A Machine Learning Approach to Predicting Calories Burned During Exercise

Umme Habiba Tamanna
Department of Computer Science and Engineering
Bangladesh University
Dhaka, Bangladesh
tamanna.active24@gmail.com

Jannatul Ferdousi Babli
Department of Computer Science and Engineering
Bangladesh University
Dhaka, Bangladesh
ferdousj2077@gmail.com

Sanjida Tabassum Toma
Department of Computer Science and Engineering
Bangladesh University
Dhaka, Bangladesh
sanjidatabssum0166@gmail.com

Muhammad Minoar Hossain
Department of Computer Science and Engineering
Bangladesh University
Dhaka, Bangladesh
minoar.hossain@bu.edu.bd

***Abstract*—A basic indicator of the amount of energy used and calories burned during exercise is crucial for tracking fitness and medical conditions. Professionals can create individualized fitness plans and help people maximize their workouts with accurate calorie estimation. Based on physiological and exercise-related factors, such as gender, age, height, weight, duration of exercise, heart rate, and body temperature, this study proposes a machine learning (ML) method for calorie burn prediction. Utilizing a variety of regression ensemble machine learning models, including Gradient Boosting Decision Trees Regression (GBDTR), Extreme Gradient Boosting Regression (XGBOOSTR), Stacked Generalization Regression (STACKINGR), Random Forest Regression (RFR), Bootstrap Aggregating Regression (BAGGINGR), and Ensemble Voting Regression (VOTINGR). With average metrics Mean Squared Error (MSE) of 14.224, Mean Absolute Error (MAE) of 2.022, R-squared (R^2) of 0.9964, Peak Signal-to-Noise Ratio (PSNR) of 37.41 dB, and Signal-to-Noise Ratio (SNR) of 29.29 dB, the XGBOOSTR model stands out as the best performer in this study's cross-validation approach. Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) in XAI are used to further interpret the research's findings, emphasizing the significance of comprehending feature contributions like duration, heart rate, and body temperature. The planned study offers insightful information to support wearable health applications, improve personalized fitness tracking, and help medical professionals encourage healthier lifestyles.**

***Index Terms*—Calories burned, Exercise monitoring, Machine Learning, Regression, Random Forest, Gradient Boosting**

I. Introduction

Maintaining fitness, avoiding chronic illnesses, and enhancing general well-being all depend on physical activity. The number of calories burned is a crucial measure of how effective exercise is, and it varies depending on a number of variables, including age, gender, body composition, heart rate, body temperature, and duration. In addition to supporting individual fitness objectives, accurate calorie expenditure estimation helps doctors prescribe customized exercise regimens.

Exercise intensity and duration have the biggest effects on calories burned. According to studies, a 30-minute session of moderate-to-intense exercise can burn 150–400 kcal [1]. However, accurate estimation is difficult due to physiological differences; overestimation can lead to nutritional imbalance, while underestimation can hamper fitness progress. This highlights the necessity of trustworthy, automated prediction techniques.

A potent tool for evaluating health data and forecasting energy use is machine learning, or ML. To increase the accuracy of calorie estimation, ML techniques have been used in a number of studies to predict energy expenditure and track exercise. For instance, ML techniques have been used in a number of studies to predict energy expenditure and track exercise. For instance, Panwar et al. [9] used regression models including SVR and XGBoost to estimate calorie burn from physiological and activity features. Alfred et al. [10] applied Random Forest models to improve prediction accuracy using demographic and biometric inputs. In a similar vein, Basavaraj et al. [11] employed ensemble learning methods (e.g. XGBoost, Random Forest) and outperformed conventional statistical techniques in calorie burn prediction.

Our system uses a dataset of 1500 exercise records with various physiological and activity-related features to develop a method that is inspired by the need for precise and customized calorie burn estimation. We use a number of regression ensemble machine learning models, such as RFR, XGBOOSTR, STACKINGR, and GBDTR, and we employ robust preprocessing techniques to guarantee data quality. Performance is assessed using cross-validation techniques, and Explainable

Artificial Intelligence (XAI) techniques like Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are used to further interpret the best-performing model. The main contributions are as follows.

- A comprehensive evaluation of ensemble regression models for calorie burn estimation.
- Exploration of model interpretability through XAI to assess the impact of key features (e.g., body temperature, heart rate, duration).

The remainder of this paper is organized as follows: Section II presents the methodology, Section III reports the results and analysis, and Section IV concludes with key findings and implications.

II. Methodology

This study's main goal is to use advanced ML techniques to predict exercise-induced calorie burn more accurately. While Sections A through G provide specific insights into the study, Figure 1 shows the general architecture of the suggested system.

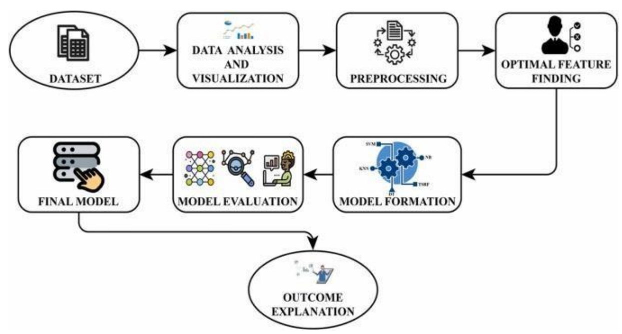

Fig. 1. Architecture of the Research

A. Dataset

A dataset [8] with 1500 entries is used in this study. Eight predictive features and one target variable make up the dataset. Demographic and physiological characteristics like gender, age, height, weight, length of exercise, heart rate, and body temperature are among the features. These factors serve as the foundation for forecasting how many calories will be burned while exercising. The dataset is explained in detail in Table I.

B. Data Analysis and Visualization

The dataset was analyzed and preprocessed in this study using a variety of visualization techniques, including heatmaps, box plots, density plots, and histograms. These techniques are essential for looking at the underlying feature distributions, finding hidden patterns, and spotting possible outliers. The quality and interpretability of the dataset are enhanced by such visual analyses, which is crucial for the development of predictive models and feature engineering later on.

TABLE I
Descriptive Features of the Dataset

Features	Description	Variable Type	Unit
User ID	Participant ID number	Numerical	–
Gender	Participant gender[a]	Nominal	–
Age	Participant age	Numerical	Years
Height	Participant height	Numerical	cm
Weight	Participant weight	Numerical	kg
Duration	Exercise duration	Numerical	Minutes
Heart Rate	Avg. heart rate	Numerical	bpm
Body Temp	Body temperature	Numerical	$^{\circ}$C
Calories	Calories burned[b]	Numerical	kcal

[a]0 = Male, 1 = Female.
[b]Target variable.

Histograms were generated for each feature, as shown in Figure 2. These histograms display the frequency distribution of values across intervals, providing insights into how the data is spread. For example, the histogram of Gender reveals a clear binary distribution (0 = Male, 1 = Female). The Age feature shows a right-skewed distribution, with most participants clustered between 20 and 50 years. Similarly, Height and Weight demonstrate near-normal distributions, whereas Duration of exercise and Calories burned exhibit right-skewed patterns with a concentration of values in the lower ranges.

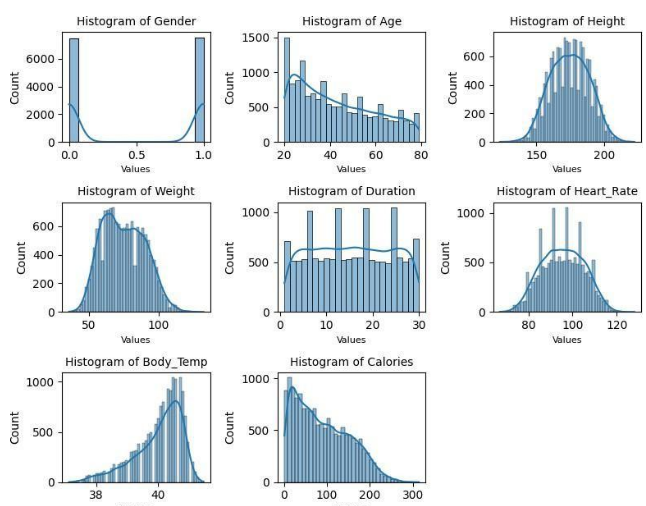

Fig. 2. Histogram illustrating the distribution of different features

A smooth estimate of the probability density function is provided by a density plot, which is a graphical depiction of a probability density function [6]. The density plots, which depict the distribution patterns of features important for calorie burn prediction, are displayed in Figure 3.Age, body temperature, and calories all have right-skewed distributions, with the majority of values concentrated in the lower ranges, as this graphic illustrates. The distributions of height, weight, and heart rate are almost normal, suggesting that participant variability is balanced. The distribution of duration is fairly uniform, indicating that workouts are spaced out over short to long periods of time. The binary nature of the Gender feature is confirmed by the two distinct peaks it displays.

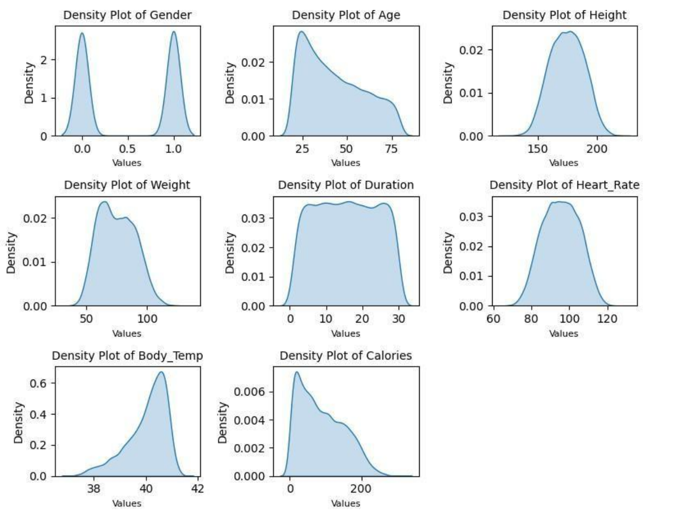

Fig. 3. Density Plot highlighting the overall concentration of the dataset.

By highlighting statistical measures like the median, quartiles, and possible outliers, a box plot, also known as a box-and-whisker plot, offers a condensed visual summary of a dataset [2]. It allows for the detection of variations that could impact prediction accuracy by illuminating the central pattern, spread, and imbalance of features in calorie burn prediction. The dataset's box plots are displayed in Figure 4, where a number of features—such as Weight, Duration, and Calories—show pronounced outliers, or instances of abnormally high or low values. There are some extreme points outside the whiskers in both body temperature and heart rate, which also exhibit some variability. Height and Age, on the other hand, show less variation from their ranges within quarters and seem more stable. Because it is binary, the Gender feature does not exhibit any significant outliers or spread.

Figure-5 The dataset includes demographic and physiological features to predict Calories burned. A heatmap shows correlations among variables. Strong positive correlations, such as Weight and Duration with Calories, highlight key predictors, while weaker ones, like Age, have less impact. By analyzing this heatmap, we can identify potential multicollinearity among variables (e.g., Height and Weight) and focus on the most influential predictors.

C. Preprocessing

To guarantee the quality of the dataset and its suitability for machine learning models, several steps were taken during the preprocessing phase. First, since the User ID column had no predictive value and only functioned as a unique identifier, it was removed. To make it compatible with machine learning algorithms, the gender attribute—which was initially represented as the categorical values "Male" and "Female"—was transformed into a binary numerical format, where 0 = Male and 1 = Female.These preprocessing procedures improved the dataset's consistency and completeness, which improved the model's performance during later training and assessment.

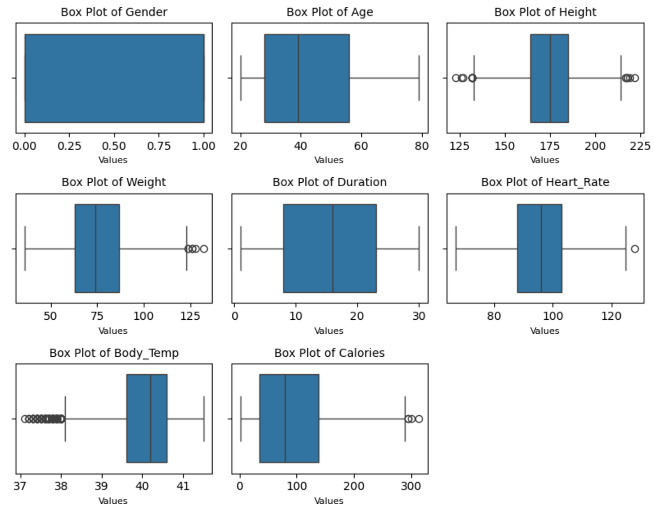

Fig. 4. The box plot represents the outlier range of different features.

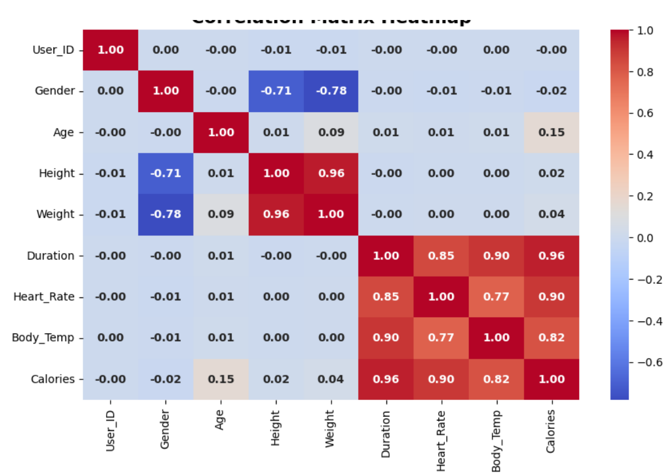

Fig. 5. Heatmap visualizing the intensity of relationships across the features.

D. Optimal Feature Finding

Feature selection was used to decrease redundancy and increase model accuracy. Weight, duration, heart rate, and body temperature were found to have the strongest correlations with calorie expenditure, whereas gender and age had less of an impact. Furthermore, the dominance of these variables was validated by feature importance from tree-based models (Gradient Boosting and Random Forest). To ensure effective model training and enhanced performance, the final feature set was honed to highlight the most predictive attributes.

E. Model Formation

In our research, we implemented several well-established ensemble regression techniques, including BAGGINGR [5], RFR [6], VOTINGR [12], XGBOOSTR [7], STACKINGR

[13], and GBDTR [14] to predict calorie expenditure during exercise. Among these, the evaluation revealed that the XGBOOSTR ensemble method achieved the highest performance and provided the most reliable predictions.

XGBoost is a gradient boosting framework designed for efficiency and scalability, capable of handling large datasets and complex feature interactions [7]. It builds an ensemble of decision trees sequentially, where each tree corrects the residual errors of the previous ones. Unlike standard boosting, XGBoost incorporates regularization techniques to reduce overfitting and improve generalization. The basic formulation of XGBoost regression is shown in Equation (1):

$$F(x) = F_{previous}(x) + \eta \cdot h(x) \tag{1}$$

Here, $F(x)$, $F_{previous}(x)$, η, and $h(x)$ are updated prediction, prediction from the previous step, learning rate, and prediction from the current decision tree respectively.

F. Model Evaluation

The metrics MSE, MAE, R2, PSNR, and SNR are used in the context of regression to evaluate the performance of all models[19]. Table II illustrates the measurement of different performance metrics. These metrics aid in determining the prediction and generalization ability of a regression model. In general, the goal is to maximize R2, PSNR, and SNR while minimizing MSE and MAE.

TABLE II
IN-DEPTH OVERVIEW OF PERFORMANCE METRICS

Name	Equation	Meaning
MSE	$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	Indicates better predictive performance
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert Y_i - \hat{Y}_i\rvert$	Provides a measure of the average squared deviation
R2	$1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$	Proportion of the variance in the dependent variable
PSNR	$10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right)$	Indicates better reconstruction quality
SNR	$10 \cdot \log_{10}\left(\frac{Signal\ Power}{Noise\ Power}\right)$	Provides a more manageable scale for human perceptio

G. Final Model

The XGBoostR showed the highest accuracy and robustness when all ensemble regression models were compared using performance evaluation metrics. As a result, using the provided dataset, XGBoostR is regarded as the best model for estimating the number of calories burned during exercise.

H. Outcome Explanation

This study also uses XAI techniques to examine how individual features affect the final model's predictions. Including interpretability guarantees more transparency and boosts trust in the model's decision-making process. LIME and SHAP Shapley Additive Explanations were used in this work to interpret the results. In order to provide a comprehensive understanding of the model's predictions, SHAP offers a unified framework for measuring the contribution of each feature. Features like body temperature, heart rate, and duration, for example, were found to have the biggest effects on calorie estimation.LIME complements this by emphasizing local interpretability, which explains individual predictions by using simpler surrogate models to approximate the complex model around particular data points. This makes it possible to comprehend the model's response to feature variations for specific exercise sessions on a deeper level.The study enhances the calorie burn prediction system's interpretability and reliability by combining SHAP and LIME, guaranteeing that the findings are not only precise but also clear and significant for fitness tracking and medical applications.

III. RESULT

The performance of every model is compiled in Table III. With the lowest MAE (1.29) and MSE (3.91), the highest R2 (0.9990), and the best PSNR (43.47 dB) and SNR (34.82 dB), the XGBOOSTR model demonstrates remarkable predictive accuracy and robustness. XGBOOSTR consistently outperforms BaggingR, RFR, and StackingR, although they all do well . VotingR, on the other hand, performs worse, having the highest errors and the lowest R2 (0.9919). According to this evaluation, XGBOOSTR is the best model for predicting calorie burn.

TABLE III
PERFORMANCE COMPARISON OF REGRESSION MODELS

Model	MAE	MSE	R^2	PSNR (dB)	SNR (dB)
BaggingR	1.7327	7.7437	0.9980	40.51	31.86
RFR	1.7279	7.6858	0.9981	40.54	31.89
GBDTR	2.7357	14.7690	0.9963	37.70	29.06
XGBOOSTR	**1.2941**	**3.9176**	**0.9990**	**43.47**	**34.82**
VotingR	3.9490	31.8454	0.9919	34.37	25.72
StackingR	1.7964	8.1578	0.9979	40.28	31.63

Table IV presents the fold-wise performance of the XGBoost model. Fold 4 is the best performer, achieving the lowest MAE (1.880), MSE (9.870), and RMSE (3.142), with the highest R^2 (0.9974), PSNR (38.78 dB), and SNR (30.70 dB), indicating superior predictive accuracy. On the other hand, Fold 7 demonstrates the weakest performance, with the highest MSE (17.561), MAE (2.177), and lowest R^2 (0.9954), suggesting slightly less accurate predictions.

Figure 6 shows the SHAP plot for our final model XGBoost. This visualization highlights the most influential features for predicting calorie expenditure. Among them, Duration and Heart Rate emerge as strong predictors, followed by Weight and Body Temperature. These features have a substantial impact on the model's predictions, with longer exercise duration and higher heart rate contributing positively to calorie burn. This concise visualization enhances our understanding of the directional influence of key variables on the model's outcomes. The LIME plot for our finished XGBoost model is shown in Figure X. Plotting the most important characteristics makes

TABLE IV
10-FOLD RESULTS FOR THE BEST MODEL XGBOOST

Fold	MAE	MSE	R^2	PSNR (dB)	SNR (dB)
1	2.082	16.684	0.9959	37.72	28.68
2	2.056	14.301	0.9964	37.26	29.27
3	1.975	13.238	0.9966	37.34	29.61
4	1.880	9.870	0.9974	38.78	30.70
5	1.917	11.470	0.9971	38.00	30.20
6	1.982	13.897	0.9963	36.84	29.30
7	2.177	17.561	0.9954	36.77	28.32
8	2.053	14.555	0.9964	37.19	29.07
9	2.031	14.235	0.9963	37.16	29.21
10	2.069	16.428	0.9957	37.00	28.54

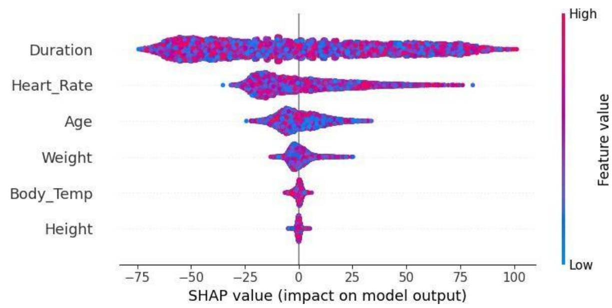

Fig. 6. The impact of various features on the model's output.

it easier to understand how each feature affects the prediction result. By making it easier to see how each attribute affects the model's decision-making process, this visualization increases model transparency and boosts the predictive framework's explainability and credibility.

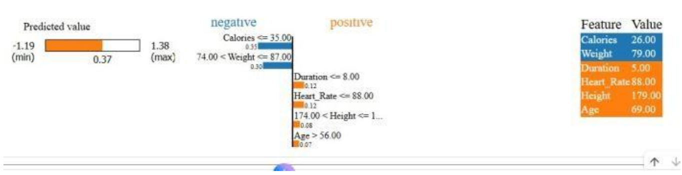

Fig. 7. LIME analysis with key features influencing a specific model prediction

IV. DISCUSSSION

Our study highlights how well EXAI and ML approaches work together to predict calorie burn accurately. Since it gives people and professionals practical advice for preserving and enhancing health, calorie estimation is essential to health-care, sports, and lifestyle management. Informed decision-making is made possible by our method, which promotes improved fitness tracking, increases confidence in AI-driven recommendations, and guarantees predictions that are both accurate and comprehensible.This study's evaluation of several ensemble regression models showed that the XGBoostR performed best, with the highest prediction accuracy. RFR, on the other hand, was found to be the least effective method. However, our study still has some shortcomings, including the lack of practical application in wearable fitness equipment and the absence of feature optimization techniques that could improve model performance even more.A comparison between the suggested method and current approaches is given in Table V. Prior research mainly used ML techniques to predict calories, but they were not interpretable. By combining XAI and XGBoostR, on the other hand, our approach not only increases prediction accuracy but also clarifies the factors that influence calorie burn. Our system gives both the predicted calorie burn and explanatory insights, which helps users and healthcare providers better understand the dynamics of calorie expenditure than traditional models that only output calorie values. Our method is more dependable and useful for real-world applications in fitness and health because of this dual capability.

TABLE V
COMPARATIVE ANALYSIS OF DIFFERENT METHODS FOR CALORIE PREDICTION

Method	Dataset		Technology	Identification	XAI
	Samples	*Features*			
I3CS et al. [1]	1200	6	ML models	Calorie burn regression	NO
JETIR et al. [2]	1500	7	XGBoost	Calorie prediction	NO
IARJSET et al. [3]	1500	8	XGBoost	Calorie burn prediction	NO
Proposed Method	**1236**	**8**	**XGBOOSTR**	**Calorie prediction**	**YES**

V. CONCLUSION

By combining XAI techniques ML ensemble models, this study effectively improves the prediction of calorie expenditure during exercise. We applied multiple regression ensembles, including GBDTR, XGBOOSTR, STACKINGR, RFR, BAGGINGR, and VOTINGR, to a dataset of 1,500 samples that included age, gender, weight, height, body temperature, heart rate, and calories burned. After thorough cross-validation, the XGBOOSTR model outperformed the other models in terms of prediction accuracy.By using LIME and SHAP, interpretable insights into feature importance were obtained, emphasizing the important roles that body temperature, heart rate, and exercise duration play in calorie expenditure. These results highlight how ML and XAI can enhance individualized fitness tracking and assist medical professionals in creating focused wellness and exercise regimens. In order to improve personalized health monitoring, future research will concentrate on growing the dataset, adding more physiological characteristics, and implementing the models in wearable applications that operate in real time.

REFERENCES

[1] Harvard Health. (2021, March 8). Calories burned in 30 minutes for people of three different weights. https://www.health.harvard.edu/diet-and-weight-loss/calories-burned-in-30-minutes-for-people-of-three-different-weights

[2] Wikipedia contributors. (2025, August 23). Probability density function. Wikipedia. https://en.wikipedia.org/wiki/Probability$_d$ensity$_f$ unction

[3] McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1), 12-16.

[4] Bu¨hlmann, P. (2012). Bagging, boosting and ensemble methods. *Handbook of Computational Statistics: Concepts and Methods*, 985-1022.

[5] Parmar, A., Katariya, R., & Patel, V. (2019). A review on random forest: An ensemble classifier. In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI)* 2018 (pp. 758-763). Springer.

[6] Pavlyshenko, B. (2018, August). Using stacking approaches for machine learning models. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)* (pp. 255-258). IEEE.

[7] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... and Zhou, T. (2015). XGBoost: extreme gradient boosting. *R package version*, 0.4-2, 1(4), 1-4.

[8] Calories burnt prediction. (2025, January 20). Kaggle. https://www.kaggle.com/datasets/ruchikakumbhar/calories-burnt-prediction

[9] Panwar, P., Bhutani, K., Sharma, R., & Saini, R. (2023). A study on calories burnt prediction using machine learning. *ITM Web of Conferences*, 54, 01010. https://doi.org/10.1051/itmconf/20235401010

[10] Alfred, T., Zarina, C., Sharma, R., & Normaiza, R. (2020). A study on Comparison of Machine Learning Methods for Calories Burn Prediction. *Journal of Informatics and Web Engineering*, 3(1):182-191. https://doi.org/10.33093/jiwe.2024.3.1.12

[11] A, S. D. K., S, G., Basavaraj, M., & S, M. V. (2025). Hybrid model analysis for calorie prediction using ensemble learning techniques: XGBOOST and Random Forest. *theaspd.com*. https://doi.org/10.64252/1zx66k89

[12] GeeksforGeeks. (2025, July 23). Voting regressor. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/voting-regressor/

[13] https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/. (2018).

[14] Gradient boosted decision trees. (2020). Google for Developers. https://developers.google.com/machine-learning/decision-forests/intro-to-gbdt