1

Predictive Modeling for Heatstroke Risk Forecasting Integrating Physiological Features Using Ensemble Classifier

Md Mamun Sheikh* [1], Shahera Hossain [2], Md Atiqur Rahman Ahad [3]
[1]University of Dhaka, Bangladesh,[2]University of Asia Pacific, Bangladesh,
[3]University of East London, UK fluctuate

Abstract

Heatstroke is life-threatening, with rising mortality rates attributed to increasing global temperatures. Acute discernment of heatstroke symptoms remains imperative for effective interventions. We proposed an advanced post-processing technique to forecast heatstroke risk, enhancing machine learning algorithms based on physiological features. We implemented an ensemble model that amalgamates the statistical and forecasting models' outputs. Utilizing statistical models such as Random Forest, Ada-Boost, and Support Vector Machine, we attained an accuracy of 96.19±1.6%. Employing forecasting models like Auto-Regressive Integrated Moving Average (ARIMA), Seasonal ARIMA, and Prophet yielded 96.69±1.45% classification accuracy. The ensemble of statistical and forecasting models exhibited outstanding metrics, including precision of 99.16±0.15%, recall of 98.93±0.14%, and F1-score of 98.68±0.10% in thermal comfort forecasting. This groundbreaking endeavor holds promise for advancing the state-of-the-art in heatstroke prevention, augmenting the accuracy and reliability of the predictive models, and consequently, benefiting public health initiatives.

Keywords— Heatstroke; Thermal Comfort Forecasting; Physiological Features; Machine Learning; Model Ensemble.

[1]mamunsheikhdueee52@gmail.com
[2]shaherahossain@gmail.com
[3]atiqahad@gmail.com

1 Introduction

Thermal comfort, the subjective experience of feeling comfortable and satisfied with one's surrounding thermal environment, plays a pivotal role in human well-being and productivity. Achieving an optimal thermal comfort level is crucial for maintaining physical comfort, mental well-being, and overall health. Individuals exposed to extreme temperatures or unsuitable thermal conditions can experience discomfort, reduced productivity, and even health issues, making the study and prediction of thermal comfort a vital area of research [1, 2, 3].

In recent years, there has been a concerning escalation in heat-related fatalities in the United Kingdom. As per the Office for National Statistics, the number of deaths attributed to heat-related illnesses since 2017 has reached a troubling figure of 10,064. The devastating impact of the summer heat waves in 2022 was evident in the staggering death toll of 56,303 in England and Wales. This troubling trend is further exacerbated by the mounting impact of global warming, leading to the intensification of heat waves. Projected estimates indicate a potential rise of over 5°C in external air temperatures in the UK by 2070, posing a significant risk of overheating in buildings. Consequently, the mortality rate due to overheating is anticipated to surge by 260% by the 2050s [14, 15, 16].

Over the years, extensive research has been conducted in the field of thermal comfort, aiming to understand the factors that influence human perception and response to varying thermal conditions. The heatstroke challenge dataset [4] allows for predicting personal thermal comfort using physiological features. This study aims to develop robust models for accurate forecasts, benefiting public health and urban planning.

In this research, we have proposed a robust post-processing technique that leverages an ensemble of different classifiers to predict heatstroke risk. By combining the outputs of diverse classifiers, our ensemble approach enhances predictive accuracy and reliability, contributing to more effective heatstroke risk prediction. The main contribution of our proposed methodology are:

- Leveraging statistical models like Random Forest and Ada-Boost for non-linear and complex feature interactions and feature importance insight
- Employing forecasting models like ARIMA and Prophet for effective short-term forecasting and capturing temporal patterns and seasonality
- Finally, employing an ensemble approach to capitalize on individual strengths, offset weaknesses, minimize biases, and avoid over-fitting

The wealth of research in this area has led to the development of promising techniques and methods, enabling more accurate and efficient heatstroke risk prediction and thermal comfort forecasting. The authors of [5] conducted a comprehensive investigation employing various algorithms, including K-Nearest Neighbors (KNN), LightGBM, and Convolutional Neural Networks (CNN), for heatstroke risk prediction or thermal comfort forecasting. Remarkably, these models yielded an impressive accuracy of 97.6%. In the papers [6, 7], the authors explored the implementation of various classical machine learning models using the dataset [4], achieving a promising level of accuracy.

On the other hand, numerous deep learning frameworks have been employed for heat-

stroke forecasting or thermal comfort prediction. These powerful techniques, such as Neural Networks, CNN, and Recurrent Neural Networks (RNNs), have shown great potential in capturing intricate patterns and relationships in physiological and environmental data. In [8], the author introduced an artificial neural network (ANN) based model that achieved a notable forecasting accuracy of 91%. The research highlights the potential of artificial neural networks as a powerful tool for improving forecasting accuracy and enhancing our understanding of thermal comfort dynamics. In [9], the author employed a bidirectional long short-term memory (Bi-LSTM) model, achieving a promising accuracy of 94.82%.

In the domain of thermal comfort prediction, traditional time series forecasting models like ARIMA, SARIMA (Seasonal ARIMA), and modern techniques like Prophet have been widely employed to capture and forecast the temporal patterns of thermal comfort data. In [10], the author introduced a fusion model that combines the strengths of ARIMA, SARIMA, and LSTM for thermal comfort prediction. This fusion approach aims to capitalize on the unique capabilities of each model to improve the overall forecasting performance. The author of [11] proposed a SARIMA and support vector machine (SVM) based forecasting model that can capable of predicting the thermal comfort in buildings. By incorporating these models into the prediction pipeline, researchers and practitioners can gain valuable insights into thermal comfort dynamics and make informed decisions to enhance comfort and well-being in various environments.

In addition to individual forecasting models, ensemble techniques serve as a powerful strategy to further improve the precision and robustness of thermal comfort and heatstroke forecasting. Ensemble methods amalgamate the predictions from multiple diverse models. By leveraging model diversity and combining their outputs, ensemble techniques effectively mitigate the limitations and biases of individual models, leading to enhanced overall predictive performance. The author in [12] introduced a bagging-based ensemble technique, demonstrating its reliability and high accuracy in thermal perception prediction. In their publication [13], Park et al. introduced an ensemble transfer learning model that combines pre-trained deep learning and machine learning models to forecast thermal comfort using physiological data. This novel approach leverages the benefits of transfer learning, where knowledge learned from one domain is transferred and adapted to a related domain, improving the performance and efficiency of thermal comfort prediction.

2 Methodology

The essence of any real-world problem-solving endeavor resides in the methodology adopted. A superior technique has the potential to significantly enhance productivity across diverse contexts. In this section, we expound upon our procedure employed for precise forecasting of personal thermal comfort and prediction of heatstroke risk in the healthcare domain, leveraging physiological and extracted features. Our paradigm is structured into several integral sections, each contributing to the comprehensive approach of our study.

- Data pre-processing

- Features extraction
- Model evaluation and training
- Post-processing

The previously expounded methodology is illustrated in Figure 1, where it takes the form of a meticulously designed process diagram.

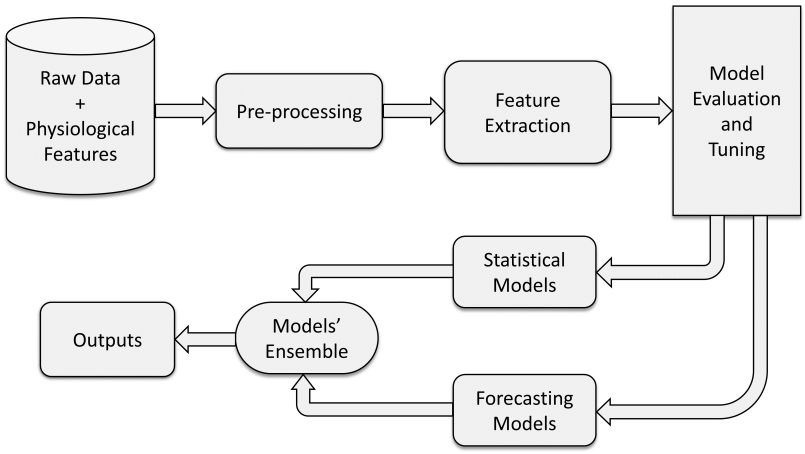


FIGURE 1: The workflow diagram of our proposed methodology

2.1 Data pre-processing

The originality and effectiveness of a methodology largely rest on the manner in which raw data is pre-processed. A superior pre-processing technique contributes significantly to the performance of the models. The dataset employed in this study encompasses training, test sets, and submission files, serving as the fundamental elements for our analytical endeavors. Regarding the training and test datasets, we are furnished with timestamps, and accelerometer sensor data alongside extracted time domain and frequency domain heart rate variability (HRV) features, complemented by essential subject-related information and data collection environment details.

The primary challenge posed by this dataset lies in handling time-series data. The participants' rest periods after task performance introduce periodicity and time-hops within the data, resulting in ambiguous patterns. To address this, we first mitigate the effects of time-hops and re-sample the data at uniform frequency, thereby establishing a uniform temporal basis for further analysis. The back-fill technique was employed to re-sample the dataset uniformly at a frequency of 1Hz. This approach ensures that missing data points are filled with the most recent available values, thereby maintaining data continuity and enabling consistent analysis. To prepare the dataset for training, we discard duplicate timestamps and eliminate any missing data points. This data-cleaning process ensures the integrity of the dataset, providing a clean and reliable foundation for model training and analysis.

2.2 Features extraction

Handling time-series data poses significant challenges mainly due to the inherent limitations of raw data, making it challenging to achieve desirable performance metrics. Time-series data consists of complex temporal dependencies and patterns that require a more comprehensive and sophisticated approach for accurate analysis and prediction. Extracting meaningful insights from raw time-series data demands thoughtful data pre-processing, feature engineering, and the selection of appropriate forecasting models.

Indeed, in this particular challenge, we are provided with extracted physiological features. However, these alone may not suffice for effective forecasting models. Forecasting models necessitate the ability to account for seasonal patterns, which are crucial components in time-series data analysis. These periodic fluctuations in data can significantly impact predictions, and therefore, incorporating timestamp features capturing seasonal patterns becomes imperative to enhance the forecasting model's accuracy and effectiveness. By incorporating these timestamp features, the model gains the capability to adapt and discern patterns across time, leading to more reliable and robust predictions. In our endeavor to address seasonal patterns effectively, we conducted a meticulous extraction of timestamp features. This encompassed incorporating critical time elements such as hour, minute, and second, along with day and the number of days. These intricate temporal attributes provide our forecasting model with the necessary context to discern and adapt to the recurring patterns inherent in time-series data.

The utilization of raw accelerometer data may expose our model to a high variance problem due to the influence of sensor placement on the underlying patterns of the dataset. The location of the sensor can significantly impact the data distribution and characteristics, resulting in varying and unpredictable outcomes. To mitigate this challenge, we adopted a prudent approach by extracting essential statistical features, including the mean, median, standard deviation, and entropy. These statistical measures serve to encapsulate vital information from the data, offering valuable insights into its central tendencies, variability, and complexity. By incorporating these derived features, our model attains enhanced performance, providing more reliable and stable forecasting.

The process of selecting optimal features significantly shapes the model's performance outcome. Among the array of features, certain attributes may lack relevance or have a strong correlation with predicting personal comfort levels. Leveraging the Random Forest Classifier (RFC), we accentuate pivotal features that enhance predictive accuracy. The visual representation of feature significance, conducted using the Gini impurity technique of random forest classifiers (RFCs), is elegantly illustrated in Figure 2.

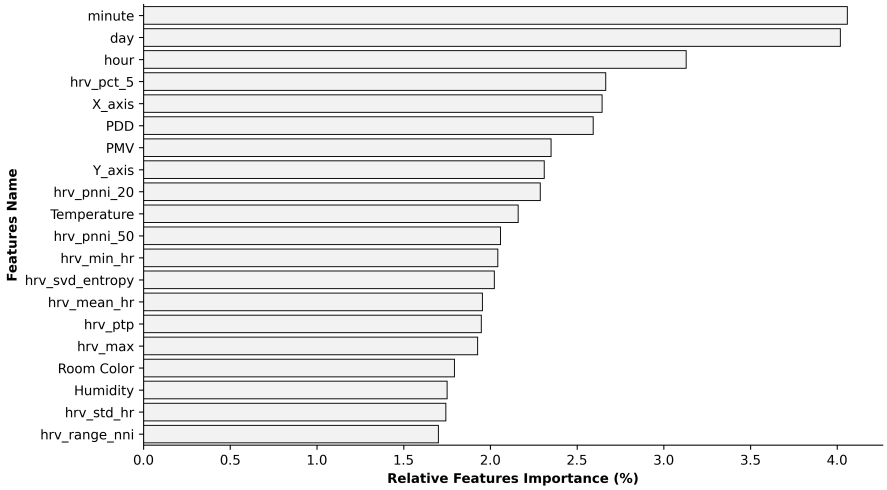


FIGURE 2: Top 20 Relative Feature Importance by RFC

2.3 Model evaluation and training

The essence of achieving accurate forecasting lies in the meticulous process of model selection and training. This critical endeavor involves making astute decisions regarding the most fitting model architecture and fine-tuning its parameters to effectively capture the underlying patterns enmeshed within the data, ultimately yielding reliable predictions for future events.

In our attempt to predict heatstroke risk and personal thermal comfort, we have employed a comprehensive approach, integrating both statistical and forecasting models. The synergistic interplay between these two model categories imparts a holistic perspective to our analysis.

Statistical models (i.e., RFC, SVM, and AdaBoost) assume a pivotal role in our data exploration, leveraging their innate capability to discern intricate correlations between the predicted label and other relevant variables. By delving into the underlying patterns and relationships within the data, these models provide crucial insights into the factors influencing variations in heatstroke risk and personal thermal comfort.

On the other hand, forecasting models(i.e., ARIMA, SARIMA, and Prophet) are adept at discerning and incorporating seasonal patterns that are inherent in time-series data. By accounting for the cyclic nature of heatstroke risk and personal thermal comfort variations, these models generate precise predictions for future trends and fluctuations, effectively capturing the temporal dynamics of the phenomena.

After that, models are trained to check their robustness and performance. In our methodology, we implemented two distinct dataset-splitting approaches for robust model training and evaluation. The first method involved a k folds cross-validation (CV) technique, where 4 folds of the training dataset were allocated for model training, while the remaining fold constituted the validation set. This random split enabled us to assess the models' generalization performance and identify potential

over-fitting.

Additionally, we leveraged the leave-one-subject-out cross-validation technique as our second dataset-splitting approach. This technique ensures that each subject's data is retained as the validation set while the rest of the subjects' data are used for training. This approach is particularly useful when dealing with time-series data from multiple subjects, as it aids in assessing model performance across diverse individuals, mitigating biases, and offering valuable insights into its robustness.

Once the models were trained on the respective datasets, we proceeded to fine-tune the hyper-parameters of each model. Hyper-parameter tuning is a critical step that involves searching for the optimal parameter configurations using a grid search technique that yield superior performance. This process was carried out meticulously to achieve enhanced outcomes and fine-tuned models with superior predictive capabilities.

2.4 Post-processing

After individual model training, we developed a strategic model ensemble approach to harness the collective power of multiple models. This ensemble technique amalgamates the predictions from statistical and forecasting models, each with its unique strengths and capabilities, to form a consolidated and more accurate forecast.

In our work, we implemented a voting classifier as our ensemble model. The classifier aggregates the outputs of the base estimators based on the voting technique. There are two main types of voting: hard voting and soft voting.

In the scenario of hard voting, each base estimator casts a single vote for a class label c. The final prediction (y) is obtained by selecting the class (c) with the most votes:

$$y = \text{argmax}_c \sum_{i=1}^{n} 1 \cdot \{y_i = c\} \tag{1.1}$$

Here, n represents the number of base estimators, and y_i is the prediction of i^{th} base estimator.

On the other hand, the soft voting classifier considers the probability estimates provided by each base estimator. It computes weighted averages of these probabilities for each class and selects the class with the highest average probability as the final prediction:

$$y = \text{argmax}_c \sum_{i=1}^{n} w_i \cdot p_i(c) \tag{1.2}$$

Here, w_i is the weight of the i^{th} base estimators, and $p_i(c)$ is the probability that the i^{th} base estimator predicts class c.

The weights w_i are chosen to reflect the confidence that we have in each base classifier. The weights can be chosen in a variety of ways, such as equal weights, variance-based weights, or error-rate-based weights. For variance-based weights, weights are inversely proportional to the variance of the base estimators. The weights are inversely proportional to the error rate of the base classifier for error-rate-based weights. The weighted voting classifier can be more accurate than the simple voting classifier because it takes into account the confidence that we have in each base estimator. However, it is also more complex to train and tune. For this reason, we

have assigned equal weights for each base estimator.

By combining the outputs of the aforementioned models, the ensemble approach effectively mitigates the limitations of individual models and leverages their complementary aspects. This collaborative effort leads to a more robust and reliable forecasting system, as it minimizes the impact of potential errors or biases inherent in any single model. This strategic integration exemplifies our commitment to maximizing forecasting performance and delivering valuable insights for public health initiatives.

3 Results and Analysis

In this crucial section, we conduct a thorough evaluation of the diverse models employed in our forecasting endeavor. We meticulously analyze their general performance, taking into account their strengths and limitations in predicting personal thermal comfort. To ensure a comprehensive assessment, we employ a wide range of performance metrics, encompassing precision, recall, and F1-score. These metrics enable us to gain valuable insights into the models' overall effectiveness in making precise and reliable forecasts.

To initiate our analysis, we embarked on training the baseline statistical and forecasting models using their default hyper-parameters, employing both random split and leave-one-subject-out cross-validation techniques. The encouraging outcomes from these models are vividly depicted in Table 1.1 and Table 1.2, affirming their promising performance. The successful results attained in this preliminary phase served as a robust basis for our research, paving the way for further optimization and meticulous refinement of the models in subsequent stages. Tahera Hossain reported an

TABLE 1.1: Performance of statistical models with default hyper-parameters.

Model's Name	**Splitting Technique**	**Precision**	**Recall**	**F1-Score**
RFC	5 Folds CV	93.31	90.01	92.15
SVM		90.79	88.23	90.00
AdaBoost		92.47	92.31	91.89
RFC	Leave-one-subject-out CV	92.10	91.51	90.76
SVM		85.21	87.32	83.54
AdaBoost		90.63	90.27	90.31

optimal accuracy of 96.41% with K-Nearest Neighbor classifiers [7]. Analysis of Table 1.1 and Table 1.2 revealed the highest precision, 93.31% from RFC and 95.79% from the SARIMA model, without hyper-parameter fine-tuning. This leaves room for improved performance and further enhancement.

To attain superior performance in our forecasting endeavor, we recognized the significance of fine-tuning the hyper-parameters of each model. This crucial step involved an exhaustive and systematic exploration of a range of hyper-parameter

TABLE 1.2: Performance of forecasting models with default hyper-parameters.

Model's Name	Splitting Technique	Precision	Recall	F1-Score
ARIMA	5 Folds CV	95.71	92.35	94.32
SARIMA		95.79	93.45	93.00
Prophet		93.91	93.25	91.00
ARIMA	Leave-one-out-CV	93.13	92.07	94.76
SARIMA		93.01	92.69	91.35
Prophet		90.64	89.32	88.75

values through a novel grid search cross-validation technique.

During the fine-tuning process, we iteratively assessed various combinations of hyper-parameters for each model, evaluating their impact on forecasting accuracy and precision. The grid search approach enabled thorough exploration of hyper-parameter space for optimal configurations.

Table 1.3 and Table 1.4 present the performance metrics obtained after meticulous fine-tuning of hyper-parameters for both statistical and forecasting models. These tables provide a comprehensive and detailed overview of the models' enhanced predictive capabilities, showcasing the significant improvements achieved through the fine-tuning process.

TABLE 1.3: Performance of statistical models after fine-tuning hyper-parameters.

Model's Name	Splitting Technique	Precision	Recall	F1-Score
RFC	5 Folds CV	98.02	97.67	96.37
SVM		95.45	93.90	92.47
AdaBoost		96.91	96.88	93.40
RFC	Leave-one-subject-out CV	97.45	96.98	96.31
SVM		93.09	92.71	91.39
AdaBoost		95.03	94.58	93.19

TABLE 1.4: Performance of forecasting models after fine-tuning hyper-parameters.

Model's Name	Splitting Technique	Precision	Recall	F1-Score
ARIMA	5 Folds CV	98.27	97.46	98.03
SARIMA		97.46	96.78	97.09
Prophet		95.43	95.15	93.88
ARIMA	Leave-one-out CV	97.51	97.10	96.38
SARIMA		96.66	97.31	95.21
Prophet		94.12	93.81	92.76

In the conclusive phase of our forecasting process, we implemented a pivotal ensemble technique that played a central role in refining and consolidating the model predictions. Through this post-processing approach, we fine-tuned and calibrated

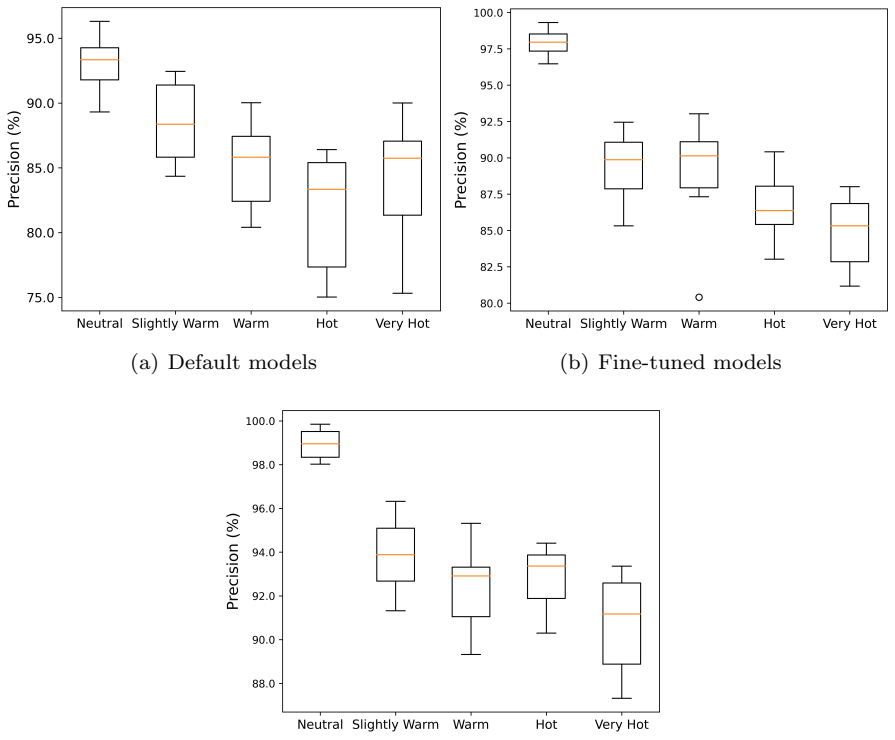
the individual model predictions, ensuring greater alignment with the ground truth data. This ensemble strategy facilitated a more comprehensive and robust analysis, resulting in improved forecasting precision. The ensemble is performed by tuning the voting process (hard voting and soft voting) according to Equation 1.1 and Equation 1.2. The performance of the ensemble classifier is depicted in Table 1.5. By referring to Table 1.5, it becomes evident that the ensemble approach significantly enhances performance by addressing the limitations of individual base estimators. Following the fine-tuning of voting parameters, the soft voting technique outperforms the hard voting method in terms of predictive accuracy and effectiveness. Figure 3 showcases

TABLE 1.5: Performance of ensemble model

Model's Name	**Splitting Technique**	**Precision**	**Recall**	**F1-Score**
Ensemble Classifier	5 Folds CV	99.31	99.07	98.78
	Leave-one-subject-out CV	99.01	98.88	98.57

the class-wise performance of the default hyper-parameters models, fine-tuned models, and ensemble models in a boxplot style. These graphical representations comprehensively overview the models' performance metrics across different classes. Upon analyzing the boxplot 1.3(a), significant variance in the box size is evident, signifying that models without hyper-parameter tuning struggle to predict certain classes accurately. Conversely, as depicted in Boxplot 1.3(b), the proximity of the median and quartiles suggests a higher degree of confidence in the model's predictions for most of the classes, even though performance may be less optimal for certain other classes.

In conclusion, Boxplot 1.3(c) illustrates the ensemble model's notable confidence across the majority of classes, resulting in an exceptional precision of $99.16 \pm 0.15\%$. This performance surpasses that of the models discussed in the papers [5, 6, 7].



(a) Default models

(b) Fine-tuned models

(c) Ensemble model

FIGURE 3: Class-wise performance of different models

4 Conclusion and Future Works

In this research, accurate prediction of personal thermal comfort level necessitates knowledge of the environmental thermal sensation. Therefore, we meticulously forecasted the heatstroke risk label under diverse environmental setups. By leveraging classical and forecasting matching learning models, we successfully predicted personal thermal comfort, ensuring a comprehensive understanding of individuals' thermal well-being in varying conditions.

In our investigation, we conducted a comprehensive evaluation of various models such as RFC, Ada-Boost, SVM, ARIMA, SARIMA, and Prophet. The initial performance of statistical and forecasting models, with default parameters, averaged 90.25% and 93.7%, respectively. However, upon hyper-parameter optimization, these figures improved significantly to 96.19% and 96.69%. It is important to note that exploring an extensive parameter space incurs substantial time complexity, necessitating a trade-off between performance gains and computational efficiency. The hyper-parameter space, best parameters, and the temporal evaluation of the whole procedure are listed

in the Appendix section. Finally, we merged the statistical and forecasting models using the voting classifier, achieving an average precision of 99.16%. This ensemble technique outperformed all the models discussed in [5, 6, 7]. The combined strength of both statistical and forecasting models mitigated the biases of individuals and prevented the model from over-fitting.

Throughout this research, we encountered several challenges that demanded thoughtful consideration and strategic solutions. Foremost, the fluctuating levels of personal thermal comfort provided in the dataset posed a significant obstacle, impacting the performance of our models. Moreover, the presence of numerous duplicate timestamp values required meticulous removal to ensure the integrity of the data. Lastly, the dataset encompassed a diverse range of subjects with varying ages and genders, potentially influencing the patterns of the forecasting labels. Despite these challenges, our research persevered, employing innovative methodologies to address these issues and ultimately paving the way for more accurate and reliable predictions of personal thermal comfort.

In our study, we focused on classical machine learning for heatstroke risk forecasting. However, future research endeavors will explore deep learning frameworks and generative models, such as GANs, to augment data and enhance performance. Additionally, data synthesis through GANs will bolster the dataset's size and diversity, improving model generalization. These innovations will contribute to the continual advancement of heatstroke forecasting, ultimately supporting public health initiatives.

Bibliography

[1] Takashi Akimoto, Shin ichi Tanabe, Takashi Y anai,and Masato Sasaki. Thermal comfort and productivity- evaluation of workplace environment in a task conditioned office. Building and Environment, 45(1):45–50, 2010. International Symposium on the Interaction between Human and Building Environment Special Issue Section.

[2] Li Lan, Zhiwei Lian, and Li Pan. The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings. Applied Ergonomics, 42(1):29–36, 2010.

[3] Li Lan, Pawel Wargocki, and Zhiwei Lian. Quantitative measurement of productivity loss due to thermal discomfort. Energy and Buildings, 43(5):1057–1062, 2011. Tackling building energy consumption challenges - Special Issue of ISHVAC 2009, Nanjing, China.

[4] Tahera Hossain, Kizito Nkurikiyeyezu, Imane El Messaoudi, Kazuki Honda , Christina Garcia, Tahia Tazin, Guillaume Lopez, 2023. 5TH ABC Challenge: Forecasting Thermal Comfort Sensations for Heatstroke Prevention: Leveraging Physiological Data for Better Outcomes. Available at: https://dx.doi.org/10.21227/ys7g-6t64.

[5] Tahera Hossain, Yusuke Kawasaki, Kazuki Honda, Kizito Nkurikiyeyezu, Guillaume Lopez, 'Toward Human Thermal Comfort Sensing: New Dataset and Analysis of Heart Rate Variability (HRV) Under Different Activities', The 4th International Conference on Activity and Behavior Computing (ABC 2022), pp. 27 pages, Taylor & Francis, October, 2022.

[6] Tahera Hossain, Yusuke Kawasaki, Kazuki Honda, Anna Yokokubo, Guillaume Lopez, 'Heat Comfort Prediction Using Pulse Variation', DICOMO2022 Symposium, July, 2022.

[7] Tahera Hossain, Yusuke Kawasaki, Kizito Nkurikiyeyezu, Guillaume Lopez, 'Toward the Prediction of Environmental Thermal Comfort Sensation using Wearables', 18th International Conference on Intelligent Environments (IE2022) (Workshop), pp. 10 pages, June, 2022.

[8] Sulzer, M., Christen, A., & Matzarakis, A. (2023). Predicting indoor air temperature and thermal comfort in occupational settings using weather forecasts, indoor sensors, and artificial neural networks. Building and Environment, 234, 110077. https://doi.org/10.1016/j.buildenv.2023.110077

[9] Zhang, W., Cui, G., Wang, Y. et al. A human comfort prediction method for indoor personnel based on time-series analysis. Build. Simul. 16, 1187–1201 (2023). https://doi.org/10.1007/s12273-023-1010-8

[10] S. S. Kumar, A. Kumar, S. Agarwal, M. Syafrullah and K. Adiyarta, "Forecasting indoor temperature for smart buildings with ARIMA, SARIMAX, and

LSTM: A fusion approach," 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Jakarta, Indonesia, 2022, pp. 186-192, doi: 10.23919/EECSI56542.2022.9946498.

[11] Bourdeau, M., Zhai, X. Q., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustainable Cities and Society, 48, 101533. https://doi.org/10.1016/j.scs.2019.101533

[12] Wu, Z., Li, N., Peng, J., Cui, H., Liu, P., Li, H., & Li, X. (2018). Using an ensemble machine learning methodology-Bagging to predict occupants' thermal comfort in buildings. Energy and Buildings, 173, 117-127. https://doi.org/10.1016/j.enbuild.2018.05.031

[13] Park, H., & Park, D. Y. (2022). Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors. Building and Environment, 207, 108492. https://doi.org/10.1016/j.buildenv.2021.108492

[14] ONS. (2023, August 11). Excess mortality during heat-periods: 1 June to 31 August 2022. Retrieved from https://www.ons.gov.uk/people population and community/births deaths and marriages/deaths/articles/excess mortality during heat periods/england and wales 1 june to 31 august 2022

[15] TakingCare Personal Alarms. (2023, January 20). The Main Causes of UK Heatwave Deaths. Retrieved from https://taking.care/blogs/resources-advice/the-main-causes-of-uk-heatwave-deaths

[16] GOV.UK. (2023, February 14). Heat mortality monitoring report: 2022. Retrieved from https://www.gov.uk/government/publications/heat-mortality-monitoring-reports/heat-mortality-monitoring-report-2022

1 Appendix A

Definition of different performance metrics:

$$Precision = \frac{TP}{TP + FP} \tag{.3}$$

$$Recall/Sensitivity = \frac{TP}{TP + FN} \tag{.4}$$

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{.5}$$

The definition of TP, TN, FP and FN are illustrated in Table .7.

TABLE .6: List of hyper-parameters

Model's Name	Hyper-parameter's Name	Hyper-parameter's Values	Best Parameter
RFC	n_estimators	[50, 100, 500, 1000]	500
	criterion	["gini", "entropy", "log_loss"]	"gini"
	max_depth	[20, 40, None]	40
	max_features	["sqrt", "log2", None]	None
	bootstrap	[True, False]	True
	class_weight	["balanced", "balanced_subsample"]	"balanced"
Ada-Boost	estimator	[None, RFC()]	None
	n_estimators	[10, 20, 50, 100]	100
	learning_rate	[0.1, 0.5, 1, 5, 10]	0.5
	algorithm	["SAMME", "SAMME.R"]	"SAMME.R"
SVM	C	[0.1, 1, 5, 10, 100]	10
	kernel	["linear", "poly", "rbf", "sigmoid"]	"rbf"
	degree	[2, 3, 5, 10]	10
	gamma	["scale", "auto"]	"auto"
	class_weight	["balanced", None]	"balanced"
ARIMA	q (AR Order)	[0, 1, 2, 5, 10]	5
	d (Integration Order)	[0, 1, 2, 5, 10]	1
	q (MA Order)	[0, 1, 2, 5, 10]	10
SARIMA	Q (Seasonal AR Order)	[0, 1, 2, 5, 10]	10
	D (Seasonal Integration Order)	[0, 1, 2, 5, 10]	5
	Q (Seasonal MA Order)	[0, 1, 2, 5, 10]	0
	s (Seasonal Periodicity)	[0, 1]	0
Prophet	changepoint_prior_scale	[0.01, 0.1, 1, 5, 10]	10
	seasonality_prior_scale	[0.01, 0.1, 1, 5, 10]	0.1
	holidays_prior_scale	[0.01, 0.1, 1, 5, 10]	0.01
	seasonality_mode	["additive", "multiplicative"]	"additive"
Ensemble	voting	["hard", "soft"]	"soft"

TABLE .7: Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

TABLE .8: Temporal Evaluation Across Phases of Model Development

Phases	Evaluation Time
Data Pre-processing	$\leq$ 2 minutes (Training Dataset)
	$\leq$ 1 minute (Test Dataset)
Features Extraction	$<$ 30 seconds (Both Training and Test Dataset)
5 Folds-CV Training	Approximately 5 minutes / Classifier; Total 45 minutes.
Leave-one-out-CV Training	Approximately 15 minutes / Classifier; Total 2 hours.
Hyper-parameters Searching	Approximately 3 hours / Classifier; Total 23 hours.
Post-processing (Ensemble)	Approximately 45 minutes.
Prediction and Submission Generation	$<$ 30 seconds (Test Dataset).

TABLE .9: Miscellaneous Information

Name	**Description**
Dataset Used	Heatstroke Challenge Dataset [4]
Model Types	Ensemble of Classical and Forecasting models.
Classification Model	RFC, Ada-Boost, SVC, ARIMA, SARIMA, Prophet, and Ensemble Model
Post-processing	Soft Voting
Performance Metrics	Precision, Recall, and F1-Score
Device Specification	Google Colab (Free Version)
Programming Languages	Python
Library Used	Numpy, Pandas, Matplotlib, Scipy, sklearn, statsmodels and and fbprophet