MACHINE LEARNING APPLICATIONS FOR MONITORING HEAT STRESS IN LIVESTOCK

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Michael Thaddeus Gorczyca

May 2019

ABSTRACT

Heat stress is harmful to the health and productivity of livestock. Several models have been proposed for assessing heat stress, but these models assume a particular relationship, chosen by the researcher, between environment conditions and a physiological measure of heat stress. These assumptions may not accurately represent the true, underlying relationship. To account for realistic relationships, we employed machine learning algorithms to (1) rank the effect of environmental heat stressors (air temperature, solar radiation, relative humidity and wind speed) on physiological responses (skin temperature, core, body temperature and respiration rate) of dairy cows, and (2) predict core, skin, and hair-coat temperatures of piglets. The advantage of using machine learning algorithms is that they are data-driven procedures that have greater expressive power than previous modeling procedures considered (mechanistic and linear models).

This thesis is organized such that Chapter 1 demonstrates an application of machine learning algorithms to predict physiological responses (skin temperature, core-body temperature, and respiration rates) of dairy cows from environmental heat stressors (air temperature, relative humidity, solar radiation, and wind speed) as well as their interaction terms and rank the effect of these on each physiological response. Chapter 2 demonstrates an application of machine learning algorithms to predict physiological responses of piglets from environmental heat stressors for piglets.

Chapter 1 demonstrates that neural networks consistently produced the lowest root mean square error, RMSE, in predicting skin temperature, core-body temperature and respiration rate of dairy cows. The RMSE for skin temperature was 0.38 °C; for core-body temperature was 0.41 °C; and for respiration rate was 12 respirations per minute. Ranking of environmental heat stressors

showed that air temperature has the largest effect on each physiological response, followed by solar radiation, and thirdly by the interaction of air temperature and relative humidity. Wind speed and relative humidity were inconsequential heat stressors. Chapter 2 demonstrates that neural networks, gradient boosted machines, and random forests were the best algorithms, based on the lowest mean squared error on the testing dataset, to predict rectal, skin-surface, and hair coat-surface temperatures, respectively. This supports the use of machine learning algorithms to predict the physiological temperatures of piglets.

BIOGRAPHICAL SKETCH

Michael Thaddeus Gorczyca was born in Utica, New York. After completing his schoolwork at Pittsford-Mendon High School in Rochester, New York in June of 2013, Michael entered Cornell University in Ithaca, New York. He received Bachelor of Arts degrees with majors in mathematics as well as statistical science from Cornell University in May of 2017. In August of 2017, he entered the information science graduate program at Cornell University. In March of 2018, he transferred into the biological and environmental engineering graduate program at Cornell University.

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Kifle G. Gebremedhin, for his continuous support and the inspiration he has provided me during my time in the biological and environmental engineering department. He has given me a great deal of freedom in choosing both coursework and research projects. I could not have imagined having a better advisor and mentor for my M.S. study. I would also like to thank Professor Christopher M. De Sa for reviewing this work.

Ph.D. student Hugo Milan has played a major role in both transferring me into the biological engineering department as well as getting me acclimated to the department. My parents have been supportive through the process of receiving this graduate degree. They have done what they could to ensure that I had the best education possible. I owe them much gratitude.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DNN	Deep neural network
GBM	Gradient boosted machine
kg	Kilogram
MSE	Mean squared error
MNL	Minimum number of observations in a leaf for a decision tree
min.	Minute
NN	Neural network
#Breaths	Number of breaths
#Hidden	Number of hidden layers in a neural network
#Neurons	Number of neurons in a hidden layer for a neural network
#Trees	Number of trees in a random forest or gradient boosted machine
NVS	Number of variables considered in a split for a decision tree
PLM, GLM	Penalized linear model
RF	Random forest
ReLU	Rectified linear unit
RH	Relative humidity
RMSE	Root mean squared error
RR	Respiration rate
SEM	Standard error measurement
SR	Solar radiation
SHSP	Supplemental heat source power

LIST OF SYMBOLS

°C	Degrees Celsius
m^3	Cubic meter
m/s	meters per. second
T_a	Air temperature
T_c	Core temperature
T_g	Black globe temperature
T_h	Hair-coat temperature
T_s	Skin temperature
R^2	Coefficient of determination
u	Wind speed
W	Weight matrix for neural network hidden layer
W	Watt
W/m^2	Watts per. square meter
x_i	Input vector from the i^{th} observation of data
y_i	Response variable from the i^{th} observation of data
ε	Learning rate annealing (during initial training) and momentum (at later stages) hyper-parameter for ADADELTA optimizer
α	Penalty distribution parameter in the elastic net for generalized linear models
β	Regression coefficient for linear model
β_0	Intercept term for linear model
f	Activation function for neural network
h	Output of hidden layer from neural network

λ	Penalty parameter in the elastic net for generalized linear models
ρ	Momentum (memory to prior weight updates) hyper-parameter for ADADELTA optimizer
θ	Vector of offsets (bias terms) in neural network hidden layer
x	Interaction between two input variables
$\|\| \cdot \|\|_1$	L_1 norm
$\|\| \cdot \|\|_2$	L_2 norm
(a, b)	Uniform distribution from a to b
□$_d$(a, b)	Uniform discrete distribution from a to b

Chapter 1

RANKING OF ENVIRONMENTAL HEAT STRESSORS FOR DAIRY COWS USING MACHINE LEARNING ALGORITHMS

Abstract

Heat stress is harmful to the health and productivity of dairy cows. While several models have been developed to assess heat stress conditions of dairy cows, many of these models assume a particular relationship, chosen by the researcher, between environmental conditions and their physiological responses to heat stress. These assumptions may not accurately represent the true underlying effects. This study uses machine learning algorithms to evaluate how environmental heat stressors (relative humidity, RH; wind speed, u; air temperature, T_a; and solar radiation, SR) influence physiological responses, core temperature (T_c), skin temperature (T_s), and respiration rate (RR) of dairy cows. The advantage of this approach is that many machine learning algorithms automatically consider nonlinearity in data, which removes subjectivity from researchers choosing the relationship between the predictor and response variables. Four algorithms were considered in this study: penalized linear regression, random forests, gradient boosted machines, and neural networks. Random forest models were consistently the most accurate in predicting the three physiological responses. The root mean squared error, RMSE, for core temperature, T_c, was 0.359 °C; skin temperature, T_s, was 0.379 °C; and respiration rate, RR, was 10.655 respirations per minute. Air temperature had the highest ranking for effect on T_c and RR, while solar radiation had the highest ranking for effect on T_s. Wind speed and relative humidity displayed much lower effects as environmental heat stressors. Ranking of environmental heat stressors would help farmers to intervene before an anticipated stressing

environmental condition occurs. Early intervention could result in improving animal comfort and consequently increase production and reduce maintenance costs.

Introduction

Heat stress proves to be a serious issue in the dairy industry, as it negatively affects the productivity and health of dairy cows (Barash et al., 2001; West et al., 2003; Bohmanova et al., 2007). To aid in the process of detecting heat stress in dairy cows, several researchers have developed heat stress indices, which provide relationships between physiological responses (core temperature, T_c; skin temperature, T_s; or respiration rate, RR) and environmental heat stressors (relative humidity, RH; wind speed, u; air temperature, T_a; and solar radiation, SR). The underlying issue with the heat stress indices developed thus far is that they assume relationships between the environmental heat stressors and physiological responses to be linear (Bouraoui et al., 2002; Dikmen and Hansen, 2009) with a limited order of interaction terms (Schoen, 2005; Dikmen and Hansen, 2009; Wang et al., 2018a), or consider additive effects (Yano et al., 2014). In reality, the relationships among the variables are unknown, complex, and nonlinear (Hastie et al., 2003). In other words, assumptions of linearity or additivity for environmental heat stressors may be too restrictive, and models built from these assumptions may not correctly represent how environmental heat stressors affect the physiological responses of dairy cows. Furthermore, an understanding of the true relationships in heat-stress data would help to trigger intervention by farmers before an anticipated problem arises. Early intervention could improve animal comfort and consequently increase production and reduce maintenance costs.

This study develops machine learning models to predict physiological responses (T_c, T_s, RR) of dairy cows and ranks the influence of the environmental heat stressors (RH, u, T_a, and SR) on each physiological response. The advantage of this methodology is that machine learning models are data driven in determining the nonlinear relationships in the data, which removes the assumption of linearity or additivity between the environmental heat stressors and the physiological responses. To the best of our knowledge, this is the first study in which machine learning models are used to rank the effects of environmental variables on physiological responses of dairy cows.

Materials and Methods

Dataset Overview

The dataset used for this study was obtained from a study conducted at the University of California-Davis dairy facility with 19 Holstein-Friesian cows (Chen et al., 2015). Environmental heat stressors and physiological responses were recorded every 5 minutes while the cows were restrained outdoors for 1 hour each day in headlocks at an unshaded, clean feed bunk. The entire experimental trial lasted for 21 days. The environmental heat stressors were recorded with a portable weather station (WS-16; Novalynx Corp., Auburn, CA). For the physiological responses, T_c was measured using vaginally indwelling loggers (Minilog12-TX, Vemco Ltd., Bedford, NS, Canada); T_s was recorded using loggers (Thermochron iButton DS1921H, Embedded Data Systems, Lawrenceburg, KY, USA) taped to shaved skin on the "side", "upper leg", "lower leg", and "shoulder"; and RR was measured by timing flank movements and converting to breaths per. minute. The skin temperatures measured from the side, upper leg, and lower leg were averaged to define T_s for this study. The shoulder temperature was not included in the averaging because the shoulder is directly exposed to solar radiation (Chen et al., 2015; Wang et al., 2018a). The purpose of the Chen et al. (2015) study was to determine the effects of sprinkler flow rate and droplet size on heat stress abatement of cattle in a Mediterranean climate. The data used in this study was, however, from the control group, which was not sprayed with water. The data processing procedure reduced the size of the dataset from

3591 to 513 measurements. Table 1 shows the summary statistics of the dataset used in this study.

Table 1. Summary statistics of the dataset (Chen et al., 2015) prior to data normalization. Number of samples = 513. Relative humidity is denoted by RH; wind speed by u; air temperature by T_a; and solar radiation by SR; core temperature by T_c; surface-skin temperature by T_s; and respiration rate by RR.

Heat Stressor	**Mean**	**Standard Deviation**	**Minimum Value**	**Maximum Value**
RH (%)	30.9	6.9	17.0	50.0
u (m/s)	1.7	0.9	0.0	4.8
T_a (°C)	31.6	3.53	21.2	39.5
SR (W/m^2)	800.0	85.3	194.0	981.0
Physiological Measures	**Mean**	**Standard Deviation**	**Minimum Value**	**Maximum Value**
T_c (°C)	39.5	0.6	38.4	41.1
T_s (°C)	37.6	0.9	35.0	40.0
RR (breaths per min.)	95.6	17.4	48.0	140.0

After data processing, the dataset was augmented to include interaction terms between the environmental heat stressors. The environmental heat stressors and their interaction terms were then normalized to account for different magnitudes of measurement because different scales of measurement may distort the significance of each heat stressor and interaction term in the ranking system (Jolliffe, 2002).

Model Development

Machine Learning Model Development

In order to develop a ranking system for environmental heat stressors, it was necessary to develop models that predict physiological responses from environmental data. The machine learning algorithms considered for developing these models were penalized linear regression with the elastic net penalty (Zou and Hastie, 2005), random forests (Breiman, 2001), gradient boosted machines (Friedman, 2001), and neural networks (Goodfellow et al., 2016). These algorithms are discussed in detail in Gorczyca et al. (2018). Models were developed with the R statistical software (R Core Team, 2017) using the h2o package (H2O.ai Team, 2017).

Each algorithm considered has hyperparameters, which influence how a model learns relationships from data. Table 2 provides a summary of the hyperparameter space considered for optimization. To develop machine learning models for this study, a random search for hyperparameter optimization was performed (Bergstra and Bengio, 2012). For the penalized linear regression models, random forests, and gradient boosted machines, 1000 random search iterations were performed for each of these algorithms. For neural networks, 2000 random search iterations were performed, as neural networks have a larger number of hyperparameters (5000 models were developed overall).

Table 2. Hyperparameter space configured for model development.

Hyperparameter	Sampling Distribution
Penalized Linear Regression	
α	$\mathcal{U}(0, 1)$[a]
λ	$10^{\mathcal{U}(-12, 1)}$
Random Forest	
#Trees	$\mathcal{U}_d(1, 250)$[a]
MNL[b]	$\mathcal{U}_d(1, 30)$
NVS[c]	$\mathcal{U}_d(1, 10)$
Max. Tree Depth	$\mathcal{U}_d(1, 100)$
Gradient Boosted Machine	
#Trees	$\mathcal{U}_d(1, 200)$
MNL[b]	$\mathcal{U}_d(1, 30)$
Max. Tree Depth	$\mathcal{U}_d(1, 100)$
Learning rate	$\mathcal{U}(0.0001, 0.5000)$
Annealing	$\mathcal{U}(0.001, 1)$
Neural Networks	
Hidden Layers	$\mathcal{U}_d(1, 3)$
#Neurons	$\mathcal{U}_d(1, 1500)$[d]
Activation function	ReLU or Hyperbolic Tangent
Dropout rates[e]	$\mathcal{U}(0, 0.33)$
Epochs	$\mathcal{U}_d(10, 10{,}000)$
ρ[f]	$\mathcal{U}(0.75, 0.999)$
ε[f]	$10^{\mathcal{U}(-12, -3)}$

(a) $\mathcal{U}(a, b)$ denotes uniform continuous distribution from *a* to *b*, $\mathcal{U}_d(a, b)$ denotes uniform discrete distribution from *a* to *b.*
(b) MNL: minimum number of observations in a leaf.
(c) NVS: number of variables used in each split.
(d) If the number of hidden layers is greater than 1, the upper bound of the sampling distribution was changed to 1000 to decrease memory constraints.
(e) From Srivastava et al. (2014).
(f) Hyperparameters from the AdaDelta optimizer (Zeiler, 2012).

Model Selection and Assessment

A total of 15000 models were developed, 5000 for each physiological response. Every model was developed from the same training dataset, which consisted of 307 random samples (60%) from the entire dataset. An out-of-sample validation dataset consisting of 103 random samples (20%) was used to select a model developed from each machine learning algorithm that had the best performance metrics. Model selection for each machine learning algorithm was based on which model had the lowest root mean squared error on the validation dataset (RMSE; Hastie et al., 2003). Finally, an out-of-sample test dataset consisting of 103 random samples (20%) was used to gather the performance metrics of these selected models. The performance metrics gathered on the test dataset are RMSE, mean absolute error (MAE), and coefficient of determination (R^2). For RMSE and MAE, smaller values indicate better predictive performance. For R^2, larger values indicate better predictive performance.

The model with the lowest RMSE on each test dataset was used to rank the effects of each environmental heat stressor on a physiological response. If the best performing model was a penalized linear regression model, then the ranking is the magnitude of a weight for an input variable divided by the largest weight in magnitude for all the input variables. For a dataset with *p* input variables, the rank of the jth input variable is defined as

$$\text{rank}(w_j) = \frac{|w_j|}{\max_{i \in 1,\dots,p} |w_i|}$$

where, w_j represents the weight of the jth input variable.

The ranking procedure for neural networks is an extension of the ranking procedure for penalized linear regression, which is the summation of all the weights for each input variable divided by the summation of the absolute value of all the weights. This procedure is described in detail by Gedeon (1997), and can be used with the h2o software (H2O.ai Team, 2017).

If the best performing model is a random forest or a gradient boosted machine, then the ranking is the amount each input variable improved the mean squared error (MSE) of the model divided by the total amount every input variable improved the MSE (Breiman, 2001). The rankings for the random forest and gradient boosted machine models are also gathered using the h2o software (H2O.ai Team, 2017).

Results and Discussion

Model Performance

Tables 3-5 show the performance metrics of the selected model from each machine learning algorithm for T_c, T_s, and RR, respectively (the model from each machine learning algorithm that minimized RMSE on the validation dataset). Random forest models consistently had the best performance metrics for each physiological response. Furthermore, the random forest models achieved R^2 values similar to those currently reported in literature (Dikmen and Hansen, 2009; Wang et al., 2018a; Wang et al., 2018b) as empirically indicated by the scatterplots of the prediction outputs from each selected random forest models on the test datasets (Figure 1). While this gives evidence to the strength of using nonlinear machine learning algorithms, this does not mean that a random forest model will consistently attain the best performance metrics for predicting heat stress. The underlying reasoning for this is that the complexity of relationships between variables in a dataset is not understood unless a variety of machine learning models are developed and are assessed, even if the relationships in a dataset appear similar to those in another dataset (Wolpert and Macready, 1997). If a dataset does not contain much complexity, then, simple modelling procedures such as linear regression will demonstrate strong predictive performance (MacKay, 2003). However, as data becomes more complex, simple modelling procedures would require extensive feature engineering (transforming input variables to improve model performance) to capture the nonlinear relationships in data. Employing feature engineering does not, however, guarantee that a linear regression model would be as good a predictor as a machine learning model.

Table 3. Model performance on the test dataset when predicting core temperature. The best performance metric attained for each dataset is shown in bold, with the standard error of each performance metric in parentheses. RMSE represents root mean squared error, MAE is mean absolute error, and R^2 is coefficient of determination, PLM is linear regression with elastic net penalty, RF is random forest, GBM is gradient boosted machine, and NN is neural network.

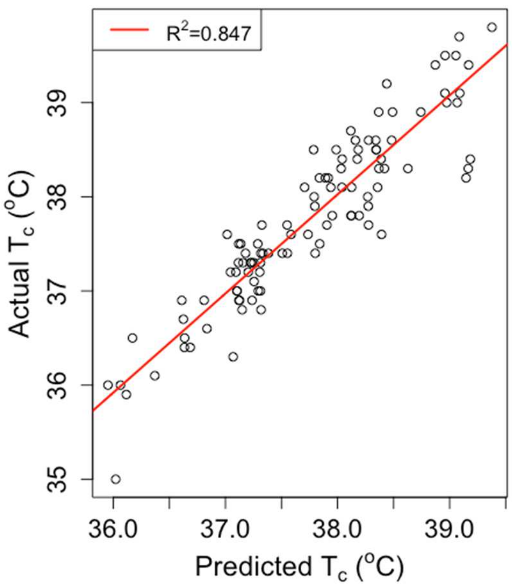
Core Temperature			
Model	**RMSE**	**MAE**	**R^2**
PLM	0.426 (0.032)	0.351 (0.025)	0.784 (0.033)
RF	**0.359 (0.033)**	**0.283 (0.019)**	**0.847 (0.034)**
GBM	0.362 (0.036)	0.285 (0.018)	0.844 (0.033)
NN	0.381 (0.033)	0.293 (0.013)	0.827 (0.036)

Table 4. Model performance on the test dataset when predicting skin-surface temperature. The best performance metric attained for each dataset is shown in bold, with the standard error of each performance metric in parentheses. RMSE represents root mean squared error, MAE is mean absolute error, and R^2 is coefficient of determination, PLM is linear regression with elastic net penalty, RF is random forest, GBM is gradient boosted machine, and NN is neural network.

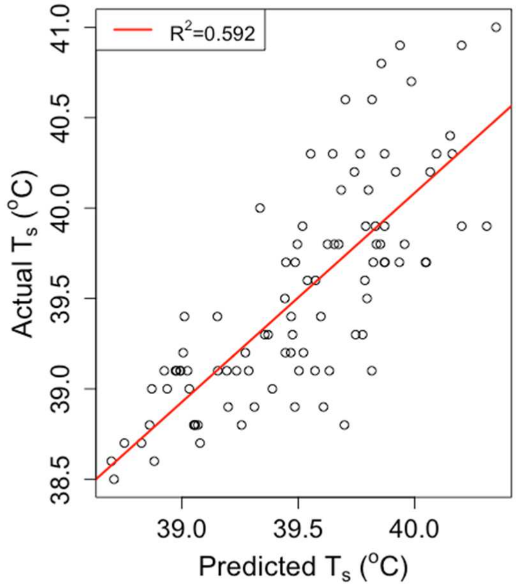
Skin-Surface Temperature			
Model	**RMSE**	**MAE**	**R^2**
PLM	0.443 (0.037)	0.344 (0.023	0.452 (0.016)
RF	**0.379 (0.031)**	**0.299 (0.022)**	**0.592 (0.015)**
GBM	0.409 (0.028)	0.326 (0.026)	0.533 (0.017)
NN	0.408 (0.035)	0.325 (0.025)	0.535 (0.015)

Table 5. Model performance on the test dataset when predicting respiration rate. The best performance metric attained for each dataset is shown in bold, with the standard error of each performance metric in parentheses. RMSE represents root mean squared error, MAE is mean absolute error, and R^2 is coefficient of determination, PLM is linear regression with elastic net penalty, RF is random forest, GBM is gradient boosted machine, and NN is neural network.

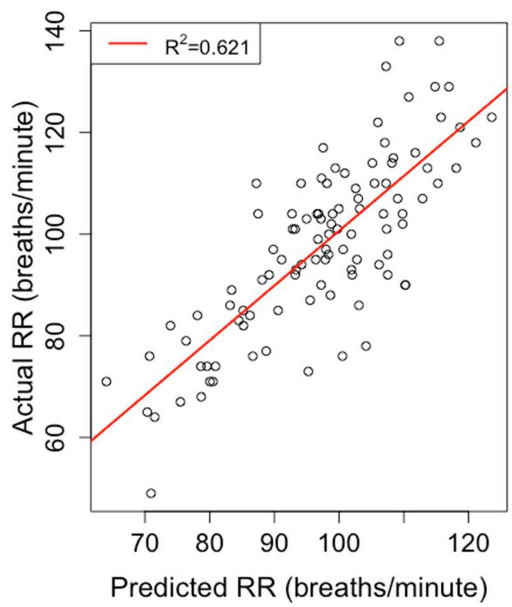
Respiration Rate			
Model	**RMSE**	**MAE**	**R^2**
PLM	11.780 (0.574)	9.656 (0.548)	0.536 (0.022)
RF	**10.655 (0.498)**	**8.368 (0.459)**	**0.621 (0.023)**
GBM	11.250 (0.512)	8.927 (0.478)	0.577 (0.027)
NN	11.033 (0.493)	9.019 (0.466)	0.593 (0.024)



(a)



(b)



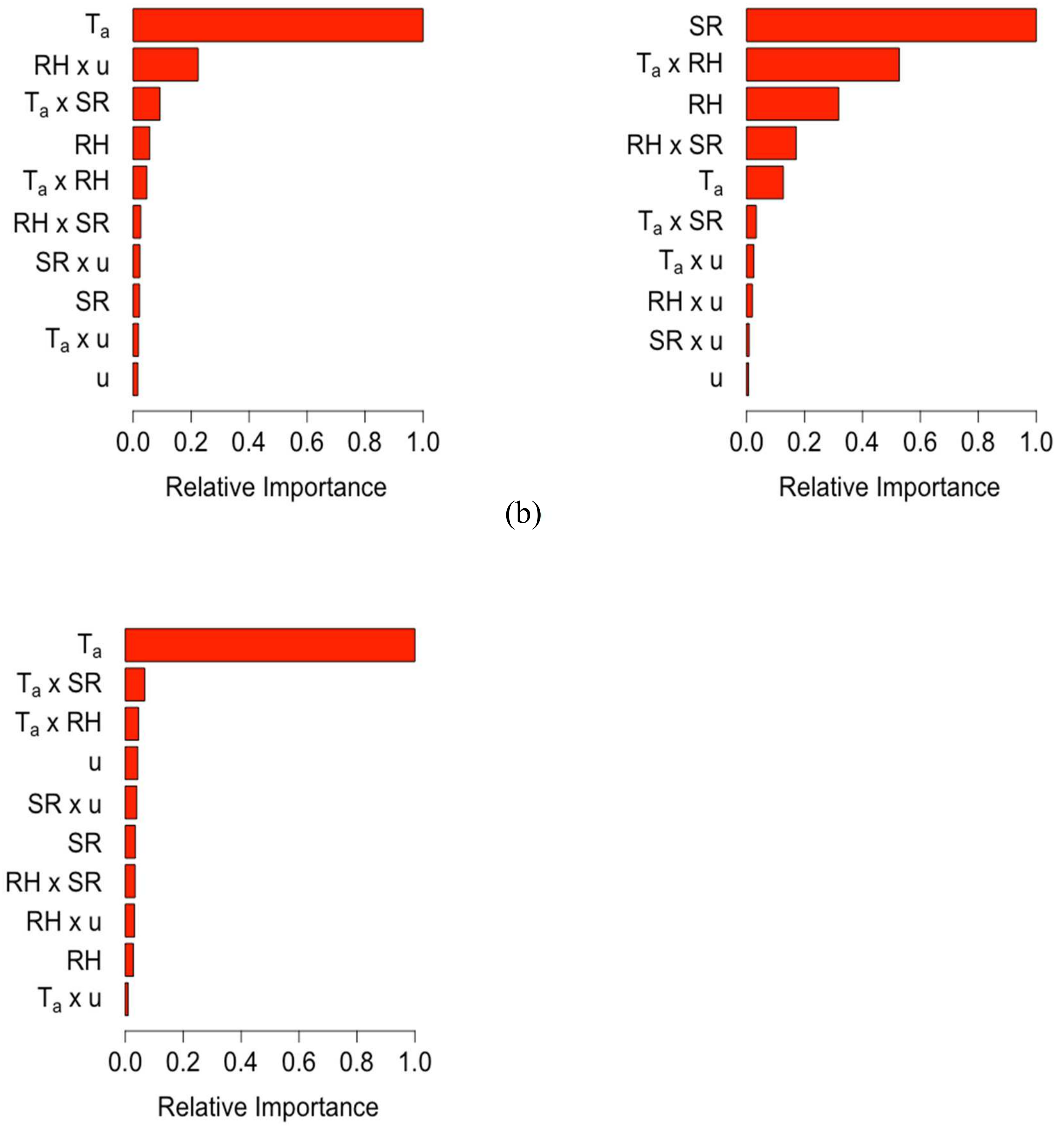
(c)

Figure 1. Predictions of core temperature, T_c (a); skin-surface temperature, T_s (b); and respiration rate, RR (c) from selected random forest models on out-of-sample data from the test dataset. These figures show that the predictions of these models are highly correlated.

Ranking of Environmental Heat Stressors and Relationship Evaluation

Interaction terms are commonly used for developing heat stress indices, and ranking these interaction terms can provide better understanding of how environmental heat stressors affect physiological responses (Schoen, 2005; Dikmen and Hansen, 2009; Wang et al, 2018a). Ranking of each physiological response for selected random forest models is given in Figure 2. For core temperature and respiration rate, air temperature (T_a) has the highest effect whereas solar radiation (SR) has the highest effect for skin-surface temperature. These effects are notably higher than the second highest ranked environmental parameter. Relative humidity (RH) and wind speed (u) had much lower effect on the physiological responses.

The rankings developed herein are valid for the dairy cows studied, and the results are consistent with findings obtained from THI models developed for different environmental conditions and different livestock species (Dikmen and Hansen, 2009; Wang et al., 2018b). This suggests that the rankings are valid for assessing effects of environmental heat stressors for dairy cows in similar environmental conditions. In this study, wind speed ranked lower than the other parameters. It should be noted, however, that the cows in this study were exposed to high air temperature and low wind speed (Chen et al., 2015). Wind speed may have a different level of effect if the dairy cows were instead exposed to high wind speed (Fournel et al., 2017).



(a) (b) (c)

Figure 2. Ranking of the effects of environmental heat stressors and their interaction terms on core-body temperature (a); skin-surface temperature (b); and respiration rate (c). T_a represents air temperature, SR is solar radiation, RH is relative humidity, and u is wind speed. An (x) represents interaction between two parameters.

Strength and Limitations of Machine Learning Models

This study is not the first to use machine learning to predict physiological responses of dairy cows from environmental heat stressors (Brown-Brandl et al., 2005). What is new in this study is that machine learning is used to rank the effects of environmental heat stressor on physiological responses. The benefit of this approach is that nonlinear machine learning algorithms have greater expressive power and better describe the underlying relationships in data than linear regression models. As a result, machine learning models are able to consider the effect of every environmental heat stressor and their interaction terms when developing the ranking without sacrificing model performance. Previous THI models did not consider such a combination of input variables because they were developed from linear or additive regression (Collier et al., 1981; Orihuela, 2000; Dikmen and Hansen, 2009; Wang et al., 2018a), and these modeling procedures do not have hyper-parameters that control overfitting to the data. It is important to note that machine learning algorithms can be applied to large dataset obtained from different livestock species and environmental conditions. The main drawback of using nonlinear machine learning models is computational time. On average, it took approximately 0.6 seconds to develop each penalized linear model, 3.6 seconds to develop each random forest, 5.4 seconds to develop each gradient boosted machine, and 8 seconds to develop each neural network.

Conclusions

The following conclusions can be drawn from this study:

(1) Machine learning models (penalized linear regression with elastic net penalty, random forests, gradient boosted machines, and neural networks) were used to predict core-body temperature, skin-surface temperature, and respiration rates of dairy cows.

(2) Random forest models consistently produced the lowest mean square error, RMSE, in predicting skin-surface temperature, core-body temperature and respiration rate of dairy cows.

The RMSE for core-body temperature was 0.359 °C; skin-surface temperature was 0.379 °C; and respiration rate was 10.655 respirations per minute.

(3) The effects of the environmental stressors on the three physiological responses (skin-surface temperature, core-body temperature and respiration rate) were ranked. The result showed that air temperature had the highest effect on core-body temperature and respiration rate, whereas solar radiation had the highest effect on skin-surface temperature. Wind speed and relative humidity displayed minimal effects as heat stressors on the three physiological responses considered in this study.

REFERENCES

Barash, H., Silanikove, N., Shamay, A., Ezra, A., 2001. Interrelationships among ambient temperature, day length and milk yield in dairy cows under a Mediterranean climate. J. Dairy Sci. 84, 2314-2320. doi: 10.3168/jds.S0022-0302(01)74679-6.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281-305.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5-32. doi:10.1023/A:1010933404324.

Brown-Brandl, T.M., Jones, D.D., Woldt, W.E., 2005. Evaluating modelling techniques for cattle heat stress prediction. Biosys. Eng. 91, 513-524.

Bohmanova, J., Misztal, I., Cole, J.B., 2007. Temperature-humidity indices as indicators of milk production losses due to heat stress. J. Dairy. Sci. 90, 1947-1956. doi: 10.3168/jds.2006-513.

Bouraoui, R., Lahmar, M., Majdoub, A., Djemali, M., Belyea, R., 2002. The relationship of temperature-humidity index with milk production of dairy cows in a Mediterranean climate. Animal Res. 51, 479-491.

Chen, J.M., Schütz, K.E., Tucker, C.B., 2015. Cooling cows efficiently with sprinklers: physiological responses to water spray. J. Dairy Sci. 98, 6925-6938.

Collier, R.J., Eley, R.M., Sharma, A.K., Pereira, R.M., Buffington, D.E., 1981. Shade management in subtropical environment for milk yield and composition in Holstein and Jersey cows. J. Dairy Sci. 64, 844-849.

Dikmen, S., Hansen, P.J., 2009. Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? J. Dairy Sci. 92, 109-116.

Friedman, J., 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist., 29, 1189-1232.

Fournel, S., Ouellet, V., Charbonneau, E., 2017. Practices for alleviating heat stress of dairy cows in humid continental climates: A literature review. Animals. 7.

Gedeon, T.D., 1997. Data mining of inputs: Analysis magnitude and functional measures. Int. J. Neural Syst. 8, 209-218.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, MA.

Gorczyca, M.T., Milan, H.F.M., Maia, A.S.C., Gebremedhin, K.G., 2018. Machine learning algorithms to predict core, skin, and hair-coat temperatures of piglets. Comput. Electron. Agric. 151, 286-294.

H2O.ai team, 2017. h2o: R Interface for H2O, version 3.16.0.2.

Hastie, T., Tibshirani, R., Friedman. J., 2003. The Elements of Statistical Learning. Springer, NY.

Jolliffe, I.T., 2002. Principal Component Analysis. Springer, NY.

MacKay, D.J.C., 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press, Cambridge, England.

Orihuela, A., 2000. Some factors affecting the behavioral manifestation of oestrus in cattle: A review. Appl. Anim. Behav. Sci. 70, 1-16.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schoen, C., 2005. A new empirical model of the temperature–humidity index. J. Appl. Meteorol. 44, 1413-1420.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929-1958.

Wang, X., Gao, H., Gebremedhin, K., Bjerg, B.S., Van Os, J., Tucker, C.B., Zhang, G., 2018. A predictive model of equivalent temperature index for dairy cattle (ETIC). J. Therm. Bio. 76, 165-170.

Wang, X., Bjerg, B.S., Choi, C.Y., Zong, C., Zhang, G., 2018. A review and quantitative assessment of cattle-related thermal indices. J. Therm. Bio. 77, 24-37.

West, J.W., Mullinix, B.G., Bernard, J.K., 2003. Effects of hot, humid weather on milk temperature, dry matter intake, and milk yield of lactating dairy cows. J. Dairy. Sci. 86, 232-242. doi: 10.3168/jds.S0022-0302(03)73602-9.

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE. Trans. Evo. Comp.

Yano, M., Shimadzu, H., Endo, T., 2014. Modelling temperature effects on milk production: a study on Holstein cows at a Japanese farm. Springer Plus. 3.

Zeiler, M.D., 2012. Adadelta: An adaptive learning rate method. arXiv:1212.5701.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. Stat. Soc. Ser. B. 67, 301-320.

Chapter 2

MACHINE LEARNING ALGORITHMS TO PREDICT CORE, SKIN, AND HAIR-COAT TEMPERATURES OF PIGLETS

Abstract

Internal-body (core) and surface temperatures of livestock are important information that indicate heat stress status and comfort of animals. Previous studies focused on developing mechanistic and empirical models to predict these temperatures. Mechanistic models based on bioenergetics of animals often require parameters that may be difficult to obtain (e.g., thickness of internal tissues). Empirical models, on the other hand, are data-based and often assume linear relationships between predictor (e.g., air temperature) and response (e.g., internal-body temperature) variables although, from the theory of bioenergetics, the relationship between the predictor and the response variables is non-linear. One alternative to consider non-linearity is to use machine learning algorithms to predict physiological temperatures. Unlike mechanistic models, machine learning algorithms do not depend on biophysical parameters, and, unlike linear empirical models, machine learning algorithms automatically select the predictor variables and find non-linear functions between predictor and response variables. In this paper, we tested four different machine learning algorithms to predict rectal (Tr), skin-surface (Ts), and hair-coat surface (Th) temperatures of piglets based on environmental data. From the four algorithms considered, deep neural networks provided the best prediction for Tr with an error of 0.36%, gradient boosted machines provided the best prediction for Ts with an error of 0.62%, and random forests provided the best predictions for Th with an error of 1.35%. These three algorithms were robust for a wide range of inputs. The fourth algorithm, generalized linear

regression, predicted at higher errors and was not robust for a wide range of inputs. This study supports the use of machine learning algorithms (specifically deep neural networks, gradient boosted machines, and random forests) to predict physiological temperature responses of piglets.

Introduction

One of the current challenges in agriculture is to increase food production to feed the world's growing population while considering environmental responsibilities and the comfort of the biological object (livestock; Hunter et al., 2017). In animal production, the challenge is in developing precision livestock farming techniques (Van Hertem et al., 2017, Wathes et al., 2008) to increase animal comfort and production. These techniques (Guarino et al., 2017) are focused on continuous monitoring of animal health, comfort, and production indicators, such as internal-body and skin-surface temperature. These temperatures indicate the health status and production levels of animals (Da Silva and Maia, 2013, Soerensen and Pedersen, 2015), as well as their heat stress level, estimated to cost the swine industry $300 million each year (St-Pierre et al., 2003).

Heat stress is a major issue that decreases animal welfare (Silanikove, 2000), production (Nienaber et al., 1999), reproduction (Wolfenson et al., 2000), and growth potential (Collin et al., 2001). To cope with heat stress, pigs rely on behavioral (Vasdal et al., 2009) and physiological (Brown-Brandl et al., 2001, Brown-Brandl et al., 2014, Robertshaw, 2006) responses. Because of the importance of monitoring heat stress of pigs (Shao and Xin, 2008), and the difficulty of measuring the necessary parameters that indicate heat stress (McCafferty et al., 2015), two classical approaches are used to estimate heat stress of animals: (1) mechanistic modelling, and (2) empirical modelling.

Mechanistic models are based on the biophysical understanding of conservation of energy, momentum, and mass in live animals (Collier and Gebremedhin, 2015, DeShazer, 2009). Using conservation equations, a governing equation for the problem is formulated and solved analytically or numerically. The limitations of analytical and numerical models are the assumption that internal and/or superficial temperatures are known, or a simple mathematical relationship exists between them, and/or some of the parameters are also difficult to obtain (e.g., thickness of internal tissues, etc.). Furthermore, mechanistic models reveal that the relationship between environmental and physiological responses are non-linear (Hensley et al., 2013, Milan and Gebremedhin, 2016a, Milan and Gebremedhin, 2016b, McArthur, 1981).

Empirical models are data-based and usually assume a linear relationship between predictor variables (e.g., air temperature) and the response variable (e.g., internal-body temperature). These relationships are chosen by the researcher and has a considerable impact on the accuracy of the model (Mostaço et al., 2015, Pathak et al., 2009, Ramirez, 2017, Soerensen and Pedersen, 2015).

A third approach that is receiving increased attention from swine researchers are machine learning and computer vision algorithms (Kamilaris and Prenafeta-Boldú, 2018). Recent applications include monitoring animal behavior (Cross et al., in press, Lao et al., 2016, Nasirahmadi et al., 2017, Shao and Xin, 2008), and weight (Kashiha et al., 2014, Shi et al., 2016, Wongsriworaphon et al., 2015). In this paper, we propose the use of machine learning algorithms to predict internal-body temperature, skin-surface temperature, and hair-coat surface temperature of piglets from environmental variables. The advantage of this approach compared to mechanistic models is that it does not rely on biophysical parameters. The advantage of this approach compared to empirical models is that it automatically finds a non-linear function from the data, removing the subjectivity from the researcher choosing the relationship between predictor and response variables. To the best of our knowledge, this is the first study that applies machine learning algorithms to predict physiological temperatures of swine.

Materials and Methods

Experimental Measurements

Animal use and research protocol were approved by the Animal Care and Use Committee from São Paulo State University (FAPESP Proc. 17.519/14). The experiment was conducted in Jaboticabal, São Paulo, Brazil (21°15′40″ South Latitude and 595 m elevation) for five consecutive days. Ten 5-days-old piglets (weight = 3.76 ± 0.41 kg, mean ± SEM) from the commercial lineage "Large White" were randomly selected from the same farrowing. The farrowing was not provided with supplemental heat. The selected piglets were randomly separated into 5 groups (2 piglets in each group) and managed inside a brooder (1.0 × 1.0 × 1.0 m3) from 3 a.m. to 8 a.m. Physiological measurements were performed hourly and started one hour after the piglets were inside the brooder (i.e., from 4 a.m. to 8 a.m.) to allow for adaptation to the environment. Four of the five groups were provided with supplemental heat (lamps) with intensities of 60 W, 100 W, 160 W, or 200 W. The fifth group (control) was not provided with supplemental heat.

Skin-surface temperature (Ts, °C) at the upper leg of the animal was measured with a skin-temperature probe (MLT422/AL, ADInstruments, accuracy ± 0.2 °C) and rectal temperature (Tr, °C) was measured with a rectal temperature probe (MLT1403, ADInstruments, accuracy ± 0.2 °C). These probes were connected to thermistor pods (ML309, ADInstruments), and the pods were connected to a data acquisition system (PL3516/P, PowerLab 16/35 and LabChart Pro, ADInstruments) that recorded data every second for approximately 5 min. Hair-coat-surface temperature (Th, °C) at the upper leg was measured with an infrared thermometer (Model 568, Fluke, accuracy ± 1 °C). Air temperature (Ta, °C) and relative humidity (RH, %) inside the brooder were measured every minute (HOBO U12 Temp/RH, Onset, accuracy ± 0.35 °C and ± 2.5%). Black globe temperature (Tg, °C) inside the brooder was measured using a 15-cm dia.

black globe installed 10 cm above the ground (thermocouple TMC20-HD, datalogger U12-013, accuracy ± 0.35 °C, Onset).

Model Development

Data Processing

The experiment was designed to provide 200 data points. Each individual data point contained the time of measurement (in hours), intensity of the supplemental heat, Ta, RH, Tg, Tr, Ts, and Th. Time of measurement, intensity of supplemental heat, Ta, and Tg were used as predictors of Tr, Ts, and Th. RH was not used as a predictor variable because 22% of the data was lost due to sensor failure. Further technical problems led to a reduction in the number of collected datapoints from 200 to 173. Correlations of the variables, mean and standard error of the mean were calculated. The univariate number of the outliers in the dataset was calculated using the z-score method at 2.5 standard deviations above or below the mean (Cousineau and Chartier, 2010).

The dataset was divided into training and testing datasets (Hastie et al., 2003). The training dataset was used to develop the machine learning models and the testing dataset was used to evaluate the predictive performance of the models. The training dataset consisted of 130 data points (75% of the dataset) and the testing dataset consisted of 43 points (25% of the dataset). The testing dataset was first obtained using stratified random sampling for each combination of time of measurement/intensity of supplemental heat (strata). This approach ensured that the testing dataset contained at least two data points from each stratum. Mean values were calculated for each strata of the dataset (yielding 20 data points) to determine the mean percentage error of each model for every stratum.

Overview of Machine Learning models

The machine learning algorithms used in this study were generalized linear regression model with elastic net regularization (GLM; Zou and Hastie, 2005), random forests (RF; Breiman, 2001), gradient boosted machines (GBM; Natekin and Knoll, 2013), and deep neural networks (feedforward neural networks) with the ReLU activation function (DNN; Goodfellow et al., 2016). Each algorithm has hyperparameters that influence the model learned from the data.

GLM is ordinary linear regression with penalty terms in the (sum of magnitudes) and (sum of squares) norms of the linear regression coefficients. The penalties shrink irrelevant regression coefficients and limit the impact of collinearity between the predictor variables (Zou and Hastie, 2005). The objective function of the GLM model is described as the following:

$$min_{\beta,\beta_0} \frac{1}{2N}\sum_{i=1}^{N} (x_i{}^T\beta + \beta_0 - y_i)^2 + \lambda[\alpha||\beta||_1 + (1-\alpha)||\beta||_2{}^2]$$

where β, β_0 are the regression coefficients, the summation represents the squared residual errors, x_i is the predictor variables from the i^{th} row of data, y_i is the predicted variable from the i^{th} row of data, λ is the severity of the penalty applied, and α distributes the penalty between the $L_1$and L_2 norms of the regression coefficients. The hyperparameters λ and α.

The RF and GBM models rely on decision trees, which are simple predictive models that stratify the input data space into output areas. The output-area prediction of decision trees is the mean of the response variables from the training dataset that fall in that output area (Fig. 5). For RF, several decision trees are developed independently from different subsets of the training dataset as well as from the different predictor variables. The prediction of the RF is the average of the predictions from all decision trees. The hyperparameters for RF are number of decision trees, minimum number of observations in a leaf, number of variables used to develop each split in a decision tree, and the maximum depth of the decision trees. For the GBM model, decision trees are developed sequentially, where each new decision tree is designed to improve on the predictive performance of the previous decision trees. The hyperparameters of the GBM are nearly the same as the hyperparameters for the RF, except GBM uses all predictor variables in a dataset for each split. GBM also has the learning rate of the sequential trees and an annealing rate (the influence of sequential trees on the final prediction output) as hyperparameters.

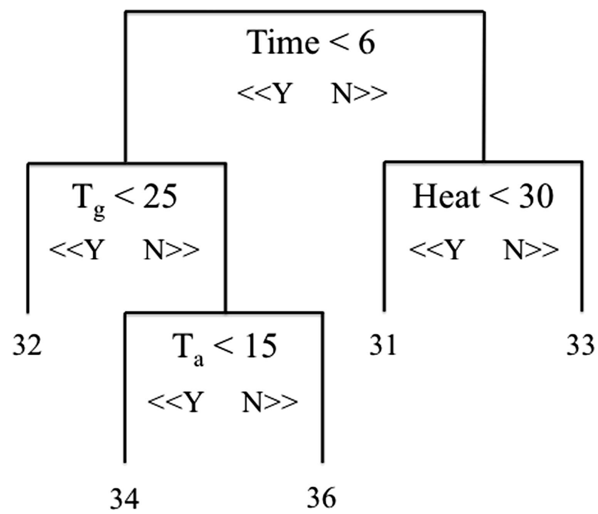


Figure 3: Example of a decision tree for predicting hair-coat surface temperature. A decision tree is developed by segmenting the input space into structured outputs. Each decision (e.g., Time < 6) represents a split of the tree. A leaf is the end node of the tree (e.g., the node with the value of 31 for Time ≥ 6 and Heat < 30). Random forests are based on creating several decision trees and averaging their output. Gradient boosted machines are based on creating several sequential decision trees, where new trees focus on improving the prediction accuracy of previous trees, and linearly combining the predictions of these trees. Time: time of measurement (hours); Heat: intensity of supplemental heat (W); Ta: air temperature (°C); Tg: black globe temperature (°C).

DNN algorithms provide flexible and robust approaches to develop nonlinear machine learning models. Feedforward neural networks, the type of DNN used in this study, consist of an input layer, hidden layers of unobserved variables, and an output layer (Fig. 4). Given an input vector x, the output of hidden layer h is computed as follows:

$$h = f(\theta + Wx)$$

where θ is a vector of offsets, W is a matrix of weights, and f is an user selected activation (nonlinear) function (ReLU was used in this study). The output from f is an input for the next layer. This process is repeated until the output layer is reached. The variable calculated in the output layer, the prediction of the feedforward neural network, is calculated as h but with a different activation function. In this study, the activation function for the output layer was the identity function, which is equivalent to linear regression with the variables of the last hidden layer. The hyperparameters for DL are the number of hidden layers, number of neurons in each hidden layer, mini-batch size (the number of observations used in each iteration in the model optimization process), epochs (the number of times the whole training dataset is used in training), dropout percentage (the percentage of neurons not used in a training epoch to avoid overfitting; Srivastava et al., 2014), and ρ and ε (hyperparameters of the ADADELTA optimization framework; Zeiler, 2012).

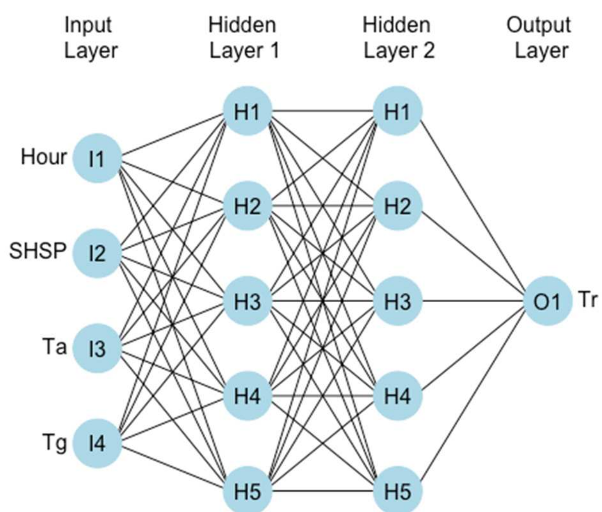


Figure 4: Feedforward neural network. Each input variable represents one neuron (In) that connects to every hidden neuron in the first hidden layer (H1m). Each hidden neuron is a non-linear function (activation function), where the outputs of the hidden neurons in the previous hidden layer are inputs to the hidden neurons in the next hidden layer. The outputs of the last hidden layer are inputs to the output neuron (O), which provides the prediction of the neural network. Time: time of measurement (hours); Heat: intensity of supplemental heat (W); Ta: air temperature (°C); Tg: black globe temperature (°C); In: input neuron n; Hnm: hidden neuron m of hidden layer n; O: output neuron; Tr: rectal temperature (°C).

Training and Testing Machine Learning Models

The objective of this paper was to develop machine learning models to predict Tr, Ts, and Th using Ta, Tg, time of measurement, and intensity of supplemental heat as predictors. The machine learning models were trained in R (R Core Team, 2017) using the H2O package (The H2O.ai Team, 2017) with modular 5-fold cross-validation (Hastie et al., 2003). To develop the machine learning models, a random search for hyperparameter optimization (Bergstra and Bengio, 2012) was performed on the hyperparameter space described in Table 1. For GLM, RF, and GBM, 1000 random searches were performed (resulting in 1000 trained models for each of these algorithms). For DNN, because of its inherently larger hyperparameter space, 2000 random searches were performed (resulting in 2000 trained deep neural network models). Computations were performed on an Oryx Pro from System76, with Pop-OS 17.10, 512 GB PCIe M.2 SSD, 64 GB DDR4 RAM memory (2133 MHz), i7-6820HK (3.6 GHz), 8 GB GeForce GTX 980 M.

Table 6: Hyperparameter space used to sample hyperparameters for training the machine learning algorithms.

Hyperparameter	Distribution[1]	Hyperparameter	Distribution
Random Forests		**Generalized Linear Model**	
#Trees	□$_d$(10, 250)	α	□(0, 1)
MNOL	□$_d$(1, 30)	λ	$10^{□(-10, 0)}$
NVS	□$_d$(1, 4)	**Deep Learning**	
Max. Tree Depth	□$_d$(1, 100)	#Hidden layers	□$_d$(1, 10)
Gradient Boosted Machines		#Neurons	□$_d$(1, 250)
Min. #Rows	□$_d$(1, 20)	Dropout percentage	□(0, 0.33)
#Trees	□$_d$(1, 100)	Epochs	□$_d$(1, 10^4)
Max. Tree Depth	□$_d$(1, 100)	Mini-batch size	□$_d$(1, 130)
Learning rate	□(0.001, 1)	□	□(0.75, 0.999)
Annealing	□(0.8, 1)	ε	$10^{□(-10, -6)}$

1: $\mathcal{U}_d(a, b)$ stands for uniform discrete random distribution from a to b; $\mathcal{U}(a, b)$ stands for uniform random distribution from a to b.

The mean squared error (MSE) was used as the evaluation metric (Hastie et al., 2003). We used cross-validation MSE to select the best performing model from each algorithm. Of these best performing models, the overall best model was the one that minimized MSE on the testing dataset.

Robustness and generalization of the best models were tested using partial dependence plots (Friedman, 2001) for 5 artificial datasets. Each dataset was designed to test how the models would perform under different conditions. Four artificial datasets had T_a = [-20, 100] °C, T_g = [-20, 100] °C, Hour = [0, 24] h, or intensity of supplemental heat = [0, 1000] W, while keeping the remaining predictor variables at their mean values. The fifth artificial dataset consisted of 10,000 random combinations of these artificial values to further test how change in the predictor variables would affect the prediction from the machine learning models.

Results and Discussion

Environmental data

Figure 6 shows the measured environmental data from the dataset stratified for the different time of measurement and intensity of supplemental heat while Table 2 shows the coefficients of correlation, mean, standard error of the mean, and number of outliers. As expected (Monteith and Unsworth, 2013), T_a and T_g increased when the intensity of the supplemental heat increased while RH decreased.

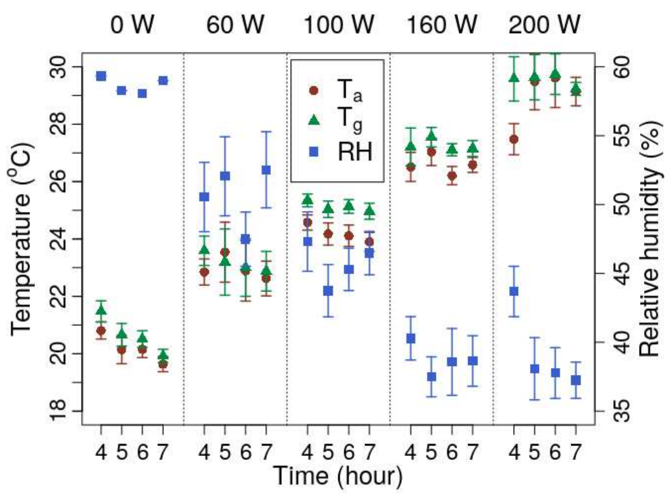


Figure 5: Experimental data (mean +/- standard error of the mean) for air temperature (T_a), black-globe temperature (T_g), and relative humidity (RH) separated by time and intensity of supplemental heat.

Table 7: Correlation coefficients, mean and standard error of the mean, and number of univariate outliers of the measured data. The number of outliers is displayed on the rightmost column, the mean and standard error of each data variable is displayed on the main diagonal of the table, and

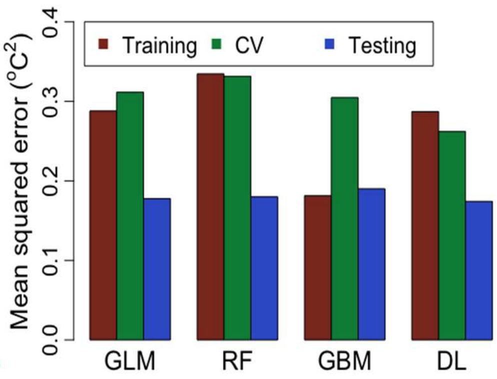
the correlation coefficients are displayed on the remaining entries of the table. No outliers were removed from training and testing datasets.

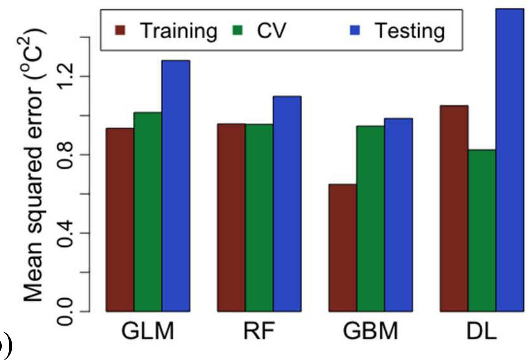
	Hour	SHSP[1]	T_a	T_g	RH[2]	T_r	T_s	T_h	#Outliers
Hour	5.490 ± 1.149	-0.009	-0.020	-0.075	-0.110	-0.497	-0.699	-0.214	0
SHSP[1]	-0.009	101.850 ± 72.441	0.891	0.912	-0.706	0.330	0.225	0.642	0
T_a	-0.020	0.891	24.455 ± 3.424	0.969	-0.590	0.287	0.258	0.743	11
T_g	-0.075	0.912	0.969	25.068 ± 3.464	-0.596	0.306	0.282	0.740	13
RH[2]	-0.110	-0.706	-0.590	-0.596	43.888 ± 7.833	-0.063	0.071	-0.275	17
T_r	-0.497	0.330	0.287	0.306	-0.063	37.917 ± 0.637	0.493	0.418	18
T_s	-0.699	0.225	0.258	0.282	0.071	0.493	32.803 ± 1.506	0.428	11
T_h	-0.214	0.642	0.743	0.740	-0.275	0.418	0.428	33.836 ± 1.864	16

[1]Variables: Hour: time of measurement (hour); Heat: intensity of supplemental heat (W); T_a: air temperature (°C); T_g: black globe temperature (°C); RH: relative humidity (%); T_r: rectal temperature (°C); T_s: skin-surface temperature (°C); T_h: hair-coat surface temperature (°C).
[2]Number of samples for RH was 132.

Performance of machine learning models

Figure 6 shows the MSE of the best performing machine learning models (that minimized cross-validation MSE) using cross-validation, training dataset, and testing dataset. Table 3 shows the hyperparameters for these models. Figure 7 shows the training and testing MSE of these models for the training iterations. The GLM model converged at one iteration, but the other models required more than ten iterations to converge. The best overall model (based on minimum testing MSE), was DNN for T_r, GBM for T_s, and RF for T_h.

(a) 

(b) 

(c)

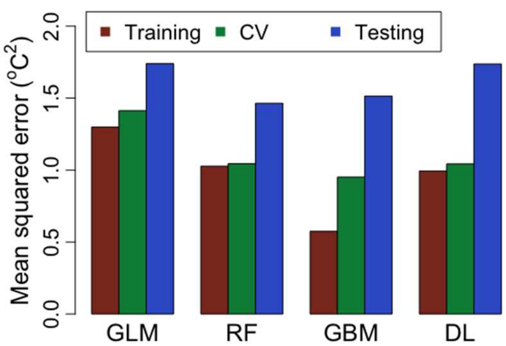


Figure 6: Performance of the best machine learning models for predicting rectal (T_r; a), skin-surface (T_s; b), and hair-coat surface (T_h; c) temperatures. GLM: generalized linear regression model with elastic net regularization; RF: random forests; GBM: gradient boosted machines; DNN: deep neural network with ReLU activation function.

Table 8: Hyperparameters of the best machine learning models.

Hyperparameter[a]	T_r	T_s	T_h
Generalized Linear Model			
λ	1.632×10^{-10}	0.240	8.749×10^{-7}
α	0.244	0.453	0.409
Random Forests			
MNOL[b]	4	6	2
#Trees	61	46	236
Max. Tree Depth	72	26	82
NVS[c]	1	3	2
Gradient Boosted Machines			
MNOL[b]	20	12	10
#Trees	80	60	25
Max. Tree Depth	29	2	41
Learning rate	0.351	0.504	0.398
Annealing	0.976	0.882	0.808
Deep Learning			
#Hidden layers[d]	2	4	2
#Neurons[d]	(242, 190)	(11, 53, 241, 230)	(65, 20)
Dropout Percentage	(0.13, 0.19)	(0.06, 0.19, 0.17, 0.06)	(0.03, 0.04)
Epochs	14	14.567	12.129
Mini-Batch Size	53	128	82
ρ	0.876	0.946	0.914
ε	2.855×10^{-7}	5.604×10^{-7}	5.646×10^{-8}

a Hyperparameters of the best machine learning models to predict rectal temperature (T_r, °C), skin-surface temperature (T_s, °C), and hair-coat surface temperature (T_h, °C).
b. MNOL: minimum number of observations in a leaf.
c. NVS: number of variables used in each split.
d. The numbers in parenthesis represent the value used for each hidden layer.

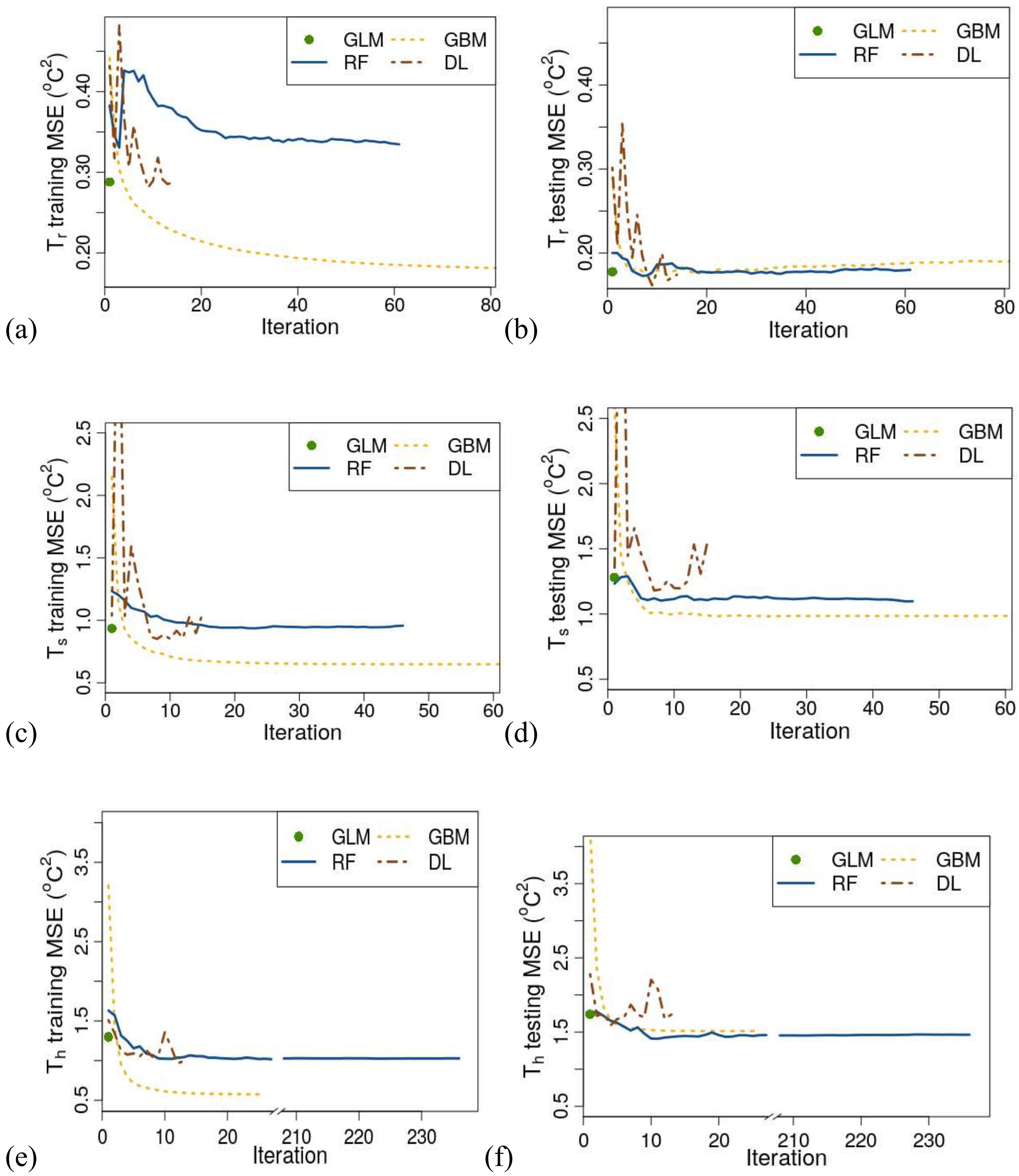
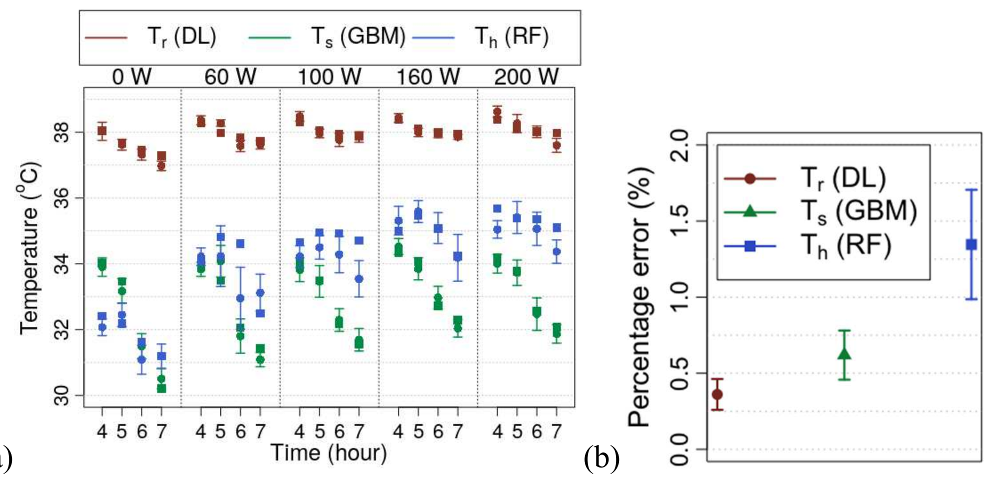


Figure 7: Mean squared error (MSE) on the training (a, c, e) and testing (b, d, f) datasets for predicting rectal (T_r; a, b), skin-surface (T_s; c, d), and hair-coat surface (T_h; e, f) temperatures using the best performing machine learning models.

Figure 8a shows the prediction output from the best machine learning algorithms using the mean dataset and Figure 8b shows the absolute percentage error. The model predictions are very close to the measured values. The observed mean absolute errors of T_r, T_s, and T_h, were 0.36%, 0.62% and 1.35%, respectively (Figure 8b). These errors are lower than those previously reported from either statistical or mechanistic models. (Mostaço et al., 2015) predicted rectal temperatures of pigs with 2.5% error using multiple linear regression for air enthalpy and tympanic temperature (known to be correlated with internal body temperature; Korthals et al., 1995). (Costa et al., 2010) predicted surface temperature of piglets with 5.5% error using a linear regression model. (Loughmiller et al., 2001) predicted mean body-surface temperature of pigs with 3.5% error using a linear regression model. (Turnpenny et al., 2000a; Turnpenny et al., 2000b) developed a mechanistic model and the resulting error was 7% for predicting skin-surface temperature of pigs.



(a) (b)

Figure 8: Measured (●) and predicted (■) rectal (T_r), skin-surface (T_s), and hair-coat surface (T_h) temperatures for the mean dataset stratified by (a) time of measurement and intensity of supplemental heat, and (b) absolute percentage errors of the predicted temperatures. Measured values and absolute percentage errors are presented as mean ± standard error of the mean. Temperatures were predicted from the best performing machine learning models. RF: random forests; GBM: gradient boosted machines; DNN: deep neural network with ReLU activation function.

Test of robustness and generalization of the best machine learning models

Figure 9, Figure 10, and Figure 11 show the partial dependence plots (Friedman, 2001) from the effect of changing one predictor variable (while keeping the remaining predictor variables at their mean values) on T_r, T_s, and T_h, respectively. These figures show, with the exception of GLM, that the machine learning models were robust with respect to the input variables because they did not produce unexpected predictions. GLM, which fits linear functions for the predictor variables, however, produced relationships that are counter to expectations, such as decreasing decreasing T_s and T_r while increasing T_g.

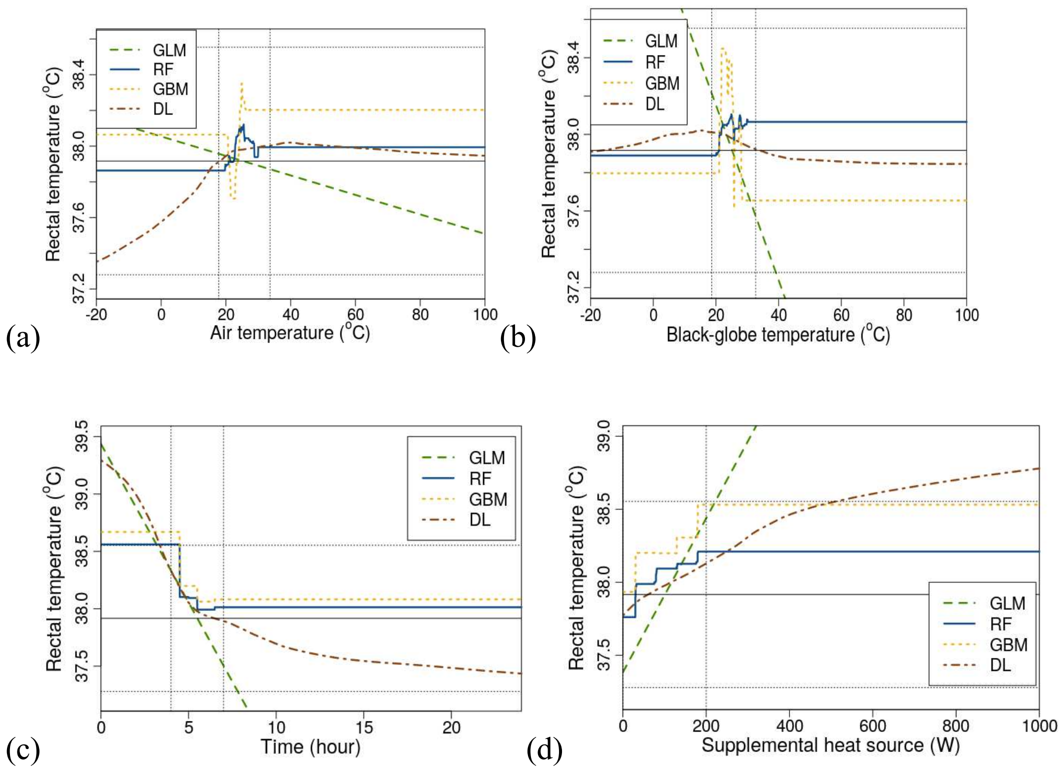


Figure 9: Test of robustness and generalization of the best machine learning models in predicting rectal temperature when changing (a) air temperature, (b) black-globe temperature, (c) time of measurement, or (d) intensity of supplemental heat, while keeping the remaining predictor variables at their mean values. The vertical dashed lines represent the range of the measured predictor variable. The horizontal solid line represents the mean rectal temperature, and the horizontal dashed lines represent the mean rectal temperature ± one standard deviation from the mean.

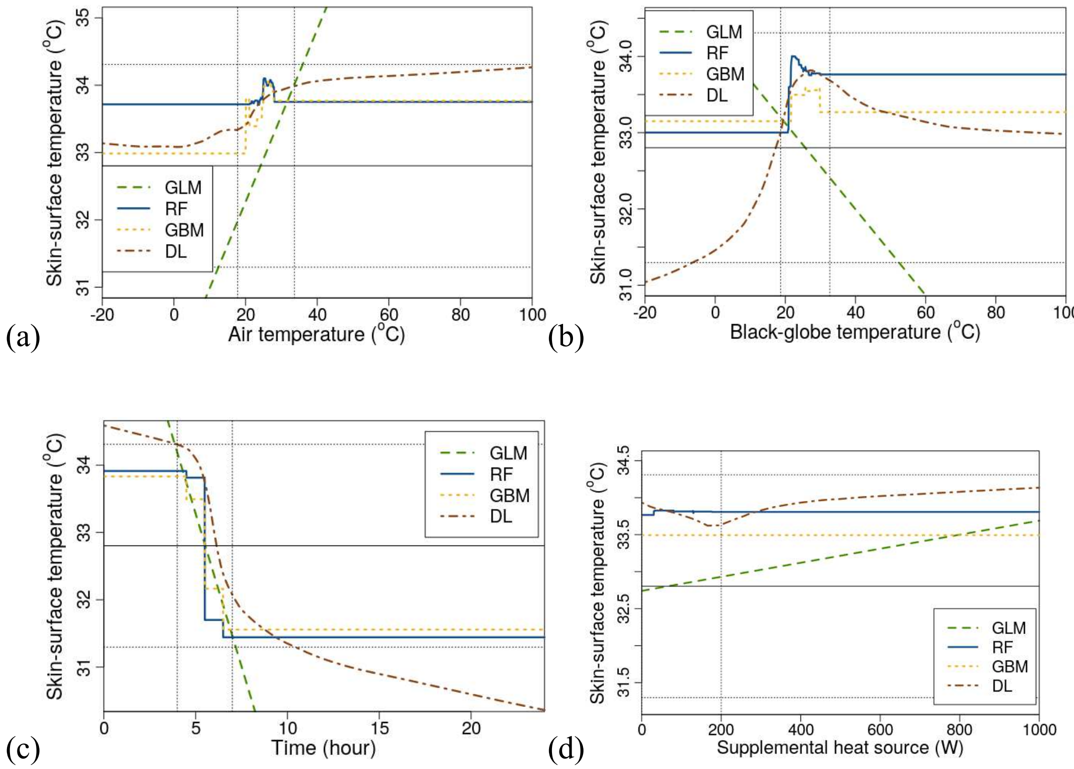


Figure 10: Test of robustness and generalization of the best machine learning models in predicting skin-surface temperature when changing (a) air temperature, (b) black-globe temperature, (c) time of measurement, or (d) intensity of supplemental heat, while keeping the remaining predictor variables at their mean values. The vertical dashed lines represent the range of the measured predictor variable. The horizontal solid line represents the mean skin-surface temperature, and the horizontal dashed lines represent the mean skin-surface temperature ± one standard deviation from the mean.

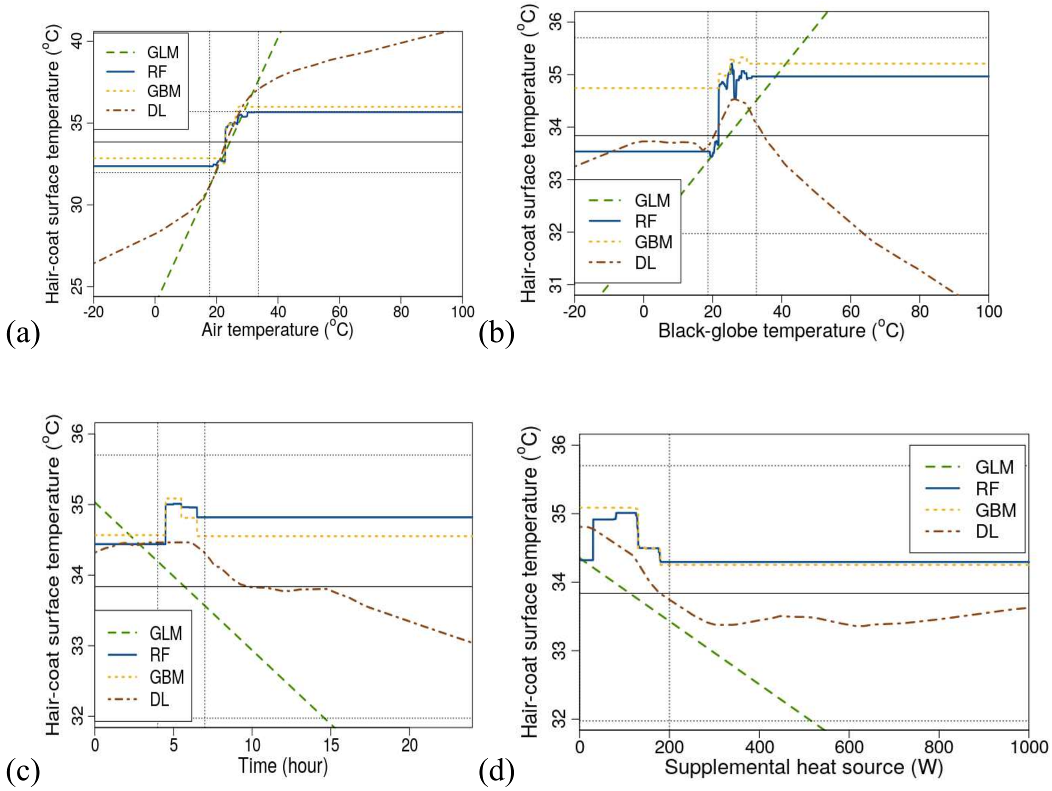


Figure 11: Test of robustness and generalization of the best machine learning models in predicting hair-coat surface temperature when changing (a) air temperature, (b) black-globe temperature, (c) time of measurement, or (d) intensity of supplemental heat, while keeping the remaining predictor variables at their mean values. The vertical dashed lines represent the range of the measured predictor variable. The horizontal solid line represents the mean hair-coat surface temperature, and the horizontal dashed lines represent the mean hair-coat surface temperature ± one standard deviation from the mean.

Figure 12 shows the effect of randomly changing all predictor variables on T_r, T_s, and T_h, which are predicted by the best performing machine learning models. This figure shows that temperature predictions using GLM resulted in higher variance, which means that GLM is not robust to changes in the predictor variables. The predictions from RF, GBM, and DNN were, however, closer to the mean measured values and the variance of their predictions was lower, which means that these algorithms are robust to changes in the predictor variables.

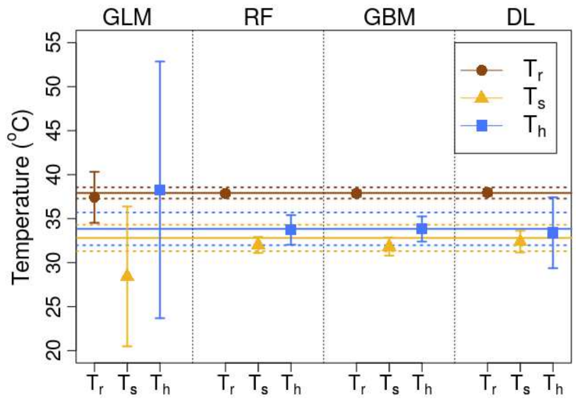


Figure 12: Test of robustness and generalization of the best machine learning models in predicting rectal (T_r), skin-surface temperature (T_s), and hair-coat surface (T_h) temperatures when randomly changing air temperature, black-globe temperature, time of measurement, and intensity of supplemental heat. Points represent mean ± one standard deviation of the mean (10,000 samples). Horizontal solid lines represent mean temperatures, and horizontal dashed lines represent mean ± one standard deviation of the mean.

Limitations and potential applications of machine learning models

The main limitations of machine learning models are that they are data-based as well as time-consuming and computationally expensive to train. In addition, if the training dataset is noisy or the model is trained inappropriately, then, the model may "learn" noise instead of the non-linear relationships that may exist between the predictor variables and the response variable (Natekin and Knoll, 2013). We showed in Section 3.3 that all algorithms considered in this study, except GLM, were robust to changes in the predictor variables. It should be noted, however, that the models were trained and tested from the same data population. This means that the models proposed in this study should not be applied to different data sets obtained from other livestock species. If a model is, however, trained with a larger dataset obtained from several livestock species, it would provide accurate predictions within the population represented by the dataset. It is also important to note that in this study, T_a and T_g were the only environmental predictor variables. Future studies may include other environmental predictor variables (e.g., relative humidity and heat stress indices) and spatio-temporal parameters (e.g., time of the year), which could improve model performance.

Training and validation of the four machine learning models considered in this study took ~ 9.5 h to complete. Most of this time was spent on training the models (~8 h in total; GLM = 50 min; RF = 35 min; GBM = 30 min; DNN = 6 h). The time to compute one prediction was ~0.3 ms, which is faster than the computing time required for analytical or numerical models (Milan and Gebremedhin, 2016b).

Our results suggest that machine learning algorithms, particularly RF, GBM, and DNN were found to be accurate in predicting rectal temperature (T_r), skin-surface temperature (T_s), and hair coat- surface (T_h) temperature, but not GLM. The main advantage of machine learning models is that only data is needed to train the non-linearity of the data. For mechanistic models, the non-linearity comes from the assumptions made in solving the conservation equations. Since machine learning algorithms predict temperatures that are necessary to solve mechanistic models, one possible application of machine learning algorithms would be to provide inputs to mechanistic models.

Conclusions

The following conclusions can be drawn from this study:

(1) Machine learning algorithms were trained to predict rectal temperature, skin-surface temperature, and hair-coat surface temperature of piglets based on environmental data.

(2) Deep neural networks, gradient boosted machines, and random forests were the best algorithms, based on the lowest mean squared error on the testing dataset, to predict rectal, skin-surface, and hair coat-surface temperatures, respectively.

(3) The data supports the use of machine learning algorithms to predict the physiological temperatures of livestock, and these temperature predictions can be used as inputs to mechanistic models.

REFERENCES

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J Mach Learn Res 13:281-305.

Bjerg, B., Pedersen, P., Morsing, S., Zhang, G., 2017. Modeling skin temperature to assess the effect of air velocity to mitigate heat stress among growing pigs. ASABE Ann Int Meet Paper 1700631, 2017.

Breiman, L., 2001. Random forests. Machine Learning 45:5-32.

Brown-Brandl, T.M., Eigenberg, R.A., Nienaber, J.A., Kachman, S.D., 2001. Thermoregulatory profile of a newer genetic line of pigs. Livest Prod Sci 71:253-260.

Brown-Brandl, T.M., Hayes, M.D., Xin, H., Nienaber, J.A., Li, H., Eigenberg, R.A., Stinn, J.P., Shepherd, T., 2014. Heat and moisture production of modern swine. ASHRAE Trans 120(1):469-489.

Collier, R.J., Gebremedhin, K.G., 2015. Thermal biology of domestic animals. Ann Rev Anim Biosci 3:513-532.

Collin, A., van Milgen, J., Dubois, S., Noblet, J., 2001. Effect of high temperature on feeding behaviour and heat production in group-housed young pigs. Br J Nutr 86:63-70.

Costa, L.N., Redaelli, V., Magnani, D., Cafazzo, S., Amadori, M., Razzauoli, E., Verga, M., Luzi, F., 2010. Preliminary study of the relationship between skin temperature of piglets measured by infrared thermography and environmental temperature in a vehicle in transit. LXIV Ann Meet Ital Soc Vet Sci pp. 193-197.

Cross, A.J., Rohrer, G.A., Brown-Brandl, T.M., Cassady, J.P., Keel, B.N., 2018. Feed-forward and generalised regression neural networks in modelling feeding behaviour of pigs in the grow-finish phase. Biosys Eng *In Press*.

Da Silva, R.G., Maia, A.S.C., 2013. Principles of animal biometeorology. Springer, New York.

DeShazer, J.A., 2009. Livestock energetics and thermal environmental management. ASABE.

Fei, F., Skidmore, A.K., Venus, V., Wang, T., Schlerf, M., Toxopeus, B., van Overjijk, S., Bian, M., Liu, Y., 2012. A body temperature model for lizards as estimated from the thermal environment. J Therm Biol 37:56-64.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann Stat 1189-1232.

Gebremedhin, K.G., Ni, H., Hillman, P.E. 1997. Modeling temperature profile and heat flux through irradiated fur layer. Transaction of ASAE 40(5):1441-1447.

Gebremedhin, K.G., Wu, B.X., 2003. Characterization of flow field in a ventilated space and simulation of heat exchange between cows and their environment. J Therm Biol 28:301-319.

Gebremedhin, K.G., Wu, B., 2016. Modeling heat loss from the udder of a dairy cow. J Therm Biol 59:34-38.

Gebremedhin, K.G., Wu, B., Perano, K., 2016. Modeling conductive cooling for thermally stressed dairy cows. J Therm Biol 56:91-99.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Massachusetts.

Guarino, M., Norton, T., Berckmans, D., Vranken, E., Berckmans, D., 2017. A blueprint for developing and applying precision livestock farming tools: a key output of the EU-PLF project. Anim Front 7(1):12-17.

Hastie, T., Tibshirani, R., Friedman, J., 2003. The Elements of Statistical Learning. Springer, New York.

Hensley, D.W., Mark, A.E., Abella, J.R., Netscher, G.M., Wissler, E.H., Diller, K.R., 2013. 50 years of computer simulation of the human thermoregulatory system. J Biomech Eng 135:021006.

Hunter, M.C., Smith, R.G., Schipanski, M.E., Atwood, L.W., Mortense, D.A., 2017. Agriculture in 2050: recalibrating targets for sustainable intensification. BioScience 67(4):386-391.

Jiang, M., Gebremedhin, K.G., Albright, L.D., 2005. Simulation of skin temperature and sensible and latent heat losses through fur layers. Transactions of ASAE 48(2):767-775.

Kashiha, M., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2014. Automatic weight estimation of individual pigs using image analysis. Comput Electron Agric 107:38-44.

Korthals, R.L., Eigenberg, R.A., Hahn, G.L., Nienaber, J.A., 1995. Measurements and spectral analysis of tympanic temperature regulation in swine. Trans ASAE 33(3):905-909.

Lao, F., Brown-Brandl, T., Stinn, J.P., Liu, K., Teng, G., Xin, H., 2016. Automatic recognition of lactating sow behaviors through depth image processing. Comput Electron Agric 125:56-62.

Li, H., Rong, L., Zhang, G., 2016. Study on convective heat transfer from pig models by CFD in a virtual wind tunnel. Comput Electron Agric 123:203-210.

Loughmiller, J.A., Spire, M.F., Dritz, S.S., Fenwick, B.W., Hosni, M.H., Hogge, S.B., 2001. Relationship between mean body surface temperature measured by use of infrared thermography and ambient temperature in clinically normal pigs and pigs inoculated with Actinobacillus pleuropneumoniae. Am J Vet Res 62(5):676-681.

McArthur, A.J., 1981. Thermal resistance and sensible heat loss from animals. J Therm Biol 6:43-47.

McArthur, A.J., 1987. Thermal interaction between animal and microclimate: a comprehensive model. J Ther Biol 126:203-238.

McCafferty, D.J., Gallon, S., Nord, A., 2015. Challenges of measuring body temperatures of free-ranging birds and mammals. Anim Biotelemetry 3:33.

Milan, H.F.M., Carvalho Jr, C.A.T., Maia, A.S.C., Gebremedhin, K.G., 2014. Graded meshes in bio-thermal problems with transmission-line modeling method. J Therm Biol 45:43-53.

Milan, H.F.M., Gebremedhin, K.G., 2016a. Triangular node for Transmission-Line Modeling (TLM) applied to bio-heat transfer. J Therm Biol 62:116-122.

Milan, H.F.M., Gebremedhin, K.G., 2016b. Tetrahedral node for Transmission-Line Modeling (TLM) applied to Bio-heat Transfer. Comp Biol Med 79:243-249.

Milan, H.F.M., Gebremedhin, K.G., 2016c. Solving bioenergetics' problems with the transmission-line modeling (TLM) method. ASABE Annual International Meeting, Orlando, FL, Paper 162420443.

Milan, H.F.M., Gebremedhin, K.G., 2017. General formulation of transmission-line modeling (TLM) method applied to bio-energetics of endotherms. 2017 ASABE Ann Int Meet, Paper 1700180.

Milan, H.F.M., Gebremedhin, K.G., 2018. General node for transmission-line modeling (TLM) method applied to bio-heat transfer. Int J Num Model El *Accepted Manuscript*.

Mondaca, M., Choi, C.Y., 2016. An evaluation of simplifying assumptions in dairy cow computational fluid dynamic models. T ASABE 59(6):1575-1584.

Monteith, J., and Unsworth, M., 2013. Principles of environmental physics: plants, animals, and the atmosphere. Academic Press, Massachusetts.

Mostaço, G.M., Miranda, K.O.S., Condotta, I.C.F., Salgado, D.D.A., 2015. Determination of piglets' rectal temperature and respiratory rate through skin surface temperature under climatic chamber conditions. J Braz Assoc Agric Eng 35(6):979-989.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front Neurorobot 7.

Nasirahmadi, A., Edwards, S.A., Sturm, B., 2017. Implementation of machine vision for detecting behaviour of cattle and pigs. Livest Science 202:25-38.

Nienaber, J.A., Hahn, G.L., Eigenberg, R.A., 1999. Quantifying livestock responses for heat stress management: a review. Int J Biometeorol 42:183-188.

Pathak, M., Parkhurst, A.M., Arias, R.A., Mader, T.L., 2009. Comparative study of time series and multiple regression for modeling dependence of cattle body temperature on environmental variables during heat stress. Ann 21st Conf Appl Statist Agric.

R Core Team. R: A language and environment for statistical computing, 2017. R Foundation for Statistical Computing, Vienna, Austria.

Radon, J., Bieda, W., Lendelova, J., Pogran, S., 2014. Computational model of heat exchange between dairy cow and bedding. Comput Electron Agric 107:29-37.

Ramirez, B.C., 2017. A novel approach to measure, understand, and assess the thermal environment in grow-finish swine facilities. Iowa State University. Graduate Theses and Dissertations. 16201.

Robertshaw, D., 2006. Mechanisms for the control of respiratory evaporative heat loss in panting animals. J App Physiol 101:664-668.

Seo, I., Lee, I., Moon, O., Hong, S., Hwang, H., Bitog, J.P., Kwon, K., Ye, Z., Lee, J., 2012. Modelling of internal environmental conditions in a full-scale commercial pig house containing animals. Biosyst Eng 111:91-106.

Shao, B., Xin, H., 2008. A real-time computer vision assessment and control of thermal comfort for group-housed pigs. Comput Electron Agric 62:15-21.

Shi, C., Teng, G., Li, Z., 2016. An approach of pig weight estimation using binocular stereo system based on LabVIEW. Comput Electron Agric 129:37-43.

Soerensen, D.D., Pedersen, J., 2015. Infrared skin temperature measurements for monitoring health in pigs: a review. Acta Vet Scand 57:5.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, RR., 2014. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929-1958.

St-Pierre, N.R., Cobanov, B., Schnitkey, G., 2003. Economic losses from heat stress by US livestock industries. J Dairy Sci 86:E52-E77.

Silanikove, N., 2000. Effects of heat stress on the welfare of extensively managed domestic ruminants. Livest Prod Sci 67:1-18.

The H2O.ai team, 2017. h2o: R Interface for H2O, version 3.16.0.2.

Turner, L.W., Blandford, G.E., Loewer, O.J., Taul, K.L., 1987a. Finite element model of heat transfer in the bovine. Part 1: theory. Transactions of the ASAE 30(3):768-774.

Turner, L.W., Loewer, O.J., Taul, K.L., Munifering, R.B., Gay, N., 1987b. Finite element model of heat transfer in the bovine. Part 2: validation. Transactions of the ASAE 30(3):775-781.

Turnpenny, J.R., McArthur, A.J., Clark, J.A., Wathes, C.M., 2000a. Thermal balance of livestock 1. A parsimonious model. Agr Forest Meteorol 101:15-27.

Turnpenny, J.R., Wathes, C.M., Clarck, J.A., McArthur, A.J., 2000b. Thermal balance of livestock 2. Application of a parsimonious model. Agr Forest Meteorol 101:29-52.

Van Hertem, T., Rooijakkers, L., Berckmans, D., Peña Fernández, A., Norton, T., Berckmans, D., Vranken, E., 2017. Appropriate data visualization is key to precision livestock farming acceptance. Comput Electron Agric 138:1-10.

Vasdal, G., Wheeler, E.F., Boe, K.E., 2009. Effect of infrared temperature on thermoregulatory behaviour in usckling piglets. Animal 3(10):1449-1454.

Wathes, C.M., Kristensen, H.H., Aerts, J-M., Berckmans, D., 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Comput Electron Agric 64:2-10.

Wolfenson, D., Roth, Z., Meidan, R., 2000. Impaired reproduction in heat-stressed cattle: basic and applied aspects. Anim Reprod Sci 60:535-547.

Wongsriworaphon, A., Arnonkijpanich, B., Pathumnakul, S., 2015. An approach based on digital image analysis to estimate the live weight of pigs in farm environments. Comput Electron Agric 115:26-33.

Zeiler, M.D., 2012. Adadelta: An adaptive learning rate method. arXiv:1212.5701.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J of the Royal Stat Soc Ser B 67(2): 301–320.