Dear Editors,

We would like to thank the Editors and Reviewers for their very positive feedback and insightful comments, which helped us to improve the paper and clarify the details of our study.

We have modified the text according to their comments, and we hope that you will find our revised manuscript adequately addressing all raised issues.

Please find our detailed response to each comment below. We cite the reviewers' comments in blue and indicate manuscript text by quoting and using the font "Times New Roman" showing changed text in red.

In addition, we verified all cited references concerning their relevance and performed some language editing as requested for improving the style of the manuscript. In response to the Editor's request, we have also added a Graphical Abstract to our submission. All modifications are marked as tracked changes in the revised manuscript.

I herewith, with approval of my co-authors, submit the revised version of our manuscript, and we hope, that you will consider it suitable for publication in Atmosphere.

Sincerely,

Peter Bröde

Reviewer 1

The proposed investigation mainly used several machine learning algorithms to analyze and predict the UTCI. The study extended the application of machine learning algorithms. The detailed comments are presented as below:

We are grateful to this reviewer for the positive attitude towards our study and for the constructive comments, which we have addressed as detailed below.

1) In the common machine learning approaches, it's highly recommended to do the pre-analysis of data, including data cleaning, independence analysis, and data preprocessing. Especially for data preprocessing, it can significantly affect the machine learning results. Please explain the reason for not using data preprocessing.

Thank you for pointing us to the relevant topic on data preprocessing. In fact, we had made limited use of data preprocessing by applying log-transformation to one variable. We have added corresponding information to Figures 2 and A1, and a paragraph at the beginning of the "Data analysis" section 2.2, which reads as:

> "As the data were simulated by the UTCI-Fiala model [27], data cleaning was not necessary. To restrict (data) expert input to the algorithms, we applied only limited data preprocessing by log-transforming the percentage skin blood flow data (VblSk)."

2) The training data is important to the performance of machine learning algorithms. Besides focusing on the total number of training data in the manuscript, it's recommended to show the total distribution of the whole data. In addition, what are the resources of the data (from experiment, observation, or simulation)?

We agree with this comment concerning the importance of showing the data distribution and had therefore included Figure 2 showing the distribution of the training data and Figure A1 showing the distribution of the combined training (reference) and test (non-reference) data, respectively. We now include a corresponding text referring to the distribution in the caption and the manuscript at the beginning of section 3, which reads:

> "Figure 2 illustrates the distribution of the training data, i.e., the 48-dimensional set of twelve physiological output variables at four points in time predicted by the UTCI-Fiala model [27] for the UTCI reference conditions with air temperature ranging from -55 °C to +55 °C."

In addition, we state that the data arose as output from simulations performed using the UTCI-Fiala model, e.g., in section 2 'Methods and materials':

> "Our approach was to apply selected SL algorithms deeming representative for recent applications [5-11,35-41] to the multi-dimensional data simulated over a comprehensive grid of relevant climatic conditions by the UTCI-Fiala model [27] at stage 2 of the UTCI development (Figure 1)."

3) It's not enough to evaluate the prediction performance by only using RMSE. It will be more convictive by showing the combination of different performance metrics, such as R-square and RMSE.

In response to this comment, we have added the new supplemental Figure A2 to Appendix A showing that utilizing the squared correlation (R^2) and the mean absolute error (MAE) as alternative metrics confirmed the results obtained with RMSE as presented in Figure 3A. We refer to this new Figure expanding the results section 3.1 as follows:

> "UTCI and the equivalent temperatures predicted by the diverse algorithms were highly correlated with squared correlation (R^2) exceeding 0.95 for any dataset and with any method as shown in Appendix A by the supplemental Figure A2.
>
> […]
>
> The results for RMSE were confirmed by the corresponding analyses utilizing the squared correlation coefficient (R^2) and the mean absolute error (MAE) as alternative performance metrics, as shown in Figure A2."

4) The regression results in Figure A1 is not fully analyzed. What are the contributions of the regression analysis in Figure A1 for the proposed investigation?

In response to this comment, we have expanded the explanation of Figure A1 in Appendix A, which now reads:

> "Figure A1 illustrates the distribution and correlation of air temperature with the twelve physiological responses at four points in time predicted by the UTCI-Fiala model [27] for the combined reference and non-reference datasets, i.e. training and test data, including the individual regression lines. The latter were used by the KDM algorithm [58,59] for ensemble modelling of the equivalent temperature as averaged predictions from the inverted regression lines weighted by the squared correlation coefficients, as discussed in section 4.2."

5) The title of this manuscript is " ... compared the expert judgement ...". However, the expert judgement is only used as the training data for the machine learning algorithms. The comparison between the machine learning and the expert judgement is not clear. It's suggested to present the detailed comparison or to revise the title of the manuscript.

In response to this comment (and to a related comment from reviewer 2 concerning novelty aspects), we have re-phrased the second para of the introduction referring to 'expert judgement', as follows:

> "The rising number of SL applications trigger a demand for quantitatively assessing the skills of statistical learning in comparison to results involving expert judgement. Our study aimed to contribute to such an assessment utilizing as a testbed the development of the Universal Thermal Climate Index (UTCI)…"

6) It's still not clear about the reasons of low accuracy for PCA and t-SNE analysis. It may attract more research interests. It's suggested to further explain the results to attract the potential readers.

Adhering to this hind, we have expanded the corresponding para in the discussion, which stresses on the different approaches of clustering algorithms looking for patterns in the data, vs limit values derived from external sources involving expert judgement in thermos-physiology and ergonomics, and now reads:

> "The dimensionality-reduction results by PCA and t-SNE in Figure 4A both suggested a one-dimensional structure of thermos-physiological strain concurring with the original analyses presented as stage 4 of Figure 1 [25]. In addition, the clustering algorithms were able to identify a trend from extreme heat to extreme cold. However, they did not reliably discriminate between the intermediate categories, with only marginal differences between the diverse algorithms. This demonstrates the discrepancy between clustering algorithms searching for patterns in the data and forming groups of data of comparable size on one hand [57] and the definition of UTCI stress categories based on thermo-physiological knowledge and ergonomics reasoning on the other hand [25,34]."

In addition, we have added a subsection including an outlook, which may further trigger further research:

> "4.3 Limitations and outlook
>
> As a limitation concerning generalizability, the database underlying the UTCI development had been generated by a deterministic model of human thermoregulation. Thus, it lacks random variation induced by, e.g., inter-individual or day-to-day variability of the human physiological responses, which might have complicated the analyses of this study. However, the best performing SL algorithms for equivalent temperature prediction, i.e., concerning stage 4 in Figure 1, like random forests and k-nearest neighbors had already been suggested and successfully applied in other fields pertaining to climatic change and thermal stress [35-41].
>
> While the KDM approach as outlined in section 4.2 might facilitate the development of equivalent temperature indices by avoiding the necessity for specifying reference conditions (stage 3 in Figure 1), assessment scales (stage 5 in Figure 1) automatically derived by clustering algorithms may still require

expert knowledge for their refinement. Future enhancements at this stage may be possible by including further artificial intelligence tools such as expert systems or large language models working with knowledge databases [62,63]."

7) The conclusions are highly suggested to be revised. For example, the statement "the potential supportive role for AI and the utilized SL algorithms." is not clear. How is AI used and discussed in the manuscript?

Besides, the conclusions are not fully covered the main research results in the manuscript

We are grateful to this reviewer for pointing out that the manuscript deals with statistical learning (SL) as one pillar of AI application, and not with AI per se. Thus, in response to this comment, we do not longer refer to 'AI' in the conclusion. In addition, we now recapitulate the major outcomes in the conclusion section, which reads as:

"In summary, a potential supportive role for the utilized SL algorithms when analyzing high dimensional input in thermal index development can be concluded from their adequate performance in equivalent temperature prediction. On the other hand, the low agreement of the assessment scale defined by the UTCI expert group with clustering-based thermal stress categories suggested that statistical learning algorithms will not (yet) fully replace the knowledgeable expert in biometeorological and inter-disciplinary research and thermal risk assessment."

Please also refer to our response to the similar last comment of reviewer 2.

In addition, please refer to the detailed information in the track changes attached to our responses.

Reviewer 2

The work is devoted to the topic of modeling and statistical assessment of weather conditions. This topic may be of interest to a wide range of researchers.

We very much appreciate that this reviewer recognizes the value of our study, and are gratefully acknowledging the detailed constructive comments, which we have addressed as outlined below.

Figure 1 – Scheme is very interesting but hardly could be read. Please, rework the size. Also, highlighted stages can't be seen clearly.

Following this hint, we have reworked Figure 1 by enlarging the figure, increasing the text size and changing the color used for highlighting; and we hope that this will increase the legibility of Figure 1.

Abstract should be reworked. First sentence is not suitable. Also, this sentence is duplicated in line 37-28 and it does not sound well. Please, reformulate it.

In response to this comment, we have rephrased the corresponding sentence and re-structured the abstract, which now reads:

"The Universal Thermal Climate Index (UTCI), a complex tool for the assessment of outdoor thermal stress created by an international expert group, is an equivalent temperature index based on the 48-dimensional output of an advanced model of human thermoregulation formed by 12 variables at four consecutive 30-minute intervals, which were calculated for 105,642 thermal conditions from extreme cold to extreme heat. This study assessed the skills of statistical learning (SL) in supplementing or even replacing the expert knowledge inherent to UTCI by comparing the performance of SL algorithms to the results accomplished by an international endeavor involving more than 40 experts from 23 countries, We found that random forests and k-nearest neighbors closely predicted UTCI values, but that clustering applied after dimension reduction algorithms (principal component analysis and t-distributed stochastic neighbor embedding) were inadequate for risk assessment in relation to the UTCI stress categories. This indicates a potential supportive role for SL, as it will not (yet) fully replace the bio-meteorological expert knowledge."

We have also re-phrased the second para of the introduction, as outlined below in our response to the next comment.

What is the novelty of the work? Please, highlight novelty in the text in the Introduction section.

In response to this comment, we have re-phrased the second paragraph of the introduction focusing on the assessment of SL skills, as announced by the title. The text now reads:

"The rising number of SL applications trigger a demand for quantitatively assessing the skills of statistical learning in comparison to results involving expert judgement. Our study aimed to contribute to such an assessment utilizing as a testbed the development of the Universal Thermal Climate Index (UTCI)…"

Conclusion section is too short. It must be obtained. It is necessary to clearly summarize all the observations established in the work. It is important to write whether the tasks of the work have been solved and to what extent, as well as whether there are prospects for this topic or it has been exhausted.

We appreciate this comment, which is similar to comment (7) of reviewer 1. Thus, we have not only updated the conclusion section accordingly, but have also included a new subsection 4.3 "Limitations and outlook", which now includes the prospected utility of our findings, where the whole section reads as follows:

"4.3 Limitations and outlook

As a limitation concerning generalizability, the database underlying the UTCI development had been generated by a deterministic model of human thermoregulation. Thus, it lacks random variation induced by, e.g., inter-individual or day-to-day variability of the human physiological responses, which might have complicated the analyses of this study. However, the best performing SL algorithms for equivalent temperature prediction, i.e., concerning stage 4 in Figure 1, like random forests and k-nearest neighbors had already been suggested and successfully applied in other fields pertaining to climatic change and thermal stress [35-41].

While the KDM approach as outlined in section 4.2 might facilitate the development of equivalent temperature indices by avoiding the necessity for specifying reference conditions (stage 3 in Figure 1), assessment scales (stage 5 in Figure 1) automatically derived by clustering algorithms may still require expert knowledge for their refinement. Future enhancements at this stage may be possible by including further artificial intelligence tools such as expert systems or large language models working with knowledge databases [62,63].

5. Conclusions

In summary, a potential supportive role for the utilized SL algorithms when analyzing high dimensional input in thermal index development can be concluded from their adequate performance in equivalent temperature prediction. On the other hand, the low agreement of the assessment scale defined by the UTCI expert group with clustering-based thermal stress categories suggested that statistical learning algorithms will not (yet) fully replace the knowledgeable expert in biometeorological and inter-disciplinary research and thermal risk assessment."

In addition, please refer to the detailed information in the track changes attached to our responses.

atmosphere

Article

Skills of statistical learning algorithms in thermal stress assessment compared with the expert judgement inherent to the Universal Thermal Climate Index (UTCI)

Peter Bröde [1,*], Dusan Fiala [2] and Bernhard Kampmann [3]

[1] Leibniz Research Centre for Working Environment and Human Factors at TU Dortmund (IfADo), Ardeystrasse 67, D-44139 Dortmund, Germany; broede@ifado.de
[2] ErgonSim–Human Thermal Modelling, Robert-Bosch-Str.20, 72469 Messstetten; dfiala@ergonsim.de
[3] School of Mechanical Engineering and Safety Engineering, Department of Occupational Health Science, University of Wuppertal, Germany; kampmann@uni-wuppertal.de
* Correspondence: broede@ifado.de; Tel.: +49 231 1084 225

Abstract: ~~.~~ ~~The objective of this paper was to verify the applicability of statistical learning (SL) compared to human reasoning with respect to the Universal Thermal Climate Index (UTCI), a complex tool for the assessment of outdoor thermal stress~~ The Universal Thermal Climate Index (UTCI), a complex tool for the assessment of outdoor thermal stress created by an international expert group, ~~. UTCI~~ is an equivalent temperature index based on the 48-dimensional output of an advanced model of human thermoregulation formed by 12 variables at four consecutive 30-minute intervals, which were calculated for 105,642 thermal conditions from extreme cold to extreme heat. This study assessed the skills of statistical learning (SL) in supplementing or even replacing the expert knowledge inherent to UTCI by c~~C~~omparing the performance of SL algorithms to the results accomplished by an international endeavor involving more than 40 experts from 23 countries, ~~w~~We found that random forests and k-nearest neighbors closely predicted UTCI values, but that clustering applied after dimension reduction algorithms (principal component analysis and t-distributed stochastic neighbor embedding) were inadequate for risk assessment in relation to the UTCI stress categories. This indicates a potential supportive role for SL, as it will not (yet) fully replace the bio-meteorological expert knowledge.

Keywords: bio-meteorological index; heat stress; cold stress; high dimensional data; artificial intelligence; machine learning; UTCI

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

Statistical or machine learning (SL) is central to artificial intelligence (AI) applications [1-3] with potential relevance to environmental risk assessment, especially in settings with high dimensional input as for thermal stress indices [4]. There, they may assist or even attempt replacing the bio-meteorological expert judgement, as indicated by the increasing number of recent applications in diverse fields of biometeorology concerning, e.g., indoor and outdoor thermal comfort, the impact assessment of climate change related heat stress, urban planning, the adaptation of buildings and human behavior to changing climatic conditions, or establishing a link between human physiology and thermal sensation [5-11].

The rising number of SL applications trigger a demand for quantitatively assessing the skills of statistical learning in comparison to results involving expert judgement. Our study aimed to contribute to such an assessment utilizing as a testbed the development of ~~objective of this paper was to verify the applicability of SL compared to human reasoning with respect to~~ the Universal Thermal Climate Index (UTCI), a complex assessment tool for the physiological strain related to the outdoor thermal environment [12], which soon after its release had been widely adopted and applied in core fields of biometeorology covering weather services, public health research and epidemiology, precautionary planning of heat health warning systems and urban environments including applications to tourism and health resorts, and climate impact research [13-24].

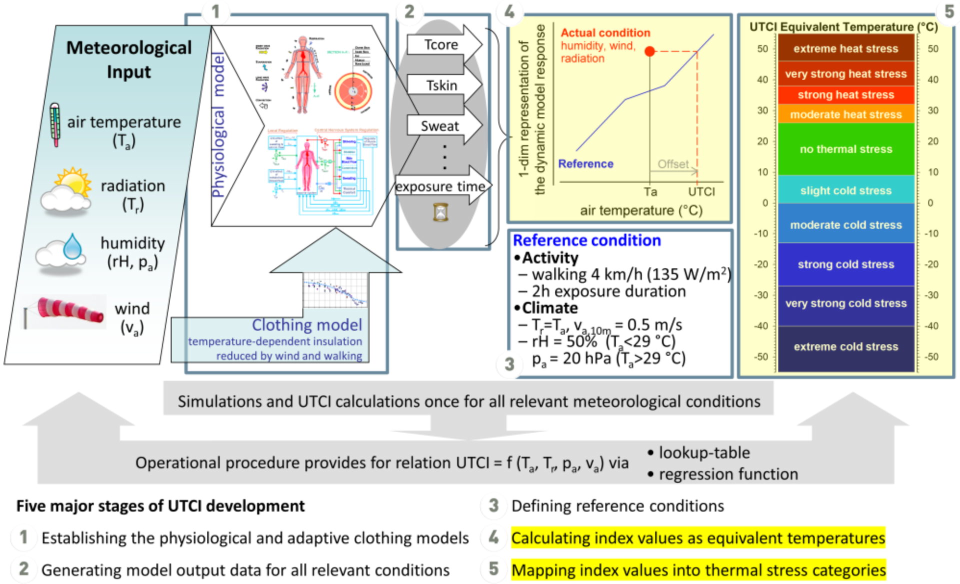

Figure 1. Stages of developing the elements of the operational procedure for calculating UTCI equivalent temperatures with thermal stress categories, modified from [25], where the highlighted stages 4 and 5 are of major concern in this study.

The development of UTCI was accomplished by an international inter-disciplinary endeavor [26] involving more than 40 experts from 23 countries [12]. Figure 1 visualizes the concept and the major stages of UTCI development [25].

UTCI was conceptualized as an equivalent temperature (in °C) defined as air temperature of the reference condition with the same dynamic physiological response as the actual condition. The physiological response to thermal stress was derived at stage 1 from the output of an advanced human model of thermoregulation [27], which had been coupled with an adaptive clothing model with clothing insulation changing depending on air temperature [28]. Extensive simulation runs were performed at stage 2, which also included the validation of predicted physiological responses against experimental laboratory and field data [29-32]. After the experts reached consensus about the definition of reference conditions in stage 3, stage 4 derived UTCI on an equivalent temperature scale. This involved a multivariate approach [25,33] including a dimension reduction step of the multidimensional model output to a one-dimensional strain indicator by principal component analysis. Then, UTCI values for non-reference conditions concerning wind speed, humidity as well as solar and thermal radiation were identified by searching the reference condition with the same strain indicator value (Figure 1). At stage 5, for assessment purposes, a scale classifying the UTCI values into ten categories of thermal stress was added to the operational procedure [25]. The assessment scale was established by consensus and expert judgement involving the comparison of the model output values concerning the thermal state of the human body including thermal sensation and effectors of thermoregulation to established ergonomics limit criteria regarding cold and heat stress [25,34].

Thus, UTCI integrates human expert reasoning with the high-dimensional model output from numerous simulations performed for a huge grid of relevant conditions representing the reference and non-reference climates ranging from the extreme cold to extreme heat. Therefore, UTCI and its underlying data represent a suitable test case for benchmarking the skills of SL algorithms against expert knowledge concerning the comprehensive assessment of the human responses to thermal stress in biometeorological applications.

Focusing on the data-driven stages 4 and 5 of the UTCI development (Figure 1), the aim of this study was to compare the performance of SL algorithms to the outcome from the UTCI development performed by the international expert group [12].

2. Materials and Methods

Our approach was to apply selected SL algorithms deeming representative for recent applications [5-11,35-41] to the multi-dimensional data simulated over a comprehensive grid of relevant climatic conditions by the UTCI-Fiala model [27] at stage 2 of the UTCI development (Figure 1).

2.1. UTCI data

The operational procedure [25] was based on the dynamic physiological response characterized by the 48-dimensional model output formed by 12 variables at 4 consecutive 30-minute intervals (Table 1). This output was generated for 1,051 reference conditions, which were characterized by low wind speed (0.5 m/s measured 10 m above ground level), by relative humidity of 50%, but water vapor pressure capped at 20 hPa for air temperatures above 29 °C, and by shadowed conditions with mean radiant temperature equaling air temperature, with the latter covering the temperature range from -110 °C to +100 °C with 0.2 K increment. Notably, for reference conditions, UTCI values are known and equal to air temperature by definition (Figure 1).

Another set of 104,591 non-reference conditions with varying levels of air temperature (-50 °C to +50 °C with 1 K increment), wind speed (0.5 to 30.3 m/s 10 m above ground), relative humidity (5% to 100%, but with maximum water vapor pressure of 50 hPa) and solar and thermal radiation (mean radiant temperature from 30 K below to 70 K above air temperature with 5 K increment) had to be valued in UTCI equivalent temperatures (°C) and classified in 10 stress categories ranging from extreme cold to extreme heat, as indicated by stages 4 and 5 in Figure 1.

For independent test purposes, these analyses were repeated with an external set of 1,000 non-reference conditions, which had been simulated previously [25,42], and served as non-reference test data in this study.

Table 1. Physiological output variables obtained from the UTCI-Fiala model [27].

Physiological Variable[1]	Abbreviation[2]	Unit
rectal temperature	Tre	°C
mean skin temperature	Tskm	°C
facial skin temperature	Tskfc	°C
hand skin temperature	Tskhn	°C
total net heat loss	Qsk	W
evaporative (latent) heat loss	Esk	W
sweat rate	Mskdot	g/min
metabolic heat production	Metab	W
heat generated by shivering	Shiv	W
skin wettedness	wettA	% of body area
skin blood flow	VblSk	% of basal value

cardiac output	sVbl	% of basal value

[1] In addition, scores of the dynamic thermal sensation (DTS) on the 7-unit ASHRAE [43] scale (-3:cold; -2:cool; -1:slightly cool; 0:neutral; +1:slightly warm; +2:warm; +3:hot) were estimated from the physiological output [44].
[2] Each output variable was calculated for 30, 60, 90 and 120 min of simulated exposure duration.

2.2. Data analysis

As the data were simulated by the UTCI-Fiala model [27], data cleaning was not necessary. To restrict (data) expert input to the algorithms, we applied only limited data preprocessing by log-transforming the percentage skin blood flow data (VblSk).

Splitting the reference conditions in sets of 840 training and 211 test data and using the external 1,000 non-reference conditions as additional non-reference test data, we compared the results of the UTCI expert group to diverse SL algorithms in predicting UTCI equivalent temperature values from the 48 predictors, using the root-mean squared error (RMSE) as performance metric with error defined as deviation of UTCI to the equivalent temperatures predicted by the SL algorithms. The supervised learning techniques comprised linear regression (MLR: multiple linear regression; LASSO: least absolute shrinkage and selection operator), tree-based and ensemble methods (CART: classification and regression trees; RF: random forests; XGBoost: extreme gradient boosting) as well as support vector machines (SVN), and the non-parametric k-nearest neighbors (KNN) [1-3,45-47]. More specifically, because UTCI equivalent temperature equals air temperature for reference conditions, we could label UTCI equivalent temperatures for the reference training data and did utilize the above-mentioned regression algorithms for predicting equivalent temperature by the 48 predictors (12 variables at 4 points in time) listed in Table 1.

For comparison to the UTCI assessment scale, UTCI values were then categorized by hierarchical and k-means clustering [1] after applying principal component analysis (PCA) and t-distributed stochastic neighbor embedding t-SNE [48], respectively, to the high-dimensional UTCI model output for dimensionality reduction. In accordance with the approach of the UTCI expert group~~approach~~ [25], we searched for ten categories using the UTCI reference data set limited to the 926 values with air temperatures between -110 °C and +75 °C covering the range of UTCI values obtained for non-reference conditions [25]. This enabled ranking the resulting clusters by the intra-cluster mean air temperatures, which were equal to UTCI for the reference conditions according to the equivalent temperature definition (Figure 1).

Metrics of agreement between the UTCI stress categories and the classification found by clustering were derived from the confusion matrix by calculating the overall accuracy, defined as proportion of the UTCI stress categories that were classified correctly, and by Cohen's kappa, defined as $(p_o-p_e)/(1-p_e)$, with p_o denoting the observed proportion of agreement and p_e the hypothetical proportion of agreement due to chance [49]. The calculations were performed with the statistical software R version 4.3.3 [50] using the packages caret [51], xgboost [47], yardstick [52], tidyverse [53], and cowplot [54].

3. Results

Figure 2 illustrates the distribution of the training data, i.e., the 48-dimensional set of twelve physiological output variables at four points in time predicted by the UTCI-Fiala model [27] for the UTCI reference conditions with air temperature ranging from -55 °C to +55 °C. Similarly, Figure A1 depicts the corresponding data for the combined reference and non-reference conditions depending on air temperature.

While several parameters, like rectal, mean skin, and facial temperatures, as well as skin blood flow showed a time-dependent pattern of values increasing from cold to heat stress, other parameters showed non-monotonous relationships, e.g., metabolic heat production increasing both in the cold due to shivering, and in the heat due to the so-called Q_{10}-effect, i.e., the rise in metabolic rate associated with increasing body core temperature.

The UTCI-Fiala model had implemented an increase of metabolic rate by about 7% per 1 K rise in core temperature in accordance with the results of climate chamber experiments [27,55]. Such non-linear and non-monotonous patterns of stress-strain relationships with respect to the air temperature-depending physiological responses, as well as the shivering response only occurring in cold conditions, posed a particular challenge to the learning algorithms.

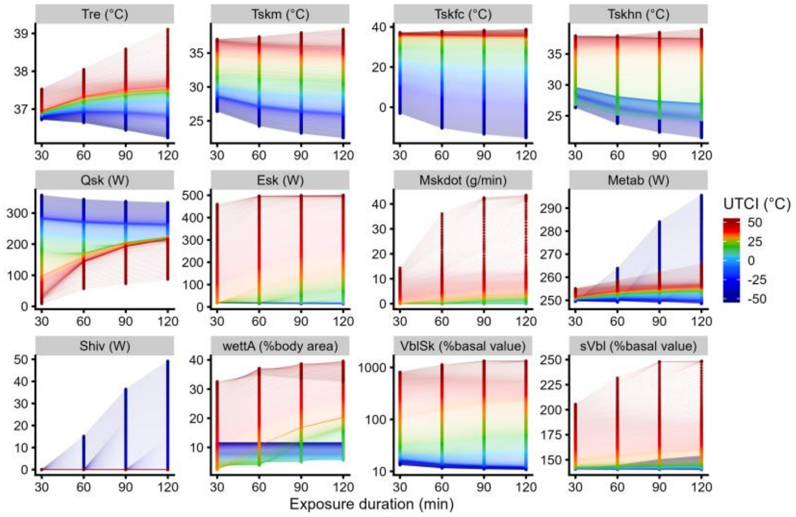

Figure 2. Set of twelve physiological output variables (abbreviations in Table 1) at four points in time from the simulated reference conditions plotted representing the distribution of the training data with UTCI ranging from -55 °C to +55 °C. Note the log-scale applied to the skin blood flow data (VblSk) shown in the third column of the bottom row.

3.1. UTCI equivalent temperature calculation

UTCI and the equivalent temperatures predicted by the diverse algorithms were highly correlated with squared correlation (R^2) exceeding 0.95 for any dataset and with any method as shown in Appendix A by the supplemental Figure A2.

However, performance differences became obvious in Figure 3A showing the typical prediction error RMSE obtained from the separate algorithms for the three sets of training and test data, respectively. While multiple linear regression and LASSO worked well for UTCI reference conditions, RMSE increased to more than 7 °C for the non-reference data, approximately corresponding to one heat stress category deviation (Figure 1). RMSE values of similar magnitude for the non-reference test data occurred with CART and SVM, whereas gradient boosting lowered the prediction error. The best performance was observed with RF and KNN, both showing an excellent fit for the reference data with RMSE<0.2 °C, and a good agreement with RMSE of about 3 °C for non-reference UTCI, where RMSE for KNN was below the prediction error obtained for calculating UTCI from meteorological input by the polynomial regression function provided by the operational procedure (Figure 1) [25]. Among the tree-based algorithms, CART performance was inferior to RF and gradient boosting (XGBoost) in all cases in accordance with earlier observations in different application areas [46]. The results for RMSE were confirmed by the corresponding analyses utilizing the squared correlation coefficient (R^2) and the mean absolute error (MAE) as alternative performance metrics, as shown in Figure A2.

As an essential requirement to thermal stress indices, identical index values should indicate equivalent thermal strain [12]. Thus, Figure 3B evaluates this criterion for the equivalent temperatures determined by the diverse SL algorithms in comparison to the original UTCI concerning the dynamic thermal sensation (DTS) from the non-reference test data averaged over the four simulated points in time (30, 60, 90, and 120 min). The resulting pattern for the best performing algorithms RF and KNN was similar to the original UTCI showing a small variation around the reference values, whereas the other algorithms, especially CART, showed larger and partly systematic deviations (Figure 3B).

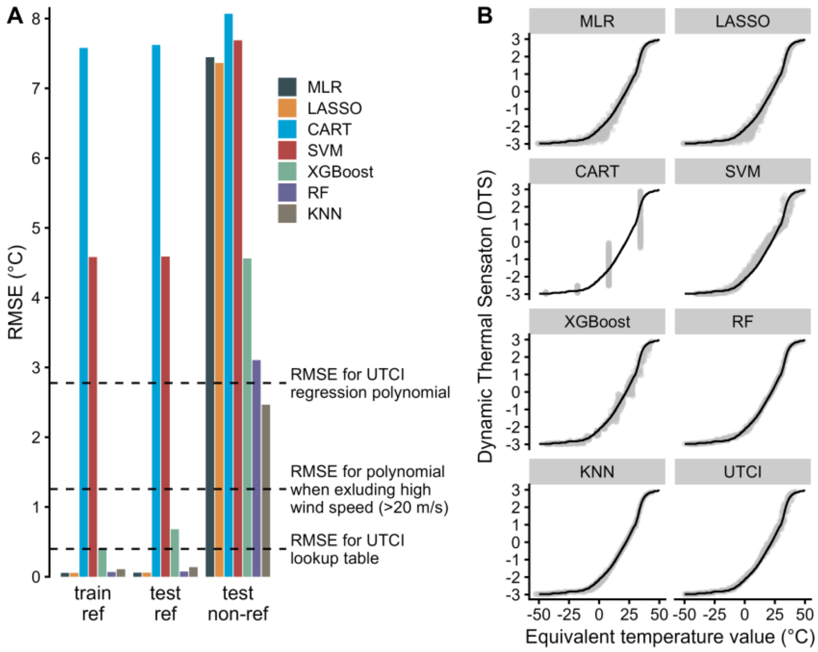

Figure 3. (**A**) Root-mean squared error (RMSE) for predicting UTCI equivalent temperatures from 48 physiological variables by different supervised SL algorithms (MLR: multiple linear regression; LASSO: least absolute shrinkage and selection operator; CART: classification and regression trees; SVM: support vector machines; XGBoost: extreme gradient boosting; RF: random forests; KNN: k-nearest neighbors) for the training reference (train ref) data and two test datasets for reference (ref) and non-reference (non-ref) UTCI conditions. For comparison, horizontal reference lines indicate the RMSE for calculating UTCI from meteorological input using the polynomial regression function or the lookup table provided by the operational procedure [25]. (**B**) Dynamic thermal sensation votes from the non-reference test data depending on equivalent temperature predicted by the SL algorithms from (A) in comparison to the original UTCI (lower right panel), with lines indicating the values for the reference conditions.

3.2 UTCI assessment scale and thermal stress categories

As illustrated by Figure 4A, the UTCI-Fiala model output projected to the first two dimensions by PCA, which explained about 90% of total variance, exhibited a one-dimensional structure along the UTCI categories from extreme cold to extreme heat, which followed an increasing course concerning the first principal component (dim1) in accordance with the original analysis [25]. Similarly, a one-dimensional structure also emerged with t-SNE, however, the resulting curve followed a non-monotonous pattern in each of the two dimensions (dim1 and dim2).

Figure 4B shows the confusion matrices comparing the groups formed by hierarchical or k-means clustering following dimensionality reduction by PCA and t-SNE, respectively, to the categories of the UTCI assessment scale. Although correct classifications followed a trend from extreme cold to heat, discrepancies concerning the intermediate UTCI categories were obvious.

Consequently, the quantification of agreement in Figure 4C resulted in low to fair levels of agreement according Cohen's kappa varying between 0.1 and 0.4 [49]. The different methods as defined by the combinations of the dimension-reduction (PCA vs t-SNE) with the clustering algorithms (hierarchical vs k-means) performed similarly with k-means clustering following t-SNE yielding the highest scores for both the accuracy (0.395) and the kappa coefficient (0.33), respectively.

Figure 4. (**A**) Two-dimensional projections from principal component analysis (PCA), and t-distributed stochastic neighbor embedding (t-SNE), respectively, of the reference data (48 physiological variables plus 4 calculated dynamic thermal sensations) colored by the ten UTCI categories. (**B**) Confusion matrices showing the proportion (Prop) of conditions in the original ten UTCI stress categories (y-axis) assigned to the corresponding number of groups by hierarchical (upper panels) and k-means-clustering (lower panels) applied to the projections from PCA and t-SNE, respectively. (**C**) Metrics (accuracy and Cohen's kappa) on the agreement of the UTCI stress categories with the classifications by the different clustering algorithms applied to PCA and t-SNE projections.

4. Discussion

4.1 UTCI approach compared to SL algorithms

Our results concerning the equivalent temperature index UTCI indicate that recent machine learning algorithms such as random forests or k-nearest neighbors were capable to produce equivalent temperatures deviating from UTCI with RMSE of about 3 °C for the general, non-reference use case. For comparison, this deviation was similar to, or in case of the k-nearest neighbor regression algorithm even below the RMSE associated with calculating UTCI from the meteorological input variables using the polynomial regression function [25]. The non-parametric k-nearest neighbor regression algorithm might be better suited to capture non-monotonous stress-strain relationships like metabolic rate increasing under both cold and heat stress conditions (Figure 2, Figure A1), which could explain the superior performance of k-nearest neighbor regression compared to the other SL algorithms. The deviations found for the two best performing algorithms might be considered acceptable given that uncertainties in the measurement or estimation of the mean radiant temperature in application scenarios could lead to absolute deviations above 6 K concerning the UTCI calculation by the polynomial regression function, which might even increase to more than 8 K when taking the other meteorological input variables (air temperature, humidity, wind speed) into account [56]. In addition, the values of dynamic thermal sensation (DTS) showed only limited variability at equivalent temperatures produced by random forest and k-nearest neighbors (Figure 3B), which was of similar magnitude as observed for the original UTCI. As identical equivalent temperature values should indicate identical thermal strain [12], this supports the potential usefulness of random forest and k-nearest neighbor regression for thermal index development.

~~As a limitation concerning generalizability, the database underlying the UTCI development had been generated by a deterministic model of human thermoregulation. Thus, it lacks random variation induced by, e.g., inter-individual or day-to-day variability of the human physiological responses, which might have complicated the analyses of this study. However, the best performing algorithms like random forests and k-nearest neighbors had already been suggested and successfully applied in other fields concerning climatic change and thermal stress [35-41].~~

In summary, our results suggest that selected SL algorithms may be helpful tools to derive summarizing metrics for complex environmental assessment tasks characterized by high dimensional data describing stress and strain. Notably, these useful features concerned a narrow, but important data-analytical task within the UTCI development (stage 4 in Figure 1), which itself had been based on multivariate statistical methods [25] bearing some similarities with the SL algorithms applied in this study. From the opposite perspective, the close similarities of UTCI with equivalent temperatures calculated by modern machine learning algorithms suggest that the UTCI approach still appears appropriate considering recent data analytic developments.

~~Though~~ The dimensionality-reduction results by PCA and t-SNE in Figure 4A both suggested a one-dimensional structure of thermos-physiological strain concurring with the original analyses presented as stage 4 of Figure 1 [25]. In addition, the clustering algorithms were able to identify a trend from extreme heat to extreme cold.~~,~~ However, they did not reliably discriminate between the intermediate categories, with only marginal differences between the diverse algorithms. This demonstrates the discrepancy between clustering algorithms searching for patterns in the data and forming groups of data of comparable size on one hand [57] and the definition of UTCI stress categories based on thermo-physiological knowledge and ergonomics reasoning on the other hand [25,34]. ~~Future enhancements at this stage (stage 5 in Figure 1) may be possible by including further artificial intelligence tools such as expert systems or large language models working with knowledge databases [58,59].~~

4.2 Alternative approach by ensemble modelling

Ensemble modelling techniques such as random forests and gradient boosting, which had shown good performance in this study (Figure 3), aim at enhancing the predictive capabilities of simple models by averaging single predictions of numerous models and have been successfully applied to the assessment of human biological features changing with aging by a single quantity termed 'biological age' [58,59]. Biological age is defined as the chronological age of a population reference associated with identical levels of single or composite aging biomarkers, where the population reference is usually taken as means of a comprehensive sample obtained from large-scale clinical or epidemiological studies [58]. Thus, the concept of biological age bears close similarities with equivalent temperature based thermal indices [12] with age taking the role of temperature [60].

Therefore, as additional analysis, we applied the widely used Klemera-Doubal method (KDM) [59], which predicts the outcome of interest, here air temperature replacing age, as weighted average of the predictions obtained by inverting the individual regression lines, as shown in Figure A1 for the predicted 48 physiological responses from the combined reference and non-reference datasets. The weights are chosen by the KDM algorithm proportional to the squared correlation coefficients. We calculated KDM equivalent temperatures using the R package BioAge [58], and compared them in Figure 5 to the UTCI values determined for the external test data [42]. Both types of equivalent temperatures were highly correlated with UTCI values above KDM equivalent temperatures in the heat and below KDM in the cold (Figure 5).

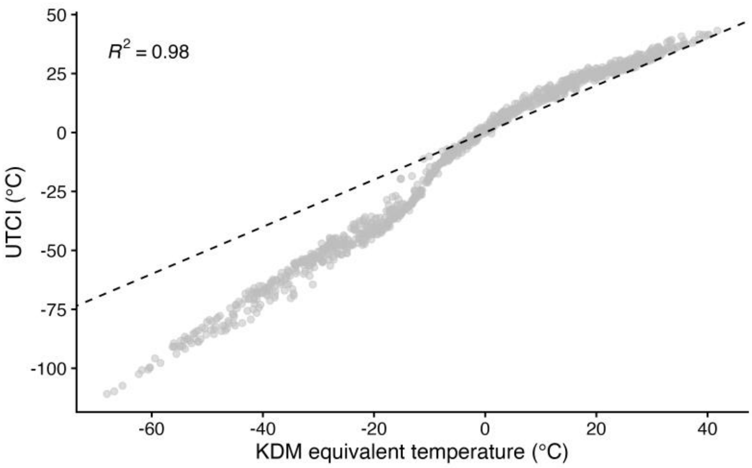

Figure 5. UTCI values and KDM equivalent temperatures for the external test data comprising 1,000 non-reference conditions with squared correlation coefficient (R^2) and dashed line of identity (y=x).

It is important to note that the KDM algorithms allows for the determination of equivalent temperatures without defining reference conditions, thus omitting stage 3 from Figure 1. In KDM, the reference conditions are unknown, but implicitly referring to those conditions representing the 'population average' in the underlying data, with air temperature equaling the predicted equivalent temperature after the complex ensemble modelling algorithm involving inverting and weighting individual regression functions. Concerning the combined reference and non-reference data, this population average will be characterized by higher values of wind speed and mean radiant temperature compared to the UTCI reference conditions. This explains, at least partly, the pattern observed in Figure 5, because the non-reference test data will be further away from the UTCI reference than from the implicit population average in KDM in terms of wind speed and mean radiant temperature. This will increase the offset, i.e., the shift of equivalent to air temperature as shown in Figure 1 for UTCI compared to KDM, and lead to higher UTCI values in the heat and lower values in the cold.

As the KDM procedure removes the necessity for finding consensus about the definition of the reference conditions, it could simplify the development process of an equivalent temperature index. On the other hand, it will make the resulting equivalent temperature less interpretable due to the unspecified reference conditions. Consequently, the utilization of the KDM algorithm for the development of an equivalent temperature index will require the most careful design of the underlying database, which will implicitly also define the (unknown) reference to which all conditions will be compared.

Hence, we believe that the UTCI approach with a specified reference will promote results that are more interpretable in terms of thermal stress and physiological strain, which forms an essential requirement for assessment procedures concerning the human responses to the thermal environment [12,61].

Of course, the KDM algorithm could also be used to derive predictive equations for the equivalent temperatures from the reference data only and then apply those to the non-reference data, in the same manner as it was accomplished for the other SL algorithms in section 3.1. Implementing this analysis, KDM showed inferior performance with RMSE exceeding 23 °C for the reference and 42 °C for the non-reference conditions, far above the deviations reported for the other SL algorithms in Figure 3A. According to this result, the application of KDM for the development of equivalent temperature indices will be limited to the special case when the definition of reference conditions will not be possible or warranted.

4.3 Limitations and outlook

As a limitation concerning generalizability, the database underlying the UTCI development had been generated by a deterministic model of human thermoregulation. Thus, it lacks random variation induced by, e.g., inter-individual or day-to-day variability of the human physiological responses, which might have complicated the analyses of this study. However, the best performing SL algorithms for equivalent temperature prediction, i.e., concerning stage 4 in Figure 1, like random forests and k-nearest neighbors had already been suggested and successfully applied in other fields pertaining to climatic change and thermal stress [35-41].

While the KDM approach as outlined in section 4.2 might facilitate the development of equivalent temperature indices by avoiding the necessity for specifying reference conditions (stage 3 in Figure 1), assessment scales (stage 5 in Figure 1) automatically derived by clustering algorithms may still require expert knowledge for their refinement. Future enhancements at this stage may be possible by including further artificial intelligence tools such as expert systems or large language models working with knowledge databases [62,63].

5. Conclusions

In summary, ~~our results indicate the~~ a potential supportive role for ~~AI and~~ the utilized SL algorithms when analyzing high dimensional input in thermal ~~stress assessment and equivalent temperature~~ index development can be concluded from their adequate performance in equivalent temperature prediction.~~,~~ ~~although~~ On the other hand, the low

hat formatiert: Schriftfarbe: Schwarz

hat formatiert: Schriftfarbe: Schwarz

hat formatiert: Nicht unterstrichen, Schriftfarbe: Schwarz agreement of the assessment scale defined by the UTCI expert group with clustering-based thermal stress categories suggested that statistical learning algorithms~~it~~ will not (yet) fully replace the knowledgeable expert in biometeorological and inter-disciplinary research and thermal risk assessment.

Supplementary Materials: ~~Not applicable~~Supplementary figures are included in Appendix A.

Author Contributions: Conceptualization, P.B.; D.F. and B.K.; methodology, P.B.; software, D.F.; validation, P.B.; formal analysis, P.B.; investigation, P.B.; D.F. and B.K.; data curation, P.B.; writing—original draft preparation, P.B.; writing—review and editing, D.F. and B.K.; visualization, P.B.; funding acquisition, B.K. All authors have read and agreed to the published version of the manuscript.

Funding: UTCI was developed within COST Action 730; the COST office is supported by the EU framework program Horizon Europe.

Institutional Review Board Statement: Not applicable for this study not involving humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are partly available as supplemental material to previous publications [25] or stored in public repositories [42]. Further raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors acknowledge and thank all members of the former COST Action 730 for support and contributions to fruitful discussions.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

This appendix contains ~~a~~ supplemental Figures A1 and A2.

Figure A1 ~~showing~~ illustrates the distribution and correlation of air temperature with the twelve physiological responses at four points in time predicted by the UTCI-Fiala model [27] for the combined reference and non-reference datasets, i.e. training and test data, including the individual regression lines. ~~, as they~~The latter –were used by the KDM algorithm [58,59] for ensemble modelling of the equivalent temperature as averaged predictions from the inverted regression lines weighted by the squared correlation coefficients, as discussed in section 4.2.

Figure A2 presents the squared correlation coefficients (R^2) and the mean absolute error (MAE) as goodness-of-fit metrics complementing the root mean squared error (RMSE) presented in Figure 3A for the UTCI equivalent temperature predictions from diverse SL algorithms.

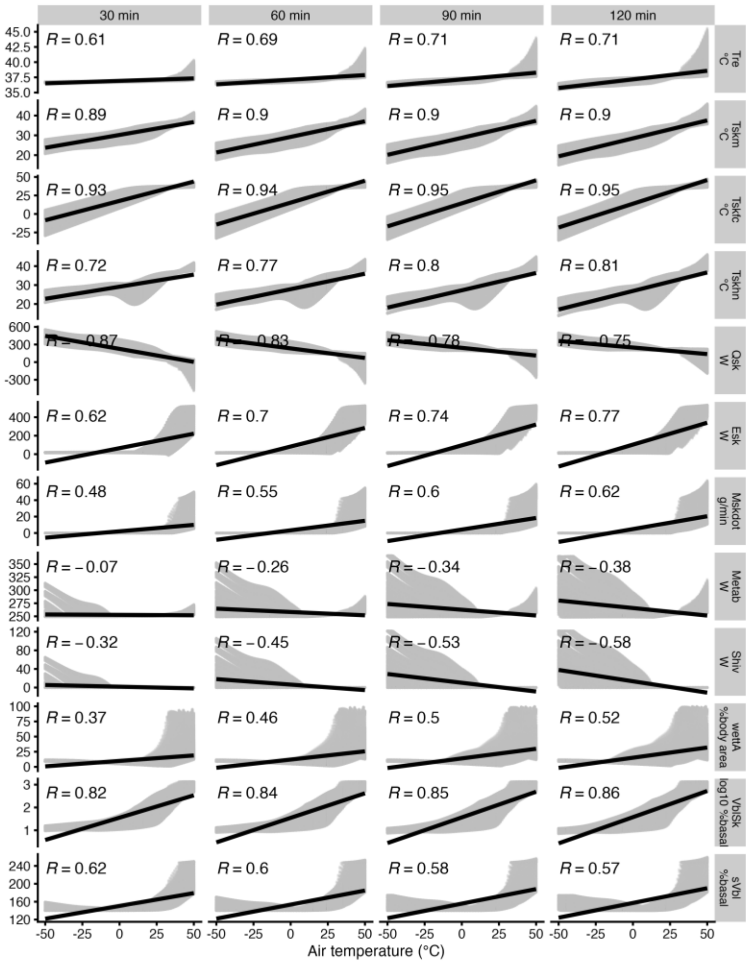

Figure A1. Set of twelve physiological output variables (abbreviations in Table 1) at four exposure times from the combined reference and non-reference conditions plotted related to air temperatures ranging from -50°C to +50 °C with correlation coefficients (R) and linear regression lines. Note the log-transformation applied to the skin blood flow data (VblSk) shown in the second row from the bottom up.

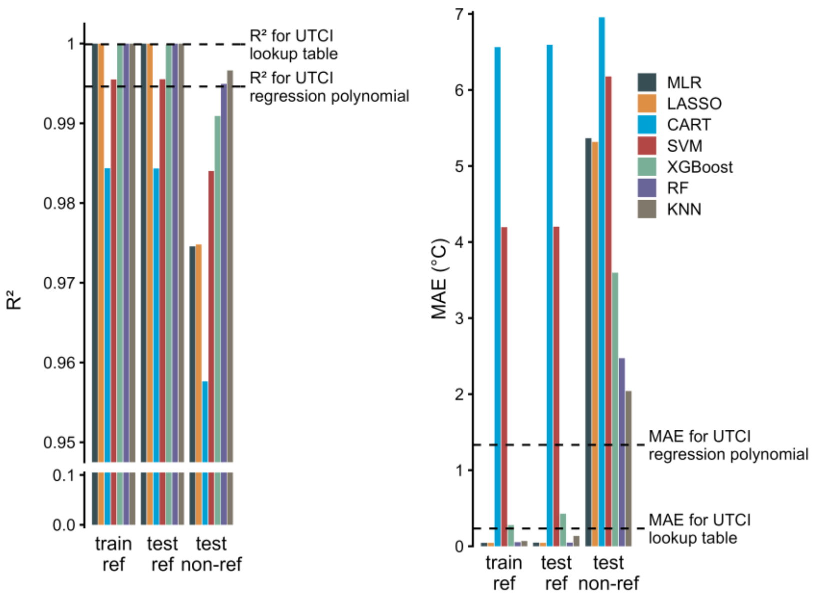

Figure A2. Squared correlation coefficients (R^2, left panel) and mean absolute error (MAE, right panel) as goodness-of-fit metrics in addition to the RMSE presented in Figure 3A for predicting UTCI equivalent temperatures from 48 physiological variables by different supervised SL algorithms (MLR: multiple linear regression; LASSO: least absolute shrinkage and selection operator; CART: classification and regression trees; SVM: support vector machines; XGBoost: extreme gradient boosting; RF: random forests; KNN: k-nearest neighbors) for the training reference (train ref) data and two test datasets for reference (ref) and non-reference (non-ref) UTCI conditions. For comparison, horizontal reference lines indicate the corresponding metrics for calculating UTCI from meteorological input using the polynomial regression function or the lookup table provided by the operational procedure [25].

References

1. Berk, R.A. *Statistical Learning from a Regression Perspective*; Springer International Publishing: Cham, 2020, doi:10.1007/978-3-030-40189-4.
2. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer New York: New York, NY, 2009, doi:10.1007/978-0-387-84858-7.
3. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*; Springer New York: New York, NY, 2013, doi:10.1007/978-1-4614-7138-7.
4. Matzarakis, A. Curiosities about Thermal Indices Estimation and Application. *Atmosphere* **2021**, *12*, 721, doi:10.3390/atmos12060721.
5. Ali, A.; Jayaraman, R.; Azar, E.; Maalouf, M. A comparative analysis of machine learning and statistical methods for evaluating building performance: A systematic review and future benchmarking framework. *Building and Environment* **2024**, *252*, 111268, doi:10.1016/j.buildenv.2024.111268.
6. Aparicio-Ruiz, P.; Barbadilla-Martín, E.; Guadix, J.; Muñuzuri, J. Predicting the clothing insulation through machine learning algorithms: A comparative analysis and a practical approach. *Building Simulation* **2024**, doi:10.1007/s12273-024-1114-9.
7. Hamed, M.M.; Al-Hasani, A.A.J.; Nashwan, M.S.; Sa'adi, Z.; Shahid, S. Assessing the growing threat of heat stress in the North Africa and Arabian Peninsula region connected to climate change. *Journal of Cleaner Production* **2024**, *447*, 141639, doi:10.1016/j.jclepro.2024.141639.

Formatiert: MDPI_7.1_References, Einzug: Links: 0 cm, Erste Zeile: 0 cm

8. Guo, R.; Yang, B.; Guo, Y.; Li, H.; Li, Z.; Zhou, B.; Hong, B.; Wang, F. Machine learning-based prediction of outdoor thermal comfort: Combining Bayesian optimization and the SHAP model. *Building and Environment* **2024**, *254*, 111301, doi:10.1016/j.buildenv.2024.111301.
9. Wang, J.; Li, Q.; Zhu, G.; Kong, W.; Peng, H.; Wei, M. Recognition and prediction of elderly thermal sensation based on outdoor facial skin temperature. *Building and Environment* **2024**, *253*, 111326, doi:10.1016/j.buildenv.2024.111326.
10. Wang, M.; Gou, Z. Gaussian Mixture Model based classification for analyzing longitudinal outdoor thermal environment data to evaluate comfort conditions in urban open spaces. *Urban Climate* **2024**, *53*, 101792, doi:10.1016/j.uclim.2023.101792.
11. Choi, Y.; Seo, S.; Lee, J.; Kim, T.W.; Koo, C. A machine learning-based forecasting model for personal maximum allowable exposure time under extremely hot environments. *Sustainable Cities and Society* **2024**, *101*, 105140, doi:10.1016/j.scs.2023.105140.
12. Jendritzky, G.; de Dear, R.; Havenith, G. UTCI - Why another thermal index? *International Journal of Biometeorology* **2012**, *56*, 421-428, doi:10.1007/s00484-011-0513-7.
13. Geletič, J.; Lehnert, M.; Krč, P.; Resler, J.; Krayenhoff, E.S. High-Resolution Modelling of Thermal Exposure during a Hot Spell: A Case Study Using PALM-4U in Prague, Czech Republic. *Atmosphere* **2021**, *12*, 175, doi:10.3390/atmos12020175.
14. Dimitriadou, L.; Nastos, P.; Zerefos, C. Defining Heatwaves with Respect to Human Biometeorology. The Case of Attica Region, Greece. *Atmosphere* **2021**, *12*, 1100, doi:10.3390/atmos12091100.
15. Zeng, D.; Wu, J.; Mu, Y.; Deng, M.; Wei, Y.; Sun, W. Spatial-Temporal Pattern Changes of UTCI in the China–Pakistan Economic Corridor in Recent 40 Years. *Atmosphere* **2020**, *11*, 858, doi:10.3390/atmos11080858.
16. Basarin, B.; Lukić, T.; Matzarakis, A. Review of Biometeorology of Heatwaves and Warm Extremes in Europe. *Atmosphere* **2020**, *11*, 1276, doi:10.3390/atmos11121276.
17. Staiger, H.; Laschewski, G.; Matzarakis, A. Selection of Appropriate Thermal Indices for Applications in Human Biometeorological Studies. *Atmosphere* **2019**, *10*, 18, doi:10.3390/atmos10010018.
18. Di Napoli, C.; Pappenberger, F.; Cloke, H.L. Verification of Heat Stress Thresholds for a Health-Based Heat-Wave Definition. *Journal of Applied Meteorology and Climatology* **2019**, *58*, 1177-1194, doi:10.1175/jamc-d-18-0246.1.
19. Potchter, O.; Cohen, P.; Lin, T.-P.; Matzarakis, A. A systematic review advocating a framework and benchmarks for assessing outdoor human thermal perception. *Science of The Total Environment* **2022**, *833*, 155128, doi:10.1016/j.scitotenv.2022.155128.
20. Urban, A.; Di Napoli, C.; Cloke, H.L.; Kyselý, J.; Pappenberger, F.; Sera, F.; Schneider, R.; Vicedo-Cabrera, A.M.; Acquaotta, F.; Ragettli, M.S., et al. Evaluation of the ERA5 reanalysis-based Universal Thermal Climate Index on mortality data in Europe. *Environmental Research* **2021**, *198*, 111227, doi:10.1016/j.envres.2021.111227.
21. Di Napoli, C.; Barnard, C.; Prudhomme, C.; Cloke, H.L.; Pappenberger, F. ERA5-HEAT: A global gridded historical dataset of human thermal comfort indices from climate reanalysis. *Geoscience Data Journal* **2020**, *8*, 2-10, doi:10.1002/gdj3.102.
22. Di Napoli, C.; Pappenberger, F.; Cloke, H.L. Assessing heat-related health risk in Europe via the Universal Thermal Climate Index (UTCI). *International Journal of Biometeorology* **2018**, *62*, 1155-1165, doi:10.1007/s00484-018-1518-2.
23. Ioannou, L.G.; Tsoutsoubi, L.; Mantzios, K.; Vliora, M.; Nintou, E.; Piil, J.F.; Notley, S.R.; Dinas, P.C.; Gourzoulidis, G.A.; Havenith, G., et al. Indicators to assess physiological heat strain – Part 3: Multi-country field evaluation and consensus recommendations. *Temperature* **2022**, *9*, 274-291, doi:10.1080/23328940.2022.2044739.
24. Brimicombe, C.; Di Napoli, C.; Cornforth, R.; Pappenberger, F.; Petty, C.; Cloke, H.L. Borderless Heat Hazards With Bordered Impacts. *Earth's Future* **2021**, *9*, e2021EF002064, doi:10.1029/2021EF002064.
25. Bröde, P.; Fiala, D.; Blazejczyk, K.; Holmér, I.; Jendritzky, G.; Kampmann, B.; Tinz, B.; Havenith, G. Deriving the operational procedure for the Universal Thermal Climate Index (UTCI). *International Journal of Biometeorology* **2012**, *56*, 481-494, doi:10.1007/s00484-011-0454-1.
26. McGregor, G.R. Special issue: Universal Thermal Climate Index (UTCI). *International Journal of Biometeorology* **2012**, *56*, 419-419, doi:10.1007/s00484-012-0546-6.
27. Fiala, D.; Havenith, G.; Bröde, P.; Kampmann, B.; Jendritzky, G. UTCI-Fiala multi-node model of human heat transfer and temperature regulation. *International Journal of Biometeorology* **2012**, *56*, 429-441, doi:10.1007/s00484-011-0424-7.
28. Havenith, G.; Fiala, D.; Blazejczyk, K.; Richards, M.; Bröde, P.; Holmér, I.; Rintamaki, H.; Ben Shabat, Y.; Jendritzky, G. The UTCI-clothing model. *International Journal of Biometeorology* **2012**, *56*, 461-470, doi:10.1007/s00484-011-0451-4.
29. Psikuta, A.; Fiala, D.; Laschewski, G.; Jendritzky, G.; Richards, M.; Blazejczyk, K.; Mekjavic, I.B.; Rintamäki, H.; de Dear, R.; Havenith, G. Validation of the Fiala multi-node thermophysiological model for UTCI application. *International Journal of Biometeorology* **2012**, *56*, 443-460, doi:10.1007/s00484-011-0450-5.
30. Kampmann, B.; Bröde, P.; Fiala, D. Physiological responses to temperature and humidity compared to the assessment by UTCI, WGBT and PHS. *International Journal of Biometeorology* **2012**, *56*, 505-513, doi:10.1007/s00484-011-0410-0.
31. Bröde, P.; Krüger, E.L.; Rossi, F.A.; Fiala, D. Predicting urban outdoor thermal comfort by the Universal Thermal Climate Index UTCI - a case study in Southern Brazil. *International Journal of Biometeorology* **2012**, *56*, 471-480, doi:10.1007/s00484-011-0452-3.
32. Bröde, P.; Kampmann, B. Temperature–Humidity-Dependent Wind Effects on Physiological Heat Strain of Moderately Exercising Individuals Reproduced by the Universal Thermal Climate Index (UTCI). *Biology* **2023**, *12*, 802, doi:10.3390/biology12060802.

33. Bröde, P.; Fiala, D.; Kampmann, B.; Havenith, G.; Jendritzky, G. Der Klimaindex UTCI - Multivariate Analyse der Reaktion eines thermophysiologischen Simulationsmodells. In Proceedings of 55. Kongress der Gesellschaft für Arbeitswissenschaft, Dortmund, 2009 04.-06.03.; pp. 705-708.
34. Bröde, P.; Blazejczyk, K.; Fiala, D.; Havenith, G.; Holmér, I.; Jendritzky, G.; Kuklane, K.; Kampmann, B. The Universal Thermal Climate Index UTCI Compared to Ergonomics Standards for Assessing the Thermal Environment. *Industrial Health* **2013**, *51*, 16-24, doi:10.2486/indhealth.2012-0098.
35. Aguilera, J.J.; Korsholm Andersen, R.; Toftum, J. Prediction of Indoor Air Temperature Using Weather Data and Simple Building Descriptors. *International Journal of Environmental Research and Public Health* **2019**, *16*, 4349, doi:10.3390/ijerph16224349.
36. Benita, F.; Tunçer, B. Exploring the effect of urban features and immediate environment on body responses. *Urban Forestry & Urban Greening* **2019**, *43*, 126365, doi:10.1016/j.ufug.2019.126365.
37. Berrang-Ford, L.; Sietsma, A.J.; Callaghan, M.; Minx, J.C.; Scheelbeek, P.F.D.; Haddaway, N.R.; Haines, A.; Dangour, A.D. Systematic mapping of global research on climate and health: a machine learning review. *The Lancet Planetary Health* **2021**, *5*, e514-e525, doi:10.1016/S2542-5196(21)00179-0.
38. Kim, J.; Schiavon, S.; Brager, G. Personal comfort models – A new paradigm in thermal comfort for occupant-centric environmental control. *Building and Environment* **2018**, *132*, 114-124, doi:10.1016/j.buildenv.2018.01.023.
39. Liu, K.; Nie, T.; Liu, W.; Liu, Y.; Lai, D. A machine learning approach to predict outdoor thermal comfort using local skin temperatures. *Sustainable Cities and Society* **2020**, *59*, 102216, doi:10.1016/j.scs.2020.102216.
40. Mantzios, K.; Ioannou, L.G.; Panagiotaki, Z.; Ziaka, S.; Périard, J.D.; Racinais, S.; Nybo, L.; Flouris, A.D. Effects of Weather Parameters on Endurance Running Performance: Discipline Specific Analysis of 1258 Races. *Medicine & Science in Sports & Exercise* **2021**, *54*, 153-161, doi:10.1249/mss.0000000000002769.
41. Shin, J.-Y.; Min, B.; Kim, K.R. High-resolution wind speed forecast system coupling numerical weather prediction and machine learning for agricultural studies — a case study from South Korea. *International Journal of Biometeorology* **2022**, *66*, 1429-1443, doi:10.1007/s00484-022-02287-1.
42. Bröde, P. UTCI-Test-Data. Zenodo: 2021. Availabe online: https://doi.org/10.5281/zenodo.5503968 (accessed on 2022-10-04).
43. ASHRAE. *Standard 55 - Thermal Environmental Conditions for Human Occupancy*; ASHRAE Inc.: Atlanta, 2004.
44. Fiala, D.; Lomas, K.J.; Stohrer, M. First principles modeling of thermal sensation responses in steady-state and transient conditions. *ASHRAE Transactions* **2003**, *109*, 179-186.
45. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5-32, doi:10.1023/A:1010933404324.
46. Genuer, R.; Poggi, J.-M. *Random Forests with R*; Springer International Publishing: Cham, 2020, doi:10.1007/978-3-030-56485-8.
47. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA; pp. 785–794.
48. van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579-2605.
49. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **1960**, *20*, 37-46, doi:10.1177/001316446002000104.
50. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2024. Availabe online: https://www.R-project.org/ (accessed on 2024-03-09).
51. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **2008**, *28*, 1 - 26, doi:10.18637/jss.v028.i05.
52. Kuhn, M.; Vaughan, D.; Hvitfeldt, E. yardstick: Tidy Characterizations of Model Performance – R package version 1.3.1. 2024. Availabe online: https://CRAN.R-project.org/package=yardstick (accessed on 2024-04-19).
53. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.A.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J., et al. Welcome to the Tidyverse. *Journal of Open Source Software* **2019**, *4*, 1686, doi:10.21105/joss.01686.
54. Wilke, C.O. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2' – R package version 1.1.3. 2024. Availabe online: https://CRAN.R-project.org/package=cowplot (accessed on 2024-04-18).
55. Kampmann, B.; Bröde, P. Heat Acclimation Does Not Modify Q10 and Thermal Cardiac Reactivity. *Frontiers in Physiology* **2019**, *10*, 1524, doi:10.3389/fphys.2019.01524.
56. Weihs, P.; Staiger, H.; Tinz, B.; Batchvarova, E.; Rieder, H.; Vuilleumier, L.; Maturilli, M.; Jendritzky, G. The uncertainty of UTCI due to uncertainties in the determination of radiation fluxes derived from measured and observed meteorological data. *International Journal of Biometeorology* **2012**, *56*, 537-555, doi:10.1007/s00484-011-0416-7.
57. Härdle, W.; Simar, L. *Applied Multivariate Statistical Analysis*, 2nd ed.; Springer-Verlag: Berlin, 2007.
58. Kwon, D.; Belsky, D.W. A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. *GeroScience* **2021**, *43*, 2795-2808, doi:10.1007/s11357-021-00480-5.
59. Klemera, P.; Doubal, S. A new approach to the concept and computation of biological age. *Mechanisms of Ageing and Development* **2006**, *127*, 240-248, doi:10.1016/j.mad.2005.10.004.
60. Bröde, P.; Claus, M.; Gajewski, P.D.; Getzmann, S.; Wascher, E.; Watzl, C. From Immunosenescence to Aging Types–Establishing Reference Intervals for Immune Age Biomarkers by Centile Estimation. *International Journal of Molecular Sciences* **2023**, *24*, 13186, doi:10.3390/ijms241713186.

61. Notley, S.R.; Mitchell, D.; Taylor, N.A.S. A century of exercise physiology: concepts that ignited the study of human thermoregulation. Part 3: Heat and cold tolerance during exercise. *European Journal of Applied Physiology* **2024**, *124*, 1-145, doi:10.1007/s00421-023-05276-3.
62. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y., et al. A Survey on Evaluation of Large Language Models. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, Article 39, doi:10.1145/3641289.
63. Asemi, A.; Ko, A.; Nowkarizi, M. Intelligent libraries: a review on expert systems, artificial intelligence, and robot. *Library Hi Tech* **2021**, *39*, 412-434, doi:10.1108/LHT-02-2020-0038.