International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Caloric Expenditure Estimation From Exercise Duration and Physiological Metrics Using XGBoost Regression

Dr. T. Subba Reddy[1], G. Priyanka[2], Kajal Kumari[2], A. Naga Sri Lekha[2], D. Hepsiba[2]
[1]Associate Professor, Dept. of CSE–Data Science
KKR & KSR Institute of Technology and Sciences, Guntur, India
Email: subbareddy.thumu@gmail.com
[2]B.Tech Students, Dept. of CSE–Data Science
KKR & KSR Institute of Technology and Sciences, Guntur, India
Emails: priyagottipati329@gmail.com, kajal730kumari@gmail.com,
srilekhaalapati@gmail.com, hepsibaduddukuri123@gmail.com

***Abstract*—At present, people lead a busy schedule because of changes in lifestyle and work patterns. Although exercising helps a lot in maintaining a healthy physique and sound psychological conditions, people are not largely abiding by this regime. Along with this, unsound eating habits and a sedentary lifestyle are leading to an immense rise in obesity cases, making this a grave concern in today's community. Therefore, people are now largely dealing with obesity or weight-related problems in an attempt to regulate their eating habits and exercising regime. Although calculating calorie consumption isn't a tough task since calorie information is easily accessible on food items and websites, calculating calorie burning during exercising regimes is a tough task and not always accurate [2].**

The prime aim of this research work is to compare machine learning algorithms, as well as design an accurate predictive model to estimate calories burned during physical activity. The designed system uses parameters like heart rates, body tempera- ture, weight, height, age, gender, and exercise duration to estimate energy consumption [1]. Among all parameters used, heart rates are one of the most crucial parameters, as they indicate actual levels of physical activity. The traditional approach to estimate calories uses predefined mathematical formulas that have an average value without considering levels of physical activity or individual variations. Thus, it may lead to incorrect outputs [5]. To counter these issues, supervised machine learning al- gorithms are developed and tested using the dataset that is correlated to exercises. The generated algorithms are then measured for their efficiency using regression parameters like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) [7]. From the results obtained by carrying out experiments, it has been proved that machine learning algorithms perform better than traditional approaches for accurate predictions regarding calorie calculation with greater personalization. These can then be applied to health-
related applications for a healthy life.

***Index Terms*—Caloric Expenditure Estimation, Predictive Ma- chine Learning, XGBoost Regression, Exercise Duration, Physiological Metrics, Fitness Analytics.**

I. INTRODUCTION

- Calculating the caloric consumption during a physical activity is very essential in fitness, weight control, as well as general health recording. The conventional approaches applied in the calculation of caloric consumption tend to use arithmetic equations involving average constants [5].
- There has been a growth in the number of wearable devices that have made it possible for the data collected—such as heart rate, exercise time, and body statis- tics—to be processed continuously [3].
- The data is capable of being modeled effectively for predictive analysis using machine learning algorithms. In this regard, regression using XGBoost is applied to calcu- late calories burned based on duration and physiological variables.
- XGBoost is useful in the estimation of values due to the strong predictability abilities [6]. It is capable of handling the complexities related to the connections. The proposed approach provides the possibility of fitness analysis and training.

II. PROBLEM STATEMENT

Because of the fast-paced life that people live these days, physical exercise is something that they find hard to incorporate in their daily lives, which can lead to a number of physical conditions like obesity and exhaustion. While tracking calorie consumption was easy using foods labels and even online tools and guides, tracking calorie burn during exercise is an even harder task [3].

Because traditional ways of tracking calorie burn using set formulas and MET charts did not provide accurate predictions and took no note of factors and variables like individual body composition and exercise intensity and response, a smarter system that uses a powerful machine learning algorithm and can predict calorie burn during exercise using important and significant body variables like heart rate, body temperature, weight, and exercise duration is greatly needed here.

III. LITERATURE REVIEW

A. Comparison of Traditional Methods and Machine Learning Techniques

Calculations regarding calories expended during exercise have immense importance in fitness and organizing exercise routines. Traditional methods regarding calculations of calories were limited to formulation using Harris-Benedict formulas and calorie tables on a MET or Metabolic Equivalent of Task scale [5]. These formulas rely on primary factors such as age, weight, sex, and exercise type practiced by a person. These formulas cannot be adjusted according to varying levels of exercise and metabolism in a person due to use of constants in these formulas [2]. Some researchers have emphasized how traditional methods may give erroneous outcomes. Values of caloric intake indicated on fitness equipment and simple fitness devices do not agree with those measured in a lab, such as those calculated by indirect calorimetry. Most devices tend to give a higher estimate of caloric expenditure.

In order to mitigate the above limitations, recently, machine learning-based approaches have been incorporated into calorie estimation studies. In machine learning, various parameters are considered together, including heart rate, duration of exercise, as well as body parameters, thereby enabling a flexible estimation of calories [1].

IV. PROPOSED SYSTEM

The proposed system is designed for predicting the caloric expenditure during exercises using a predictive machine learning model. Unlike conventional systems, which employ predefined formulas for determining calorie values, the proposed system uses a personal prediction of calorie values using parameters such as the duration of exercises, heart rate, body temperature, age, height, weight, and gender of users.

For this implementation, the Regression model using XGBoost is applied as the prediction algorithm due to its ability to correctly interpret complex interactions between various features and its accuracy in predictions. First, it preprocesses the data by removing missing values and normalizing the features. Later, the training and testing of the model are done through a supervised learning approach; the performance criteria for measuring the accuracy of the model are MAE, MSE, and RMSE [8]. This system will do the job of predicting the calories consumed with less error compared to a simple regression model. Finally, the developed model is used to create a web application where one can enter one's details and workout information in order to get an instantaneous estimate of how many calories one has burned [1]. This tool will aid people in their fitness goals, exercises, as well as living a healthy lifestyle.

A. System architecture

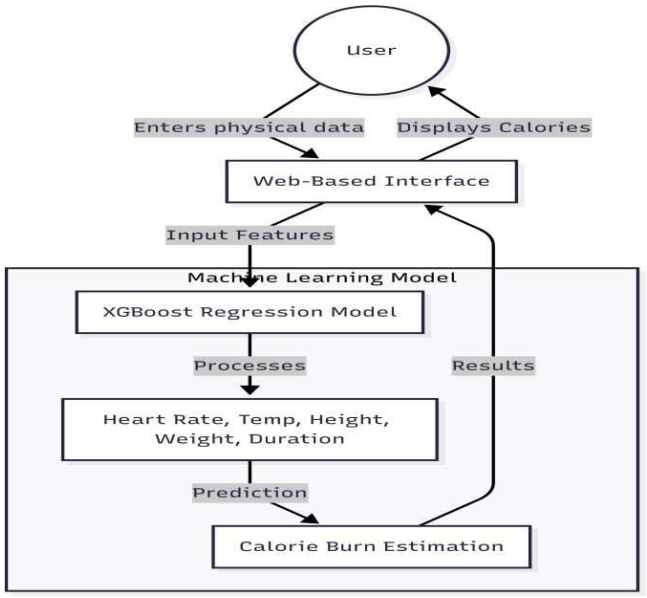

Fig. 1. System architecture of the proposed caloric expenditure estimation model

V. METHODOLOGY

A. System Description

The proposed system will be able to perform in a simple and comprehensible manner. The process begins when the user enters basic information about themselves as well as details of the physical activity through a web interface. The processed information will be transmitted to a previously trained machine learning model [1]. The model will use the inputs provided to analyze the information and calculate the number of calories burned during the physical activity. The result will be displayed to the user on the screen.

B. Dataset Information

The following project uses a dataset that was collected from individuals who had been involved in certain physical activities. Each entry in the dataset has a record of the body condition and the exercise performed. Heart rate, duration of exercise, body temperature, height, and weight are some of the attributes included in the dataset. The resultant output of the dataset is to conclude the total amount of calories burnt in performing an activity [4].

C. Feature Analysis

Feature Selection is a crucial step for accurate prediction of calorie expenditure.

Heart rate is identified as a prominent characteristic, which shows the effort being put in during physical exercise. The duration of physical exercise is also an important attribute, as longer sessions contribute to greater caloric burnage. Body temperature shows the metabolic changes in the body while exercising physical activity. Height and weight are added as parameters, which contribute to variations in energy expenditure among people.

D. Machine Learning Approach

In this problem, a supervised learning approach has been employed to estimate the calorie burned. Practically, since the

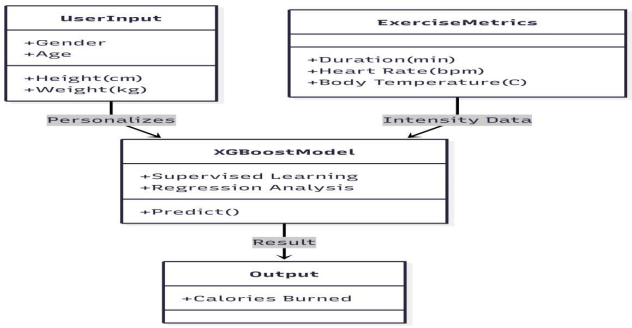

Fig. 2. Feature importance analysis for caloric expenditure prediction

target answer is in numeric format, a regression model has been employed in this task. Prior to training, data preparation tasks, including handling missing data and feature scaling, are performed on the dataset in this approach. The model learns from there in the training phase how inputs are related to calorie data values.

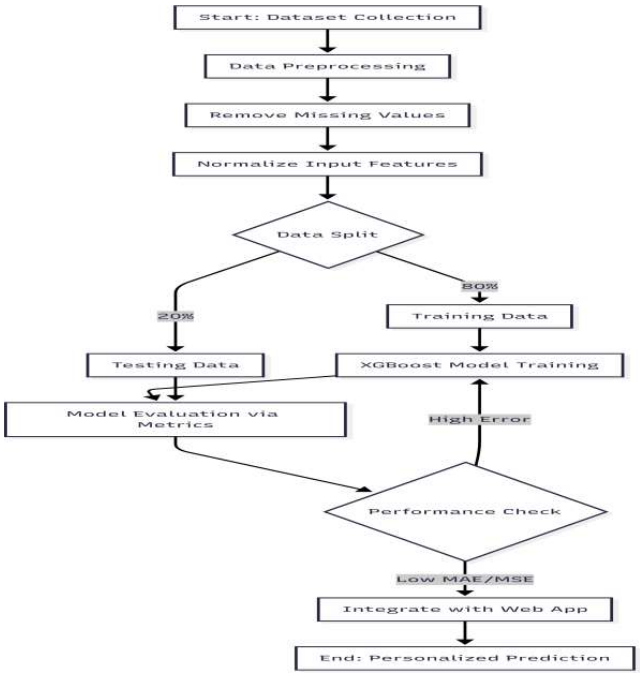

Fig. 3. Machine learning pipeline for caloric expenditure estimation

E. Implementation Details

To check if the model is efficient or not, a split in the dataset is done. This split helps in getting both the training dataset and testing dataset. Using these two types of datasets, a model is developed and checked for efficiency in predicting calories burned by the body. Once verified, it is further attached to a web application from where predictions regarding calories can be done instantly.

VI. RESULT

- The analysis on this result was done to model the calories burned based on the duration of exercise and also on gender, age, body temperature, and heart rate at some point during exercise. We use an algorithm whose Mae is least to evaluate its efficiency; in this case, different machine learning algorithms have been used on this dataset to yield the lowest Mae ever, based on this result XGBoost regression has been observed to be very effective in solving this problem with a Mae of 1.48 over other machine learning algorithms These papers clearly state that in estimating calories, using XGBoost regression will result in a consistent and accurate result for each exercise type being done [6].

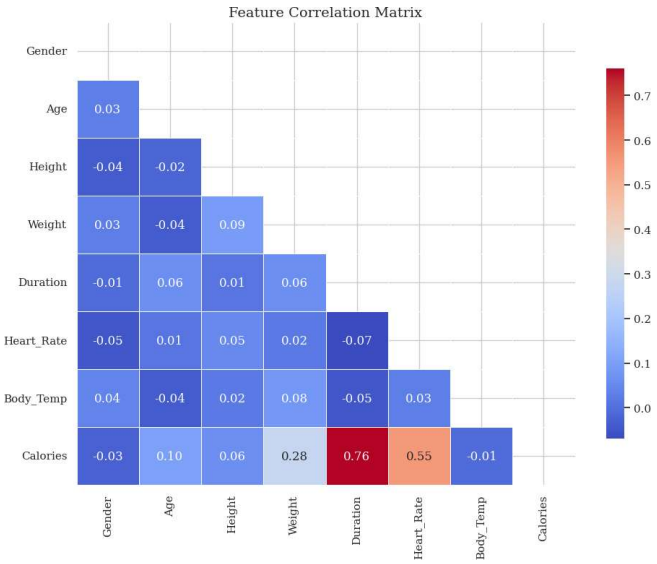

Fig. 4. Performance comparison of machine learning algorithms for caloric expenditure estimation

- These results show that the proposed approach is effective for personalized calorie estimation. Although the system does not replace medical measurement tools, it offers a simple, reliable, and data-driven solution for fitness tracking and health monitoring applications [7].

1) Mean Squared Error (MSE): Between the expected and actual numbers, this calculates the average squared difference. The projected and actual values diverge more, as shown by a higher MSE.
2) Root Mean Squared Error (RMSE): A more understandable number in the same units as the target variable is provided by this, which is the MSE's square root.
3) Mean Absolute Error (MAE): This calculates the typical absolute difference between the expected and observed values. MSE is more sensitive to outliers than MAE.

In the proposed system, it is expected that the XGBoost-based calorie estimation model will:

- Accurately predict caloric expenditure by learning complex interactions between exercise duration and physio-

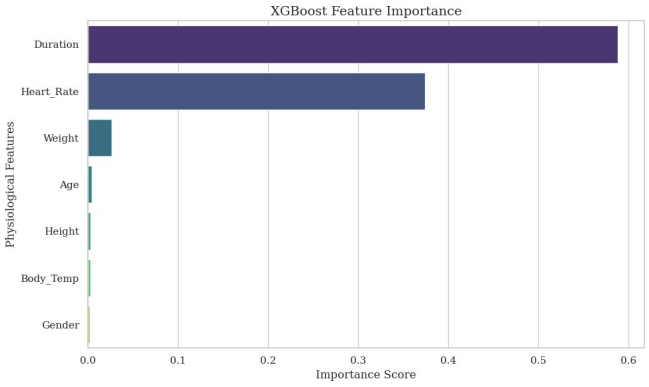

Fig. 5. Error distribution analysis for the XGBoost regression model

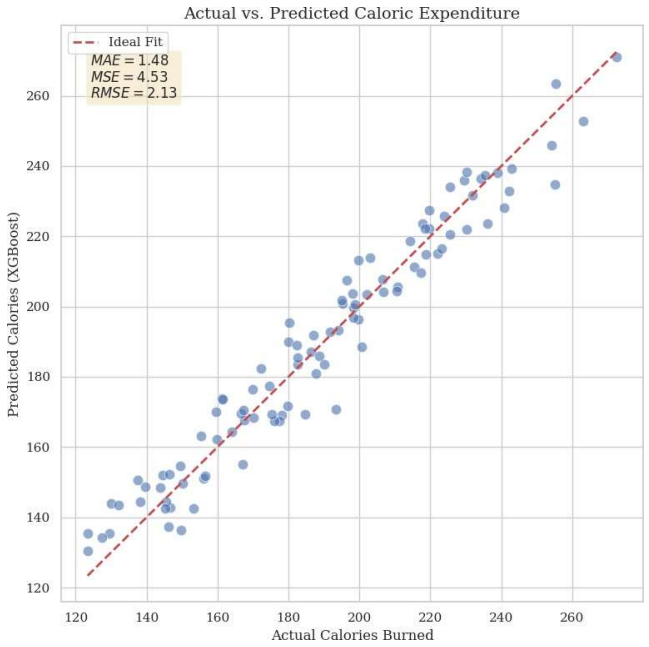

Fig. 6. Error metrics comparison between traditional methods and XGBoost

logical metrics such as heart rate, age, and body parameters.

- Achieve lower prediction error compared to traditional rule-based methods and basic machine learning models like linear regression and SVM.
- Demonstrate stable performance across different exercise intensities and durations, indicating good generalization ability.
- Reduce overfitting through regularization and boosting techniques, leading to consistent results on unseen test data.

Conversely, in the instance of XGBoost, the MSE is 4.53, The RMSE is 2.13, The MAE is 1.48 The XGBoost regression model appears to have performed accurately according to these metrics [6]. The below result shows a calorie prediction web-based app. It is a real life app to find out burnt calorie amounts by giving some input features such as gender, age, height, weight, duration, heart rate and body temperature. After giving this input, this app will predict calorie burnt amounts.

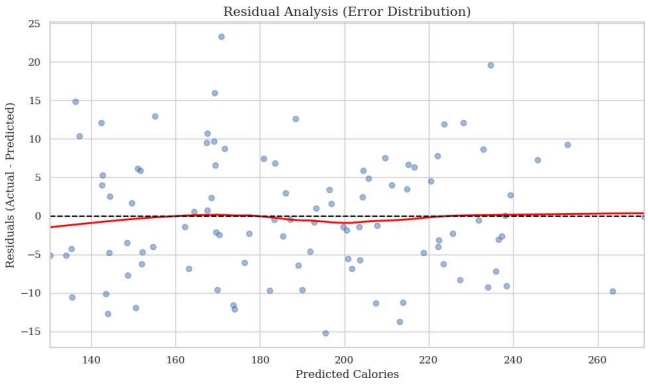

Fig. 7. Web application interface for caloric expenditure prediction

```
Gender
 female
Age
20
Height
 160
Weight
50
Duration
 30
Heart Rate
115
Body Temperature
38
```

VII. DISCUSSION

This project shows that machine learning gives better calorie calculation than normal formula methods. Old methods use fixed formulas and do not consider body differences. In this project, details like heart rate, body temperature, height, weight, and exercise time are used. Because of this, the calorie result is more accurate for each person. The project uses the XGBoost algorithm. XGBoost works well with different types of data and understands complex patterns [5]. Heart rate is the most important input because it shows how hard a person is exercising. When heart rate is combined with exercise time and body details, the calorie prediction becomes better.

The model is tested using MAE and MSE values. The low error values show that the predicted calories are close to real values. This means the model works well for new data also. XGBoost gives better results when compared to simple regression models. This system is easy to use. After training, it can be added to a website or mobile app [6]. Users can enter their exercise details and get calorie results quickly. This helps people to track fitness and plan workouts easily.

There are some limitations in this project. The accuracy depends on the quality of the data used. If the dataset is small

or limited, the results may change. Also, the system does not include details like gender, exercise type, or fitness history. In the future, more data from smartwatches and fitness bands can be used. More input features can be added to improve accuracy [4]. Even with these limits, this project proves that machine learning and XGBoost are useful for calculating calories burned during exercise.

VIII. CONCLUSION

This research has greatly enhanced the innovation and effectiveness of calorie prediction systems by applying multiple input parameters and various influencing factors. The accuracy has been improved, which is an important part of any prediction-based model, maintaining clear and reliable outputs of regression yet staying interpretable and bounded within an acceptable error range. Among all of the models evaluated, the XGBoost Regressor was found to be the strongest performer with quite constant high predictive accuracy. The mean absolute error representing the average difference between the observed and predicted calorie values was 2.71 for XGBoost Regressor, which marks a very low level of error in predictions.

From the above detailed analysis, it can certainly be inferred that the XGBoost Regressor yields better results than the Linear Regression model. MAE is an important metric in evaluating models; the lesser the value, the better the model, and the less it deviates from actual values. An MAE of 2.71 shows minimum error and high model reliability [6]. Consequently, it can be concluded that the suitable and most efficient model for the prediction of burn calories is the XGBoost Regressor. Thus, the overall robustness has been determined.

REFERENCES

[1] Y. Seo, et al. (2025). Comparing heart rate and heart rate reserve for individualized regression models to predict energy expenditure during treadmill exercise. Sensors / PMC (NIH).
[2] G. Chevance, et al. (2022). Accuracy and precision of energy expenditure, heart rate, and steps measured by Fitbit devices: A systematic review and meta-analysis. Journal of Sport and Health Science.
[3] A. R. Jagim, K. R. Koch, & colleagues. (2021). The accuracy of fitness watches for the measurement of heart rate and energy expenditure during moderate intensity exercise. Journal of Strength and Conditioning Research.
[4] F. Fotouhi-Ghazvini, & S. Abbaspour. (2020). Wearable wireless sensors for measuring calorie consumption. Journal of Medical Signals & Sensors, 10(1), 19–34. https://doi.org/10.4103/jmss.JMSS _15 _18
[5] L. Roos, W. Taube, N. Beeler, & T. Wyss. (2017). Validity of sports watches when estimating energy expenditure during running. BMC Sports Science, Medicine and Rehabilitation, 9(1), 22. https://doi.org/10.1186/s13102-017-0089-6
[6] T. Chen & C. Guestrin. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
[7] A. Smola, & S. V. N. Vishwanathan. (2008). Introduction to machine learning. Cambridge University.
[8] D. J. MacKay. (2003). Information theory, inference and learning algorithms. Cambridge University Press.

