Oral Temperature Prediction with Infrared Thermography Using Tree-based Regression Models

Zhixin Mao, Ziyi Ding

Introduction

Background

- Body temperature is a very common measurement for many diseases such as Severe Acute Respiratory Syndrome (SARS) in 2003 [1] and Coronavirus in 2019 (COVID-19) [2]. There are many important body sites to measure body temperature, but some of them are not convenient to access. Infrared thermography (IRT) is a technology that detects infrared energy emitted from an object, converts it to temperature, and displays an image of temperature distribution. There are many factors that can impact the measurement, such as gender, skin humidity, or hair density. It is not practical to control for all factors, but analyzing their relationships with IRT is very important for enhancing our understanding and application of this technology in a clinical setting.

Introduction

Motivation

- The application and interpretation for this new clinical dataset was not fully covered in previous works.
- Based on the complex structure of this new data, the utilization of more flexible tree-based approaches is necessary.

Objectives

- How effective are tree-based methods in identifying the significant variables for accurately predicting oral temperature?
- What are the differences in the results of predicted oral temperatures when comparing fast mode and monitor mode?

State-of-the-art

- In the previous study of Wang et al [4], a large-scale clinical analysis of fever-screening thermography was done by the Health Center of the University of Maryland (UMD) and they collected clinical information of more than 1000 subjects measured within a wide room temperature range of 20-29 degrees Celsius. They have found a high correlation and non-linear relationship between their variables, but they only considered three regression methods, which are weighted linear regression, binning method and piecewise regression. Their analysis indicated that there was no clear optimal method that could improve all clinical accuracy metrics.
- Another similar work to predict non-contact core body temperature by testing regression models has been done by Chayabhan et al [8]. They proposed a trained model prediction using IR-measured facial feature temperatures to predict core body temperature. Even though their study has concluded that the linear regression model showed the lowest minimum-root-mean-square error compared with reference temperatures, tree-based models also had good performances on temple plane and eye regions of interest.
- Due to the limitation of linear regression approaches, testing more flexible tree-based methods on the prediction of oral temperature by using this new IRT dataset is very important.

Materials and Methods

Dataset

- The raw data was collected by a clinical study at the Health Center of the University of Maryland (UMD) at College Park from November 2016 to May 2018. There were two IRTs used in this clinical study– one was FLIR (IRT-1) and the other was ICI (IRT-2). In this analysis, we only focused on FLIR measurement in round 1.
- This dataset contained 1020 subjects and 38 variables. There are a total of 26 facial temperature variables and 6 demographic information including gender, age, ethnicity, ambient temperature, humidity, and measuring distance. The targets are **average oral temperatures** under two operation modes (fast mode and monitor mode) of the oral thermometer. For modeling, we did not include subject, cosmetic, time, and date in our analysis.
- In data processing, we dropped those subjects containing missing values and combined the two age groups from 21-25 and 26-30 to one age group from 21-30. In total, there are 1001 subjects left for further analysis.
- To evaluate our modeling, the subjects were randomly separated into training (80%, 600 subjects) and testing (20%, 401 subjects) sets.

Materials and Methods

Methods

- Bagging
- Random Forest
- Gradient-boosting Trees
- Bayesian Additive Regression Trees

Evaluation

- Perform 5-fold cross-validation to optimize parameters for those four tree-based models.
- Apply the testing data to each trained model and calculate root mean squared error (RMSE) as our metrics for comparing their performances.

Bagging

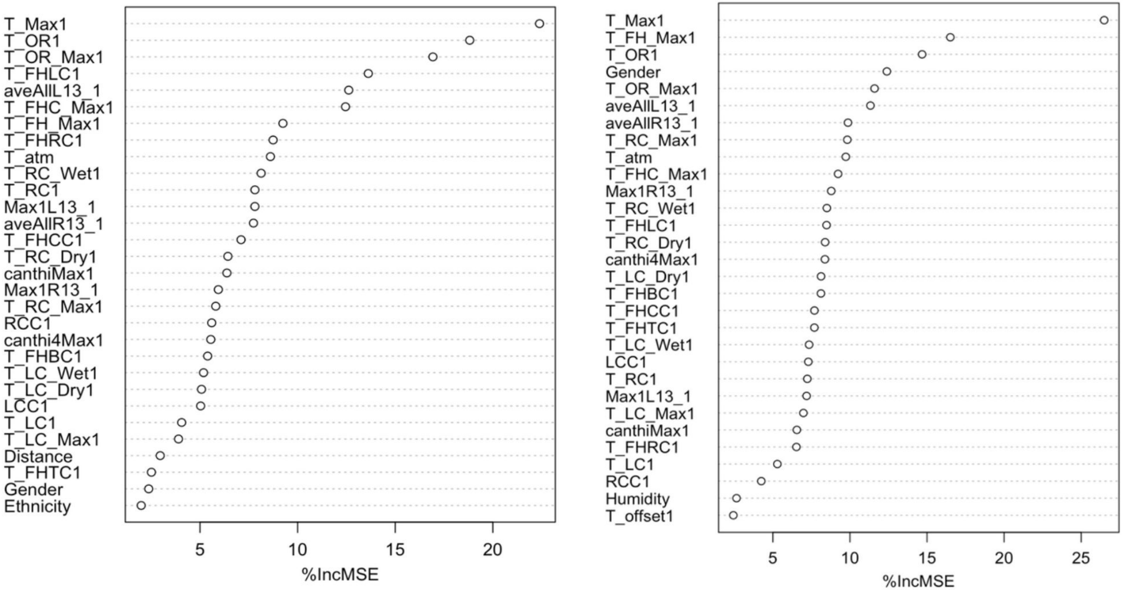

Figure 1. The mean decrease of accuracy in predictions on the OOB samples for each target. The left plot is for aveOralF. The right plot is for aveOralM.

Higher mean decrease in accuracy values suggest that predictors are more important to the model.

Most Important

T_OR1, T_Max1

Strong Effects

T_OR_Max1, aveAllL13_1, T_atm

Only Influential on aveOralF

T_FHLC1, T_FHRC1, T_FHC_Max1

Only Influential on aveOralM

Gender, T_FH_Max1, aveAllR13_1

Random Forest

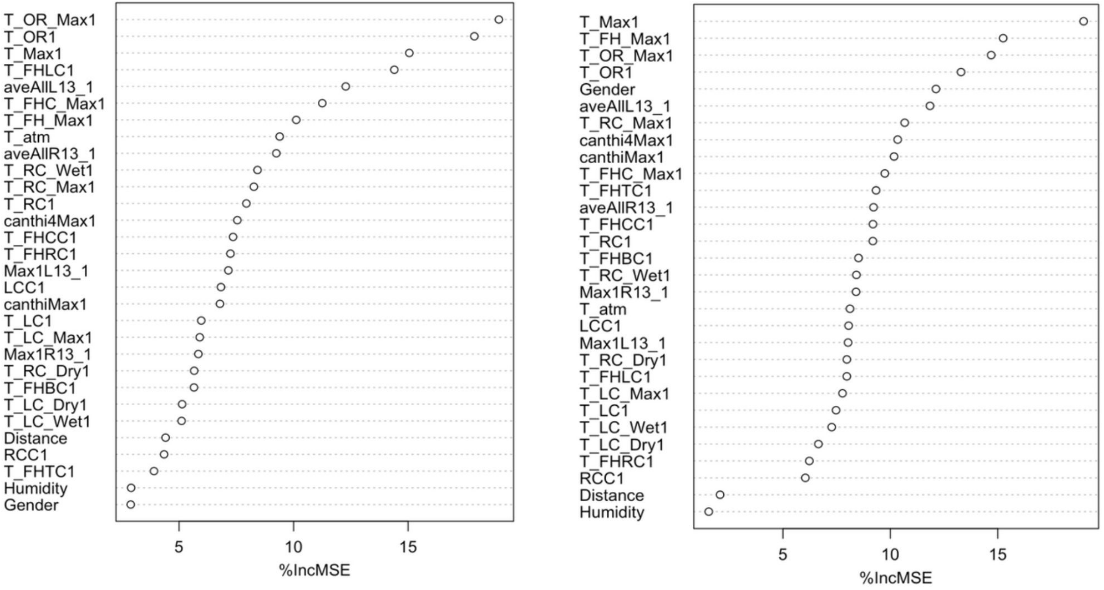

Figure 2. The mean decrease of accuracy in predictions on the OOB samples for each target. The left plot is for aveOralF. The right plot is for aveOralM.

Most Important

T_OR_Max1, T_OR1, T_Max1

Only Influential on aveOralF

T_FHLC1, aveAllL13_1

Only Influential on aveOralM

Gender, T_FH_Max1

A separate modeling is necessary for those two responses.

Gradient-boosting Trees

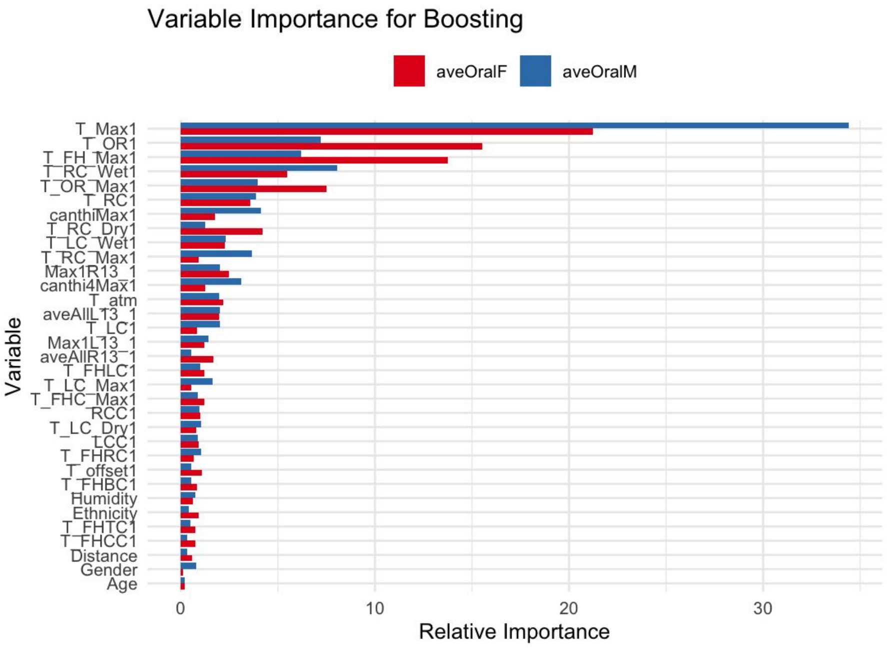

Figure 3. The relative influence of each variable on the training set. Red bars represent aveOralF, and the blue bars represent aveOralM

Most Important

T_Max1, with a higher impact on aveOralM

Strong Effects

T_OR1, with a higher impact on aveOralF

Only Influential on aveOralF

T_RC_Dry1, aveAllL13_1, T_LC_Wet1

Only Influential on aveOralM

canthiMax1, canthi4Max1, T_RC1, Max1L13_1

Bayesian Additive Regression Trees

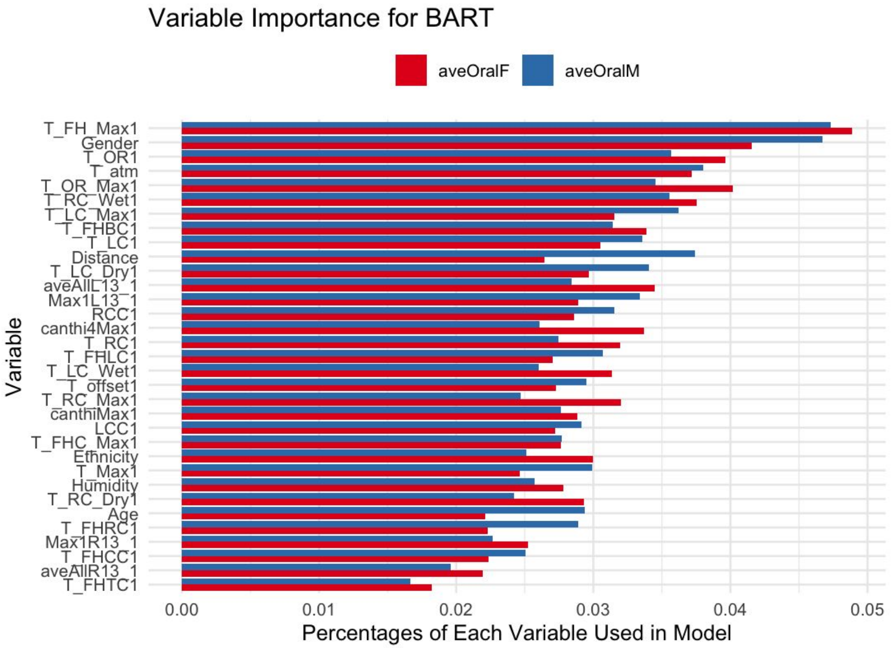

Figure 4. The percentages of each variable used in the BART model on the training set. Red bars represent aveOralF, and the blue bars represent aveOralM

Most Important

T_PH_Max1, Gender

Strong Effects

T_OR1, T_atm, T_OR_Max1

More Influential on aveOralF

T_RC1 and T_RC_Wet1

More Influential on aveOralM

canthiMax1

Best Model

Table 1. The performances of four thee-based regression methods on the test set

	Bagging	RF	**Boosting**	BART
RMSE for aveOralF	0.234	0.234	**0.046**	0.129
RMSE for aveOralM	0.259	0.259	**0.063**	0.195

The boosting method outperforms Bagging, RF, and BART for both aveOralF and aveOralM responses, with the lowest RMSE of 0.046 and 0.063 respectively.

Comparison

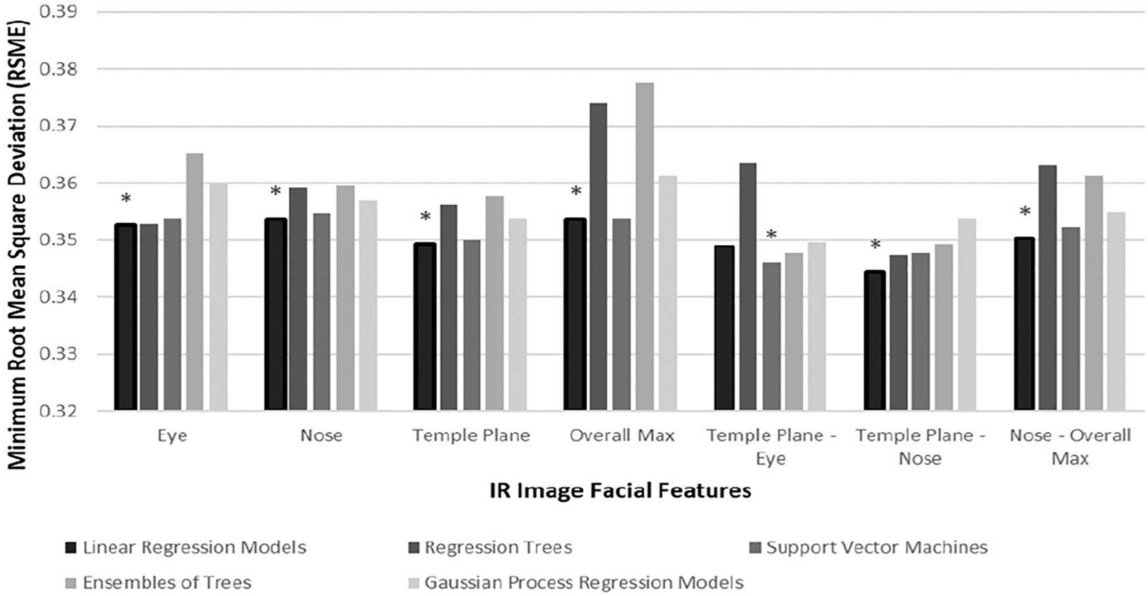

Figure 5. Identification of the best regression model for each IR thermograph input feature, frame, and pair using RMSE from Chayabhan's paper

Chayabhan's paper

Input Variables

Facial feature temperatures from IR images

Data Splite

Training with 90%, testing with 10%

Best Performance

Linear regression models, with RMSE of 0.353

Our Tree-based Model

Boosting model, with RMSE of 0.046

Conclusions and Future Work

Conclusion

- Across tree-based methods, certain predictors like *T_Max1*, *T_OR1*, and *canthiMax1* were consistently influential for the responses *aveOralF* and *aveOralM*.
- With cross-validation, the boosting model with 5000 trees and 4 depth of each tree, significantly preform better the other models in terms of RMSE, making it the best method for our dataset.

Future Work

- The next steps could involve a deeper analysis of the threshold effects observed in the boosting model, particularly for the *T_Max1* variable, which may inform more comprehensive predictive modeling.
- Experimenting with other forms of boosting algorithms, and applying the model to different datasets (different cameras and group) to validate its robustness.
- Furthermore, exploring the integration of the identified influential predictors into a more complex ensemble model, which can perhaps combine the strengths of boosting and BART.

Reference

[1] Chiu W, Lin P, Chiou HY, et al. Infrared Thermography to Mass-Screen Suspected Sars Patients with Fever. *Asia Pacific Journal of Public Health*. 2005;17(1):26-28.

[2] Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020, 395, 497–506.

[3] Zissis G.J., Wolfe W.L. The Infrared Handbook. Technical report, DTIC document. 1978

[4] Ismael Fernández-Cuevas, Joao Carlos Bouzas Marins, Javier Arnáiz Lastras, Pedro María Gómez Carmona, Sergio Piñonosa Cano, Miguel Ángel García-Concepción, Manuel Sillero-Quintana, Classification of factors influencing the use of infrared thermography in humans: A review, Infrared Physics & Technology, Volume 71, 2015

[5] Wang, Q., Zhou, Y., Ghassemi, P., McBride, D., Casamento, J.P., & Pfefer, T.J. (2021). Infrared Thermography for Measuring Elevated Body Temperature: Clinical Accuracy, Calibration, and Evaluation. *Sensors (Basel, Switzerland), 22*

[6] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Springer series in statistics[J]. New York, NY, USA, 2001.

[7] Breiman L. Random forests[J]. Machine learning, 2001, 45: 5-32.

[8] Wang, Q., Zhou, Y., Ghassemi, P., Chenna, D., Chen, M., Casamento, J., Pfefer, J., & Mcbride, D. (2023). Facial and oral temperature data from a large set of human subject volunteers (version 1.0.0). *PhysioNet*.

Reference

[9] Wikipedia contributors. "Bootstrap aggregating." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 3 Feb. 2024. Web. 21 Mar. 2024.

[10] Breiman, L. Bagging predictors. *Mach Learn* **24**, 123–140 (1996).

[11] Wikipedia contributors. "Random forest." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 6 Mar. 2024. Web. 22 Mar. 2024.

[12] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2001: 1189-1232.

[13] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)[J]. The annals of statistics, 2000, 28(2): 337-407.

[14] Chipman H A, George E I, McCulloch R E. BART: Bayesian additive regression trees[J]. 2010.

[15] Hastie T, Tibshirani R. Bayesian backfitting (with comments and a rejoinder by the authors[J]. Statistical Science, 2000, 15(3): 196-223.

[16] Liaw A, Wiener M (2002). “Classification and Regression by randomForest.” *R News*, **2**(3), 18-22.

[17] Ridgeway G. Generalized Boosted Models: A guide to the gbm package[J]. Update, 2007, 1(1): 2007.

[18] Sparapani R, Spanbauer C, McCulloch R. Nonparametric machine learning and efficient computation with Bayesian additive regression trees: the BART R package[J]. Journal of Statistical Software, 2021, 97: 1-66.