NSAID and antimicrobial treatment for calf fever caused by pneumonia.* Veterinary record, 2017. **181**(2): p. 45-45.

79. Lee, Y., Bok, J.D., Lee, H.J., Lee, H.G., Kim, D., Lee, I., Kang, S.K., Choi, Y.J., *Body Temperature Monitoring Using Subcutaneously Implanted Thermo-loggers from Holstein Steers.* Asian-australasian journal of animal sciences, 2016. **29**(2): p. 299-306.

80. Iwasaki, W., Ishida, S., Kondo, D., Ito, Y., Tateno, J., Tomioka, M., *Monitoring of the core body temperature of cows using implantable wireless thermometers.* Computers and electronics in agriculture, 2019. **163**: p. 104849.

81. Chung, H., Li, J., Kim, Y., Van Os, J.M.C., Brounts, S.H., Choi, C.Y., *Using implantable biosensors and wearable scanners to monitor dairy cattle's core body temperature in real-time.* Computers and electronics in agriculture, 2020. **174**: p. 105453.

82. Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., Li, B., *Modelling of heat stress in a robotic dairy farm. Part 2: Identifying the specific thresholds with production factors.* Biosystems engineering, 2020. **199**: p. 43-57.

83. Hill, T.M., Bateman, H.G., Suarez-Mena, F.X., Dennis, T.S., Schlotterbeck, R.L., *Short communication: Changes in body temperature of calves up to 2 months of age as affected by time of day, age, and ambient temperature.* Journal of dairy science, 2016. **99**(11): p. 8867-8870.

84. Kou, H., Zhao, Y., Ren, K., Chen, X., Lu, Y., Wang, D., *Automated measurement of cattle surface temperature and its correlation with rectal temperature.* PLoS ONE, 2017. **12**(4): p. e0175377.

85. Peng, D., Chen, S., Li, G., Chen, J., Wang, J., Gu, X., *Infrared thermography measured body surface temperature and its relationship with rectal temperature in dairy cows under different temperature-humidity indexes.* International Journal of Biometeorology, 2019. **63**(3): p. 327-336.

86. Kaufman, J.D., Saxton, A.M., Ríus, A.G., *Short communication: Relationships among temperature-humidity index with rectal, udder surface, and vaginal temperatures in lactating dairy cows experiencing heat stress.* Journal of dairy science, 2018. **101**(7): p. 6424-6429.

87. Jorquera-Chavez, M., Fuentes, S., Dunshea, F.R., Warner, R.D., Poblete, T., Jongman, E.C., *Modelling and Validation of Computer Vision Techniques to Assess Heart Rate, Eye Temperature, Ear-Base Temperature and Respiration Rate in Cattle.* Animals, 2019. **9**(12): p. 1089.

88. Piccione, G., Caola, G., Refinetti, R., *Daily and estrous rhythmicity of body temperature in domestic cattle.* BMC Physiology, 2003. **3**(1): p. 7-7.

89. Burfeind, O., Suthar, V.S., Heuwieser, W., *Effect of heat stress on body temperature in healthy early postpartum dairy cows.* Theriogenology, 2012. **78**(9): p. 2031-2038.

90. Burdick, N.C., Carroll, J.A., Dailey, J.W., Randel, R.D., Falkenberg, S.M., Schmidt, T.B., *Development of a self-contained, indwelling vaginal temperature probe for use in cattle research.* Journal of thermal biology, 2012. **37**(4): p. 339-343.

91. Burfeind, O., Von Keyserlingk, M.A.G., Weary, D.M., Veira, D.M., Heuwieser, W., *Short communication: Repeatability of measures of rectal temperature in dairy cows.* Journal of dairy science, 2010. **93**(2): p. 624-627.

92. Lee, C., Gebremedhin, K., Parkhurst, A., Hillman, P., *Placement of temperature probe in bovine vagina for continuous measurement of core-body temperature.* International Journal of Biometeorology, 2015. **59**(9): p. 1201-1205.

93. Kendall, P.E., Tucker, C.B., Dalley, D.E., Clark, D.A., Webster, J.R., *Milking frequency affects the circadian body temperature rhythm in dairy cows.* Livestock science, 2008. **117**(2): p. 130-138.

94. Vickers, L.A., Burfeind, O., Von Keyserlingk, M.A.G., Veira, D.M., Weary, D.M., Heuwieser, W., *Technical note: Comparison of rectal and vaginal temperatures in lactating dairy cows.* Journal of dairy science, 2010. **93**(11): p. 5246-5251.

95. Alzahal, O., Alzahal, H., Steele, M.A., Van Schaik, M., Kyriazakis, I., Duffield, T.F., McBride, B.W., *The use of a radiotelemetric ruminal bolus to detect body temperature changes in lactating dairy cattle.* Journal of dairy science, 2011. **94**(7): p. 3568-3574.

96. Hicks, L.C., Hicks, W.S., Bucklin, R.A., Shearer, J.K., Bray, D.R., Soto, P., Carvalho, V., *Comparison of Methods of Measuring Deep Body Temperatures of Dairy Cows*, in *6th International Symposium*. 2001, American Society of Agricultural and Biological Engineers: Louisville, Kentucky, USA. p. 432–438.

97. Prendiville, D.J., Lowe, J., Earley, B., Spahr, C., Kettlewell, P., *Radiotelemetry systems for measuring body temperature*. 2002, Grange Research Centre: Tegasc, Ireland.

98. Bewley, J., Einstein, M., Grott, M., Schutz, M., *Comparison of Reticular and Rectal Core Body Temperatures in Lactating Dairy Cows.* Journal of Dairy Science, 2008. **91**(12): p. 4661-72.

99. Ipema, A.H., Goense, D., Hogewerf, P.H., Houwers, H.W.J., van Roest, H., *Pilot study to monitor body temperature of dairy cows with a rumen bolus.* Computers and electronics in agriculture, 2008. **64**(1): p. 49-52.

100. Lees, A.M., Lees, J.C., Lisle, A.T., Sullivan, M.L., Gaughan, J.B., *Effect of heat stress on rumen temperature of three breeds of cattle.* International Journal of Biometeorology, 2018. **62**(2): p. 207-215.

101. Jonsson, N.N., Kleen, J.L., Wallace, R.J., Andonovic, I., Michie, C., Farish, M., Mitchell, M., Duthie, C.-A., Jensen, D.B., Denwood, M.J., *Evaluation of reticuloruminal pH measurements from individual cattle: Sampling strategies for the assessment of herd status.* The veterinary journal (1997), 2019. **243**: p. 26-32.

102. Han, C.S., Kaur, U., Bai, H., Roqueto dos Reis, B., White, R., Nawrocki, R.A., Voyles, R.M., Kang, M.G., Priya, S., *Invited review: Sensor*

technologies for real-time monitoring of the rumen environment. Journal of Dairy Science, 2022. **105**(8): p. 6379-6404.

103. Sharpe, K.T., Heins, B.J., Buchanan, E.S., Reese, M.H., *Evaluation of solar photovoltaic systems to shade cows in a pasture-based dairy herd.* Journal of Dairy Science, 2021. **104**(3): p. 2794-2806.

104. Knauer, W.A., Godden, S.M., McDonald, N., *Technical note: Preliminary evaluation of an automated indwelling rumen temperature bolus measurement system to detect pyrexia in preweaned dairy calves.* Journal of dairy science, 2016. **99**(12): p. 9925-9930.

105. Ammer, S., Lambertz, C., Gauly, M., *Is reticular temperature a useful indicator of heat stress in dairy cattle?* Journal of dairy science, 2016. **99**(12): p. 10067-10076.

106. Vázquez-Diosdado, J.A., Miguel-Pacheco, G.G., Plant, B., Dottorini, T., Green, M., Kaler, J., *Developing and evaluating threshold-based algorithms to detect drinking behavior in dairy cows using reticulorumen temperature.* Journal of Dairy Science, 2019. **102**(11): p. 10471-10482.

107. Cantor, M., Costa, J., Bewley, J., *Impact of Observed and Controlled Water Intake on Reticulorumen Temperature in Lactating Dairy Cattle.* Animals, 2018. **8**(11): p. 194.

108. Ammer, S., Lambertz, C., Gauly, M., *Comparison of different measuring methods for body temperature in lactating cows under different climatic conditions.* The Journal of Dairy Research, 2016. **83**(2): p. 165-172.

109. Hahn, G.L., Eigenberg, R.A., Nienaber, J.A., Littledike, E.T., *Measuring physiological responses of animals to environmental stressors using a microcomputer-based portable datalogger.* Journal of animal science, 1990. **68**(9): p. 2658-2665.

110. Mader, T., Holt, S., Hahn, G., Davis, M., Spiers, D., *Feeding strategies for managing heat load in feedlot cattle.* Journal of Animal Science, 2002. **80**(9): p. 2373-82.

111. Davis, M.S., Mader, T.L., Holt, S.M., Parkhurst, A.M., *Strategies to reduce feedlot cattle heat stress: effects on tympanic temperature.* Journal of animal science, 2003. **81**(3): p. 649-661.

112. Jara, I.E., Keim, J.P., Arias, R.A., *Behaviour, tympanic temperature and performance of dairy cows during summer season in southern Chile.* Archivos de medicina veterinaria, 2016. **48**: p. 113-118.

113. Bergen, R.D., Kennedy, A.D., *Relationship between vaginal and tympanic membrane temperature in beef heifers.* Canadian Journal of Animal Science, 2000. **80**(3): p. 515-518.

114. McCorkell, R., Wynne-Edwards, K., Windeyer, C., Schaefer, A., *Limited efficacy of Fever Tag(®) temperature sensing ear tags in calves with naturally occurring bovine respiratory disease or induced bovine viral diarrhea virus infection.* The Canadian veterinary journal, 2014. **55**(7): p. 688-690.

115. Richeson, J.T., Powell, J.G., Kegley, E.B., Hornsby, J.A., *Evaluation of an ear-mounted tympanic thermometer device for bovine respiratory disease diagnosis*. 2012, Arkansas Agricultural Experiment Station Division of Agriculture University of Arkansas System Fayetteville: Arkansas. p. 40-42.

116. Reid, E.D., Fried, K., Velasco, J.M., Dahl, G.E., *Correlation of rectal temperature and peripheral temperature from implantable radio-frequency microchips in Holstein steers challenged with lipopolysaccharide under thermoneutral and high ambient temperatures.* Journal of animal science, 2012. **90**(13): p. 4788-4794.

117. Theurer, M.E., Anderson, D.E., White, B.J., Miesner, M.D., Larson, R.L., *Effects of weather variables on thermoregulation of calves during periods of extreme heat.* American journal of veterinary research, 2014. **75**(3): p. 296-300.

118. Pohl, A., Heuwieser, W., Burfeind, O., *Technical note: Assessment of milk temperature measured by automatic milking systems as an indicator of body temperature and fever in dairy cows.* Journal of dairy science, 2014. **97**(7): p. 4333-4339.

119. Poikalainen, V., Praks, J., Veermäe, I., Kokin, E., *Infrared temperature patterns of cow's body as an indicator for health control at precision cattle farming.* Agronomy research, 2012. **10**: p. 187-194.

120. Berckmans, D., *Precision livestock farming technologies for welfare management in intensive livestock systems.* Revue scientifique et technique, 2014. **33**(1): p. 189-196.

121. Norton, T., Berckmans, D., *Developing precision livestock farming tools for precision dairy farming.* Animal frontiers, 2017. **7**(1): p. 18-23.

122. Godyń, D., Herbut, P., Angrecka, S., *Measurements of peripheral and deep body temperature in cattle – A review.* Journal of thermal biology, 2019. **79**: p. 42-49.

123. Schaefer, A.L., Cook, N.J., Bench, C., Chabot, J.B., Colyn, J., Liu, T., Okine, E.K., Stewart, M., Webster, J.R., *The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography.* Research in veterinary science, 2012. **93**(2): p. 928-935.

124. Hoffmann, G., Schmidt, M., Ammon, C., *First investigations to refine video-based IR thermography as a non-invasive tool to monitor the body temperature of calves.* 2016. **10**(9): p. 1542-1546.

125. Lowe, G., McCane, B., Sutherland, M., Waas, J., Schaefer, A., Cox, N., Stewart, M., *Automated Collection and Analysis of Infrared Thermograms for Measuring Eye and Cheek Temperatures in Calves.* Animals, 2020. **10**(2): p. 292.

126. Wang, Y., Kang, X., Chu, M., Liu, G., *Deep learning-based automatic dairy cow ocular surface temperature detection from thermal images.* Computers and Electronics in Agriculture, 2022. **202**: p. 107429.

127. Cuthbertson, H., Tarr, G., González, L., *Methodology for data processing and analysis techniques of infrared video thermography used to measure cattle temperature in real time.* Computers and Electronics in Agriculture, 2019. **167**: p. 105019.

128. Legrand, A., Schütz, K.E., Tucker, C.B., *Using water to cool cattle: Behavioral and physiological changes associated with voluntary use of cow showers.* Journal of dairy science, 2011. **94**(7): p. 3376-3386.

129. Bach, A.J.E., Stewart, I.B., Disher, A.E., Costello, J.T., *A Comparison between Conductive and Infrared Devices for Measuring Mean Skin Temperature at Rest, during Exercise in the Heat, and Recovery.* PLoS ONE, 2015. **10**(2): p. e0117907.

130. Bar, D., Kaim, M., Flamenbaum, I., Hanochi, B., Toaff-Rosenstein, R.L., *Technical note: Accelerometer-based recording of heavy breathing in lactating and dry cows as an automated measure of heat load.* Journal of dairy science, 2019. **102**(4): p. 3480-3486.

131. Gaughan, J., Sm, H., Hahn, G., Mader, T., Ra, E., *Respiration rate—is it a good measure of heat stress in cattle?* Asian-Australas J Anim Sci, 2000. **13**.

132. Brown-Brandl, T.M., Eigenberg, R.A., Nienaber, J.A., Hahn, G.L., *Dynamic Response Indicators of Heat Stress in Shaded and Non-shaded Feedlot Cattle, Part 1: Analyses of Indicators.* Biosystems engineering, 2005. **90**(4): p. 451-462.

133. Milan, H.F.M., Maia, A.S.C., Gebremedhin, K.G., *Technical note: Device for measuring respiration rate of cattle under field conditions 1.* Journal of Animal Science, 2016. **94**(12): p. 5434-5438.

134. Eigenberg, R.A., Hahn, G.L., Nienaber, J.A., Brown-Brandl, T.M., Spiers, D.E., *Development of a new respiration rate monitor for cattle.* Transactions of the ASAE, 2000. **43**(3): p. 723-728.

135. Atkins, I.K., Cook, N.B., Mondaca, M.R., Choi, C.Y., *Continuous Respiration Rate Measurement of Heat-Stressed Dairy Cows and Relation*

to Environment, Body Temperature, and Lying Time. Transactions of the ASABE, 2018. **61**(5): p. 1475-1485.

136. Davison, C., Michie, C., Hamilton, A., Tachtatzis, C., Andonovic, I., Gilroy, M., *Detecting Heat Stress in Dairy Cattle Using Neck-Mounted Activity Collars.* Agriculture, 2020. **10**(6): p. 210.

137. Pastell, M., Aisla, A.M., Hautala, M., Poikalainen, V., Praks, J., Veermäe, I., Ahokas, J., *Contactless measurement of cow behavior in a milking robot.* Behavior research methods, 2006. **38**(3): p. 479-486.

138. Sachs, J., Helbig, M., Kmec, M., Herrmann, R., Schilling, K., Plattes, S., Fritsch, H. *Remote heartbeat capturing of high yield cows by uwb radar*. in *International Radar Symposium*. 2015. IEEE.

139. Tuan, S.-A., Rustia, D.J.A., Hsu, J.-T., Lin, T.-T., *Frequency modulated continuous wave radar-based system for monitoring dairy cow respiration rate.* Computers and Electronics in Agriculture, 2022. **196**: p. 106913.

140. Song, H., Wu, D., Yin, X., Jiang, B., He, D., *Respiratory behavior detection of cow based on Lucas-Kanade sparse optical flow algorithm.* Transactions of the Chinese Society of Agricultural Engineering, 2019. **35**(17): p. 215-224.

141. Wu, D., Yin, X., Jiang, B., Jiang, M., Li, Z., Song, H., *Detection of the respiratory rate of standing cows by combining the Deeplab V3+ semantic segmentation model with the phase-based video magnification algorithm.* Biosystems Engineering, 2020. **192**: p. 72-89.

142. Strutzke, S., Fiske, D., Hoffmann, G., Ammon, C., Heuwieser, W., Amon, T., *Technical note: Development of a noninvasive respiration rate sensor for cattle.* Journal of dairy science, 2019. **102**(1): p. 690-695.

143. de Carvalho, G.A., Salman, A.K.D., da Cruz, P.G., de Souza, E.C., da Silva, F.R.F., Schmitt, E., *Technical note: An acoustic method for assessing the respiration rate of free-grazing dairy cattle.* Livestock science, 2020. **241**.

144. de Melo Costa, C.C., de Melo Costa, C.C., Maia, A.S.C., Maia, A.S.C., Nascimento, S.T., Nascimento, S.T., Nascimento, C.C.N., Nascimento, C.C.N., Neto, M.C., Neto, M.C., de França Carvalho Fonsêca, V., de França Carvalho Fonsêca, V., *Thermal balance of Nellore cattle.* International journal of biometeorology, 2018. **62**(5): p. 723-731.

145. Fuentes, S., Gonzalez Viejo, C., Tongson, E., Dunshea, F.R., Dac, H.H., Lipovetzky, N., *Animal biometric assessment using non-invasive computer vision and machine learning are good predictors of dairy cows age and welfare: The future of automated veterinary support systems.* Journal of Agriculture and Food Research, 2022. **10**: p. 100388.

146. Mader, T., Davis, M., Brown-Brandl, T., *Environmental factors influencing heat stress in feedlot cattle.* Journal of Animal Science, 2006. **84**(3): p. 712-9.

147. Gaughan, J., Mader, T., Holt, S., Lisle, A., *A new heat load index for feedlot cattle.* Journal of Animal Science, 2008. **86**(1): p. 226-34.

148. Lees, J.C., Lees, A.M., Gaughan, J.B., *Developing a heat load index for lactating dairy cows.* Animal production science, 2018. **58**(8): p. 1387.

149. Islam, M.A., Lomax, S., Doughty, A.K., Islam, M.R., Clark, C.E.F., *Automated Monitoring of Panting for Feedlot Cattle: Sensor System Accuracy and Individual Variability.* Animals, 2020. **10**(9): p. 1518.

150. Laister, S., Stockinger, B., Regner, A.-M., Zenger, K., Knierim, U., Winckler, C., *Social licking in dairy cattle—Effects on heart rate in performers and receivers.* Applied animal behaviour science, 2011. **130**(3): p. 81-90.

151. Stewart, M., Wilson, M.T., Schaefer, A.L., Huddart, F., Sutherland, M.A., *The use of infrared thermography and accelerometers for remote monitoring of dairy cow health and welfare.* Journal of dairy science, 2017. **100**(5): p. 3893-3901.

152. Lowe, G., Sutherland, M., Waas, J., Schaefer, A., Cox, N., Stewart, M., *Infrared Thermography—A Non-Invasive Method of Measuring Respiration Rate in Calves.* Animals, 2019. **9**(8): p. 535.

153. Wu, D., Han, M., Song, H., Song, L., Duan, Y., *Monitoring the respiratory behavior of multiple cows based on computer vision and deep learning.* Journal of Dairy Science, 2023. **106**(4): p. 2963-2979.

154. Tresoldi, G., Schütz, K.E., Tucker, C.B., *Assessing heat load in drylot dairy cattle: Refining on-farm sampling methodology.* Journal of dairy science, 2016. **99**(11): p. 8970-8980.

155. Becker, C.A., Aghalari, A., Marufuzzaman, M., Stone, A.E., *Predicting dairy cattle heat stress using machine learning techniques.* Journal of dairy science, 2021. **104**(1): p. 501-524.

156. Schütz, K.E., Rogers, A.R., Cox, N.R., Webster, J.R., Tucker, C.B., *Dairy cattle prefer shade over sprinklers: Effects on behavior and physiology.* Journal of dairy science, 2011. **94**(1): p. 273-283.

157. Grinter, L.N., Mazon, G., Costa, J.H.C., *Voluntary heat stress abatement system for dairy cows: Does it mitigate the effects of heat stress on physiology and behavior?* Journal of Dairy Science, 2023. **106**(1): p. 519-533.

158. Herbut, P., Hoffmann, G., Angrecka, S., Godyń, D., Vieira, F.M.C., Adamczyk, K., Kupczyński, R., *The effects of heat stress on the behaviour of dairy cows – a review.* Annals of Animal Science, 2021. **21**(2): p. 385-402.

159. McDonald, P.V., von Keyserlingk, M.A.G., Weary, D.M., *Hot weather increases competition between dairy cows at the drinker.* Journal of Dairy Science, 2020. **103**(4): p. 3447-3458.

160. Heinicke, J., Hoffmann, G., Ammon, C., Amon, B., Amon, T., *Effects of the daily heat load duration exceeding determined heat load thresholds on activity traits of lactating dairy cows.* Journal of Thermal Biology, 2018. **77**: p. 67-74.

161. Sammad, A., Luo, H., Qiu, W., Galindez, J.M., Wang, Y., Guo, G., Huang, X., Wang, Y., *Automated monitoring of seasonal and diurnal variation of rumination behaviour: Insights into thermotolerance management of Holstein cows.* Biosystems Engineering, 2022. **223**: p. 115-128.

162. Porto, S.M.C., Arcidiacono, C., Anguzza, U., Cascone, G., *A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns.* Biosystems Engineering, 2013. **115**(2): p. 184-194.

163. Porto, S.M.C., Arcidiacono, C., Anguzza, U., Cascone, G., *The automatic detection of dairy cow feeding and standing behaviours in free-stall barns by a computer vision-based system.* Biosystems Engineering, 2015. **133**: p. 46-55.

164. Mammi, L., Cavallini, D., Palmonari, A., Concolino, A., Ghiaccio, F., Buonaiuto, G., Visentin, G., Formigoni, A., *504 Late-Breaking: Automatic Monitoring Systems to Detect Behavioral and Productive Variations during Heat Stress in Dairy Cows.* Journal of Animal Science, 2021. **99**: p. 180-181.

165. Laurindo, G., Ribeiro, G., Damasceno, F., Ferraz, P., Ferraz, G., Nascimento, J., *Thermal Environment and Behavior Analysis of Confined Cows in A Compost Barn.* Animals, 2022. **12**: p. 2214.

166. Fuentes, A., Yoon, S., Park, J., Park, D.S., *Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information.* Computers and Electronics in Agriculture, 2020. **177**: p. 105627.

167. Tsai, Y.-C., Hsu, J.-T., Ding, S.-T., Rustia, D.J.A., Lin, T.-T., *Assessment of dairy cow heat stress by monitoring drinking behaviour using an embedded imaging system.* Biosystems engineering, 2020. **199**: p. 97-108.

168. Wu, D., Wang, Y., Han, M., Song, L., Shang, Y., Zhang, X., Song, H., *Using a CNN-LSTM for basic behaviors detection of a single dairy cow in a complex environment.* Computers and electronics in agriculture, 2021. **182**: p. 106016.

169. Chapinal, N., Veira, D.M., Weary, D.M., von Keyserlingk, M.A.G., *Technical Note: Validation of a System for Monitoring Individual Feeding and Drinking Behavior and Intake in Group-Housed Cattle.* Journal of Dairy Science, 2007. **90**(12): p. 5732-5736.

170. McDonald, P.V., von Keyserlingk, M.A.G., Weary, D.M., *Technical note: Using an electronic drinker to monitor competition in dairy cows.* Journal of Dairy Science, 2019. **102**(4): p. 3495-3500.

171. Siegford, J.M., Steibel, J.P., Han, J., Benjamin, M., Brown-Brandl, T., Dórea, J.R.R., Morris, D., Norton, T., Psota, E., Rosa, G.J.M., *The quest to develop automated systems for monitoring animal behavior.* Applied Animal Behaviour Science, 2023. **265**: p. 106000.

172. Foris, B., Thompson, A.J., von Keyserlingk, M.A.G., Melzer, N., Weary, D.M., *Automatic detection of feeding- and drinking-related agonistic behavior and dominance in dairy cows.* Journal of Dairy Science, 2019. **102**(10): p. 9176-9186.

173. Li, G., Chen, S., Chen, J., Peng, D., Gu, X., *Predicting rectal temperature and respiration rate responses in lactating dairy cows exposed to heat stress.* Journal of dairy science, 2020. **103**(6): p. 5466-5484.

174. Chapman, N.H., Chlingaryan, A., Thomson, P.C., Lomax, S., Islam, M.A., Doughty, A.K., Clark, C.E.F., *A deep learning model to forecast cattle heat stress.* Computers and Electronics in Agriculture, 2023. **211**: p. 107932.

175. Baum, E., Haussler, D., *What size net gives valid generalization?* Advances in neural information processing systems, 1988. **1**.

176. Abu-Mostafa, Y.S., *Hints.* Neural Computation, 1995. **7**(4): p. 639-671.

177. Fuentes, S., Gonzalez Viejo, C., Tongson, E., Lipovetzky, N., Dunshea, F.R., *Biometric Physiological Responses from Dairy Cows Measured by Visible Remote Sensing Are Good Predictors of Milk Productivity and Quality through Artificial Intelligence.* Sensors, 2021. **21**(20): p. 6844.

178. Gorczyca, M.T., Gebremedhin, K.G., *Ranking of environmental heat stressors for dairy cows using machine learning algorithms.* Computers and electronics in agriculture, 2020. **168**: p. 105124.

179. Pacheco, V.M., Sousa, R.V.d., Rodrigues, A.V.D.S., Sardinha, E.J.d.S., Martello, L.S., *Thermal imaging combined with predictive machine learning based model for the development of thermal stress level classifiers.* Livestock science, 2020. **241**.

180. Sousa, R.V.d., Rodrigues, A.V.d.S., Abreu, M.G.d., Tabile, R.A., Martello, L.S., *Predictive model based on artificial neural network for assessing beef cattle thermal stress using weather and physiological variables.* Computers and Electronics in Agriculture, 2018. **144**: p. 37-43.

181. Hernández-Julio, Y.F., Yanagi, T., de Fátima Ávila Pires, M., Aurélio Lopes, M., Ribeiro de Lima, R., *Models for Prediction of Physiological Responses of Holstein Dairy Cows.* Applied artificial intelligence, 2014. **28**(8): p. 766-792.

182. George, W.D., Godfrey, R., Ketring, R.C., Vinson, M.C., Willard, S.T., *Relationship among eye and muzzle temperatures measured using digital infrared thermal imaging and vaginal and rectal temperatures in hair sheep and cattle.* Journal of animal science, 2014. **92**.

183. Dallago, G.M., Figueiredo, D.M.d., Andrade, P.C.d.R., Santos, R.A.d., Lacroix, R., Santschi, D.E., Lefebvre, D.M., *Predicting first test day milk yield of dairy heifers.* Computers and Electronics in Agriculture, 2019. **166**: p. 105032.

184. Viana, C.M., Santos, M., Freire, D., Abrantes, P., Rocha, J., *Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach.* Ecological Indicators, 2021. **131**: p. 108200.

185. Grelet, C., Dardenne, P., Soyeurt, H., Fernandez, J.A., Vanlierde, A., Stevens, F., Gengler, N., Dehareng, F., *Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions.* Methods, 2021. **186**: p. 97-111.

186. Mota, L.F.M., Pegolo, S., Baba, T., Peñagaricano, F., Morota, G., Bittante, G., Cecchinato, A., *Evaluating the performance of machine learning methods and variable selection methods for predicting difficult-to-measure traits in Holstein dairy cattle using milk infrared spectral data.* Journal of Dairy Science, 2021. **104**(7): p. 8107-8121.

187. Wang, Q., Bovenhuis, H., *Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle.* Journal of Dairy Science, 2019. **102**(7): p. 6288-6295.

188. Li, G., Huang, Y., Chen, Z., Chesser, G.D., Purswell, J.L., Linhoss, J., Zhao, Y., *Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming: A Review.* Sensors, 2021. **21**(4): p. 1492.

189. Ellis, J.L., Jacobs, M., Dijkstra, J., van Laar, H., Cant, J.P., Tulpan, D., Ferguson, N., *Review: Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data.* Animal, 2020. **14**: p. s223-s237.

190. Thompson, V.A., Barioni, L.G., Rumsey, T.R., Fadel, J.G., Sainz, R.D., *The development of a dynamic, mechanistic, thermal balance model for Bos indicus and Bos taurus.* The Journal of Agricultural Science, 2014. **152**(3): p. 464-482.

191. Gebremedhin, K.G., Wu, B., *Modeling heat loss from the udder of a dairy cow.* Journal of Thermal Biology, 2016. **59**: p. 34-38.

192. Chen, E., Narayanan, V., Pistochini, T., Rasouli, E., *Transient simultaneous heat and mass transfer model to estimate drying time in a wetted fur of a cow.* Biosystems Engineering, 2020. **195**: p. 116-135.

193. Li, J., Narayanan, V., Kebreab, E., Dikmen, S., Fadel, J.G., *A mechanistic thermal balance model of dairy cattle.* Biosystems Engineering, 2021. **209**: p. 256-270.

194. Yan, G., Liu, K., Hao, Z., Li, H., Shi, Z., *Development and evaluation of thermal models for predicting skin temperature of dairy cattle.* Computers and Electronics in Agriculture, 2021. **188**: p. 106363.

195. Zhou, M., Groot Koerkamp, P.W.G., Huynh, T.T.T., Aarnink, A.J.A., *Development and evaluation of a thermoregulatory model for predicting thermal responses of dairy cows.* Biosystems Engineering, 2022. **223**: p. 295-308.

196. McArthur, A.J., *Thermal interaction between animal and microclimate: a comprehensive model.* Journal of theoretical biology, 1987. **126**(2): p. 203-238.

197. Ehrlemark, A.G., Sällvik, K.G., *A Model of Heat and Moisture Dissipation from Cattle Based on Thermal Properties.* Transactions of the ASAE, 1996. **39**(1): p. 187-194.

198. Gebremedhin, K.G., Wu, B.X., *Characterization of flow field in a ventilated space and simulation of heat exchange between cows and their environment.* Journal of Thermal Biology, 2003. **28**(4): p. 301-319.

199. McGovern, R.E., Bruce, J.M., *AP—Animal Production Technology: A Model of the Thermal Balance for Cattle in Hot Conditions.* Journal of Agricultural Engineering Research, 2000. **77**(1): p. 81-92.

200. Mitchell, D., Snelling, E.P., Hetem, R.S., Maloney, S.K., Strauss, W.M., Fuller, A., *Revisiting concepts of thermal physiology: Predicting responses of mammals to climate change.* Journal of Animal Ecology, 2018. **87**(4): p. 956-973.

201. Huang, T., Rong, L., Zhang, G., Brandt, P., Bjerg, B., Pedersen, P., Granath, S.W.Y., *A two-node mechanistic thermophysiological model for pigs reared in hot climates – Part 1: Physiological responses and model development.* Biosystems Engineering, 2021. **212**: p. 302-317.

202. Stolwijk, J.A., *A mathematical model of physiological temperature regulation in man*. 1971, NASA.

203. Wathes, C.M., Kristensen, H.H., Aerts, J.M., Berckmans, D., *Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall?* Computers and Electronics in Agriculture, 2008. **64**(1): p. 2-10.

204. Foroushani, S., Amon, T., *Thermodynamic assessment of heat stress in dairy cattle: lessons from human biometeorology.* International Journal of Biometeorology, 2022.

205. Ekine-Dzivenu, C.C., Mrode, R., Oyieng, E., Komwihangilo, D., Lyatuu, E., Msuta, G., Ojango, J.M.K., Okeyo, A.M., *Evaluating the impact of heat stress as measured by temperature-humidity index (THI) on test-day milk yield of small holder dairy cattle in a sub-Sahara African climate.* Livestock science, 2020. **242**.

206. Rejeb, M., Sadraoui, R., Najar, T., M'rad, M., *A Complex Interrelationship between Rectal Temperature and Dairy Cows' Performance under Heat Stress Conditions.* Open Journal of Animal Sciences, 2016. **06**: p. 24-30.

207. Carabaño, M.J., Logar, B., Bormann, J., Minet, J., Vanrobays, M.L., Díaz, C., Tychon, B., Gengler, N., Hammami, H., *Modeling heat stress under different environmental conditions.* Journal of dairy science, 2016. **99**(5): p. 3798-3814.

208. Benni, S., Pastell, M., Bonora, F., Tassinari, P., Torreggiani, D., *A generalised additive model to characterise dairy cows' responses to heat stress.* Animal, 2020. **14**(2): p. 418-424.

209. Dado-Senn, B., Ouellet, V., Lantigua, V., Van Os, J., Laporta, J., *Methods for detecting heat stress in hutch-housed dairy calves in a continental climate.* Journal of Dairy Science, 2023. **106**(2): p. 1039-1050.

210. Dado-Senn, B., Ouellet, V., Dahl, G.E., Laporta, J., *Methods for assessing heat stress in preweaned dairy calves exposed to chronic heat stress or continuous cooling.* Journal of dairy science, 2020. **103**(9): p. 8587-8600.

211. Kovács, L., Kézér, F.L., Póti, P., Boros, N., Nagy, K., *Short communication: Upper critical temperature-humidity index for dairy calves based on physiological stress variables.* Journal of dairy science, 2020. **103**(3): p. 2707-2710.

212. Pinto, S., Hoffmann, G., Ammon, C., Amon, T., *Critical THI thresholds based on the physiological parameters of lactating dairy cows.* Journal of thermal biology, 2020. **88**.

213. Dalcin, V.C., Fischer, V., Daltro, D.d.S., Alfonzo, E.P.M., Stumpf, M.T., Kolling, G.J., Silva, M.V.G.B.d., McManus, C., *Physiological parameters for thermal stress in dairy cattle.* Revista brasileira de zootecnia, 2016. **45**(8): p. 458-465.

214. Ouellet, V., Toledo, I.M., Dado-Senn, B., Dahl, G.E., Laporta, J., *Critical Temperature-Humidity Index Thresholds for Dry Cows in a Subtropical Climate.* Frontiers in Animal Science, 2021. **2**(28).

215. Shu, H., Guo, L., Bindelle, J., Fang, T., Xing, M., Sun, F., Chen, X., Zhang, W., Wang, W., *Evaluation of environmental and physiological indicators in lactating dairy cows exposed to heat stress.* International Journal of Biometeorology, 2022. **66**(6): p. 1219-1232.

216. Nabenishi, H., Ohta, H., Nishimoto, T., Morita, T., Ashizawa, K., Tsuzuki, Y., *Effect of the Temperature-Humidity Index on Body Temperature and Conception Rate of Lactating Dairy Cows in Southwestern Japan.* Journal of Reproduction and Development, 2011. **57**(4): p. 450-456.

217. Hillman, P.E., Lee, C.N., Willard, S.T., *Body Temperature Versus Microclimate Selection in Heat Stressed Dairy Cows* Transactions of the ASABE 2005. **48**(2): p. 795-801.

218. Yan, G., Shi, Z., Li, H., *Critical Temperature-Humidity Index Thresholds Based on Surface Temperature for Lactating Dairy Cows in a Temperate Climate.* Agriculture, 2021. **11**(10): p. 970.

219. Kim, N.Y., Moon, S.H., Kim, S.J., Kim, E.K., Oh, M., Tang, Y., Jang, S.Y., *Summer season temperature-humidity index threshold for infrared thermography in Hanwoo (Bos taurus coreanae) heifers.* Asian - Australasian Journal of Animal Sciences, 2020. **33**(10): p. 1691.

220. Toledo, I.M., Fabris, T.F., Tao, S., Dahl, G.E., *When do dry cows get heat stressed? Correlations of rectal temperature, respiration rate, and performance.* JDS Communications, 2020. **1**(1): p. 21-24.

221. Müschner-Siemens, T., Hoffmann, G., Ammon, C., Amon, T., *Daily rumination time of lactating dairy cows under heat stress conditions.* Journal of Thermal Biology, 2020. **88**: p. 102484.

222. NRC, *A Guide to Environmental Research on Animals*. 1971, National Academy Press: Washington, DC, USA. p. 374.

223. Bianca, W., *Relative Importance of Dry- and Wet-Bulb Temperatures in Causing Heat Stress in Cattle.* Nature, 1962. **195**(4838): p. 251-252.

224. Schütz, K.E., Rogers, A.R., Poulouin, Y.A., Cox, N.R., Tucker, C.B., *The amount of shade influences the behavior and physiology of dairy cattle.* Journal of dairy science, 2010. **93**(1): p. 125-133.

225. Yan, G., Liu, K., Hao, Z., Shi, Z., Li, H., *The effects of cow-related factors on rectal temperature, respiration rate, and temperature-humidity index thresholds for lactating cows exposed to heat stress.* Journal of thermal biology, 2021. **100**: p. 103041.

226. Igono, M., Johnson, H., Steevens, B., Hainen, W., Shanklin, M., *Effect of season on milk temperature, milk growth hormone, prolactin, and somatic cell counts of lactating cattle.* International Journal of Biometeorology, 1988. **32**(3): p. 194-200.

227. Amamou, H., Beckers, Y., Mahouachi, M., Hammami, H., *Thermotolerance indicators related to production and physiological responses to heat stress of holstein cows.* Journal of thermal biology, 2019. **82**: p. 90-98.

228. Kabuga, J.D., *The influence of thermal conditions on rectal temperature, respiration rate and pulse rate of lactating Holstein-Friesian cows in the humid tropics.* International Journal of Biometeorology, 1992. **36**(3): p. 146-150.

229. Martello, L., Savastano Junior, H., Silva, S., Balieiro, J., *Alternative body sites for heat stress measurement in milking cows under tropical conditions and their relationship to the thermal discomfort of the animals.* International Journal of Biometeorology, 2010. **54**(6): p. 647-652.

230. Du Preez, J.H., *Parameters for the determination and evaluation of heat stress in dairy cattle in South Africa.* Onderstepoort journal of veterinary research, 2000. **67**(4): p. 263-271.

231. Sammad, A., Wang, Y.J., Umer, S., Lirong, H., Khan, I., Khan, A., Ahmad, B., Wang, Y., *Nutritional Physiology and Biochemistry of Dairy Cattle under the Influence of Heat Stress: Consequences and Opportunities.* Animals, 2020. **10**(5): p. 793.

232. Fuquay, J.W., *Heat Stress as it Affects Animal Production.* Journal of animal science, 1981. **52**(1): p. 164-174.

233. Spiers, D.E., Spain, J.N., Ellersieck, M.R., Lucy, M.C., *Strategic application of convective cooling to maximize the thermal gradient and reduce heat stress response in dairy cows.* Journal of dairy science, 2018. **101**(9): p. 8269-8283.

234. Wang, X., Bjerg, B.S., Choi, C.Y., Zong, C., Zhang, G., *A review and quantitative assessment of cattle-related thermal indices.* Journal of thermal biology, 2018. **77**: p. 24-37.

235. Herbut, P., Angrecka, S., Walczak, J., *Environmental parameters to assessing of heat stress in dairy cattle—a review.* International journal of biometeorology, 2018. **62**(12): p. 2089-2097.

236. Li, S., Gebremedhin, K.G., Lee, C., Collier, R., *Evaluation of Thermal Stress Indices for Cattle*, in *American Society of Agricultural and Biological Engineers Annual International Meeting*. 2009: NV, USA.

237. Kovács, L., Kézér, F.L., Ruff, F., Szenci, O., Jurkovich, V., *Association between human and animal thermal comfort indices and physiological heat*

stress indicators in dairy calves. Environmental Research, 2018. **166**: p. 108-111.

238. Yan, G., Li, H., Zhao, W., Shi, Z., *Evaluation of thermal indices based on their relationships with some physiological responses of housed lactating cows under heat stress.* International journal of biometeorology, 2020. **64**(12): p. 2077-2091.

239. Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., Li, B., *Modelling of heat stress in a robotic dairy farm. Part 1: Thermal comfort indices as the indicators of production loss.* Biosystems engineering, 2020. **199**: p. 27-42.

240. Kovács, L., Kézér, F.L., Ruff, F., Jurkovich, V., Szenci, O., *Assessment of heat stress in 7-week old dairy calves with non-invasive physiological parameters in different thermal environments.* PLOS ONE, 2018. **13**(7): p. e0200622.

241. Buffington, D.E., Collazo-Arocho, A., Canton, G.H., Pitt, D., Thatcher, W.W., Collier, R.J., *Black Globe-Humidity Index (BGHI) as Comfort Equation for Dairy Cows.* Transactions of the ASAE, 1981. **24**(3): p. 711-0714.

242. Mader, T.L., Johnson, L.J., Gaughanf, J.B., *A comprehensive index for assessing environmental stress in animals.* Journal of animal science, 2010. **88**(6): p. 2153-2165.

243. Bohmanova, J., Misztal, I., Cole, J.B., *Temperature-Humidity Indices as Indicators of Milk Production Losses due to Heat Stress.* Journal of dairy science, 2007. **90**(4): p. 1947-1956.

244. Shock, D.A., LeBlanc, S.J., Leslie, K.E., Hand, K., Godkin, M.A., Coe, J.B., Kelton, D.F., *Studying the relationship between on-farm environmental conditions and local meteorological station data during the summer.* Journal of Dairy Science, 2016. **99**(3): p. 2169-2179.

245. Heins, B.J., Pereira, G.M., Sharpe, K.T., *Precision technologies to improve dairy grazing systems.* JDS Communications, 2023. **4**(4): p. 318-323.

246. Ren, K., Bernes, G., Hetta, M., Karlsson, J., *Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning.* Journal of Systems Architecture, 2021. **116**: p. 102139.

247. Xu, J., Zhou, S., Xia, F., Xu, A., Ye, J., *Research on the lying pattern of grouped pigs using unsupervised clustering and deep learning.* Livestock Science, 2022. **260**: p. 104946.

248. Gorczyca, M.T., Milan, H.F.M., Maia, A.S.C., Gebremedhin, K.G., *Machine learning algorithms to predict core, skin, and hair-coat*

temperatures of piglets. Computers and Electronics in Agriculture, 2018. **151**: p. 286-294.

249. Astudillo, R., Frazier, P.I. *Thinking inside the box: A tutorial on grey-box Bayesian optimization*. in *Winter Simulation Conference*. 2021. IEEE.

250. Mondaca, M.R., Choi, C.Y., Cook, N.B., *Understanding microenvironments within tunnel-ventilated dairy cow freestall facilities: Examination using computational fluid dynamics and experimental validation.* Biosystems engineering, 2019. **183**: p. 70-84.

251. Spiers, D., Spain, J., Leonard, M., Lucy, M. *Effect of cooling strategy and night temperature on dairy cow performance during heat stress*. in *Livestock Environment VI, Proceedings of the 6th International Symposium 2001*. 2001. American Society of Agricultural and Biological Engineers.

252. Islam, M.A., Lomax, S., Doughty, A., Islam, M.R., Jay, O., Thomson, P., Clark, C., *Automated Monitoring of Cattle Heat Stress and Its Mitigation.* Frontiers in Animal Science, 2021. **2**(60).

253. Fuentes, S., Gonzalez Viejo, C., Cullen, B., Tongson, E., Chauhan, S.S., Dunshea, F.R., *Artificial Intelligence Applied to a Robotic Dairy Farm to Model Milk Productivity and Quality based on Cow Data and Daily Environmental Parameters.* Sensors, 2020. **20**(10): p. 2975.

254. Miura, R., Yoshioka, K., Miyamoto, T., Nogami, H., Okada, H., Itoh, T., *Estrous detection by monitoring ventral tail base surface temperature using a wearable wireless sensor in cattle.* Animal reproduction science, 2017. **180**: p. 50-57.

255. Wang, L., Zhang, M., Li, Y., Xia, J., Ma, R., *Wearable multi-sensor enabled decision support system for environmental comfort evaluation of mutton sheep farming.* Computers and Electronics in Agriculture, 2021. **187**: p. 106302.

256. Bai, S., Teng, G., Du, X., Du, X., *Design and implementation on real-time monitoring system of laying hens environmental comfort based on LabVIEW.* Transactions of the Chinese Society of Agricultural Engineering, 2017. **33**: p. 237-244.

257. Tan, J.-H., Ng, E.Y.K., Rajendra Acharya, U., Chee, C., *Infrared thermography on ocular surface temperature: A review.* Infrared Physics & Technology, 2009. **52**(4): p. 97-108.

258. Halachmi, I., Guarino, M., *Precision livestock farming: a 'per animal' approach using advanced monitoring technologies.* Animal, 2016. **10**(9): p. 1482-1483.

259. Ma, S., Yao, Q., Masuda, T., Higaki, S., Yoshioka, K., Arai, S., Takamatsu, S., Itoh, T., *Development of Noncontact Body Temperature Monitoring and*

Prediction System for Livestock Cattle. IEEE Sensors Journal, 2021. **21**(7): p. 9367-9376.

260. Gloster, J., Ebert, K., Gubbins, S., Bashiruddin, J., Paton, D.J., *Normal variation in thermal radiated temperature in cattle: implications for foot-and-mouth disease detection.* BMC Veterinary Research, 2011. **7**(1): p. 73.

261. Chen, X., Ogdahl, W., Hanna, L., Dahlen, C., Riley, D., Wagner, S., Berg, E., Sun, X., *Evaluation of beef cattle temperament by eye temperature using infrared thermography technology.* Computers and Electronics in Agriculture, 2021. **188**: p. 106321.

262. Uddin, J., Phillips, C.J.C., Auboeuf, M., McNeill, D.M., *Relationships between body temperatures and behaviours in lactating dairy cows.* Applied Animal Behaviour Science, 2021. **241**: p. 105359.

263. Montanholi, Y.R., Swanson, K.C., Schenkel, F.S., McBride, B.W., Caldwell, T.R., Miller, S.P., *On the determination of residual feed intake and associations of infrared thermography with efficiency and ultrasound traits in beef bulls.* Livestock Science, 2009. **125**(1): p. 22-30.

264. Cuthbertson, H., Tarr, G., Loudon, K., Lomax, S., White, P., McGreevy, P., Polkinghorne, R., González, L.A., *Using infrared thermography on farm of origin to predict meat quality and physiological response in cattle (Bos Taurus) exposed to transport and marketing.* Meat Science, 2020. **169**: p. 108173.

265. Lowe, G., Sutherland, M., Waas, J., Schaefer, A., Cox, N., Stewart, M., *Infrared Thermography-A Non-Invasive Method of Measuring Respiration Rate in Calves.* Animals, 2019. **9**(8).

266. Zhang, X., Kang, X., Feng, N., Liu, G., *Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector.* Computers and Electronics in Agriculture, 2020. **178**: p. 105754.

267. Kim, S., Hidaka, Y., *Breathing Pattern Analysis in Cattle Using Infrared Thermography and Computer Vision.* Animals, 2021. **11**(1): p. 207.

268. Jaddoa, M.A., Gonzalez, L., Cuthbertson, H., Al-Jumaily, A., *Multiview Eye Localisation to Measure Cattle Body Temperature Based on Automated Thermal Image Processing and Computer Vision.* Infrared Physics and Technology, 2021. **119**.

269. Wang, Y., Kang, X., He, Z., Feng, Y., Liu, G., *Accurate detection of dairy cow mastitis with deep learning technology: a new and comprehensive detection method based on infrared thermal images.* animal, 2022. **16**(10): p. 100646.

270. Chu, M., Li, Q., Wang, Y., Zeng, X., Si, Y., Liu, G., *Fusion of udder temperature and size features for the automatic detection of dairy cow*

mastitis using deep learning. Computers and Electronics in Agriculture, 2023. **212**: p. 108131.

271. LeCun, Y., Bengio, Y., Hinton, G., *Deep learning.* Nature, 2015. **521**(7553): p. 436-444.

272. Lin, J., Yang, H., Chen, D., Zeng, M., Wen, F., Yuan, L. *Face parsing with roi tanh-warping*. in *the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

273. Kütük, Z., Algan, G. *Semantic segmentation for thermal images: A comparative survey*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

274. Khan, M., El Saddik, A., Alotaibi, F.S., Pham, N.T., *AAD-Net: Advanced end-to-end signal processing system for human emotion detection & recognition using attention-based deep echo state network.* Knowledge-Based Systems, 2023. **270**: p. 110525.

275. Khan, M., Saeed, M., Saddik, A.E., Gueaieb, W. *ARTriViT: Automatic Face Recognition System Using ViT-Based Siamese Neural Networks with a Triplet Loss*. in *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*. 2023.

276. Montanholi, Y.R., Lim, M., Macdonald, A., Smith, B.A., Goldhawk, C., Schwartzkopf-Genswein, K., Miller, S.P., *Technological, environmental and biological factors: referent variance values for infrared imaging of the bovine.* Journal of Animal Science and Biotechnology, 2015. **6**(1): p. 27.

277. Han, J., Siegford, J., Colbry, D., Lesiyon, R., Bosgraaf, A., Chen, C., Norton, T., Steibel, J.P., *Evaluation of computer vision for detecting agonistic behavior of pigs in a single-space feeding stall through blocked cross-validation strategies.* Computers and Electronics in Agriculture, 2023. **204**: p. 107520.

278. Ronneberger, O., Fischer, P., Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015. Cham: Springer International Publishing.

279. Simonyan, K., Zisserman, A., *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.

280. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C. *Ghostnet: More features from cheap operations*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

281. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q. *ECA-Net: Efficient channel attention for deep convolutional neural networks*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

282. Deng, J., Dong, W., Socher, R., Li, L.J., Kai, L., Li, F.-F. *ImageNet: A large-scale hierarchical image database*. in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009.

283. Long, J., Shelhamer, E., Darrell, T. *Fully convolutional networks for semantic segmentation*. in *the IEEE conference on computer vision and pattern recognition*. 2015.

284. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. *Pyramid scene parsing network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

285. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C. *Mobilenetv2: Inverted residuals and linear bottlenecks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

286. He, K., Zhang, X., Ren, S., Sun, J. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

287. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. *Encoder-decoder with atrous separable convolution for semantic image segmentation*. in *the European conference on computer vision*. 2018.

288. Chollet, F. *Xception: Deep learning with depthwise separable convolutions*. in *the IEEE conference on computer vision and pattern recognition*. 2017.

289. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., *SegFormer: Simple and efficient design for semantic segmentation with transformers.* Advances in Neural Information Processing Systems, 2021. **34**: p. 12077-12090.

290. Altman, D.G., Bland, J.M., *Measurement in Medicine: The Analysis of Method Comparison Studies.* Journal of the Royal Statistical Society. Series D (The Statistician), 1983. **32**(3): p. 307-317.

291. Muggeo, V.M., *Segmented: an R package to fit regression models with broken-line relationships.* R news, 2008. **8**(1): p. 20-25.

292. Zheng, Z., Hu, Y., Guo, T., Qiao, Y., He, Y., Zhang, Y., Huang, Y., *AGHRNet: An attention ghost-HRNet for confirmation of catch-and-shake locations in jujube fruits vibration harvesting.* Computers and Electronics in Agriculture, 2023. **210**: p. 107921.

293. Burhans, W.S., Rossiter Burhans, C.A., Baumgard, L.H., *Invited review: Lethal heat stress: The putative pathophysiology of a deadly disorder in dairy cattle.* Journal of Dairy Science, 2022. **105**(5): p. 3716-3735.

294. Zou, K., Chen, X., Wang, Y., Zhang, C., Zhang, F., *A modified U-Net with a specific data argumentation method for semantic segmentation of weed*

images in the field. Computers and Electronics in Agriculture, 2021. **187**: p. 106242.

295. Hoffmann, G., Schmidt, M., Ammon, C., Rose-Meierhöfer, S., Burfeind, O., Heuwieser, W., Berg, W., *Monitoring the body temperature of cows and calves using video recordings from an infrared thermography camera.* Veterinary Research Communications, 2013. **37**(2): p. 91-99.

296. Pacheco, V.M., Sousa, R.V., Sardinha, E.J.S., Rodrigues, A.V.S., Brown-Brandl, T.M., Martello, L.S., *Deep learning-based model classifies thermal conditions in dairy cows using infrared thermography.* Biosystems Engineering, 2022. **221**: p. 154-163.

297. Church, J.S., Hegadoren, P.R., Paetkau, M.J., Miller, C.C., Regev-Shoshani, G., Schaefer, A.L., Schwartzkopf-Genswein, K.S., *Influence of environmental factors on infrared eye temperature measurements in cattle.* Research in Veterinary Science, 2014. **96**(1): p. 220-226.

298. Muniz, P.R., Magalhães, R.d.S., Cani, S.P.N., Donadel, C.B., *Non-contact measurement of angle of view between the inspected surface and the thermal imager.* Infrared Physics & Technology, 2015. **72**: p. 77-83.

299. Wang, X., Hu, F., Yang, R., Wang, K., *An Infrared Temperature Correction Method for the Skin Temperature of Pigs in Infrared Images.* Agriculture, 2023. **13**(3): p. 520.

300. McManus, R., Boden, L.A., Weir, W., Viora, L., Barker, R., Kim, Y., McBride, P., Yang, S., *Thermography for disease detection in livestock: A scoping review.* Frontiers in Veterinary Science, 2022. **9**.

301. Wang, F.-K., Shih, J.-Y., Juan, P.-H., Su, Y.-C., Wang, Y.-C., *Non-Invasive Cattle Body Temperature Measurement Using Infrared Thermography and Auxiliary Sensors.* Sensors, 2021. **21**(7): p. 2425.

302. Charlton, M., Stanley, S.A., Whitman, Z., Wenn, V., Coats, T.J., Sims, M., Thompson, J.P., *The effect of constitutive pigmentation on the measured emissivity of human skin.* PLOS ONE, 2020. **15**(11): p. e0241843.

303. Isola, J.V.V., Menegazzi, G., Busanello, M., dos Santos, S.B., Agner, H.S.S., Sarubbi, J., *Differences in body temperature between black-and-white and red-and-white Holstein cows reared on a hot climate using infrared thermography.* Journal of Thermal Biology, 2020. **94**: p. 102775.

304. Anzures-Olvera, F., Véliz, F.G., de Santiago, A., García, J.E., Mellado, J., Macías-Cruz, U., Avendaño-Reyes, L., Mellado, M., *The impact of hair coat color on physiological variables, reproductive performance and milk yield of Holstein cows in a hot environment.* Journal of Thermal Biology, 2019. **81**: p. 82-88.

305. Soerensen, D.D., Clausen, S., Mercer, J.B., Pedersen, L.J., *Determining the emissivity of pig skin for accurate infrared thermography.* Computers and Electronics in Agriculture, 2014. **109**: p. 52-58.

306. White, B.J., Goehl, D.R., Amrine, D.E., *Comparison of a Remote Early Disease Identification (REDI) System to Visual Observations to Identify Cattle with Bovine Respiratory Diseases.* International Journal of Applied Research in Veterinary Medicine, 2015. **13**(1).

307. Fuentes, S., Gonzalez Viejo, C., Chauhan, S.S., Joy, A., Tongson, E., Dunshea, F.R., *Non-Invasive Sheep Biometrics Obtained by Computer Vision Algorithms and Machine Learning Modeling Using Integrated Visible/Infrared Thermal Cameras.* Sensors, 2020. **20**(21): p. 6334.

308. Shu, H., Wang, W., Guo, L., Bindelle, J., *Recent Advances on Early Detection of Heat Strain in Dairy Cows Using Animal-Based Indicators: A Review.* Animals, 2021. **11**(4): p. 980.

309. Massaroni, C., Lo Presti, D., Formica, D., Silvestri, S., Schena, E., *Non-Contact Monitoring of Breathing Pattern and Respiratory Rate via RGB Signal Measurement.* Sensors, 2019. **19**(12): p. 2758.

310. Wang, M., Li, X., Larsen, M.L.V., Liu, D., Rault, J.-L., Norton, T., *A computer vision-based approach for respiration rate monitoring of group housed pigs.* Computers and Electronics in Agriculture, 2023. **210**: p. 107899.

311. Mantovani, R.R., Menezes, G.L., Dórea, J.R.R., *Predicting respiration rate in unrestrained dairy cows using image analysis and Fast Fourier Transform.* JDS Communications, 2023.

312. Jorquera-Chavez, M., Fuentes, S., Dunshea, F.R., Warner, R.D., Poblete, T., Morrison, R.S., Jongman, E.C., *Remotely Sensed Imagery for Early Detection of Respiratory Disease in Pigs: A Pilot Study.* Animals, 2020. **10**(3): p. 451.

313. van Erp-van der Kooij, E., Almalik, O., Cavestany, D., Roelofs, J., van Eerdenburg, F., *Lying postures of dairy cows in cubicles and on pasture.* Animals, 2019. **9**(4): p. 183.

314. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., Montes, D., *Ultralytics/yolov5: v7. 0-YOLOv5 SotA realtime instance segmentation.* Zenodo, 2022.

315. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L. *Microsoft coco: Common objects in context*. in *European conference on computer vision*. 2014. Springer.

316. Guo, T., Lin, Q., Allebach, J., *Remote estimation of respiration rate by optical flow using convolutional neural networks.* Electronic Imaging, 2021. **2021**(8): p. 267-1-267-11.

317. Salau, J., Krieter, J., *Instance Segmentation with Mask R-CNN Applied to Loose-Housed Dairy Cows in a Multi-Camera Setting.* Animals, 2020. **10**(12): p. 2402.

318. Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., Schneider, D., *Automated cattle counting using Mask R-CNN in quadcopter vision system.* Computers and Electronics in Agriculture, 2020. **171**: p. 105300.

319. Gaughan, J., Castaneda, C., *Refinement of Heat Load Index Based on Animal Factors.* Meat and Livestock Australia: North Sydney NSW, 2003.

320. Allen, J.D., Hall, L.W., Collier, R.J., Smith, J.F., *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, 2015. **98**(1): p. 118-127.

321. Collier, R.J., Dahl, G.E., VanBaale, M.J., *Major Advances Associated with Environmental Effects on Dairy Cattle.* Journal of dairy science, 2006. **89**(4): p. 1244-1253.

322. Cook, N.B., Mentink, R.L., Bennett, T.B., Burgi, K., *The Effect of Heat Stress and Lameness on Time Budgets of Lactating Dairy Cows.* Journal of Dairy Science, 2007. **90**(4): p. 1674-1682.

323. Collier, R.J., Renquist, B.J., Xiao, Y., *A 100-Year Review: Stress physiology including heat stress.* Journal of dairy science, 2017. **100**(12): p. 10367-10380.

324. Chen, C., Zhu, W., Norton, T., *Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning.* Computers and Electronics in Agriculture, 2021. **187**: p. 106255.

325. Xiao, J., Liu, G., Wang, K., Si, Y., *Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM.* Computers and Electronics in Agriculture, 2022. **194**: p. 106738.

326. Mitloehner, F., Morrow-Tesch, J., Wilson, C., Dailey, J., McGlone, J., *Behavioral sampling techniques for feedlot cattle.* Journal of animal science, 2001. **79**: p. 1189-93.

327. Uzal Seyfi, S., *Seasonal variation of the lying and standing behavior indexes of dairy cattle at different daily time periods in free-stall housing.* Animal Science Journal, 2013. **84**(10): p. 708-717.

328. Overton, M.W., Sischo, W.M., Temple, G.D., Moore, D.A., *Using Time-Lapse Video Photography to Assess Dairy Cattle Lying Behavior in a Free-Stall Barn.* Journal of Dairy Science, 2002. **85**(9): p. 2407-2413.

329. Cook, N.B., Bennett, T.B., Nordlund, K.V., *Monitoring Indices of Cow Comfort in Free-Stall-Housed Dairy Herds.* Journal of Dairy Science, 2005. **88**(11): p. 3876-3885.

330. Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G., Wang, W. *Face Mask Recognition System with YOLOV5 Based on Image Recognition*. in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. 2020.

331. Mattachini, G., Riva, E., Provolo, G., *The lying and standing activity indices of dairy cows in free-stall housing.* Applied Animal Behaviour Science, 2011. **129**(1): p. 18-27.

332. Hales, T.C., *Jordan's proof of the Jordan curve theorem.* Studies in logic, grammar and rhetoric, 2007. **10**(23): p. 45-60.

333. Lv, J., Xu, H., Han, Y., Lu, W., Xu, L., Rong, H., Yang, B., Zou, L., Ma, Z., *A visual identification method for the apple growth forms in the orchard.* Computers and Electronics in Agriculture, 2022. **197**: p. 106954.

334. Tucker, C.B., Jensen, M.B., de Passillé, A.M., Hänninen, L., Rushen, J., *Invited review: Lying time and the welfare of dairy cows.* Journal of Dairy Science, 2021. **104**(1): p. 20-46.

335. Overton, M.W., Moore, D.A., Sischo, W.M. *Comparison of commonly used indices to evaluate dairy cattle lying behavior*. in *Fifth International Dairy Housing Conference for 2003*. 2003. American Society of Agricultural and Biological Engineers.

336. Eslamizad, M., Lamp, O., Derno, M., Kuhla, B., *The control of short-term feed intake by metabolic oxidation in late-pregnant and early lactating dairy cows exposed to high ambient temperatures.* Physiology & Behavior, 2015. **145**: p. 64-70.

337. Olofsson, J., *Competition for Total Mixed Diets Fed for Ad Libitum Intake Using One or Four Cows per Feeding Station.* Journal of Dairy Science, 1999. **82**(1): p. 69-79.

338. Zhou, M., Aarnink, A.J.A., Huynh, T.T.T., van Dixhoorn, I.D.E., Groot Koerkamp, P.W.G., *Effects of increasing air temperature on physiological and productive responses of dairy cows at different relative humidity and air velocity levels.* Journal of Dairy Science, 2022. **105**(2): p. 1701-1716.

339. Winckler, C., *Assessing animal welfare at the farm level: do we care sufficiently about the individual?* Animal welfare, 2019. **28**(1): p. 77-82.

340. Hart, K.D., McBride, B.W., Duffield, T.F., DeVries, T.J., *Effect of milking frequency on the behavior and productivity of lactating dairy cows.* Journal of Dairy Science, 2013. **96**(11): p. 6973-6985.

341. Ito, K., Chapinal, N., Weary, D.M., von Keyserlingk, M.A.G., *Associations between herd-level factors and lying behavior of freestall-housed dairy cows.* Journal of Dairy Science, 2014. **97**(4): p. 2081-2089.

342. Ranjitkar, S., Bu, D., Van Wijk, M., Ma, Y., Ma, L., Zhao, L., Shi, J., Liu, C., Xu, J., *Will heat stress take its toll on milk production in China?* Climatic change, 2020. **161**(4): p. 637-652.

343. Cockburn, M., *Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms.* Animals, 2020. **10**(9): p. 1690.

344. Piwczyński, D., Sitkowska, B., Kolenda, M., Brzozowski, M., Aerts, J., Schork, P.M., *Forecasting the milk yield of cows on farms equipped with automatic milking system with the use of decision trees.* Animal Science Journal, 2020. **91**(1): p. e13414.

345. Abreu, L.H.P., Yanagi Junior, T., Bahuti, M., Hernández-Julio, Y.F., Ferraz, P.F.P., *Artificial neural networks for prediction of physiological and productive variables of broilers.* Engenharia agrícola, 2020. **40**(1): p. 1-9.

346. Sousa, R.V.d., Canata, T.F., Leme, P.R., Martello, L.S., *Development and evaluation of a fuzzy logic classifier for assessing beef cattle thermal stress using weather and physiological variables.* Computers and electronics in agriculture, 2016. **127**: p. 176-183.

347. Brown-Brandl, T.M., Jones, D.D., Woldt, W.E., *Evaluating Modelling Techniques for Cattle Heat Stress Prediction.* Biosystems engineering, 2005. **91**(4): p. 513-524.

348. Pinto, S., Hoffmann, G., Ammon, C., Amon, B., Heuwieser, W., Halachmi, I., Banhazi, T., Amon, T., *Influence of Barn Climate, Body Postures and Milk Yield on the Respiration Rate of Dairy Cows.* Annals of animal science, 2019. **19**(2): p. 469-481.

349. Bovo, M., Agrusti, M., Benni, S., Torreggiani, D., Tassinari, P., *Random Forest Modelling of Milk Yield of Dairy Cows under Heat Stress Conditions.* Animals, 2021. **11**(5): p. 1305.

350. Chen, J.M., Schütz, K.E., Tucker, C.B., *Cooling cows efficiently with sprinklers: Physiological responses to water spray.* Journal of dairy science, 2015. **98**(10): p. 6925-6938.

351. Schauberger, G., Hennig-Pauka, I., Zollitsch, W., Hörtenhuber, S.J., Baumgartner, J., Niebuhr, K., Piringer, M., Knauder, W., Anders, I., Andre, K., Schönhart, M., *Efficacy of adaptation measures to alleviate heat stress*

in confined livestock buildings in temperate climate zones. Biosystems Engineering, 2020. **200**: p. 157-175.

352. Ji, B., Banhazi, T., Perano, K., Ghahramani, A., Bowtell, L., Wang, C., Li, B., *A review of measuring, assessing and mitigating heat stress in dairy cattle.* Biosystems engineering, 2020. **199**: p. 4-26.

353. Wildman, E.E., Jones, G.M., Wagner, P.E., Boman, R.L., Troutt, H.F., Lesch, T.N., *A Dairy Cow Body Condition Scoring System and Its Relationship to Selected Production Characteristics.* Journal of dairy science, 1982. **65**(3): p. 495-501.

354. de Andrade Ferrazza, R., Mogollón Garcia, H.D., Vallejo Aristizábal, V.H., de Souza Nogueira, C., Veríssimo, C.J., Sartori, J.R., Sartori, R., Pinheiro Ferreira, J.C., *Thermoregulatory responses of Holstein cows exposed to experimentally induced heat stress.* Journal of Thermal Biology, 2017. **66**: p. 68-80.

355. Hahn, G.L., *Dynamic responses of cattle to thermal heat loads.* Journal of Animal Science, 1999. **77**(suppl_2): p. 10-20.

356. Howell, K., Dudek, K., Soroko, M., *Thermal camera performance and image analysis repeatability in equine thermography.* Infrared Physics & Technology, 2020. **110**: p. 103447.

357. Shu, H., Li, Y., Fang, T., Xing, M., Sun, F., Chen, X., Bindelle, J., Wang, W., Guo, L., *Evaluation of the Best Region for Measuring Eye Temperature in Dairy Cows Exposed to Heat Stress.* Frontiers in Veterinary Science, 2022. **9**: p. 857777.

358. Shah, N.K., Gemperline, P.J., *A program for calculating Mahalanobis distances using principal component analysis.* TrAC Trends in Analytical Chemistry, 1989. **8**(10): p. 357-361.

359. Soyeurt, H., Grelet, C., McParland, S., Calmels, M., Coffey, M., Tedde, A., Delhez, P., Dehareng, F., Gengler, N., *A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra.* Journal of Dairy Science, 2020. **103**(12): p. 11585-11596.

360. Gareth, J., Daniela, W., Trevor, H., Robert, T., *An introduction to statistical learning: with applications in R*. 2013: Spinger.

361. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.

362. Friedman, J.H., *Greedy function approximation: A gradient boosting machine.* The Annals of Statistics, 2001. **29**(5): p. 1189-1232, 44.

363. McCulloch, W.S., Pitts, W., *A logical calculus of the ideas immanent in nervous activity.* The bulletin of mathematical biophysics, 1943. **5**(4): p. 115-133.

364. Zou, H., Hastie, T., *Regularization and variable selection via the elastic net.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005. **67**(2): p. 301-320.

365. Dunn, R.J.H., Mead, N.E., Willett, K.M., Parker, D.E., *Analysis of heat stress in UK dairy cattle and impact on milk yields.* Environmental Research Letters, 2014. **9**(6): p. 064006.

366. Armstrong, D.V., *Heat Stress Interaction with Shade and Cooling.* Journal of dairy science, 1994. **77**(7): p. 2044-2050.

367. Wang, X., Gao, H., Gebremedhin, K.G., Bjerg, B.S., Van Os, J., Tucker, C.B., Zhang, G., *A predictive model of equivalent temperature index for dairy cattle (ETIC).* Journal of thermal biology, 2018. **76**: p. 165-170.

368. Cha, G.-W., Moon, H.-J., Kim, Y.-C., *Comparison of Random Forest and Gradient Boosting Machine Models for Predicting Demolition Waste Based on Small Datasets and Categorical Variables.* International Journal of Environmental Research and Public Health, 2021. **18**(16): p. 8530.

369. Chaibub Neto, E., Pratap, A., Perumal, T.M., Tummalacherla, M., Snyder, P., Bot, B.M., Trister, A.D., Friend, S.H., Mangravite, L., Omberg, L., *Detecting the impact of subject characteristics on machine learning-based diagnostic applications.* npj Digital Medicine, 2019. **2**(1): p. 99.

370. Janni, K.A., *Modeling lactating cow respiration rates during heat stress based on dry-bulb and dew-point temperatures, daily milk production and air velocity*, in *2019 ASABE Annual International Meeting*. 2019, ASABE: St. Joseph, MI. p. 1.

371. Dado-Senn, B., Laporta, J., Dahl, G.E., *Carry over effects of late-gestational heat stress on dairy cattle progeny.* Theriogenology, 2020. **154**: p. 17-23.

372. Ahmed, B., Younas, U., Asar, T., Dikmen, S., Hansen, P., Dahl, G., *Cows exposed to heat stress during fetal life exhibit improved thermal tolerance.* Journal of Animal Science, 2017. **95**.

373. Ferreira, F.C., De Vries, A., *Effects of season and herd milk volume on somatic cell counts of Florida dairy farms.* Journal of Dairy Science, 2015. **98**(6): p. 4182-4197.

374. Choukeir, A.I., Kovács, L., Kézér, L.F., Buják, D., Szelényi, Z., Abdelmegeid, M.K., Gáspárdy, A., Szenci, O., *Evaluation of a commercial intravaginal thermometer to predict calving in a Hungarian Holstein-Friesian dairy farm.* Reproduction in Domestic Animals, 2020. **55**(11): p. 1535-1540.

375. Nordlund, K., Strassburg, P., Bennett, T., Oetzel, G., Cook, N.B., *Thermodynamics of standing and lying behavior in lactating dairy cows in*

freestall and parlor holding pens during conditions of heat stress. Journal of Dairy Science, 2019. **102**.

376. Yagis, E., Atnafu, S.W., García Seco de Herrera, A., Marzi, C., Scheda, R., Giannelli, M., Tessa, C., Citi, L., Diciotti, S., *Effect of data leakage in brain MRI classification using 2D convolutional neural networks.* Scientific Reports, 2021. **11**(1): p. 22544.

377. Tampu, I.E., Eklund, A., Haj-Hosseini, N., *Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images.* Scientific Data, 2022. **9**(1): p. 580.

378. Eelbode, T., Sinonquel, P., Maes, F., Bisschops, R., *Pitfalls in training and validation of deep learning systems.* Best Practice & Research Clinical Gastroenterology, 2021. **52-53**: p. 101712.

379. Tedde, A., Grelet, C., Ho, P.N., Pryce, J.E., Hailemariam, D., Wang, Z., Plastow, G., Gengler, N., Froidmont, E., Dehareng, F., Bertozzi, C., Crowe, M.A., Soyeurt, H., Consortium, o.b.o.t.G., *Multiple Country Approach to Improve the Test-Day Prediction of Dairy Cows' Dry Matter Intake.* Animals, 2021. **11**(5): p. 1316.

380. Rodenburg, J., *Robotic milking: Technology, farm design, and effects on work flow.* Journal of Dairy Science, 2017. **100**(9): p. 7729-7738.

381. Gaughan, J.B., Sharman, K., McGowan, M.R., *The effect of day-only versus day-plus-night cooling of dairy cows.* Journal of Dairy Science, 2023. **106**(7): p. 5002-5017.

382. Kopler, I., Marchaim, U., Tikász, I.E., Opaliński, S., Kokin, E., Mallinger, K., Neubauer, T., Gunnarsson, S., Soerensen, C., Phillips, C.J.C., Banhazi, T., *Farmers' Perspectives of the Benefits and Risks in Precision Livestock Farming in the EU Pig and Poultry Sectors.* Animals, 2023. **13**(18): p. 2868.

383. Schillings, J., Bennett, R., Rose, D.C., *Animal welfare and other ethical implications of Precision Livestock Farming technology.* CABI Agriculture and Bioscience, 2021. **2**(1): p. 17.

384. Schillings, J., Bennett, R., Rose, D.C., *Exploring the Potential of Precision Livestock Farming Technologies to Help Address Farm Animal Welfare.* Frontiers in Animal Science, 2021. **2**.

Appendices

Appendix 1. Supplementary materials

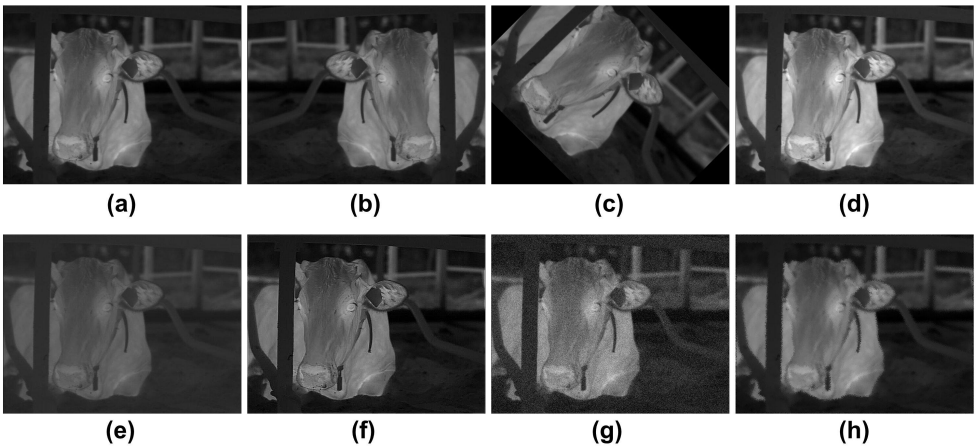


Figure A3-1: Data augmentation examples. (a) Original image, (b) flipping, (c) rotation, (d) brightness changing, (e) contrast changing, (f) sharpening, (g) Gaussian noise adding, and (h) elastic deformation.

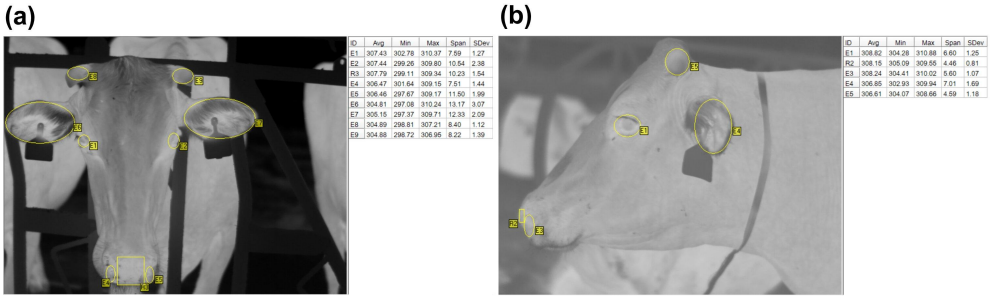


Figure A3-2: Manual temperature collection in infrared image processing software (IRBIS 3 Standard software by YSHY, Beijing, China in this case), illustrated in (a) frontal and (b) side views.

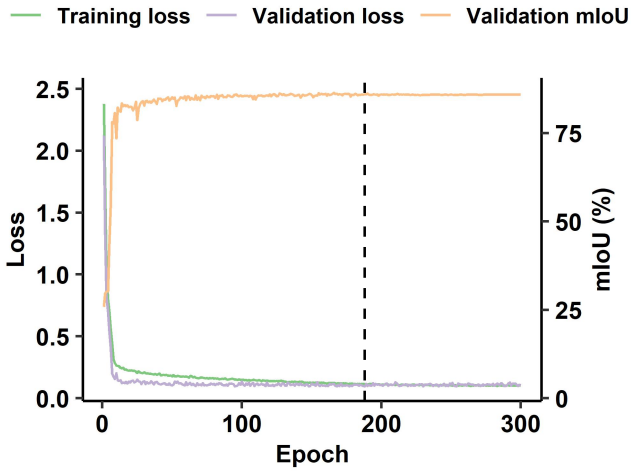


Figure A3- 3: Loss and mean Intersection over Union (mIoU) curve of the proposed architecture trained using the image-level dataset. The dashed lines show the epoch (188) with the lowest validation loss (0.096).

Table A3-1: Performance of ablation study trained using the image-level dataset (n = 100).

Model	Backbone	mIoU (%)	mPA (%)	Number of parameters (M)	Model size (MB)	FLOPs (G)	FPS
Image-level							
UNet	VGG16	83.38	90.65	24.89	96	451.81	35.4
UNet+Ghost	VGG16	84.18	92.12	17.3	66	237.24	36.1
UNet+GhostECA (proposed)	VGG16	85.07	92.01	18.43	70.3	269.63	32.6

mIoU = mean Intersection over Union; mPA = mean pixel accuracy; FLOPs = floating-point operations; FPS = frames per second.

Table A3-2: Performance of comparison study trained using the image-level dataset (n = 100).

Model	Backbone	mIoU (%)	mPA (%)	Number of parameters (M)	Model size (MB)	FLOPs (G)	FPS
FCN	VGG16	81.33	88.83	19.17	73.1	204.34	44.2
PSPNet	MobileNetV2	77.73	88.25	2.38	9.3	6.03	125.6
PSPNet	ResNet50	81.16	90.01	46.71	178	118.43	80.7
DeepLabV3+	MobileNetV2	82.36	90.9	5.82	22.4	52.9	87.2
DeepLabV3+	Xception	83.17	92.12	54.71	209	166.88	28.3
UNet	VGG16	83.38	90.65	24.89	96	451.81	35.4
UNet	ResNet50	83.03	90.88	43.93	167	184.23	53.6
SegFormer	B5	79.89	87.93	84.6	969	986.48	20.9
Proposed	VGG16	85.07	92.01	18.43	70.3	269.63	32.6

mIoU = mean Intersection over Union; mPA = mean pixel accuracy; FLOPs = floating-point operations; FPS = frames per second.

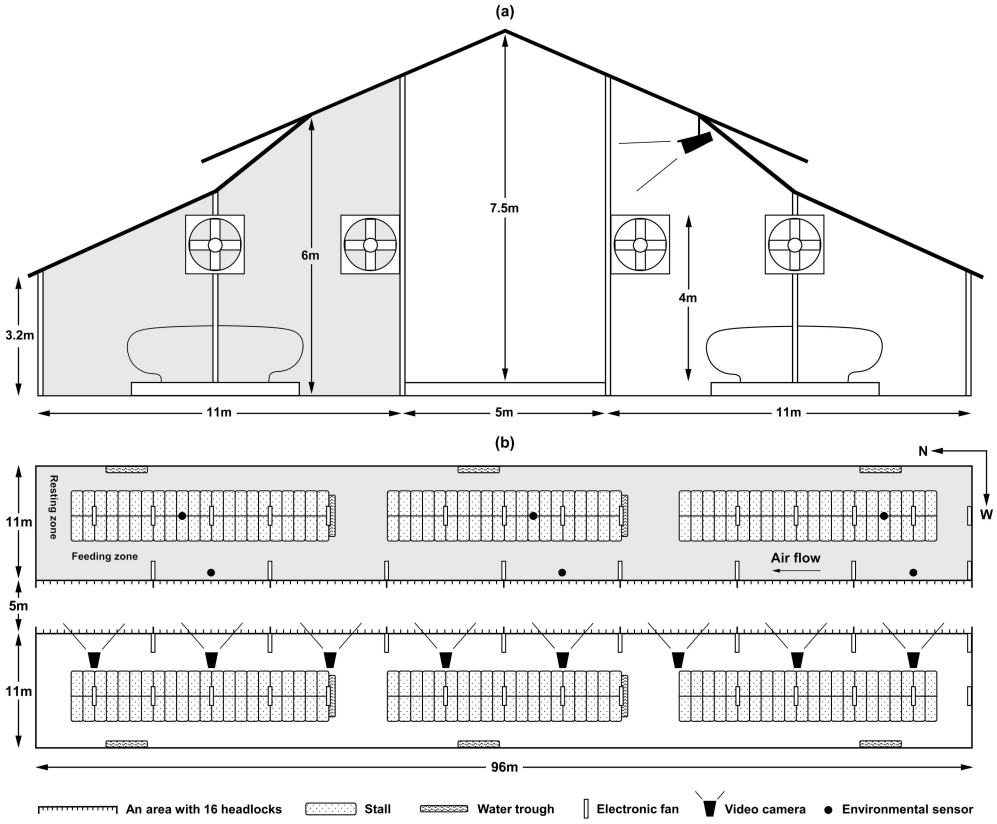


Figure A5- 1: (a) Cross-section and (b) plan of the experimental free-stall pen with side-view video cameras.

Table A5- 1: Descriptive statistics (mean ± standard deviation) of herd-level behaviour distribution and behavioural indicators obtained by applying the best performing model (YOLOv5s) on half-hourly scan samples collected during the 10-day experiment as well as the corresponding ambient temperature (Ta, °C) and temperature-humidity index (THI).

Parameter	Time of day (h)									
	10:30	**11:00**	**11:30**	**12:00**	**12:30**	**13:00**	**13:30**	**14:00**	**14:30**	**15:00**
Drinking%	4.82 ± 2.18	2.28 ± 2.72	1.12 ± 1.11	1.77 ± 1.71	1.23 ± 0.99	1.01 ± 1.00	2.75 ± 1.32	2.47 ± 2.12	2.75 ± 1.18	3.23 ± 2.08
Eating%	10.23 ± 4.01	3.82 ± 2.70	5.73 ± 3.60	8.83 ± 2.39	4.99 ± 1.70	7.68 ± 3.62	7.25 ± 4.12	12.72 ± 3.68	11.08 ± 4.17	10.63 ± 3.69
Lying%	58.34 ± 12.33	76.93 ± 4.56	74.72 ± 5.45	73.81 ± 4.04	74.48 ± 8.01	69.31 ± 7.41	65.44 ± 8.59	61.77 ± 7.40	61.09 ± 6.30	62.59 ± 13.38
Standing-in%	10.17 ± 4.75	7.50 ± 2.92	9.00 ± 3.00	8.33 ± 2.24	11.02 ± 4.63	10.57 ± 3.77	10.71 ± 3.45	9.41 ± 3.76	11.54 ± 3.71	10.66 ± 4.98
Standing-out%	16.44 ± 8.13	9.47 ± 5.39	9.42 ± 5.60	7.26 ± 5.35	8.29 ± 5.10	11.43 ± 4.95	13.85 ± 8.77	13.62 ± 7.58	13.54 ± 7.80	12.89 ± 9.39
CI	84.54 ± 8.09	91.10 ± 3.53	89.24 ± 3.46	89.94 ± 2.52	86.98 ± 5.69	86.65 ± 5.15	85.86 ± 4.80	86.80 ± 5.27	84.20 ± 4.73	84.62 ± 9.60
SUI	68.44 ± 13.46	82.08 ± 6.30	80.31 ± 6.34	82.59 ± 4.50	79.39 ± 8.26	75.91 ± 7.53	72.75 ± 9.18	72.93 ± 9.02	70.96 ± 7.13	72.54 ± 14.71
CSI	26.61 ± 10.42	16.97 ± 6.43	18.42 ± 6.17	15.59 ± 4.16	19.31 ± 7.66	22.00 ± 6.88	24.57 ± 8.43	23.03 ± 7.99	25.09 ± 6.65	23.55 ± 12.21
Ta	23.05 ± 4.33	23.45 ± 4.22	23.96 ± 4.27	24.35 ± 4.10	24.56 ± 4.27	25.10 ± 4.48	25.19 ± 4.44	25.41 ± 4.18	25.44 ± 4.36	25.71 ± 4.12
THI	70.45 ± 6.05	70.96 ± 5.77	71.55 ± 5.87	71.99 ± 5.59	72.09 ± 5.70	72.63 ± 5.83	72.56 ± 5.68	72.71 ± 5.13	72.63 ± 5.36	72.94 ± 4.96

Drinking% = percentage of cows drinking; Eating% = percentage of cows eating; Lying% = percentage of cows lying; Standing-in% = percentage of cows standing-in; Standing-out% = percentage of cows standing-out; CI = comfort index; SUI = stall-use index; CSI = cow stress index.

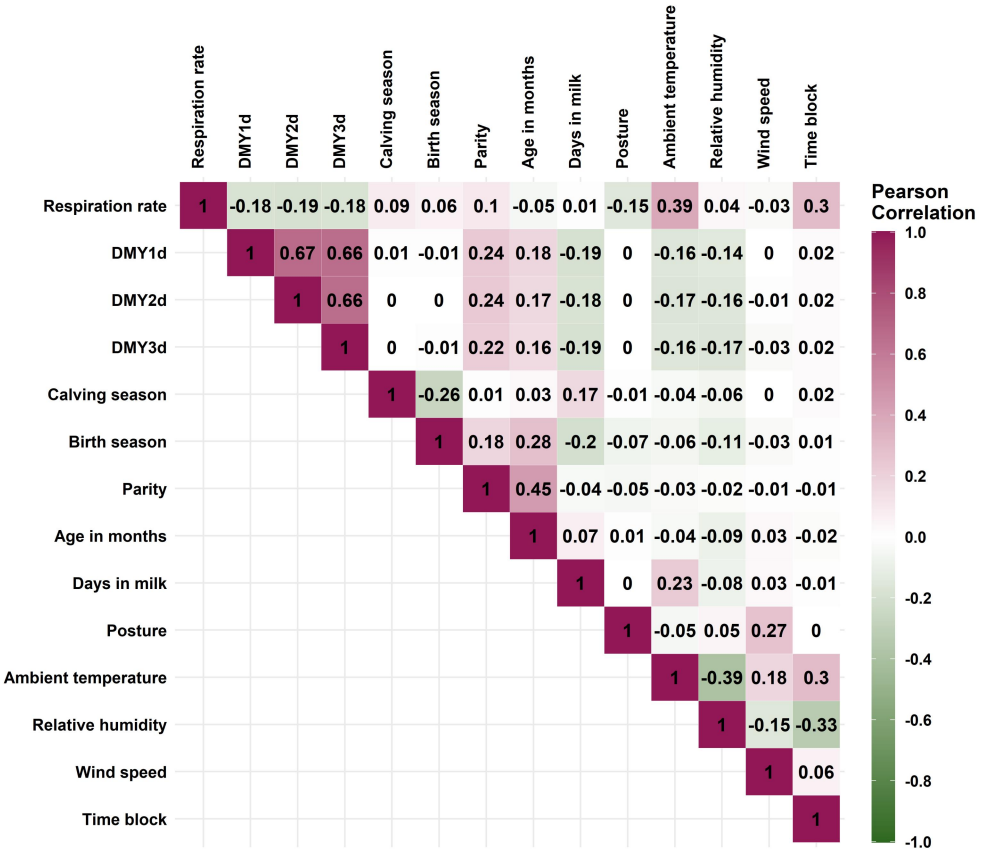


Figure A6- 1: Correlation heatmap of the respiration rate set (n = 2910). DMY1D = daily milk yield of the day before the test day; DMY2D = daily milk yield of the 2nd day before the test day; DMY3D = daily milk yield of the 3rd day before the test day.

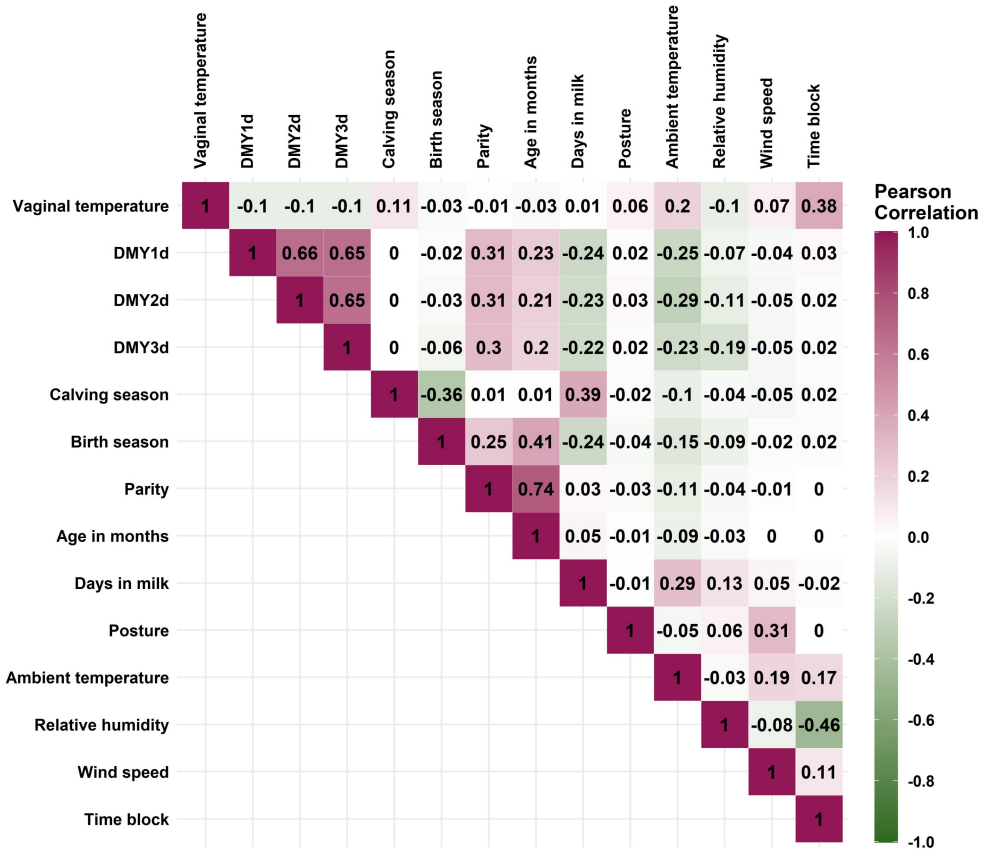


Figure A6- 2: Correlation heatmap of the vaginal temperature set (n = 1561). DMY1D = daily milk yield of the day before the test day; DMY2D = daily milk yield of the 2nd day before the test day; DMY3D = daily milk yield of the 3rd day before the test day.

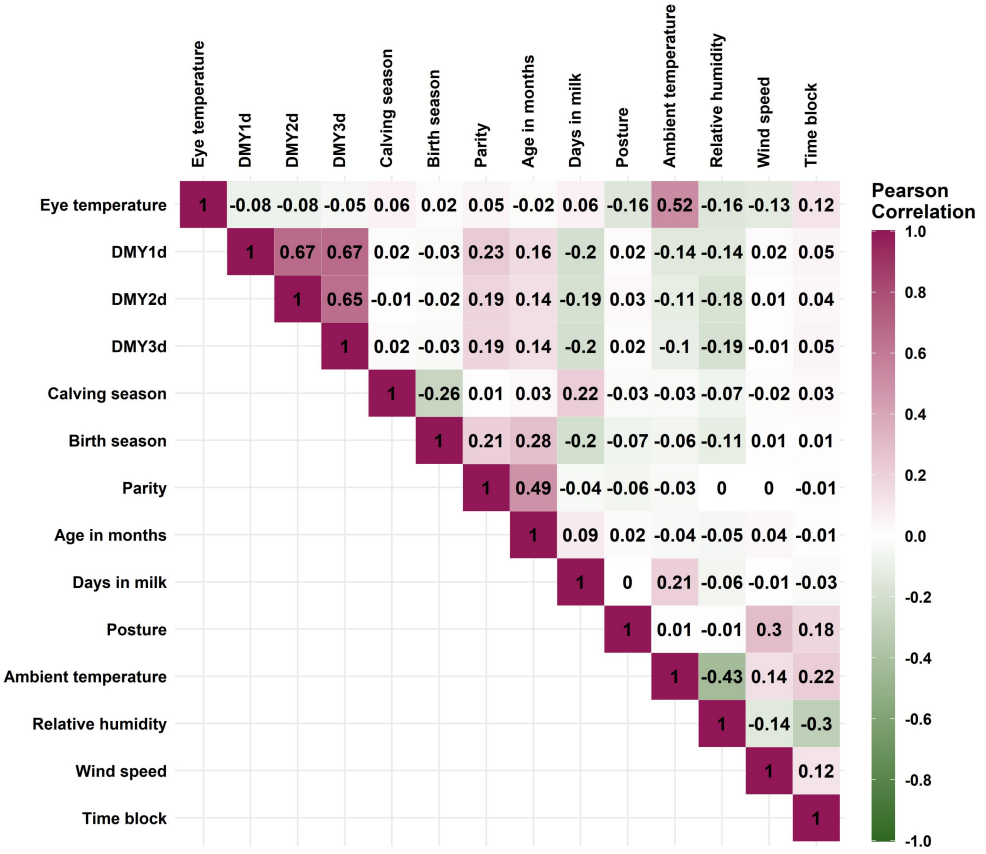


Figure A6-3: Correlation heatmap of the eye temperature set (n = 1866). DMY1D = daily milk yield of the day before the test day; DMY2D = daily milk yield of the 2nd day before the test day; DMY3D = daily milk yield of the 3rd day before the test day.

Appendix 2. List of publications

Accepted publications (peer reviewed)

1. **Shu, H.**, Wang, K., Guo, L., Bindelle, J., Wang, W. *Automated collection of facial temperatures in dairy cows via improved UNet*. Computers and Electronics in Agriculture, 2024. **220**: p. 108614.
2. **Shu, H.**, Bindelle, J., Gu, X. *Non-contact respiration rate measurement of multiple cows in a free-stall barn using computer vision methods*. Computers and Electronics in Agriculture, 2024. **218**: p. 108678.
3. **Shu, H.**, Li, Y., Bindelle, J., Jin, Z., Fang, T., Xing, M., Guo, L., Wang, W. *Predicting physiological responses of dairy cows using comprehensive variables.* Computers and Electronics in Agriculture, 2023. **207**: p. 107752.
4. **Shu, H.**, Bindelle, J., Guo, L., Gu, X. *Determining the onset of heat stress in a dairy herd based on automated behaviour recognition.* Biosystems Engineering, 2023. **226**: p. 238-251.
5. **Shu, H.**, Li, Y., Fang, T., Xing, M., Sun, F., Chen, X., Bindelle, J., Wang, W., Guo, L. *Evaluation of the Best Region for Measuring Eye Temperature in Dairy Cows Exposed to Heat Stress.* Frontiers in Veterinary Science, 2022. **9**: p. 857777.
6. **Shu, H.**, Guo, L., Bindelle, J., Fang, T., Xing, M., Sun, F., Chen, X., Zhang, W., Wang, W. *Evaluation of environmental and physiological indicators in lactating dairy cows exposed to heat stress.* International Journal of Biometeorology, 2022. **66**(6): p. 1219-1232.
7. **Shu, H.**, Gu, X. *Effect of a synthetic feline facial pheromone product on stress during transport in domestic cats: a randomised controlled pilot study.* Journal of Feline Medicine and Surgery, 2021. **24**(8): p. 691-699.
8. **Shu, H.**, Wang, W., Guo, L., Bindelle, J. *Recent Advances on Early Detection of Heat Strain in Dairy Cows Using Animal-Based Indicators: A Review.* Animals, 2021. **11**(4): p. 980.
9. Jin, Z., **Shu, H.**, Hu, T., Jiang, C., Yan, R., Qi, J., Wang, W., Guo, L. *Behavior classification and spatiotemporal analysis of grazing sheep using deep learning*. Computers and Electronics in Agriculture, 2024. **220**: p. 108894.
10. Chen, X., **Shu, H.**, Sun, F., Yao, J., Gu, X. *Impact of Heat Stress on Blood, Production, and Physiological Indicators in Heat-Tolerant and Heat-Sensitive Dairy Cows.* Animals, 2023. **13**(16): p. 2562.

11. Shi, R., **Shu, H.**, Yu, R., Wang, Y., Zhang, Z., Zhang, J., Gu, X. *Current Attitudes of Chinese Dairy Practitioners to Pain and Its Management in Intensively Raised Dairy Cattle.* Animals, 2022. **12**(22): p. 3140.
12. Li, Y., **Shu, H.**, Bindelle, J., Xu, B., Zhang, W., Jin, Z., Guo, L., Wang, W. *Classification and analysis of multiple cattle unitary behaviors and movements based on machine learning methods.* Animals, 2022. **12**(9): p. 1060.
13. Jin, Z., Guo, L., **Shu, H.**, Qi, J., Li, Y., Xu, B., Zhang, W., Wang, K., Wang, W. *Behavior Classification and Analysis of Grazing Sheep on Pasture with Different Sward Surface Heights Using Machine Learning.* Animals, 2022. **12**(14): p. 1744.
14. Chen, G., Li, C., Guo, Y., **Shu, H.**, Cao, Z., Xu, B. *Recognition of Cattle's Feeding Behaviors Using Noseband Pressure Sensor With Machine Learning.* Frontiers in Veterinary Science, 2022. **9**: p. 822621.

Presentations at international conferences (peer reviewed)

1. **Shu, H.**, Guo, L., Bindelle, J. *Predicting physiological responses of dairy cows under heat stress using machine learning methods*, in *2023 International Symposium on Animal Environment and Welfare*. 2023: Chongqing, China.
2. **Shu, H.**, Wang, K., Guo, L., Bindelle, J. *Automated collection of facial temperature in dairy cows*, in *THE 7th WORLD INTELLIGENCE CONGRESS Intelligent Agriculture Summit Forum & International Conference on Intelligent Agriculture (ICIA2023)*. 2023: Beijing, China.