https://doi.org/10.65091/icicset.v2i1.6

Calorie Burn Prediction using XGBoost with Feature Selection and SHAP Analysis

Rajat Kumar Khadka
Nepal College of Information Technology
Pokhara University, Nepal
rajat.201323@ncit.edu.np

Ayush Kayastha
Nepal College of Information Technology
Pokhara University, Nepal
ayush.201303@ncit.edu.np

Sezal Baniya
Nepal College of Information Technology
Pokhara University, Nepal
sezal.201328@ncit.edu.np

Suyog Bista
Nepal College of Information Technology
Pokhara University, Nepal
suyog.201335@ncit.edu.np

Ashim Khadka*
Nepal College of Information Technology
Pokhara University, Nepal
ashim.khadka@ncit.edu.np

*Corresponding author

***Abstract*—Calorie burn prediction plays a crucial role in fitness assessment and personalized exercise guidance. This study applies XGBoost regressor with feature selection and SHAP (SHapley Additive exPlanations) analysis to predict calories burned in an exercise session, using physiological and activity-based features for training model. Feature selection was conducted using correlation analysis and Variance Inflation Factor (VIF) screening to reduce redundancy and improve interpretability, resulting in a simplified 5-feature model that balances performance and model complexity. While the full-feature XGBoost regressor achieves the highest predictive accuracy, the proposed five-feature XGBoost regressor demonstrates similar performance with an MAE of 2.19, an MSE of 9.98, and an R^2 of 0.9972, while reducing input dimensionality. Model interpretability is further enhanced through global and local SHAP analysis, revealing the significant influence of heart rate and duration on predictions. These results indicate the potential of gradient boosting models for session calorie burn prediction while suggesting their applicability in fitness tracking systems and personalized exercise guidance system.**

***Index Terms*—Calorie burn prediction, Personalized fitness, XGBoost, VIF multicollinearity, SHAP analysis**

I. Introduction

Exercise is any form of physical activity that helps improve health, fitness, and overall well-being. It includes activities such as walking, running, cycling, swimming, or strength training. Public health organizations, such as the WHO, recommend a certain volume and intensity of exercise that can lower the risk of chronic diseases. Children and adolescents should at least do 60 minutes of moderate to high intensity exercise 3 times a week, and adults should do 150 to 300 minutes of moderate-intensity aerobic physical activity throughout the week, though these recommendations are not being followed and physical activity levels are in constant decline [1], [2].

The growing awareness of the adverse effects of sedentary lifestyles, such as obesity, diabetes, and stress, has been increasingly recognized by individuals, prompting greater attention to health awareness. Over the past decade, a significant rise in exercise trends has been observed in response to the need to maintain well-being. To effectively organize physical activities and meet caloric requirements, the accurate estimation of energy burn is deemed essential. Precise estimation of an individual's calorie burn is considered vital for tailoring personalized fitness programs, managing obesity and chronic diseases, and optimizing athletic performance [1], [2].

Traditionally, predictive equations such as Harris-Benedict and Mifflin-St Jeor helped in accounting for basal energy burn but assumed a fixed and linear relation, thus were not able to address the inter-individual heterogeneity (fitness and body composition) and intra-individual context (temperature and metabolism) [4], [5]. Doubly labeled water (DLW) and indirect calorimetry (IC) are widely regarded as the gold standard for measuring energy burn. DLW is recognized as a highly accurate method for measuring session energy burn, but is limited by cost and complexity for routine use [15]. Likewise, studies highlight IC as the preferred method for estimating resting energy burn, but it is limited in its use due to equipment, training requirements, and patient conditions [6], [16]. Both these methods show high precision, but due to the limitation of routine practical use, attention should be shifted towards sensor-based and ML algorithms to estimate session energy burn.

Wearable devices such as smartwatches utilize sensor data (accelerometer and heart rate), but systematic failure to address activities, population, and device brands necessitates the need for better methods that are more reliable and systematically evaluated in a controlled dataset [7]. A recent study developed a machine learning approach for energy burn using sensor data from smartphones, smartwatches, and data-gathering apps. The Body Mass Index (BMI) is targeted for people whose age is more than 18 years old, with BMI value of > 30, i.e., an obese population. The study demonstrated a root mean square error (RMSE) ranging from 0.28 to 0.32 across various sliding window sizes [17]–[19]. This finding supports integrating BMI into energy burn models for wearable and mobile sensor data, suggesting improved prediction accuracy, particularly for adult obese populations.

Recent studies have supported the need for standardized and validated energy burn estimation through machine learning algorithms that effectively utilize multimodal sensor data [8], [9]. ML models can capture complex, non-linear relationships between physiological signals and energy burn that traditional methods cannot represent. For instance, tree-based ensemble methods have demonstrated remarkable success in modelling the intricate interactions between heart rate, accelerometry data, and individual characteristics such as BMI, age, and gender [10], [17].

Extreme Gradient Boosting (XGBoost) is seen as a reliable algorithm for its superior performance in handling complex, high-dimensional, and non-linear relationships in medical datasets [11]. A study proposed Gradient Boosting Decision Trees on Medical Diagnosis over Tabular Data. LightGBM, Catboost, and XGBoost were compared with traditional and deep neural networks, achieving ROC AUC up to 0.98 across datasets, outperforming TabNet (0.92). This study highlights the gaps of ensemble deep-learning methods and favours gradient boosting decision trees for tabular medical data [14]. Recent work on calorie burn prediction using heart rate and duration supports the use of XGBoost regressor. Study [19] reported an MAE of 2.71, while [20] achieved an MAE of 1.480 with R^2 of 0.998. Similarly study [18] reported MAE 1.48 and R^2 of 1.00, and incorporated explainable AI technique to interpret the model prediction. However, these studies primarily focused on maximizing predictive accuracy and did not account for the impact of feature redundancy on model interpretability. Conversely, the studies cited previously [19], [20] did not attempt hyperparameter tuning or provide explanations for model predictions.

Feature selection was explored in this work primarily to reduce redundancy and improve interpretability, which is tackled through correlation analysis and Variance Inflation Factor (VIF) screening to improve model stability and interpretability. The primary objectives of this work is to construct an XGBoost-based framework incorporating feature selection based on VIF to investigate the tradeoff between model performance, input dimensionality and model interpretability. Furthermore, SHAP (SHapley Additive exPlanations) is employed to provide both global and local explanations of model predictions, enhancing transparency and insight into feature contributions [12].

The key contributions of this paper are:

- To examine the tradeoff between the predictive performance and feature dimensionality by comparing full and feature selected models.
- Model interpretation using Explainable AI (XAI) framework, SHAP.

The rest of the paper is organized as follows: Section II includes the details of the entire process and the model used. Section III discusses the detailed results. Finally, a conclusion is in section IV.

II. Methodology

The proposed model uses machine learning algorithms to predict the session calorie burn of individuals using physiological and activity-based features. The model utilizes different machine learning pipelines, which are discussed in detail.

The dataset utilized for training the predictive model was obtained from Kaggle [3]. The dataset contains 15,000 samples and nine variables. The dataset contained six numerical features (Age, Height, Weight, Duration, Heart_rate, and Body_Temp), 1 categorical feature, gender, and target variable, Calories, as shown in Table I. The categorical feature was label-encoded into a numerical feature for better interpretation of the feature relation with the target variable and model generalization. Table I presents the descriptive statistics of all features, including the mean, standard deviation, and range, providing an overview of the dataset distribution and variability.

TABLE I
Summary statistics of the dataset features used for calorie prediction.

Feature	Description	Type	Mean ± SD	Range
Gender	Binary indicator: Male(0) or Female(1)	Categorical		0 and 1
Age	Age of the individual in years (year)	Continuous	42.78 ± 16.98	20–79
Height	Height of the individual in centimeters (cm)	Continuous	174.46 ± 14.25	123–222
Weight	Weight of the individual in kilogram (kg)	Continuous	74.96 ± 15.03	36–132
Duration	Time spent exercising in minutes (min)	Continuous	15.53 ± 8.32	1–30
Heart_rate	Average heart rate during exercise session (bpm)	Continuous	95.51 ± 9.58	67–128
Body_Temp	Average body temperature during exercise session(Celsius)	Continuous	40.02 ± 0.77	37.1–41.5
Calories	Amount of calories burnt during exercise session	Continuous	89.53 ± 62.45	1–314
BMI (derived)	Calculated as weight (kg) / height (m²)	Continuous	24.34 ± 1.55	19.22–29.06

A. Feature Selection

A correlation heatmap illustrates the pairwise relationships among the features, as shown in Fig. 1. Strong correlation was observed between height and weight as well as between exercise duration, heart rate and body temperature. These relationships indicate the presence of feature redundancy within the dataset.

To gain a deeper understanding of these relationships multicollinearity was analyzed using VIF and the results are presented in Table II. The analysis reveals that height and weight exhibit VIF values exceeding 10, signifying strong multicollinearity. To reduce the feature redundancy while

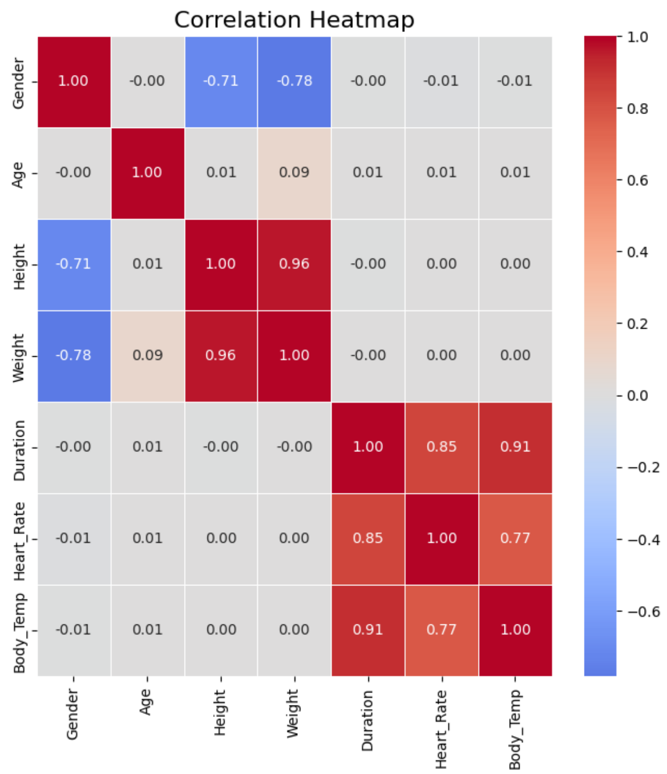

Fig. 1. Correlation heatmap of features used to predict calorie burn.

preserving relevant information, these features were combined into Body Mass Index (BMI). Duration and body temperature were also found to be strongly collinear with VIF values surpassing 5. Thus, exercise duration was retained as it directly indicates the workload while body temperature was excluded. Heart rate showed moderate collinearity, with a VIF value just below 5, so it was retained without transformation. Therefore, the final feature set includes gender, age, duration, heart rate and BMI.

TABLE II
VIF TABLE TO DETECT MULTICOLLINEARITY BETWEEN DIFFERENT FEATURES.

Feature	**VIF**
Gender	2.81
Age	1.12
Height	14.28
Weight	18.65
Duration	8.48
Heart_rate	3.42
Body_Temp	6.05

Fig. 2 presents a box plot of the features, revealing the presence of outliers in the heart rate and BMI features. The Interquartile Range (IQR), a widely recognized statistical method for outlier detection [13], was employed to identify these outliers within the dataset. Session data exceeding 1.5 times the IQR from the first quartile (Q1) and third quartile (Q3) were excluded from the dataset. Given the sufficiently large initial dataset comprising 15,000 observations. This process resulted in a reduced dataset containing 14,316 session data.

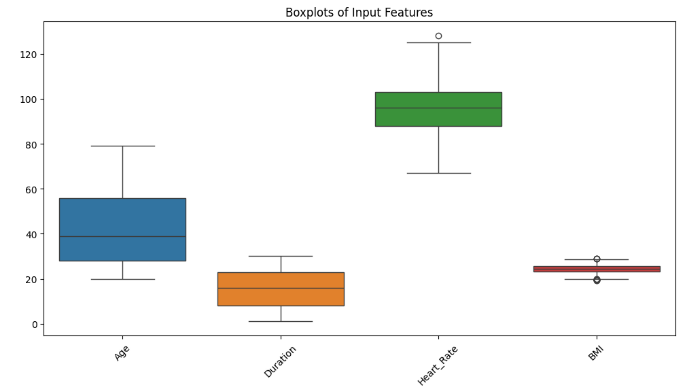

Fig. 2. Box plot of 5 selected features to detect outliers.

Standard normalization was applied to all selected continuous features to achieve a mean of 0 and standard deviation of 1, ensuring uniform feature scaling for the XGBoost regressor. The 14,316 data points were divided into training and testing sets at an 80:20 ratio, resulting in a training set comprising 11,452 data points and a testing set containing 2,864 data points.

B. Regressor Model

The primary aim of this paper is to develop an XGBoost regressor, which is recognized for its superior performance in managing complex, high-dimensional, and non-linear relationships within medical datasets [14]. For the purpose of this study, the XGBoost Regressor was evaluated across various feature sets, including a configuration with all seven original features without feature selection, a reduced set of five features comprising gender, age, duration, heart rate, and BMI, and a further reduced set of four features including gender, age, duration, and BMI.

The hyperparameters of the XGBoost regressor were optimized using GridSearchCV with 5-fold cross-validation, with the results summarized in Table III. The model selection metric used was Mean Squared Error(MSE) and the final model was retrained on training dataset using the best hyperparameters.

The model evaluation was performed on the test dataset using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2.

III. Result and Discussion

The XGBoost regressor was selected as the appropriate model for training on this dataset. The XGBoost regressor was assessed across three distinct experiments, each utilizing a different number of features: initially with all seven original features, followed by a reduced set of five features, and finally with four features.

The performance of the model was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2

TABLE III
SUMMARY OF HYPERPARAMETER TUNING OF XGBOOST REGRESSOR USING 5-FOLD CROSS VALIDATION (CV = 5).

Parameters	**Range**	**Best**
n estimators	range(0, 100, 5)	95
learning rate	[0.01, 0.1, 0.2]	0.1
max depth	range(2, 20, 2)	6
subsample	[0.8, 1.0]	0.8
colsample_bytree	[0.8, 1.0]	0.8
features	[7, 5, 4]	5

on test data. An additional parameter CV MSE reflects the variability in models performance in different folds (K=5) of training data during hyperparameter tuning. Table IV presents a comparison of the XGBoost regressor's performance across three different feature configurations: 7 features, 5 features, and 4 features. The model achieved the best performance while including all 7 original feature with MAE of 1.15, MSE of 3.25, and R^2 of 0.9990, indicating excellent predictive accuracy and variance capture. The corresponding CV MSE of 3.5541 ± 0.2825 on training data for hyperparameter tuning, suggests stable performance across folds. The model with five features (Gender, Age, Duration, Heart Rate, and BMI) shows a moderate decline in performance, with an MAE of 2.19, MSE of 9.98, and R^2 of 0.9972. The CV MSE of 11.0388 ± 0.2080 remains relatively low, indicating consistent results during training, though less accurate than the 7-feature model. The 4-feature model that excludes heart rate exhibits the poorest performance, with an MAE of 8.65, MSE of 148.86, and R^2 of 0.9587, reflecting a significant drop in accuracy. The CV variability of 149.7932 ± 3.0804 is notably higher, suggesting greater instability across training folds.

The results highlight a trade-off between performance and feature count, with the 7-feature model offering the highest accuracy, while the 5-feature model provides a balanced compromise with near-comparable predictive performance and reduced input dimensionality. The 4-feature model, while still capturing substantial variance ($R^2 = 0.9587$), underperforms, likely due to the loss of a critical predictor like Heart Rate. Consequently, the 5-feature model was selected as a practical and efficient configuration balancing performance and model complexity.

TABLE IV
COMPARISON OF PERFORMANCE PARAMETERS OF THE XGBOOST REGRESSOR WITH DIFFERENT FEATURE CONFIGURATIONS.

No. of features	**MAE**	**MSE**	R^2	**CV MSE**
7	1.15	3.25	0.9990	3.5541 ± 0.2825
5 (proposed)	2.19	9.98	0.9972	11.0388 ± 0.2080
4	8.65	148.86	0.9587	149.7932 ± 3.0804

Once the optimal number of features was determined, various models were assessed on the dataset, including the proposed XGBoost regressor, the Random Forest model, and Lasso regression with a polynomial degree of 3. As presented in Table V, the proposed XGBoost regressor emerged as the best-performing model, achieving an MAE of 2.19, an MSE of 9.98, and an R^2 of 0.9972. The Random Forest model followed with a comparable performance, recording an MAE of 2.56, an MSE of 13.93, and an R^2 of 0.9961, closely aligning with the XGBoost regressor results. Lastly, Lasso regression exhibited the lowest performance among the three, with an MAE of 3.10, an MSE of 18.99, and an R^2 of 0.9947.

TABLE V
COMPARISON OF PERFORMANCE PARAMETERS OF DIFFERENT MODELS

Model	**MAE**	**MSE**	R^2
Random Forest	2.56	13.93	0.9961
Proposed model	2.19	9.98	0.9972
Lasso Regression(polynomial degree = 3)	3.10	18.99	0.9947

Fig. 3 illustrates the actual versus predicted calorie burn values obtained using the proposed XGBoost regressor with selected 5-feature model.The close alignment of the predicted values with the identity line (slope 1 and intercept 0) indicates strong agreement between the predicted and the ground-truth values. This behavior is consistent with the quantitative evaluation on the test set, where the model achieved an MAE of 2.19, an MSE of 9.98, and an R^2 of 0.9972. The linear relationship further confirms the model's ability to accurately capture the underlying relationship between the selected physiological feature and session calorie burn, demonstrating robust predictive performance.

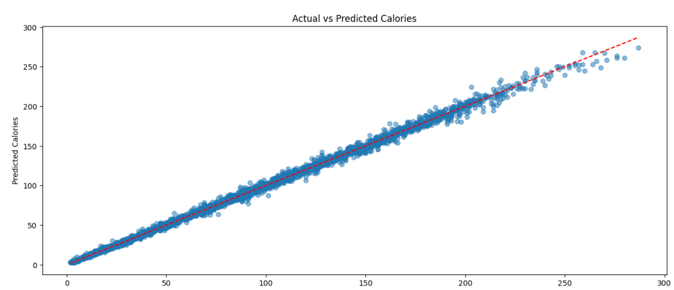

Fig. 3. Actual versus predicted values of XGBoost regressor with 5 Selected Features. The red dashed line represents the identity line.

The residual curve presented in Fig. 4 demonstrates that the majority of the data points are densely clustered within the ± 5 range, suggesting a high level of accuracy and minimal bias in the predictive model. This tight clustering around zero indicates that the model's predictions are closely aligned with the actual values, reflecting its effectiveness in capturing the underlying patterns of the dataset. A small number of data points are observed to extend beyond this dense region, sparsely distributed within the ± 10 range, which may represent outliers or instances where the model encounters greater variability. This distribution underscores the model's overall robustness, with the limited spread of residuals further supporting its reliability in estimating calorie burn, consistent with the high R^2 of 0.9972 and low MAE of 2.19 achieved with the 5-feature XGBoost regressor.

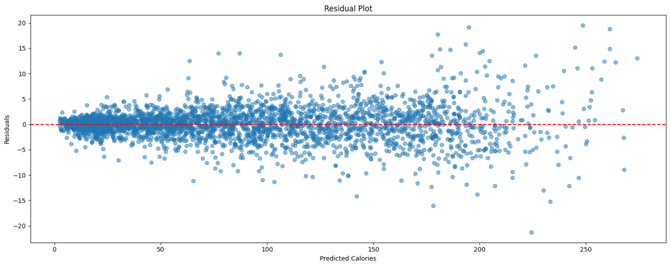

Fig. 4. Residual graph of the proposed XGBoost regressor based on 5 Selected Features

In Fig. 5, the learning curve for XGBoost regressor is presented, with RMSE plotted against the number of boosting rounds. Rapid convergence was observed within the first 20 rounds, during which the RMSE was reduced from approximately 55 to 10, indicating that key patterns were effectively captured in the initial stages of training. A minimal gap between training and validation RMSE was maintained throughout, indicating strong generalizability and that overfitting was avoided. A balanced bias-variance tradeoff was achieved, confirming that the selected hyperparameters and boosting strategy contributed to model stability and reliable performance.

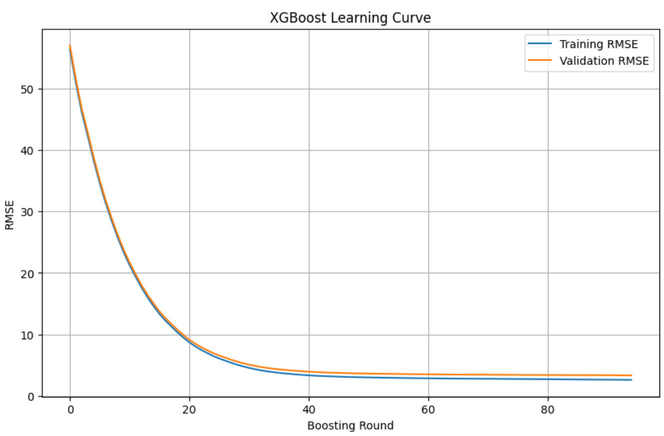

Fig. 5. Learning curve boosting round of the proposed XGBoost regressor.

The SHAP analysis was conducted to quantify and visualize each feature's contribution to the XGBoost regressor's predictions in an interpretable, global context in Fig. 6 and instance-level in Fig. 7. The color gradient further illustrates how high (red) versus low (blue) feature values correspond to positive and negative impacts, respectively. Fig. 6 presents the SHAP summary plot, which ranks features by their mean absolute SHAP values to reveal overall importance.

Based on the analysis, duration and heart rate are the most influential features. Physiologically, this behavior is expected, as calorie burn during physical activity is primarily governed by exercise duration (total work performed) and heart rate, which serves as a proxy for exercise intensity and oxygen consumption [21], [22]. Higher values of both features consistently push predictions toward increased calorie burn, indicating a strong positive association.

Age also demonstrates a positive SHAP trend, with a narrower contribution range, suggesting higher age is associated with increased predicted calorie burn within the studied population. This effect may reflect age-related differences in cardiovascular response and metabolic cost under comparable activity conditions, as reported in prior physiological studies [22], [23]. In contrast, BMI and gender exhibit SHAP values tightly clustered around zero, indicating a limited role in explaining prediction variability. Although BMI is known to influence absolute calorie burn due to increased body mass [23], its near zero SHAP importance can be attributed to the limited variability of BMI in the dataset (mean 24.34 $\pm$ 1.55), which predominantly represents a slightly overweight population. Similar observations have been reported in recent explainable AI studies on calorie burn, where activity-based features dominate model explanations in relatively homogeneous populations [24], [25].

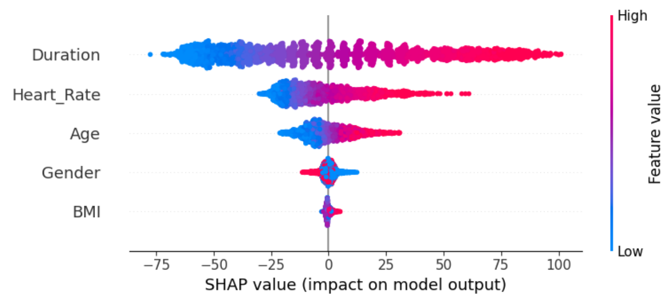

Fig. 6. Global SHAP beeswarm plot showing feature importance for calorie burn estimation using XGBoost regressor

Fig. 7 illustrates the contribution of features on the prediction of calorie burn on the individual data point. The model's base value was $E[f(x)] = 93.639$, and the final prediction of 140.386 was obtained by summing the individual SHAP contributions: duration of +51.04, heart rate of +3.76, and gender of +0.61 increased the prediction, whereas age (–7.68) and BMI (–0.97) decreased it.

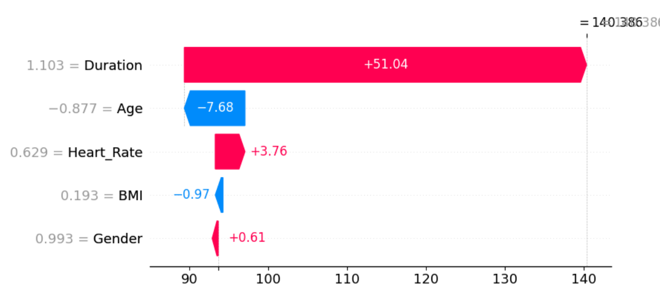

Fig. 7. Local SHAP waterfall plot illustrating feature contribution for individual prediction.

IV. Conclusion

This study introduces a machine learning framework for session calorie burn prediction using an XGBoost regressor with an emphasis on model simplicity and interpretability while maintaining competitive prediction performance. Feature selection was performed using correlation matrix and VIF screening, resulting in the combination of height and weight into BMI and the exclusion of body temperature.

Experimental results demonstrate that the 7-feature model achieved the highest predictive accuracy, while the feature selected 5-feature model attained comparable performance with an MSE of 9.98, an MAE of 2.19, and R^2 of 0.9972, despite utilizing fewer and less redundant features. Comparative analysis of proposed XGBoost regressor with Random Forest and Lasso regression confirmed the effectiveness of gradient boosting models on non-linear and tabular data.

Model interpretability was enhanced through SHAP, facilitating both global and local interpretations. Globally, the analysis highlighted activity-based data, such as heart rate and duration, as significant predictors of calorie burn, exerting a stronger influence compared to physiological data like age, gender, and BMI, which demonstrated a marginal impact. Locally, individual predictions were broken down into feature-level contributions, enabling a transparent assessment of each feature's role in the prediction process.

Overall the findings highlight a practical trade-off between predictive performance and model simplicity, where a reduced feature set offers improved interpretability. For future work, the model can be evaluated on the diverse population incorporating proper BMI range. Additionally, integration of edge computing is proposed to enable seamless deployment of the model on resource-constrained devices such as smartphones, smartwatches, and IoT sensors. This necessitates further optimization of the model to operate efficiently under limited memory and processing power, paving the way for practical, on-device calorie burn estimation.

References

[1] World Health Organization, ”Physical activity,” June 2024.

[2] World Health Organization Exercise Guidelines, ”WHO guidelines on physical activity and sedentary behaviour,” November 2020.

[3] fmendes, ”fmendes-DAT263x-demos,” June 2020.

[4] Bendavid *et al.*, “The centenary of the Harris–Benedict equations: How to assess energy requirements best? Recommendations from the ESPEN expert group,” *Clin. Nutr.*, vol. 40, no. 3, pp. 690–701, Mar. 2021.

[5] M. D. Mifflin, S. T. St Jeor, L. A. Hill, B. J. Scott, S. A. Daugherty, and Y. O. Koh, “A new predictive equation for resting energy burn in healthy individuals,” *Am. J. Clin. Nutr.*, vol. 51, no. 2, pp. 241–247, Feb. 1990.

[6] E. De Waele, J. Jonckheer, and P. Wischmeyer, “Indirect calorimetry in critical illness: A new standard of care?,” *Curr. Opin. Crit. Care*, vol. 27, no. 4, pp. 334–343, 2021.

[7] Shenglong *et al.*, “Validity of three smartwatches in estimating energy burn during outdoor walking and running,” *Front. Physiol.*, vol. 13, p. 995575, 2022.

[8] D. Chen *et al.*, “Development and validation of a smartwatch algorithm for differentiating physical activity intensity in health monitoring,” *Sci. Rep.*, vol. 14, p. 9530, 2024.

[9] C. Lendt, N. Hansen, I. Froböse, and T. Stewart, “Composite activity type and stride-specific energy burn estimation model for thigh-worn accelerometry,” *Int. J. Behav. Nutr. Phys. Act.*, vol. 21, art. no. 99, 2024.

[10] Li *et al.*, “Estimating human metabolic rate at various activities using tri-axial accelerometers,” *Build. Environ.*, vol. 277, p. 112944, 2025.

[11] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[12] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” *arXiv preprint arXiv:1705.07874*, May 2017.

[13] R. J. Hyndman, ”Outlier detection methods in statistical data analysis,” *J. Stat. Comput. Simul.*, vol. 77, no. 8, pp. 689–704, 2007.

[14] A. Y. Yıldız and A. Kalaycı, ”Gradient boosting decision trees on medical diagnosis over tabular data,” *arXiv*, 2024.

[15] K. R. Westerterp, “Doubly labelled water assessment of energy burn: principle, practice, and promise,” *Eur. J. Appl. Physiol.*, vol. 117, no. 7, pp. 1277–1285, Jul. 2017.

[16] E. De Waele and A. R. H. van Zanten, “Routine use of indirect calorimetry in critically ill patients: pros and cons,” *Crit. Care*, vol. 26, p. 123, May 2022.

[17] B. Wei, C. Romano, M. Pedram, B. Nolan, W. A. Morelli, and N. Alshurafa, ”Developing and comparing a new BMI inclusive energy burn algorithm on wrist-worn wearables,” *Sci. Rep.*, vol. 15, no. 20060, 2025.

[18] M. N. H. Likhon, F. Bhuiyan, M. S. Bhuiyan, M. A. Fahim, and A. Hossain, ”Calories Burnt Prediction: A Machine Learning Approach,” *TechRxiv*, Aug. 26, 2024.

[19] S. P. Vinoy and B. Joseph, ”Calorie Burn Prediction Analysis Using XGBoost Regressor and Linear Regression Algorithms,” *Proc. NCECA*, vol. 4, no. 1, pp. 187–191, 2022.

[20] M. T. Aziz, S. R., R. D. C. Pecho, N. U. A. Khan, A. U. H. Era, and M. D. A. Chowdhury, ”Calories Burned Prediction Using Machine Learning,” *ResearchGate*, 2023.

[21] B. E. Ainsworth *et al.*, “2011 Compendium of Physical Activities: A second update of codes and MET values,” *Med. Sci. Sports Exerc.*, vol. 43, no. 8, pp. 1575–1581, 2011.

[22] L. R. Keytel, J. H. Goedecke, T. D. Noakes, R. Hiiloskorpi, L. Laukkanen, L. van der Merwe, and E. V. Lambert, “Prediction of energy expenditure from heart rate monitoring during submaximal exercise,” *J. Sports Sci.*, vol. 23, no. 3, pp. 289–297, 2005.

[23] N. M. Byrne, A. P. Hills, G. R. Hunter, R. L. Weinsier, and Y. Schutz, “Metabolic equivalent: One size does not fit all,” *J. Appl. Physiol.*, vol. 99, no. 3, pp. 1112–1119, 2005.

[24] Y. Song *et al.*, ”Interpretable deep learning for personalized energy expenditure prediction using ECG and acceleration signals in incremental exercise,” *Sci. Rep.*, vol. 15, Art. no. 36277, 2025.

[25] Z. Guo *et al.*, ”Energy consumption analysis and prediction in exercise training based on accelerometer sensors and deep learning,” *Sci. Rep.*, vol. 15, Art. no. 19423, 2025.