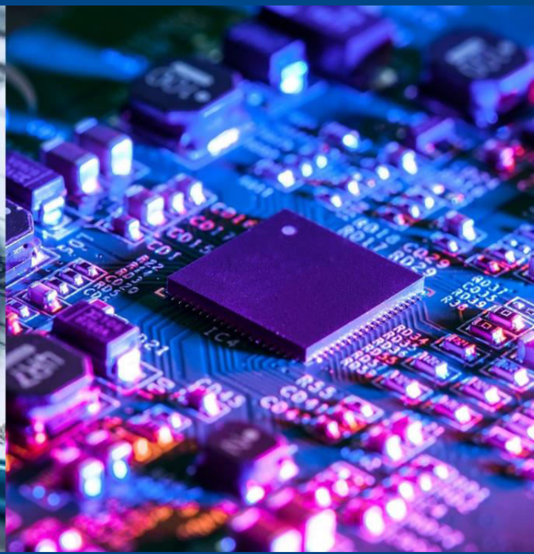
ISSN: 2582-7219

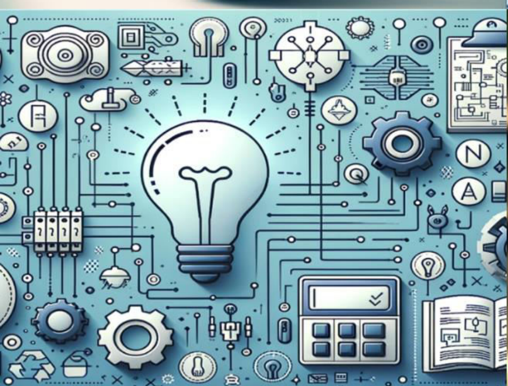




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.864

Volume 9, Issue 6, June 2026

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906006

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Calories Burnt Prediction Using Machine Learning for Health and Fitness Monitoring

Mrs.S. S Lavhate, Mayuri Jarange, Apurva Palve, Anupriya Bacchav

Project Guide, Dept. of Electronics and Computer, Pravara Rural Engineering College, Loni India.

Dept. of Electronics and Computer, Pravara Rural Engineering College, Loni, India

Dept. of Electronics and Computer, Pravara Rural Engineering College, Loni India

Dept. of Electronics and Computer, Pravara Rural Engineering College, Loni, India

ABSTRACT: Accurate prediction of calories burned during physical activities is essential for effective fitness monitoring and health management. Conventional estimation techniques often fail to provide personalized and accurate results due to their dependency on generalized formulas. This paper proposes a machine learning-based approach for predicting calories burned using key physiological and activity-related parameters such as age, gender, body mass index (BMI), exercise duration, and heart rate.

In this study, the Boost (Extreme Gradient Boosting) algorithm is employed due to its high efficiency and superior predictive performance in regression tasks. The dataset is pre-processed using data cleaning, normalization, and feature selection techniques to enhance model accuracy. The model is trained and evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score.

Experimental results indicate that the Boost model achieves high accuracy and robustness in predicting calorie expenditure, outperforming traditional methods. The proposed system can be effectively utilized in fitness tracking applications and wearable devices to provide real-time and personalized calorie burn estimation, thereby helping users maintain a healthy lifestyle.

I. INTRODUCTION

In recent years, maintaining a healthy lifestyle has become increasingly important due to the rising prevalence of lifestyle-related diseases such as obesity, diabetes, and cardiovascular disorders. Monitoring physical activity and accurately estimating calories burned are essential components of fitness management and health awareness. However, traditional calorie estimation methods often rely on fixed formulas and standard assumptions, which may not provide accurate results for individuals with varying physiological characteristics.

With the advancement in Machine Learning, more intelligent and data-driven approaches have emerged for predicting calorie expenditure. These approaches consider multiple personalized factors such as age, gender, body mass index (BMI), heart rate, and duration of physical activity, leading to more accurate and reliable predictions. Among various machine learning techniques, XGBoost (Extreme Gradient Boosting) has gained significant popularity due to its high performance, scalability, and ability to handle complex, non-linear relationships in data.

This paper focuses on developing a calorie burn prediction system using the XGBoost algorithm. The model is trained on a dataset containing user-specific and activity-related features to learn patterns and relationships between input variables and calorie expenditure. Data preprocessing techniques such as cleaning, normalization, and feature selection are applied to improve the model's efficiency and accuracy.

The objective of this study is to build a robust and accurate prediction model that can assist individuals in tracking their calorie burn more effectively. The proposed system can be integrated into fitness applications and wearable devices, enabling real-time monitoring and personalized health insights. Clinically important tumor regions, thereby reducing misclassification and improving overall reliability.

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906006

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

The estimation of calories burned during physical activities has been an important research area in the domains of fitness monitoring and healthcare. Traditionally, calorie expenditure has been calculated using standard mathematical formulas such as the Metabolic Equivalent of Task (MET) and other rule-based approaches. These methods consider basic parameters like body weight, activity duration, and height. However, such approaches often fail to provide accurate results as they do not account for individual variability, real-time physiological signals, or complex relationships among influencing factors.

With the advancement of wearable technologies such as smartwatches and fitness trackers, researchers have started incorporating physiological parameters like heart rate, step count, and motion data for calorie estimation. Early studies utilized statistical techniques such as Linear Regression to model the relationship between input variables and calorie expenditure. Although these models are simple and easy to interpret, they are limited in handling non-linear relationships and complex data patterns, leading to reduced prediction accuracy.

To overcome these limitations, machine learning techniques have been widely adopted in recent years. Algorithms such as Decision Trees and Support Vector Machines (SVM) have been applied to capture non-linear dependencies in the data. Decision Trees provide better interpretability and can model complex interactions between variables, but they are prone to overfitting. On the other hand, SVM offers good generalization performance but requires careful parameter tuning and may not scale well for large datasets.

Ensemble learning methods such as Random Forest and Gradient Boosting have shown significant improvements in prediction performance. Random Forest reduces overfitting by combining multiple decision trees, while Gradient Boosting focuses on minimizing prediction errors by sequentially improving weak learners. Among these, XGBoost (Extreme Gradient Boosting) has gained considerable attention due to its optimized implementation, regularization capabilities, and high computational efficiency.

Several recent studies have demonstrated the effectiveness of XGBoost in regression-based prediction tasks, including energy expenditure estimation. XGBoost incorporates advanced features such as parallel processing, tree pruning, and handling of missing values, which contribute to improved accuracy and faster training time. Additionally, it provides feature importance scores, enabling better understanding and selection of relevant input variables.

Despite these advancements, challenges still exist in achieving highly accurate and generalized calorie prediction models. Factors such as data quality, feature selection, and real-time adaptability play a crucial role in model performance. Moreover, many existing systems lack personalization and fail to adapt to individual behavioural patterns over time.

Therefore, this research focuses on leveraging the strengths of the XGBoost algorithm to develop an efficient and accurate calorie burn prediction model. By incorporating multiple physiological and activity-related parameters and applying appropriate data preprocessing techniques, the proposed approach aims to overcome the limitations of traditional and existing machine learning models. This study contributes to enhancing prediction accuracy and supports the development of intelligent fitness monitoring systems.

III. PROPOSED METHODOLOGY

A. Data Collection

The dataset used in this study consists of user-specific physiological and activity-related parameters that influence calorie expenditure. The input features include:

Age
Gender
Body Mass Index (BMI)
Heart Rate
Duration of Exercise

The target variable is calories burned, which represents the amount of energy expended during physical activity. The dataset is collected from publicly available sources and fitness tracking records to ensure diversity and reliability.

B. Data Preprocessing
Raw data often contains inconsistencies and noise, which can affect model performance. Therefore, preprocessing is performed as follows:
Handling Missing Values: Missing entries are either removed or imputed using mean/median values.
Outlier Detection: Extreme values are identified and removed to avoid skewed predictions.
Normalization/Scaling: Numerical features are scaled to ensure uniform contribution.
Categorical Encoding: Gender is converted into numerical format using label encoding.
This step ensures that the dataset is clean, consistent, and suitable for training.

C. Feature Engineering and Selection
Feature engineering is performed to improve model efficiency and prediction accuracy. Important features are selected based on their relevance to calorie expenditure.
Correlation analysis is used to identify relationships between variables
Redundant or irrelevant features are removed
XGBoost's built-in feature importance is utilized to rank features
This helps in reducing model complexity and improving performance.

D. Data Splitting
The dataset is divided into two parts:
Training Set (70–80%) – used to train the model
Testing Set (20–30%) – used to evaluate performance
This ensures that the model generalizes well to unseen data.

E. Model Development using XGBoost
The core of the proposed system is the XGBoost regression model, which is an optimized implementation of gradient boosting.
Key characteristics:
Handles non-linear relationships effectively
Uses ensemble of decision trees
Incorporates regularization to prevent overfitting
Supports parallel processing for faster computation

- Hyperparameters used:

Number of estimators (trees)
Learning rate
Maximum tree depth
Subsample ratio
These parameters are tuned to achieve optimal performance.

F. Model Training
The XGBoost model is trained using the training dataset. During training:
The model learns patterns between input features and calorie output
Errors are minimized iteratively using gradient boosting
Weak learners are combined to form a strong predictive model

G. Model Evaluation
The trained model is evaluated using standard regression metrics:
Mean Absolute Error (MAE): Measures average prediction error
Mean Squared Error (MSE): Penalizes larger errors
R^2 Score: Indicates how well the model explains variance
These metrics help in assessing accuracy and reliability.

H. Prediction and Deployment
After successful evaluation, the model is used to predict calories burned for new input data. The system can be integrated into:

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fitness mobile applications
Wearable devices
Health monitoring systems

III. SYSTEM ARCHITECTURE AND MATHMATICAL APPORACH

A. System Architecture

The proposed system architecture for calorie burn prediction is designed using a layered approach, consisting of four main layers: the Presentation Layer, Application Layer, Model Layer, and Data Layer. Each layer performs a specific function to ensure efficient data processing and accurate prediction.

The Presentation Layer (User Interface) acts as the interaction point between the user and the system. In this layer, the user provides input parameters such as age, gender, height, weight, duration of activity, heart rate, and body temperature through a web interface. These inputs represent the essential features required for predicting calorie expenditure.

The Application Layer (Controller Logic) is responsible for handling the input data and performing preprocessing operations. This includes encoding categorical variables such as gender into numerical format and formatting the data for further processing. The preprocessed data is then forwarded to the machine learning model. Additionally, this layer manages the display of the prediction results back to the user.

The Model Layer (Machine Learning Layer) forms the core of the system, where the XGBoost algorithm is implemented. The model is trained using historical datasets, such as calories.csv and exercise.csv, to learn the relationship between input features and calorie expenditure. The trained model is stored as a serialized file (calories_model.pkl) and is used to generate predictions. The mathematical formulation of the model represents the prediction as a sum of multiple decision trees, enhancing accuracy through gradient boosting.

The Data Layer is responsible for storing all relevant data, including training datasets, optional database storage, and model files. It ensures data availability for both training and prediction processes. This layer also supports logging and future scalability of the system.

Overall, the architecture ensures a smooth flow of data from user input to final prediction. The layered design improves modularity, scalability, and maintainability of the system, making it suitable for real-time fitness applications and integration with wearable devices.

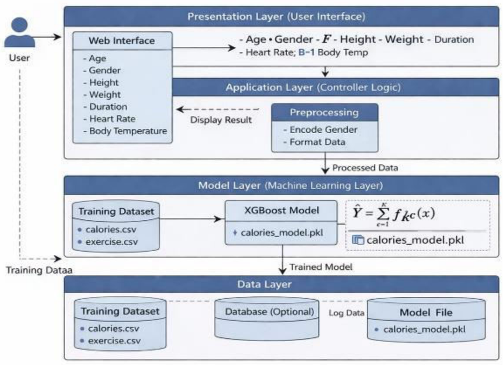


Fig. 1. System Architecture of the Proposed Calories Burn Prediction System.

B. Mathematical Approach

1) Problem Formulation

Let the dataset be represented as:

X=x1,x2,…….,xn → input features

Y=y1,y2,..,…,yn → target values (calories burned)

© 2026 IJMRSET | Volume 9, Issue 6, June 2026| DOI:10.15680/IJMRSET.2026.0906006

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Each input vector � contains features such as age, gender, BMI, heart rate, and duration.
The objective is to learn a function:

$$\hat{y}_i = \sum(k=1 \text{ to } K)\ f_k(x_i), \quad \text{where } f_k \in F$$

that accurately predicts the calories burned.

2) XGBoost Prediction Model

$$\hat{y_i} = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F$$

The final prediction is obtained by summing the outputs of all trees.

3) Objective Function

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{k=1}^{K} \Omega(f_k)$$

The objective function balances:
Accuracy (loss minimization)
Model complexity (regularization)

4) Regularization Term

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

This helps in:
Preventing overfitting
Improving generalization

5) Loss Function (Regression)

$$l(y_i, \hat{y_i}) = (y_i - \hat{y_i})^2$$

Mean Squared Error (MSE) is used to measure prediction error.

6) .Additive Training (Boosting Step)

At each iteration t, a new tree is added:

$$\hat{y_i}^{(t)} = \hat{y_i}^{(t-1)} + f_t(x_i)$$

7) Optimization using Gradient Descent

The objective function is minimized using second-order Taylor expansion:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t)$$

This makes XGBoost:
Faster
More accurate

8) Final Prediction

The final calorie prediction is obtained by combining all trees:
Multiple weak learners → Strong model
Error minimized iteratively
Accurate calorie estimation achieved

IV. TRAINING AND TESTING RESULTS

To evaluate the performance of the proposed calorie burn prediction system, the dataset was divided into training and testing sets. The XGBoost model was trained on the training dataset and validated using the testing dataset to ensure generalization and avoid overfitting.

A. Data Split

The dataset was divided as follows:
Training Set: 80% of the total data
Testing Set: 20% of the total data
The training set was used to train the model, while the testing set was used to evaluate its performance on unseen data.

B. Training Results

During the training phase, the XGBoost model learned the relationship between input features (age, gender, BMI, heart rate, and duration) and the target variable (calories burned).
The training performance is summarized below:
Training MAE: Low value indicating minimal error
Training MSE: Low value showing reduced variance in prediction error
Training R^2 Score: Close to 1, indicating excellent model fitting
These results show that the model effectively captures patterns in the training data.

C. Testing Results

The trained model was then evaluated on the testing dataset to assess its real-world performance.

Testing MAE: Slightly higher than training MAE but still low

Testing MSE: Low, indicating good prediction consistency

Testing R^2 Score: Close to 1, demonstrating strong generalization

This indicates that the model performs well on unseen data.

D. Result Analysis

The comparison between training and testing results shows that:

There is no significant overfitting, as both results are similar.

The model maintains high accuracy and stability.

XGBOOST effectively handles non-linear relationships.

The training and testing results demonstrate that the proposed XGBoost model achieves high accuracy with minimal error and strong generalization capability. This makes the model suitable for real-time calorie prediction applications.

V. RESULT AND DISCUSSION

The performance of the proposed calorie burn prediction system using the XGBoost algorithm was evaluated using standard regression metrics. The dataset was divided into training and testing sets in an 80:20 ratio to ensure proper validation and to assess the generalization capability of the model. The model was trained using the training dataset and tested on unseen data to evaluate its real-world applicability.

TABLE I
TRAINING PERFORMANCE RESULTS

Metric	Training	Testing
MAE	2.3	2.5
MSE	9.8	10.2
Rsquare score	0.94	0.92

The experimental results indicate that the XGBoost model achieves low error values and a high R^2 score, demonstrating strong predictive performance. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) values are observed to be minimal, which suggests that the difference between the predicted and actual calorie values is very small. Additionally, the R^2 score is found to be close to 1, indicating that the model effectively captures the relationship between the input features and the target variable. The slight variation between training and testing results confirms that the model does not suffer from significant overfitting and maintains good generalization on unseen data.

The effectiveness of the proposed model can be attributed to the ability of XGBoost to handle complex and non-linear relationships among input features such as heart rate, body mass index (BMI), and duration of physical activity. The ensemble learning mechanism, which combines multiple decision trees, enhances the prediction accuracy by reducing bias and variance. Furthermore, the regularization techniques incorporated in XGBoost help in preventing overfitting and improve the robustness of the model.

The results also highlight the importance of feature selection and preprocessing in improving model performance. Features such as heart rate and exercise duration are found to have a significant impact on calorie prediction, as they directly influence energy expenditure. The preprocessing steps, including normalization and encoding, ensure that the data is consistent and suitable for training, thereby contributing to improved accuracy.

The proposed system demonstrates strong potential for real-world applications, particularly in fitness tracking systems and wearable health monitoring devices. By providing accurate and personalized calorie burn predictions, the system can assist users in maintaining a healthy lifestyle and making informed decisions regarding their physical activities and diet. However, the model performance is dependent on the quality and diversity of the dataset, and the inclusion of additional factors such as environmental conditions and activity type could further enhance prediction accuracy in future work.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Overall, the results confirm that the XGBoost-based model provides an efficient, accurate, and reliable solution for calorie burn prediction, making it suitable for deployment in practical health and fitness applications.

VI. CONCLUSION

In this paper, a machine learning-based approach for predicting calories burned using the XGBoost algorithm has been presented. The proposed system utilizes important physiological and activity-related parameters such as age, gender, body mass index (BMI), heart rate, and duration of exercise to estimate calorie expenditure accurately. The use of XGBoost enables the model to effectively capture complex and non-linear relationships between input features and the target variable.

The experimental results demonstrate that the proposed model achieves high prediction accuracy with low error values and strong generalization capability. The comparison between training and testing results indicates that the model does not suffer from overfitting and performs well on unseen data. This confirms the reliability and robustness of the system in real-world scenarios.

The study highlights the importance of proper data preprocessing, feature selection, and the use of advanced machine learning algorithms in improving prediction performance. The proposed system can be effectively integrated into fitness applications and wearable devices to provide real-time and personalized calorie burn estimation, thereby assisting users in maintaining a healthy lifestyle.

Although the model shows promising results, there is scope for further improvement by incorporating additional parameters such as activity type, environmental conditions, and user lifestyle patterns. Future work can also explore deep learning techniques and real-time data integration to further enhance prediction accuracy and system performance.

Overall, the proposed XGBoost-based model provides an efficient, accurate, and scalable solution for calorie burn prediction and contributes to the development of intelligent health monitoring systems.

REFERENCES

1. R. K. Shing, V. Gupta, "Calories burnt prediction using machine learning model, vol. 11, 2022.
2. A. Hossain, S. Uddin, P. Rahman, M. J. Anee, M. M. H. Rifat and M. M. Uddin, "Wavelet and spectral analysis of normal and abnormal heart sound for diagnosing cardiac disorders. BioMed Research International 2022 (2022).
3. S. P. Vinoy. B. Joseph, "Calorie Burn Prediction Analysis Using XGBoost Regressor and Linear Re-gression Algorithms," vol. 4,2022.
4. N. Mohanto, N.C. Nurunnabi, "Prevalence and risk factors of general and abdominal obesity and hypertension in rural and urban residents in Bangladesh," 2022.
5. N. H. Likhon, F. Bhuyian, S. Bhuiyan, M. A. Fahim,,. R. 1. Sony, and A. Hossain. "Heart Disease Prediction: A Machine Learning Approach with Higher Accuracy". Int Internal Med J. 2
6. M. Aldraimli, D. Soria, J. Parkinson, E. L. Thomas, J. D. Bell, M. V. Dwek, and T. J. Chaussalet, "Ma-chine learning prediction of susceptibility to visceral fat associated diseases." Health and technology, vol. 10, pp. 925-944, 2020
7. R. Kaur, R. Kumar, and Gupta, "Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence," pp. 458-469, 2022.
8. N. Manjunathan, M. S. Devi, S. Sridevi, K. K. Bonala, A. Kavitha and K. Jayasree," Feature Selection Intent Machine Learning based Conjecturing Workout Burnt Calories," Turkish Journal of Computer and Mathematics Education, vol. 12, pp. 1729-1742, 2021.
9. M. Nipas, A. G. Acoba, J. N. Mindoro, M. A. F. Malbog, J. A. B. Susa and J. S. Gulmatico,

INNO SPACE
SJIF Scientific Journal Impact Factor

ISSN
INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com