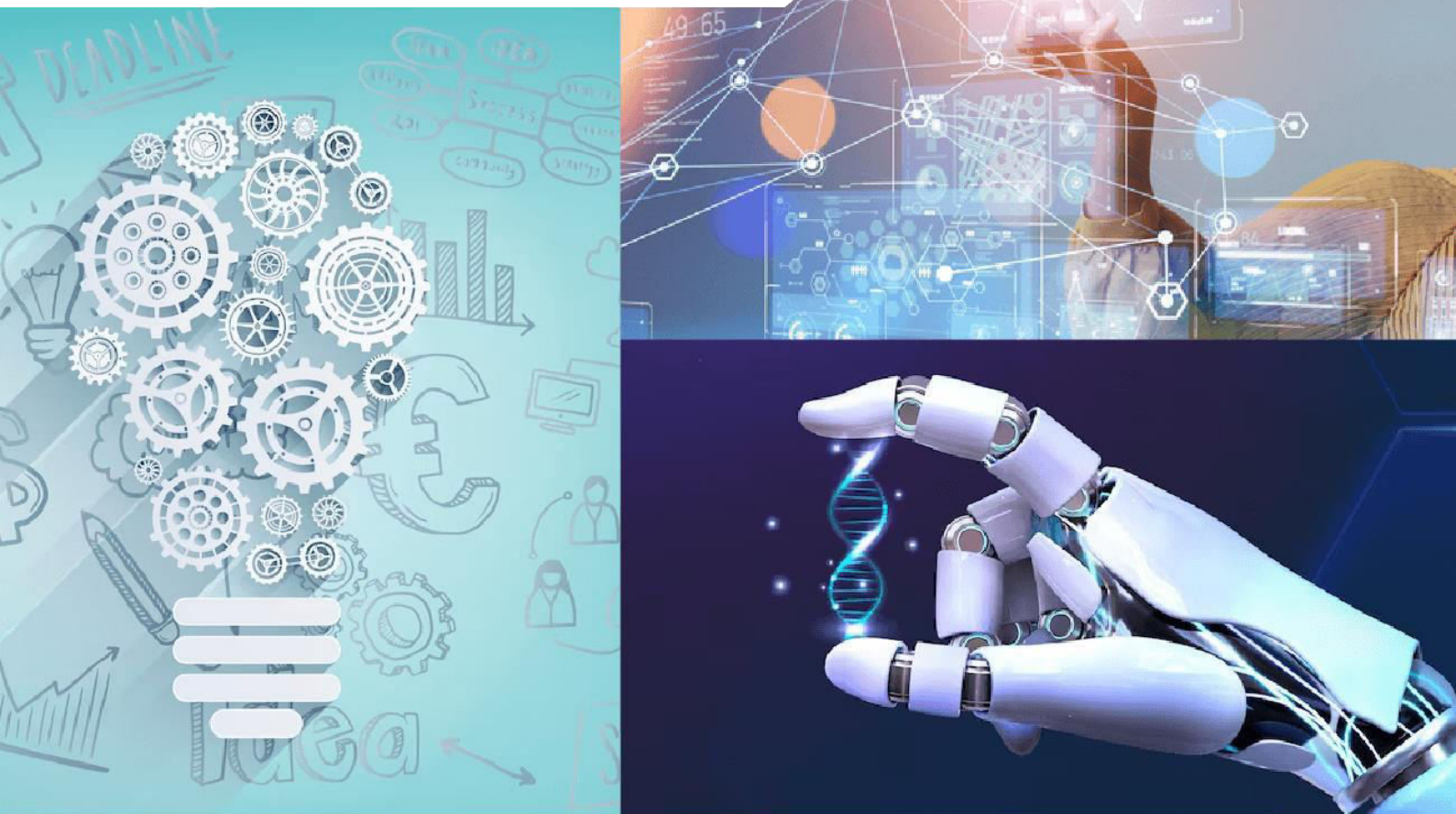




International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 1, January - February 2026

Impact Factor: 8.152



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



Calories Burn Prediction using Machine Learning

T. Kishore

Master of Computer Applications, CMR Institute of Technology, Bengaluru, India

ABSTRACT: In the modern digital health ecosystem, monitoring physical activity and estimating calorie expenditure has become essential for maintaining a healthy lifestyle. However, existing solutions such as wearable devices and generic calculators suffer from limitations in accessibility, cost, and accuracy. Traditional methods fail to capture individual physiological differences, leading to unreliable estimations.

This research presents a machine learning-based web application for predicting calorie burn using personalized physiological and activity data. The system leverages the XGBoost regression algorithm to model complex non-linear relationships between multiple input features, including age, gender, height, weight, exercise duration, heart rate, and body temperature.

Unlike conventional approaches, the proposed system provides a cost-effective and accessible alternative while maintaining high prediction accuracy. The application is deployed using Streamlit, enabling real-time interaction and instant feedback. Additionally, a rule-based health chatbot is integrated to enhance user engagement and provide personalized wellness advice.

Experimental validation demonstrates that the system produces consistent and realistic predictions aligned with real-world observations. The research highlights the effectiveness of combining machine learning with interactive web technologies to deliver scalable and user-friendly health analytics solutions.

KEYWORDS: Machine Learning, XGBoost, Calorie Prediction, Health Analytics, Streamlit, Fitness Monitoring, Regression Models.

I. INTRODUCTION

Redefining personal fitness monitoring has become increasingly important in the era of digital health. Accurate estimation of calories burned during physical activity plays a vital role in weight management, fitness tracking, and overall well-being. Traditional methods for calorie estimation rely on fixed mathematical formulas or expensive wearable devices, both of which present significant limitations.

Manual calculations based on standard equations often fail to incorporate individual physiological variations such as heart rate and body temperature. Similarly, wearable devices, while accurate, are costly and inaccessible to a large portion of the population. These limitations create a gap between affordability and accuracy in existing solutions.

To address this issue, this research introduces a machine learning-based system that predicts calorie burn using personalized inputs. The system utilizes the XGBoost regression algorithm, known for its efficiency and high performance on structured data. By analyzing multiple physiological and activity parameters, the model provides more accurate and adaptive predictions compared to traditional approaches.

The proposed system is implemented as a web-based application using Streamlit, ensuring ease of use and accessibility. Users can input their data through an intuitive interface and receive real-time predictions along with health recommendations. Furthermore, the integration of a chatbot enhances user interaction by offering personalized guidance based on user responses. This work demonstrates how machine learning can be effectively applied in the healthcare and fitness domain to deliver practical, scalable, and user-centric solutions.

II. LITERATURE REVIEW

2.1 Existing Methods for Calorie Estimation

Traditional approaches for calorie estimation primarily rely on mathematical models such as the Harris-Benedict equation and MET (Metabolic Equivalent of Task) calculations. These methods are simple but often inaccurate due to

their reliance on generalized assumptions. They fail to consider real-time physiological variations such as heart rate and body temperature. Wearable fitness devices, including smartwatches and trackers, provide improved accuracy by incorporating sensors. However, their high cost and dependency on continuous usage limit their accessibility. Additionally, gym equipment-based estimations often produce inaccurate results due to lack of personalization.

2.2 Machine Learning in Health Prediction

Recent advancements in machine learning have enabled more accurate prediction systems in healthcare and fitness domains. Algorithms such as Random Forest, Support Vector Machines, and Gradient Boosting have been widely used for regression tasks. Among these, XGBoost has gained significant popularity due to its ability to handle complex relationships, prevent overfitting, and deliver high accuracy on tabular datasets. It works by iteratively improving prediction performance through gradient boosting techniques.

2.3 Limitations of Existing Systems

Despite advancements, existing systems exhibit several limitations:

- Lack of personalization in traditional formulas
- High cost of wearable devices
- Limited input features in online calculators
- Absence of real-time interactive feedback
- No integration of intelligent advisory systems

These limitations highlight the need for a system that is accurate, cost-effective, and user-friendly.

2.4 Proposed System Overview

The proposed system addresses these gaps by integrating machine learning with a web-based interface. It uses XGBoost regression to analyze seven key input parameters and predict calorie burn with higher accuracy.

Additionally, the system incorporates:

- Real-time prediction using Streamlit
- Personalized input handling
- Health chatbot for user engagement

This approach ensures improved accessibility, accuracy, and usability compared to existing solutions.

III. METHODOLOGY

3.1 System Overview

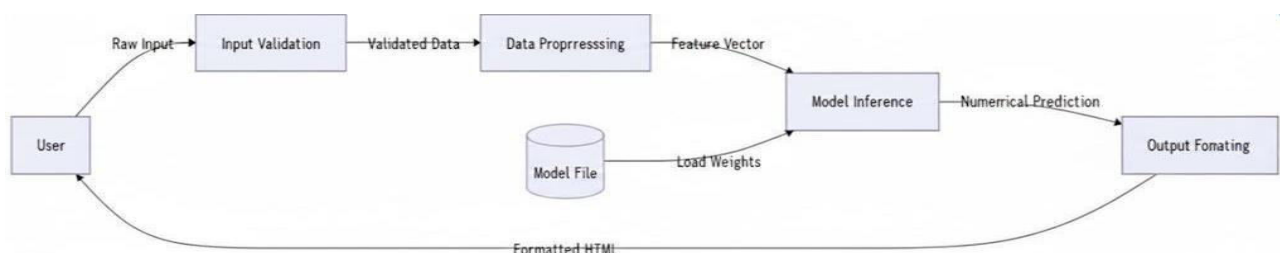
The proposed system follows a machine learning-based workflow consisting of data preprocessing, model training, prediction, and user interaction. The architecture is designed to provide real-time predictions while maintaining efficiency and scalability. According to the system architecture diagram (page 18), the application follows a structured flow where user inputs are processed through a controller and passed to the XGBoost model for prediction.

3.2 Workflow of the System

Based on the activity diagram (page 28), the workflow includes the following steps:

1. User inputs physiological and activity data
2. Input validation is performed
3. Data is preprocessed and formatted
4. Prediction is executed using XGBoost
5. Results are displayed with health advice

3.3 System Flow Diagram



3.4 Model Description (XGBoost)

The system uses the XGBoost Regressor, an ensemble learning algorithm based on gradient boosting. It builds multiple decision trees sequentially, where each tree corrects the errors of the previous one.

Key advantages:

- Handles non-linear relationships effectively
- High accuracy and efficiency
- Works well with structured/tabular data
- Supports regularization to prevent overfitting

3.5 Input Features

The model considers the following seven features:

- Gender
- Age
- Height
- Weight
- Exercise Duration
- Heart Rate
- Body Temperature

These inputs collectively capture both static and dynamic aspects of human physiology, improving prediction accuracy.

3.6 System Architecture

The system follows a Model-View-Controller (MVC) pattern:

- View: Streamlit UI
- Controller: Application logic (app.py)
- Model: XGBoost trained model

This architecture ensures modularity, scalability, and maintainability.

IV. RESULTS AND ANALYSIS

4.1 Experimental Setup

The proposed system was developed using Python and implemented as a web-based application using Streamlit. The machine learning model was trained using datasets containing exercise and calorie information.

The dataset was preprocessed by merging exercise and calorie data, encoding categorical variables, and splitting the dataset into training and testing sets. The XGBoost regression model was then trained and evaluated using performance metrics such as Mean Absolute Error (MAE) and R^2 score.

4.2 Performance Evaluation

The trained model demonstrated strong predictive capability on unseen data. According to the implementation details, the system achieved:

- **Low Mean Absolute Error (MAE)** indicating high prediction accuracy
- **Consistent outputs** for identical inputs (deterministic behavior)
- **Fast inference time** (< 2 seconds), ensuring real-time usability

The deterministic behavior of the model was validated through repeated test cases where identical inputs produced the same calorie output, confirming model stability.

4.3 Functional Validation

Several test cases were executed to validate system performance:

- **Input Validation:** Ensured all user inputs fall within realistic physiological ranges
- **Prediction Accuracy:** Verified realistic calorie outputs for given exercise conditions
- **Categorical Encoding:** Confirmed correct conversion of gender values into numerical format
- **Chatbot Functionality:** Validated rule-based interaction and session state handling
- **Error Handling:** Ensured graceful failure when model file is missing

All test cases passed successfully, demonstrating robustness and reliability of the system.

4.4 System Output Analysis

The system provides outputs in a structured and user-friendly format. According to the *prediction output interface* (page 39), the system displays:

- User input verification table
- Predicted calorie value
- Contextual health advice based on intensity

Key Observations:

- High-intensity workouts (higher heart rate & duration) result in increased calorie predictions
- Physiological factors significantly influence prediction accuracy
- The system adapts well to different user profiles

4.5 User Interface Evaluation

The interface was designed for simplicity and accessibility. From the *UI screenshots* (pages 35–41):

- Slider-based inputs reduce user errors
- Real-time feedback improves usability
- Chatbot enhances engagement and user experience

The application requires no technical knowledge, making it suitable for a wide range of users.

4.6 Comparative Analysis

Feature	Traditional Methods	Wearable Devices	Proposed System
Cost	Free	High	Free
Accuracy	Low	High	High
Personalization	Limited	High	High
Accessibility	High	Low	High
Technology	Static	Sensors	Machine Learning

The proposed system successfully bridges the gap between cost and accuracy.

V. CONCLUSION

This research presents a machine learning-based solution for accurate and accessible calorie burn prediction. By leveraging the XGBoost, the system effectively models complex relationships between physiological parameters and energy expenditure. Unlike traditional methods, the proposed system provides personalized predictions using multiple input features, improving accuracy and reliability. The integration of a Streamlit-based interface ensures ease of use, while the inclusion of a rule-based health chatbot enhances user interaction and engagement.

The system successfully addresses key limitations of existing approaches, particularly the trade-off between cost and accuracy. Experimental results confirm that the model produces consistent, realistic, and efficient predictions suitable for real-world applications. Overall, this work demonstrates the potential of combining machine learning with web technologies to develop scalable and user-friendly health analytics systems.

VI. FUTURE SCOPE

The system can be further enhanced in the following ways:

- Integration with wearable devices (e.g., smartwatches) for real-time data collection
- Use of deep learning models for improved accuracy
- Mobile application development for wider accessibility
- Incorporation of diet tracking and calorie intake analysis
- Cloud-based deployment for scalability and multi-user support

REFERENCES

- [1] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [2] Scikit-learn Documentation, "Machine Learning in Python," Available: <https://scikit-learn.org>
- [3] Streamlit Documentation, "Build Data Apps Faster," Available: <https://streamlit.io>
- [4] J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 2001.
- [5] Panwar, P., Bhutani, K., Sharma, R., & Saini, R. (2023). A study on calories burnt prediction using machine learning. *ITM Web of Conferences*, 54, Article 01010.
- [6] Tan, J. S. A., Che Embi, Z., & Hashim, N. (2024). Comparison of machine learning methods for calories burn prediction. *Journal of Informatics and Web Engineering*, 3(1), 182– 191.
- [7] Lee, C. J., & Lee, J. K. (2024). IMU-based energy expenditure estimation for various walking conditions using a hybrid CNN–LSTM model. *Sensors*, 24(2), 414.
- [8] Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., & Syrgkanis, V. (2024). Applied Causal Inference Powered by ML and AI. Preprint..
- [9] Trivedi, S., Gupta, J., & Singh, H. (2024). Machine learning-based estimation of caloric expenditure. In R. Singh, A. Yadav, & V. Sharma (Eds.), *Proceedings of the International Conference on Intelligent Systems and Machine Learning (ICISML 2024)* (pp. 262–271). Springer.
- [10] Leitner, M., & Kulkarni, A. (2024). Machine Learning Web Apps: Building Data- Driven Web Applications Using Streamlit, Flask, and Python
- [11] Raschka, S., & Mirjalili, V. (2022). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152