Core Body Temperature Estimation Using Physiological Signals: A Comparative Study of Kalman Filtering, Extended Kalman Filtering, and Machine Learning Models
CSE 593: Applied Project

Tanmai Mukku
Arizona State University
Supervised by Dr. Asif Salekin
tmukku@asu.edu

ABSTRACT

Accurate core body temperature estimation is critical for the development of personalized health monitoring applications, such as the SmartHeat Tracker app [7], which aims to provide real-time heat stress alerts. Estimating core temperature from physiological signals like heart rate and skin temperature involves addressing complex relationships. In this study, we implemented and evaluated multiple methodologies, including Kalman Filtering, Extended Kalman Filtering (EKF), and Machine Learning (ML) models, such as XGBoost, LSTM, and Deep Neural Networks, for core temperature prediction. Kalman-based approaches, including a windowed variant, were used to model dynamic physiological systems, while ML models were trained to capture non-linear patterns in the data. Comparative analysis showed that while Kalman Filters provide computational efficiency and real-time suitability, advanced ML models demonstrated higher predictive accuracy, albeit at greater computational cost. These results offer a comprehensive foundation for integrating robust and adaptive core temperature estimation techniques into the SmartHeat Tracker app, enabling reliable, personalized health insights.

KEYWORDS

Core Body Temperature Estimation, Kalman Filter, Extended Kalman Filter, Machine Learning, Physiological Signals, Heart Rate, Skin Temperature

1 INTRODUCTION

Core body temperature is a vital physiological parameter that reflects an individual's thermal state, critical for maintaining homeostasis. Accurate measurement of core temperature is essential in applications such as heat stress management, athletic performance monitoring, and healthcare. However, traditional invasive methods of core temperature measurement are impractical for continuous and non-invasive monitoring in real-world settings.

1.1 Problem Statement

The development of non-invasive techniques for estimating core body temperature presents a significant challenge. While wearable devices can measure signals such as heart rate and skin temperature, these are indirect indicators of core temperature. Translating these signals into accurate core temperature estimates involves addressing the inherent complexity and variability in physiological and environmental factors.

1.2 Importance of Core Temperature Estimation

Core body temperature estimation has widespread applications, particularly in health monitoring systems like the SmartHeat Tracker app. As shown in the figure (Figure 1), there can be a significant difference between skin temperature and core temperature, as skin temperature is heavily influenced by ambient conditions. However, accurate core temperature is essential for various applications, such as predicting if someone is on the verge of heat exhaustion. In occupational settings, timely detection of heat stress can prevent conditions like heat stroke. Similarly, athletes can use core temperature insights to optimize performance and recovery. Integrating accurate core temperature estimation into wearable technologies enhances personalized health tracking, promoting safety and well-being.

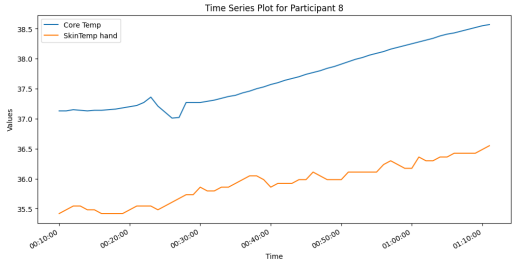

Figure 1: Time series trends for a sample participant across key features.

1.3 Challenges in Non-Invasive Estimation

Non-invasive estimation of core temperature from physiological signals involves several challenges:

- **Non-linearity:** The relationship between heart rate, skin temperature, and core temperature is non-linear and influenced by factors such as hydration and environmental conditions.
- **Noise:** Physiological measurements from wearable devices are often noisy and prone to artifacts, requiring robust filtering techniques.
- **Real-Time Constraints:** For real-world applications, the estimation process must be computationally efficient to provide real-time feedback.

1.4 Objectives of the Study

The primary objective of this study is to evaluate and compare multiple methodologies for core temperature estimation using physiological signals such as heart rate and skin temperature. Specifically, this study aims to:

- Implement Kalman Filtering and Extended Kalman Filtering for dynamic modeling of physiological systems.
- Develop Machine Learning models, including XGBoost, LSTM, and Deep Neural Networks, to capture complex non-linear relationships.
- Compare the performance of these methods in terms of accuracy, robustness, and computational efficiency.
- Lay the groundwork for integrating these techniques into the SmartHeat Tracker app for real-time, personalized health monitoring.

2 RELATED WORK

2.1 Existing Models for Core Temperature Estimation

Core temperature estimation has been extensively studied due to its critical role in health monitoring and safety. Traditional approaches rely on invasive measurements such as rectal and esophageal thermometers, which are accurate but impractical for continuous use. Non-invasive methods aim to estimate core temperature using surrogate signals like heart rate and skin temperature. For instance, Buller et al. [1] introduced a physiological strain index based on heart rate and skin temperature, leveraging regression-based models to predict core temperature in dynamic environments.

2.2 Physiological Relevance of Heart Rate and Skin Temperature

Heart rate and skin temperature are closely linked to thermoregulation. Heart rate reflects cardiovascular responses to thermal stress, while skin temperature provides insights into heat exchange with the environment. However, their relationship with core temperature is modulated by factors such as metabolic rate, hydration, and ambient conditions. Studies like [1] and [5] have demonstrated that these signals, when combined with environmental data, can provide robust core temperature estimates. Despite their relevance, the inherent variability and noise in these signals necessitate advanced modeling techniques.

2.3 Use of Kalman Filtering in Core Temperature Estimation

Kalman filtering has emerged as a prominent method for estimating core temperature in real-time scenarios. Kalman filters dynamically adjust predictions by incorporating prior knowledge and incoming observations, making them effective for noisy physiological signals. Buller et al. [1] applied Kalman filtering to fuse heart rate and skin temperature data, demonstrating its utility in dynamic environments. Extended Kalman Filters (EKF) have also been employed to handle non-linear relationships in physiological systems [6]. These methods are computationally efficient and well-suited for wearable devices, although their accuracy depends on the quality of the observation model.

2.4 Advances with Machine Learning in Physiological Signal Analysis

Recent advances in machine learning have expanded the scope of physiological signal analysis. Deep learning models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) excel at capturing temporal and spatial dependencies in physiological data. Studies like [3] have used LSTMs to predict core temperature from sequential heart rate and skin temperature data, achieving high accuracy. XGBoost, a gradient boosting algorithm, has also been employed for its ability to handle heterogeneous data and model non-linear interactions. While machine learning approaches often outperform traditional models in accuracy, they require substantial computational resources and extensive training data, posing challenges for real-time applications.

This study builds on these prior works by comparing traditional Kalman-based methods with state-of-the-art machine learning models for core temperature estimation, focusing on their integration into wearable health monitoring systems like the SmartHeat Tracker.

3 DATA PREPARATION

3.1 Dataset Details

The study utilized data collected from two trials, each involving 10 participants (n=10), resulting in a total dataset size of 20 unique participants. The trials were designed to capture physiological and environmental data under two distinct scenarios:

- **Trial 1: Outdoor Run** — Participants engaged in a normal outdoor run lasting a minimum of 1 hour, with the option to continue until voluntary termination.
- **Trial 2: Mountain Hike** — Participants participated in a hike on a mountain trail in Tempe, Arizona.

Data was collected using a wearable smartwatch (Garmin vívoactive® 5) [4], including heart rate monitors and skin temperature sensors. The core temperature (ground truth) was measured using the CORE body temperature sensor [2]. Each participant's data was labeled with key contextual features such as activity type, duration, and environmental conditions (ambient temperature, in Trial 2). All data was sampled at a frequency of 1 sample/min.

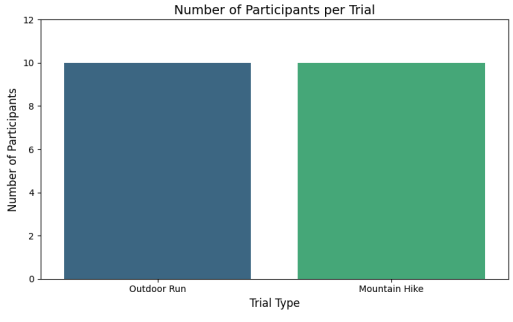

Figure 2: Overview of trial setups: outdoor run and mountain hike in varying environmental conditions.

3.2 Preprocessing

To ensure the quality of the dataset, extensive preprocessing steps were undertaken:

- **Handling Missing Values**: Missing data points were identified and filled using interpolation techniques (linear, forward-fill) based on the nature of the missing data.
- **Normalization**: Features such as heart rate and skin temperature were normalized to improve compatibility with model requirements and reduce the effects of scale differences.

Feature	Description
Heart Rate (HR)	Continuous heart rate measurements (beats per minute)
SkinTemp (Hand)	Skin temperature readings from wearable sensors (°C)
CoreTemp	Ground truth core temperature readings (°C)
Environmental Data	Ambient temperature
Condition	Experimental conditions during trials

Table 1: Key features in the dataset.

3.3 Feature Selection

The key features selected for this study were heart rate (HR), skin temperature (hand), and core temperature as the ground truth. The rationale for choosing these features was based on prior physiological studies and their relevance in core temperature estimation. Features such as environmental temperature and activity codes were also retained to enhance contextual understanding.

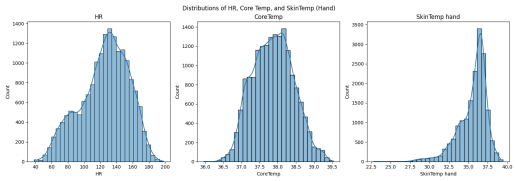

Figure 3: Distribution of heart rate, skin temperature, and core temperature across participants.

3.4 Exploratory Data Analysis

A comprehensive exploratory data analysis (EDA) was conducted to understand the distribution and relationships among the features. The following observations and insights were derived:

- **Heart Rate (HR)**: Heart rate exhibited expected variability during different activities, with higher values during exercise periods and lower values during rest periods.
- **Skin Temperature (Hand)**: Skin temperature showed a consistent correlation with ambient conditions and activity states, with minor delays due to physiological response times.
- **Core Temperature**: Core temperature remained relatively stable but showed gradual increases during extended physical activity, aligning with physiological norms.
- **Environmental Data**: Ambient temperature significantly impacted skin temperature but had limited direct influence on core temperature due to regulatory mechanisms.

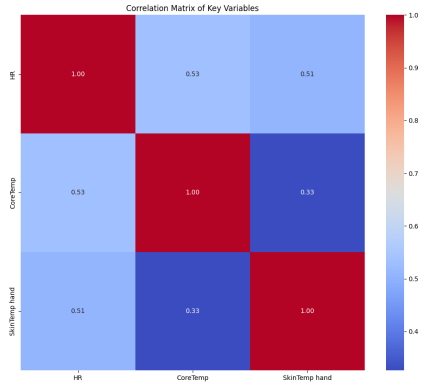

Figure 4: Correlation heatmap of key features: HR, SkinTemp, CoreTemp, and environmental variables.

In addition, statistical summaries revealed the presence of outliers in skin temperature data, particularly during transitions between activities. These outliers were addressed during preprocessing to ensure robust model performance.

Feature	Mean	Std. Dev.	Min	Max
Heart Rate (HR)	125.31	29.47	38.40	198.10
SkinTemp (Hand)	35.56	1.58	22.82	39.45
CoreTemp	37.87	0.56	35.96	39.53

Table 2: Summary statistics of key features.

These insights guided the selection of features and informed the development of models tailored to the observed data characteristics.

4 METHODOLOGY

This section outlines the methodologies applied for estimating core body temperature using various approaches, including Kalman Filter, Extended Kalman Filter, Windowed Kalman Filter, and Machine Learning models.

4.1 Kalman Filter

4.1.1 Theoretical Background. The Kalman Filter is a recursive algorithm designed for estimating the hidden states of a system from noisy observations. It operates in two steps: **Prediction** and **Update**. The prediction step forecasts the state of the system, while the update step refines this estimate based on new observations.

4.1.2 Model Description. The Kalman Filter was employed to estimate the core temperature (T_c) based on heart rate (HR) and hand skin temperature (T_{hand}) observations. The state variable is defined as:

$$\mathbf{X} = [T_c]$$

The observation vector is:

$$\mathbf{Y} = [HR, T_{hand}]$$

The state transition model is assumed to be identity, as core temperature changes are slow:

$$\mathbf{F} = \mathbf{I}$$

The observation model is linear, defined as:

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{v}$$

where $\mathbf{H}$ is the observation matrix and $\mathbf{v}$ represents observation noise.

4.1.3 Observation Matrix Derivation. The observation matrix $\mathbf{H}$ was derived using linear regression between HR, T_{hand}, and T_c from the dataset:

$$\mathbf{H} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

where h_1 and h_2 are coefficients learned from the regression.

The process and observation noise covariance matrices were set as:

$$\mathbf{Q} = \sigma_{T_c}^2 \mathbf{I}, \quad \mathbf{R} = \begin{bmatrix} \sigma_{HR}^2 & 0 \\ 0 & \sigma_{T_{hand}}^2 \end{bmatrix}$$

4.2 Extended Kalman Filter

4.2.1 Rationale for Non-Linear Dynamics. Physiological relationships, such as the dependence of HR on T_c, can be non-linear. To account for these, the Extended Kalman Filter (EKF) was used to model non-linear dynamics.

4.2.2 Adaptation of the Observation Model. The observation model for EKF is represented as:

$$\mathbf{Y} = \mathbf{h}(\mathbf{X}) + \mathbf{v}$$

where $\mathbf{h}(\mathbf{X})$ is a non-linear function:

$$h_1(T_c) = ae^{bT_c}, \quad h_2(T_c) = cT_c + d$$

The Jacobian of $\mathbf{h}(\mathbf{X})$, denoted as $\mathbf{H}$, was used for linearizing the model:

$$\mathbf{H} = \frac{\partial \mathbf{h}(\mathbf{X})}{\partial \mathbf{X}}$$

4.3 Windowed Kalman Filter

4.3.1 Sliding Window Mechanism. The Windowed Kalman Filter extends the Kalman framework by considering a sliding window of n observations for state estimation. This approach incorporates short-term trends into the estimation process.

The state vector is augmented as:

$$\mathbf{X} = [T_c^1, T_c^2, \ldots, T_c^n]^T$$

and the observation vector becomes:

$$\mathbf{Y} = [HR^1, T_{hand}^1, HR^2, T_{hand}^2, \ldots, HR^n, T_{hand}^n]^T$$

4.3.2 Adjustments to Observation Model. The observation model for the windowed approach is:

$$\mathbf{Y} = \mathbf{H}_{win}\mathbf{X} + \mathbf{v}$$

where $\mathbf{H}_{win}$ is the block matrix:

$$\mathbf{H}_{win} = \begin{bmatrix} \mathbf{H} & 0 & \ldots & 0 \\ 0 & \mathbf{H} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{H} \end{bmatrix}$$

4.4 Machine Learning Models

4.4.1 Algorithms Explored. To predict core body temperature (T_c), the following machine learning models were implemented and compared:

(1) **XGBoost**: A gradient-boosted decision tree model capable of handling non-linear relationships and feature interactions effectively.
(2) **LSTM (Long Short-Term Memory)**: A recurrent neural network model designed to capture temporal dependencies in time-series data.
(3) **Deep Neural Network (DNN)**: A fully connected feed-forward neural network for modeling complex relationships between input features and the target variable.

4.4.2 Feature Selection. The feature set was restricted to **heart rate (HR)** and **hand skin temperature**, reflecting realistic constraints where only these inputs are available in real-time through wearable devices. Time was considered an implicit variable for sequential models like LSTM.

4.4.3 XGBoost Model.

- **Input:** HR and SkinTemp hand as features; CoreTemp as the target variable.
- **Data Splitting:**
  - Training and testing datasets excluded the combination **(Participant=8)** to assess generalization.
  - The excluded subset was used for prediction evaluation.
- **Model Training:**
  - Used 100 estimators with a learning rate of 0.1 and a maximum tree depth of 5.
  - Optimized for regression using mean squared error (MSE) as the loss function.
- **Feature Importance:** XGBoost's `model.feature_importances_` confirmed significant contributions from HR and SkinTemp hand.

4.4.4 LSTM Model.

- **Input (LSTM 1):** Sequential windows of size 10 for HR and SkinTemp hand. The model predicted the 11th sample as the target (*CoreTemp*).
- **Input (LSTM 2):** Sequential windows of size 10 for HR and SkinTemp hand. The model predicted the 10th sample as the target using an LSTM.
- **Data Preparation:**
  - Features were normalized to $[0, 1]$ for stability and faster convergence.
  - Overlapping input-output pairs were created for sequential modeling.
- **Model Architecture:**
  - An LSTM layer with 64 units, followed by a dense output layer.
  - The lookback window size was set to 10 timesteps.
- **Training:**
  - The model was trained for 20 epochs with a batch size of 32.
  - Mean squared error was used as the loss function.

4.4.5 *Deep Neural Network (DNN).*

- **Input:** HR and SkinTemp hand as features; CoreTemp as the target variable.
- **Model Architecture:**
  - Two hidden layers with 64 and 32 neurons, respectively, using ReLU activations.
  - Dropout regularization (20%) was applied to prevent overfitting.
  - A final output layer with a linear activation for regression.
- **Training:**
  - The model was trained for 50 epochs with a batch size of 32.
  - Adam optimizer and mean squared error were used for optimization.

Each model's performance was evaluated on the test dataset as well as on the excluded combination **(Participant=8)**. While the results presented focus on this specific excluded combination, the models were assessed using cross-validation, where each participant was excluded in turn for evaluation.

5 EXPERIMENTAL SETUP

5.1 Performance Metrics

To evaluate the predictive performance of the models, the following metrics were used:

(1) **Mean Absolute Error (MAE):** The average of absolute differences between predicted and actual values. It provides a straightforward measure of prediction accuracy in the same units as the target variable.
(2) **Root Mean Squared Error (RMSE):** The square root of the average squared differences between predicted and actual values. RMSE penalizes larger errors more heavily, offering insight into significant deviations.
(3) **Mean Absolute Percentage Error (MAPE):** The mean of absolute percentage errors, providing a relative measure of error as a percentage of the actual values:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

(4) **R^2 (Coefficient of Determination):** Indicates the proportion of variance in the target variable explained by the model, where a value of 1 represents perfect prediction and 0 indicates no predictive capability:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

These metrics were chosen to provide a comprehensive evaluation of model performance, capturing both absolute and relative error magnitudes, as well as the explanatory power of each model.

5.2 Cross-Validation Approach

To ensure robust evaluation and minimize overfitting, the dataset was split as follows:

- **Training and Testing Split:** The dataset was divided into 80% for training and 20% for testing. Samples corresponding to **(Participant=8)** were excluded from both training and testing datasets and used solely for model prediction. While the results presented focus on **Participant 8**, cross-validation was performed by systematically excluding each participant to evaluate the models comprehensively.
- **Temporal Validation:** For sequential models like LSTM, temporal order was preserved during training and testing, ensuring that future data was never used to predict past values.
- **Evaluation Dataset:** Predictions were generated for the excluded dataset **(Participant=8)** to assess model generalization to unseen conditions.

5.3 Computational Considerations

The models were implemented using Python and trained on a system with the following specifications:

- **Processor:** Intel Core i7 with 4 cores.
- **Memory:** 16 GB RAM.
- **Software:** TensorFlow/Keras for neural network models, Scikit-learn for evaluation metrics, and XGBoost for gradient boosting.

Computational requirements were modest, with XGBoost being the most efficient and DNN requiring the longest training time.

6 RESULTS

6.1 Kalman Filter Results

The standard Kalman Filter was applied to estimate core temperature using heart rate (HR) and hand skin temperature (SkinTemp hand) as observations. The weights for these observations were derived using linear regression, allowing dynamic adaptation based on their respective contributions to core temperature estimation. The model achieved the following performance metrics:

- Mean Absolute Error (MAE): 0.17
- Root Mean Squared Error (RMSE): 0.20
- Mean Absolute Percentage Error (MAPE): 0.43%
- R^2 Score: 0.87

6.2 Extended Kalman Filter Results

The Extended Kalman Filter (EKF) incorporated nonlinear relationships between core temperature, HR, and SkinTemp hand using an exponential observation model. Despite its theoretical benefits, the EKF did not outperform the standard Kalman Filter:

- Mean Absolute Error (MAE): 0.44
- Root Mean Squared Error (RMSE): 0.50
- Mean Absolute Percentage Error (MAPE): 1.17%
- R^2 Score: -0.06

6.3 Windowed Kalman Filter Results

The Windowed Kalman Filter (WKF) aggregated multiple recent observations to improve robustness against noise. With a sliding window of 5 observations, the WKF achieved smoother estimates but did not significantly outperform the standard Kalman Filter:

- Mean Absolute Error (MAE): 0.28

- Root Mean Squared Error (RMSE): 0.32
- Mean Absolute Percentage Error (MAPE): 0.61%
- R^2 Score: 0.51

6.4 Machine Learning Models Results

Machine learning models were evaluated for their ability to predict core temperature using HR and SkinTemp hand as features.

XGBoost.

- MAE: 0.33
- RMSE: 0.43
- MAPE: 0.86%
- R^2: 0.42

LSTM. For standard LSTM predicting the next timestep:

- MAE: 0.10
- RMSE: 0.13
- MAPE: 38.27%
- R^2: 0.41

For LSTM predicting the 10th sample using the previous 10 observations:

- MAE: 0.10
- RMSE: 0.12
- MAPE: 39.23%
- R^2: 0.42

Neural Network (DNN).

- MAE: 0.11
- RMSE: 0.14
- MAPE: 24.38%
- R^2: 0.30

6.5 Comparison of All Models

Model	MAE (°C)	RMSE (°C)	MAPE (%)	R^2 Score
Kalman Filter	0.17	0.20	0.43	0.87
Extended Kalman Filter	0.44	0.50	1.17	-0.06
Windowed Kalman Filter	0.28	0.32	0.61	0.51
XGBoost	0.33	0.43	0.86	0.42
LSTM (Next Sample)	0.10	0.13	38.27	0.41
LSTM (10th Sample)	0.10	0.12	39.23	0.42
Neural Network (DNN)	0.11	0.14	24.38	0.30

Table 3: Performance Comparison of All Models for Core Temperature Estimation.

7 DISCUSSION

7.1 Analysis of Results

The results highlight key insights into the performance and behavior of different approaches for core temperature estimation.

The Kalman Filter (Figure 5) achieved strong results with an MAE of 0.17°C, RMSE of 0.20°C, MAPE of 0.43%, and R^2 of 0.87. The high R^2 score indicates the filter's ability to explain a significant proportion of the variance in the ground truth core temperature. This highlights its effectiveness in capturing the overall trends while handling noisy data. From the graph, the Kalman Filter tracks smoother changes well but introduces a slight lag during rapid transitions. This lag suggests the model prioritizes stability over responsiveness, which is acceptable for most physiological monitoring applications where abrupt changes are rare.

The Extended Kalman Filter (Figure 6) underperformed with an MAE of 0.44°C, RMSE of 0.50°C, MAPE of 1.17%, and R^2 of -0.06. The negative R^2 implies that the model performed worse than a simple mean predictor. The graph shows large deviations between the predicted and actual values, particularly during dynamic changes. Despite its potential for modeling nonlinear relationships, the poor performance can be attributed to oversimplified assumptions about the observation model and challenges in accurately approximating the Jacobian. These results highlight the need for better parameter tuning and refined observation models to make EKF competitive.

Machine learning models provided mixed results. XGBoost (Figure 7) achieved moderate performance with an MAE of 0.33°C, RMSE of 0.43°C, MAPE of 0.86%, and R^2 of 0.42. While XGBoost captured some general trends, the graph shows poor alignment during transitions and regions with high variability, indicating limitations in its ability to generalize temporal dependencies.

LSTM models excelled in terms of error metrics, achieving low MAE (0.10°C, 0.10°C) and RMSE (0.13°C, 0.12°C) for both standard LSTM (Figure 8) and the 10th-sample prediction (Figure 9). However, their high MAPE values (38.27% and 39.23%, respectively) suggest issues with scaling relative errors. The graphs reveal that LSTM models capture the general trends effectively but struggle with precise alignment during transitions, particularly in the 10th-sample prediction case. The high R^2 values (0.41 and 0.42) indicate that LSTM models explain a substantial portion of the variance in core temperature but are prone to overfitting due to their reliance on high-quality training data.

The Neural Network (DNN) (Figure 10) achieved comparable performance to LSTM with an MAE of 0.11°C, RMSE of 0.14°C, and MAPE of 24.38%, but with a lower R^2 of 0.30. This indicates that while the DNN captures some temporal patterns, it struggles to explain the variance in the data as effectively as LSTM. The graph reveals that while DNN can track trends to some extent, it is less precise in matching the amplitude of changes compared to other models.

7.2 Strengths and Weaknesses of Each Approach

The Kalman Filter's primary strength lies in its balance between accuracy and computational efficiency. Its ability to handle noisy data and smooth trends makes it highly suitable for integration into real-time applications, particularly those requiring minimal computational overhead, such as wearable devices. However, its linear nature introduces limitations during rapid or nonlinear transitions, as seen in slight prediction lags.

The Extended Kalman Filter, although theoretically capable of modeling nonlinearities, performed poorly due to oversimplified assumptions about the observation model. Its computational complexity and need for precise parameter tuning further reduce its practical applicability.

Machine learning approaches demonstrated the potential to capture complex temporal relationships. LSTM, in particular, excelled at capturing trends with high R^2 scores, making it ideal for applications requiring detailed insights. However, the high computational

cost, sensitivity to training data, and challenges in scaling errors (as indicated by high MAPE) make it less practical for real-time deployment. Similarly, the DNN showed promise but underperformed in explaining variance, likely due to its simpler architecture compared to LSTM.

7.3 Insights on Physiological Modeling

Physiological data, such as core temperature, heart rate, and skin temperature, exhibit complex relationships influenced by various factors, including activity levels, environmental conditions, and individual differences. The success of the Kalman Filter demonstrates that even simple models with well-calibrated weights can provide reliable estimates when the underlying dynamics are relatively smooth. This makes Kalman Filters an attractive choice for wearable applications that require low computational resources.

Machine learning models, while capable of capturing complex nonlinear patterns, emphasize the importance of robust preprocessing and sufficient training data. The high variance explained by LSTM models suggests their potential for integrating diverse physiological signals. However, these models' sensitivity to noise and transitions indicates that hybrid approaches, combining the strengths of Kalman Filters and machine learning, may be the most promising direction for future work.

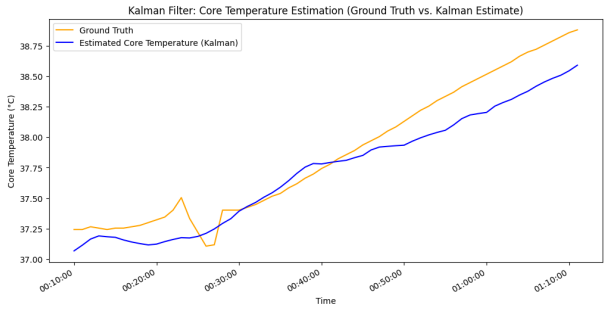

Figure 5: Kalman Filter: Core Temperature Estimation (Ground Truth vs. Predicted).

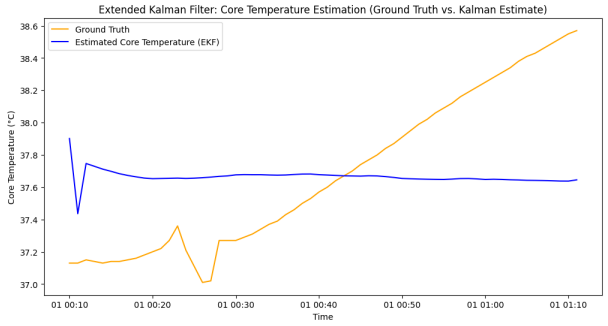

Figure 6: Extended Kalman Filter: Core Temperature Estimation (Ground Truth vs. Predicted).

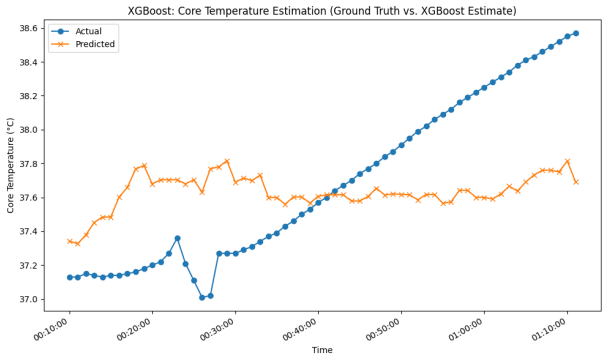

Figure 7: XGBoost: Core Temperature Estimation (Ground Truth vs. Predicted).

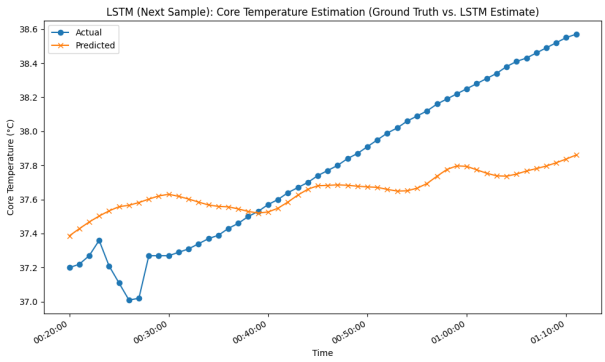

Figure 8: LSTM (Next Sample): Core Temperature Estimation (Ground Truth vs. Predicted).

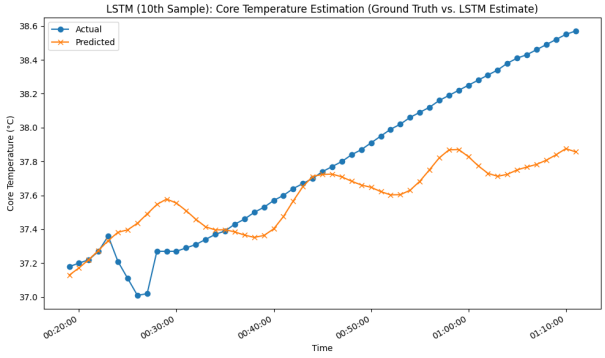

Figure 9: LSTM (10th Sample): Core Temperature Estimation (Ground Truth vs. Predicted).

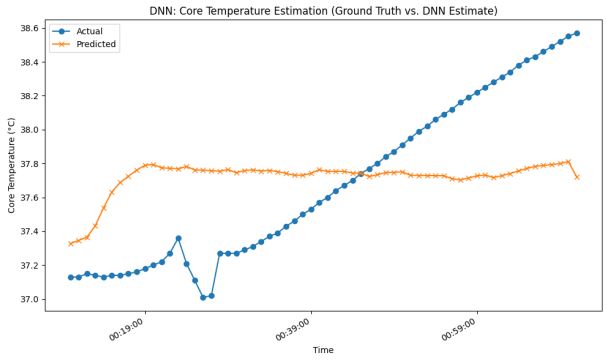

Figure 10: Neural Network (DNN): Core Temperature Estimation (Ground Truth vs. Predicted).

8 CONCLUSION AND FUTURE WORK

8.1 Summary of Findings

This study evaluated multiple methods for estimating core body temperature, including Kalman Filtering, Extended Kalman Filtering, Windowed Kalman Filtering, and machine learning models (XGBoost, LSTM, and DNN). The Kalman Filter demonstrated a compelling balance of accuracy and computational efficiency, achieving low errors (MAE: 0.17°C, RMSE: 0.20°C, MAPE: 0.43%, and R^2: 0.87). In contrast, the Extended Kalman Filter underperformed due to limitations in its observation model. Machine learning models, particularly LSTM, excelled at capturing temporal relationships, but their higher computational requirements and challenges in generalizing across participants limit their real-time applicability.

The analysis highlighted the importance of selecting a model that balances accuracy with ease of integration and computational cost, particularly for applications requiring real-time processing.

8.2 Applications in Real-World Scenarios

The ability to estimate core temperature accurately has significant applications in health monitoring, particularly for heat-related health risks. This is highly relevant for scenarios such as wearable health devices, occupational safety for workers in extreme environments, and athletes undergoing rigorous training. The Kalman Filter's low computational overhead and strong performance make it a promising candidate for integration into mobile health applications like the SmartHeat Tracker app [7]. Its ability to process noisy physiological data with minimal resource requirements aligns well with the constraints of wearable and mobile technologies.

8.3 Limitations and Areas for Improvement

Several limitations were observed in this study. While the Kalman Filter performed well, its reliance on linear relationships between variables may restrict its accuracy in more complex physiological scenarios. Machine learning models, although capable of capturing nonlinearity, require extensive training data, which may not always be feasible in real-world applications. The Extended Kalman Filter and Windowed Kalman Filter showed potential but require further optimization in parameter tuning and dynamic observation modeling.

Additionally, this work primarily evaluated models on retrospective data. Real-time performance, scalability, and robustness across diverse populations and environmental conditions remain critical areas for further validation.

8.4 Future Work: Exploring Hybrid Models

Future work will focus on exploring hybrid models that combine the strengths of Kalman Filters and machine learning approaches. For instance, machine learning models could be used to dynamically adapt the Kalman Filter's observation weights, enabling it to better capture nonlinearities while maintaining computational efficiency. Additionally, investigating Extended Kalman Filtering with more robust observation models could further enhance its performance.

Another avenue for exploration includes time-series models incorporating physiological priors to improve generalization across participants. Testing models on a larger, more diverse dataset and evaluating their performance under real-time constraints will be essential to ensure robustness.

After further exploration and analysis, one of these methods will be selected for integration into the SmartHeat Tracker app, aligning with the app's goal of delivering actionable, real-time health insights to users.

REFERENCES

[1] Mark J Buller, William J Tharion, Samuel N Cheuvront, Scott J Montain, Robert W Kenefick, John Castellani, William A Latzka, Warren S Roberts, Mark Richter, Odest Chadwicke Jenkins, and Reed W Hoyt. 2013. Estimation of human core temperature from sequential heart rate observations. *Physiological Measurement* 34, 7 (jun 2013), 781. https://doi.org/10.1088/0967-3334/34/7/781

[2] CORE. 2024. Core Body Temperature Sensor. https://corebodytemp.com/. Accessed: 2024-12-05.

[3] Patrick Eggenberger, Braid A. MacRae, Shelley Kemp, Michael Bürgisser, René M. Rossi, and Simon Annaheim. 2018. Prediction of Core Body Temperature Based on Skin Temperature, Heat Flux, and Heart Rate Under Different Exercise and Clothing Conditions in the Heat in Young Adult Males. *Frontiers in Physiology* 9 (2018), 1780.

[4] Garmin. 2024. vívoactive® 5. https://www.garmin.com/en-US/p/1057989. Accessed: 2024-12-05.

[5] Ezequiel Juarez Garcia, David Ferguson, and Nicholas Napoli. 2022. Estimating Core Body Temperature Under Extreme Environments Using Kalman Filtering. (01 2022). https://doi.org/10.2514/6.2022-1271

[6] Nicole E. Moyen, Rohit C. Bapat, Beverly Tan, Lindsey A. Hunt, Ollie Jay, and Toby Mündel. 2021. Accuracy of Algorithm to Non-Invasively Predict Core Body Temperature Using the Kenzen Wearable Device. *International Journal of Environmental Research and Public Health* 18, 24 (2021). https://doi.org/10.3390/ijerph182413126

[7] Tanmai Mukku. 2024. Smartheat Tracker App. https://sites.google.com/asu.edu/smartheat-tracker/home?authuser=1. Accessed: 2024-12-05.