International Journal Software Engineering and Computer Science (IJSECS)
6 (3), 2026, 1271-1282
Published Online December 2026 in IJSECS (http://www.journal.lembagakita.com/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v6i3.8847.

RESEARCH ARTICLE Open Access

Comparative Evaluation of Machine Learning Regression Models for Predicting Calories Burned from Exercise Data

Pandu Satria Nugroho [1*], **Alz Danny Wowor** [2]

[1*,2] Department of Informatics Engineering, Faculty of Information Technology, Universitas Kristen Satya Wacana, Salatiga City, Central Java Province, Indonesia.

Article History:

Received: September 29, 2026.
Revised: October 7, 2026.
Accepted: November 21, 2026.
Available online: November 30, 2026.
Published: December 1, 2026.

***Corresponding Author:**
panduhid@gmail.com

Abstract: This study evaluates machine learning models for estimating calorie expenditure from tabular exercise data. The analysis utilizes 20,000 records obtained from the Kaggle Calorie Burn ML dataset, incorporating eight fundamental demographic and exercise-related predictors: age, gender, weight, height, resting heart rate, average heart rate, maximum heart rate, and session duration. To ensure methodological rigor and prevent target leakage, the predictor set strictly excludes variables derived directly from calories burned as well as all post-hoc heuristic outputs. The dataset was partitioned into training and testing subsets using an 80:20 ratio, and feature standardization was fitted exclusively on the training subset to prevent data contamination. Five conventional regression algorithms were comprehensively evaluated: linear regression, ridge regression, lasso regression, random forest regression, and support vector regression. Experimental results demonstrate that lasso regression achieved a coefficient of determination of 0.6602, root mean square error of 291.12, mean absolute error of 231.11, and mean absolute percentage error of 19.62% on the held-out test set, with linear regression and ridge regression producing nearly identical outcomes. Conversely, random forest and support vector regression obtained lower coefficient values of 0.6139 and 0.4857, respectively. Furthermore, fatigue score and recovery time were deliberately excluded from empirical evaluation because the supplied dataset lacks independent observed target labels for these outcomes. Consequently, the findings describe the prediction accuracy relative to the dataset-provided calorie label and do not constitute direct physiological validation against measured energy expenditure.

Keywords: Machine Learning; Calorie Expenditure; Regression Analysis; Exercise Dataset; Predictive Modeling.

© The Author(s) 2026, corrected publication 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license unless stated otherwise in a credit line to the material. Suppose the material is not included in the article's Creative Commons license, and your intended use is prohibited by statutory regulation or exceeds the permitted use. In that case, you must obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

1. Introduction

Physical activity has become an essential component of a healthy lifestyle as public awareness of health, fitness, and disease prevention continues to increase worldwide. The growing popularity of recreational sports, marathon events, fun runs, cycling, gym training, and High-Intensity Interval Training (HIIT) encourages individuals to exercise regularly to improve cardiovascular health, muscular strength, endurance, and overall quality of life. In parallel with this trend, wearable devices and digital fitness applications have become increasingly accessible, enabling users to monitor physiological variables such as heart rate, exercise duration, energy expenditure, and physical activity intensity. Despite these technological advances, accurately estimating calories burned, fatigue level, and appropriate recovery time remains challenging because physiological responses differ considerably among individuals. Inadequate recovery after exercise may lead to fatigue accumulation, reduced athletic performance, delayed muscle adaptation, overtraining, and an increased risk of injury (Arent *et al.*, 2024; Goodall *et al.*, 2020; James *et al.*, 2022). Recovery is a fundamental phase of physical training because physiological systems require sufficient time to restore normal metabolic, cardiovascular, and neuromuscular functions following exercise. The duration of recovery is influenced by numerous physiological and demographic factors, including age, gender, body weight, body height, resting heart rate, exercise intensity, and exercise duration. Previous studies have shown that heart rate variability can be used as an effective indicator for evaluating training load, fatigue, and recovery status (Weippert *et al.*, 2022; Vaara *et al.*, 2014). Furthermore, accumulated fatigue and inadequate recovery have been associated with decreased physical performance, impaired neuromuscular function, and an increased likelihood of sports-related injuries (Minett & Duffield, 2018; Bowman *et al.*, 2024). Studies on resistance exercise, sprint exercise, and HIIT have also demonstrated that recovery characteristics differ according to exercise intensity and individual physiological adaptation (Hanová *et al.*, 2020; Goodall *et al.*, 2020; Bongiovanni *et al.*, 2020; James *et al.*, 2022). These findings indicate that generalized recovery recommendations are often insufficient and highlight the need for personalized recovery estimation.

Recent advances in machine learning and deep learning have significantly improved predictive analytics in sports science and digital health. Unlike conventional statistical approaches, machine learning and deep learning algorithms are capable of learning complex nonlinear relationships among physiological variables and producing accurate predictions from large datasets. Physiological attributes such as resting heart rate, average heart rate, maximum heart rate, age, gender, body weight, height, and exercise duration provide valuable information for predicting exercise outcomes. Consequently, predictive models have become increasingly important for intelligent sports monitoring systems, enabling data-driven decision-making for athletes, coaches, fitness practitioners, and healthcare professionals. Several previous studies have successfully applied artificial intelligence techniques for exercise-related prediction tasks. Ogunmiloro (2026) proposed an artificial intelligence-based framework for calorie burn prediction, demonstrating the potential of predictive modeling for estimating energy expenditure. Tan *et al.* (2024) compared several machine learning algorithms for calorie prediction and reported promising predictive performance across multiple regression models. Tejaswini and Rabindranath (2025) further integrated calorie prediction into an AI-based fitness chatbot to provide personalized recommendations. These studies demonstrate that machine learning algorithms can effectively estimate calorie expenditure and support intelligent fitness applications. Deep learning techniques have also been widely adopted in sports analytics. Chen (2025) proposed a hybrid Transformer-LSTM model for athlete performance prediction, while Jia and Liang (2026) investigated real-time sports performance prediction using machine learning algorithms. In addition, Xu *et al.* (2024) demonstrated that wearable sensor data can be utilized to model physiological stress and recovery patterns. Collectively, these studies indicate that physiological data collected from wearable technologies can be effectively analyzed to estimate exercise outcomes and improve decision support in sports science.

Although previous studies have examined calorie expenditure, fatigue monitoring, and recovery assessment, these tasks require different outcome measurements. A model can only be evaluated for an outcome when the dataset contains an independently observed target label. In the dataset used in this study, Calories_Burned is available as a target column, whereas no independently measured fatigue score or recovery time labels are provided. Consequently, the empirical scope of this work is restricted to calories burned prediction; fatigue and recovery are discussed only as related research topics and are not claimed as model outputs. This study develops a leakage-controlled regression workflow for estimating Calories_Burned from eight original predictors: age, gender, body height, body weight, resting heart rate, average heart rate, maximum heart rate, and exercise duration. Five conventional regression algorithms are compared under the same 80:20 train-test partition. Preprocessing is performed after the data split, and standard scaling is fitted using training data only before being applied to the held-out test data. The contribution of this revised study is a transparent re-evaluation of calorie estimation after auditing the original feature construction for target leakage. Features that directly incorporate Calories_Burned, randomly generated auxiliary values, and post-hoc fatigue or recovery calculations are excluded from the predictive input. The study also documents why

Copyright © 2026 IJSECS

long short-term memory networks are not retained as a benchmark for this cross-sectional dataset. This narrower scope avoids presenting calculated indicators as independently learned or validated physiological outcomes. The objective of this study is to evaluate and compare linear regression, ridge regression, lasso regression, random forest regression, and support vector regression for predicting the dataset-provided Calories_Burned label. Model performance is assessed using the coefficient of determination, root mean square error, mean absolute error, and mean absolute percentage error on an unseen test subset. The study does not attempt to predict fatigue score or recovery time because independent target measurements for these variables are unavailable in the supplied dataset.

2. Related Work

Machine learning has been applied in sports analytics to estimate energy expenditure and to study related physiological responses such as fatigue and recovery. This review summarizes work on these related topics; however, the empirical experiment in this paper is limited to Calories_Burned because the supplied dataset does not contain independent target labels for fatigue or recovery.

2.1 State-of-the-Art Research

Calories Burned Prediction: Several studies have investigated calories burned prediction using machine learning algorithms. Ogunmiloro (2026) reported that XGBoost outperformed Ridge regression and random forest with an R-squared value of approximately 0.95 (Ogunmiloro, 2026). Similarly, Tan *et al*. (2024) found that LightGBM achieved the lowest mean absolute error among several regression models (Tan *et al*., 2024). Tejaswini and Rabindranath (2025) further integrated gradient boosting with an AI fitness chatbot to provide personalized exercise recommendations (Tejaswini & Rabindranath, 2025). Although these approaches achieved high prediction accuracy, they focused only on calorie estimation without considering fatigue or recovery prediction. Fatigue Prediction: Recent studies have applied deep learning models to predict fatigue using physiological time-series data. Chen (2025) proposed a hybrid Transformer-LSTM model that improved fatigue prediction accuracy, while Jia and Liang (2026) employed long short-term memory networks for real-time fatigue monitoring (Chen, 2025; Jia & Liang, 2026). Despite their promising results, these studies mainly targeted athlete monitoring and did not integrate calorie expenditure or recovery estimation into a single prediction framework.

Recovery Time Prediction: Recent developments in artificial intelligence also demonstrate the potential of machine learning for personalized exercise monitoring. Ogunmiloro (2026), Tan *et al*. (2024), and Tejaswini and Rabindranath (2025) applied machine learning approaches for calorie expenditure prediction, showing that physiological and activity-related variables can be used to estimate exercise outcomes. In addition, Chen (2025) and Jia and Liang (2026) demonstrated the application of Transformer-LSTM and machine learning algorithms for athlete performance prediction, supporting the use of predictive models in sports management systems. Additional studies further support the importance of fatigue and recovery prediction in sports analytics. Arent *et al*. (2024) emphasized that fatigue and recovery monitoring should integrate physiological and psychological indicators to obtain a more comprehensive assessment of athlete readiness. Bowman *et al*. (2024) showed that excessive training load is associated with nonfunctional overreaching, highlighting the importance of monitoring training stress. Furthermore, Goodall *et al*. (2020), Bongiovanni *et al*. (2020), and James *et al*. (2022) demonstrated that exercise intensity, particularly repeated sprint and high-intensity interval training, influences neuromuscular fatigue and recovery characteristics. These findings indicate that fatigue and recovery prediction requires consideration of multiple physiological variables rather than relying on a single indicator.

Recovery prediction has mainly relied on physiological indicators, particularly heart rate variability. Weippert *et al*. (2022) identified heart rate variability as an effective indicator of recovery, whereas Xu *et al*. (2024) demonstrated the potential of wearable sensors for personalized recovery assessment. Arent *et al*. (2024) further emphasized combining physiological and psychological indicators to evaluate athlete readiness (Weippert *et al*., 2022; Xu *et al*., 2024; Arent *et al*., 2024). Nevertheless, these studies did not integrate recovery estimation with calorie and fatigue prediction using predictive machine learning models. The reviewed studies indicate that calorie expenditure, fatigue, and recovery are related but distinct prediction tasks. The present study addresses only calorie expenditure, for which a target column is available in the selected dataset. Fatigue and recovery studies provide context for possible future work, but their findings are not treated as validation of outputs generated by the current experiment.

Copyright © 2026 IJSECS

2.2 Comparison with Previous Studies

Table 1 summarizes representative studies relevant to calorie expenditure, fatigue, and recovery. The comparison is provided to position the present calorie-estimation experiment within the broader exercise analytics literature; it is not used to claim that all three outcomes are modeled in this study.

Table 1. Comparison of Previous Studies

Authors	Focus	Method	Limitation
Ogunmiloro (2026)	Calories Burned	XGBoost, RF, Ridge	Calories only
Tan *et al*. (2024)	Calories Burned	LightGBM	No fatigue or recovery
Tejaswini & Rabindranath (2025)	Calories Burned	Gradient Boosting	Single prediction target
Chen (2025)	Fatigue	Hybrid Transformer–LSTM	Athlete-oriented
Jia & Liang (2026)	Fatigue	LSTM	No recovery estimation
Weippert *et al*. (2022)	Recovery	HRV	No predictive ML model
Xu *et al*. (2024)	Recovery	Wearable Sensors	Not integrated
Arent *et al*. (2024)	Recovery	Physiological Assessment	No integrated framework

2.3 Positioning of the Proposed Research

The reviewed literature motivates data-driven analysis of exercise-related outcomes. The present research focuses on Calories_Burned prediction using a fixed set of original demographic and exercise variables, with an explicit audit for target-derived predictors. Unlike the previous implementation, fatigue and recovery calculations are not reported as machine-learning predictions because no independent labels are available for their evaluation.

2.4 Review of Technologies, Frameworks, and Algorithms

This study compares five regression algorithms: linear regression, ridge regression, lasso regression, random forest regression, and support vector regression. Linear models provide interpretable baselines, ridge and lasso apply regularization, random forest represents a tree-based ensemble approach, and support vector regression models nonlinear relationships through a kernel function. Long short-term memory is not included in the experiment because the dataset consists of independent tabular records rather than temporally ordered sequences.

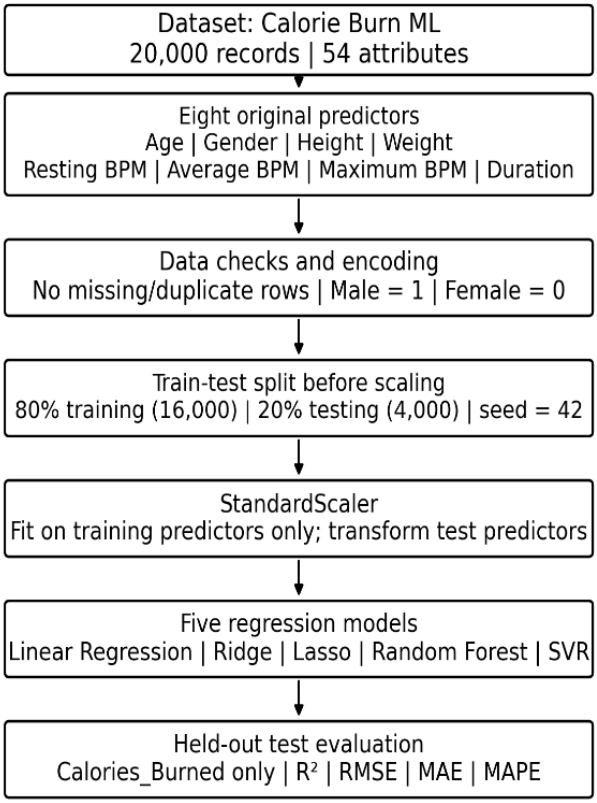
3. Methodology

This study employs a quantitative experimental approach to evaluate regression models for the Calories_Burned target in the Kaggle Calorie Burn ML dataset. The workflow consists of dataset inspection, selection of eight original predictors, train-test splitting, categorical encoding, training-only feature scaling, model fitting, and held-out performance evaluation. Eight original predictor variables are used, and Calories_Burned is the only modeled target. Features derived from the target, randomly generated auxiliary values, and post-hoc fatigue or recovery formulas are excluded to prevent leakage and unsupported claims.

3.1 Research Framework

Figure 1 presents the revised research workflow. The dataset is first inspected and restricted to eight original predictors and the Calories_Burned target. The data are then split into training and testing subsets before feature scaling. StandardScaler parameters are learned from the training subset and reused to transform the test subset. Five regression models are trained and evaluated on the same held-out records. Fatigue score and recovery time are outside the experimental scope because independent observed labels are not available.

Copyright © 2026 IJSECS



Fatigue Score and Recovery Time: no independent observed labels.
LSTM: no valid temporal sequences in the supplied records.
Figure 1. Research Framework of the Proposed Prediction Model

3.2 Dataset Description

The dataset contains demographic, anthropometric, and exercise-related attributes. For reproducibility, the revised experiment uses only the variables listed in Table 3 as model inputs. The supplied Calories_Burned column is treated as the regression label, subject to the limitation that its measurement provenance is not documented in the accompanying dataset and notebook. Dataset Source: The dataset was obtained from Kaggle under the name Calorie Burn ML. The supplied CSV file contains 20,000 records and 54 columns with numerical and categorical fields. The available notebook identifies Calories_Burned as the primary target. However, the supplied files do not document the measurement protocol, calorimetry equipment, or wearable-device validation used to establish this label. Accordingly, the present study evaluates agreement with the dataset-provided label and does not describe it as verified physiological ground truth. The original dataset includes several additional exercise, nutrition, and derived attributes. To maintain a clearly defined prediction setting, only age, gender, height, weight, resting BPM, average BPM, maximum BPM, and session duration are used as predictors. All other columns are omitted from model input, including variables whose provenance or relationship to the target is not sufficiently documented. Dataset Characteristic: The Calorie Burn ML dataset consists of 20,000 records and 54 attributes. It contains numerical and categorical variables describing participant characteristics, exercise sessions, and other associated information. Calories_Burned is the only target used in this regression experiment. Table 2 summarizes the dataset characteristics relevant to the study.

Table 2 Characteristics of the Calorie Burn ML Dataset

Characteristics	Value
Dataset name	Calorie Burn ML
Source	Kaggle
Number of records	20,000
Number of attributes	54
Data type	Numerical and categorical
Target variable	Calories_Burned (regression)
Workout types in source data	HIIT, cardio, strength, yoga

Copyright © 2026 IJSECS

Although the source file contains additional attributes, the revised model uses only the eight original predictors listed in Table 3. No fatigue or recovery variables are constructed and then reused as model targets or predictors.

3.3 Feature Description

The selected inputs were chosen to represent demographic, anthropometric, and heart-rate/activity information that is available before the target is evaluated. The target Calories_Burned is kept separate from the predictor matrix. Table 3 lists the exact model variables and their units.

Table 3. Description of Features Used in This Study

Feature	Data Type	Description
Age	Numerical	Age of the participant (years)
Gender	Categorical	Gender encoded as Male = 1 and Female = 0
Weight (kg)	Numerical	Body weight of the participant in kilograms
Height (m)	Numerical	Body height of the participant in meters
Resting_BPM	Numerical	Heart rate before exercise (beats per minute)
Avg_BPM	Numerical	Average heart rate during exercise (beats per minute)
Max_BPM	Numerical	Maximum heart rate during exercise (beats per minute)
Session_Duration (hours)	Numerical	Duration of the exercise session in hours
Calories_Burned	Numerical (target)	Dataset-provided calorie expenditure label

The eight variables above form the complete predictor set for all five models. Calories_Burned is used only as the dependent variable and is never included in any feature transformation or predictor column.

3.4 Data Preprocessing

Preprocessing comprises data-quality checks, binary encoding of gender, an 80:20 train-test split, and feature standardization. The data split is performed before fitting the scaler so that information from the test subset cannot influence the training transformation. Data Cleaning: The supplied CSV file was checked for duplicate rows and inconsistent data types in the selected columns. No duplicate rows were found in the 20,000-record file. The selected predictor fields and target were numeric or could be encoded as numeric values. Missing Value Handling: Missing-value inspection found no missing entries in the supplied 54-column dataset. Therefore, no imputation or row deletion was required for the revised experiment. Label Encoding: Gender was converted to a binary variable, with Male encoded as 1 and Female encoded as 0. The mapping was applied consistently to the training and testing records before scaling. Feature Scaling: StandardScaler was used for all eight predictors. The data were first divided into training and testing subsets; the scaler was fitted only on the 16,000 training records. The resulting training mean and standard deviation were then applied unchanged to the 4,000 test records. Min-Max scaling was not used in the revised experiment. The training-only StandardScaler transformation is defined as follows:

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

Where:
z= standardized feature value,
x= original feature value,
μ_{train}= mean of the feature calculated from training data only,
σ_{train}= standard deviation of the feature calculated from training data only.

The test data are transformed using the training-set parameters without refitting the scaler. This procedure prevents the test distribution from influencing the preprocessing stage.

3.5 Predictor Selection and Target Leakage Control

The original notebook generated multiple derived variables, including Calories_Per_Hour, Calories_Per_BPM, and Calorie_Efficiency. These quantities directly use Calories_Burned in their calculation. In the original modeling matrix, Calories_Per_BPM and Calorie_Efficiency remained among the 69 numeric predictors because they were not removed by the exclusion list. The feature-importance screening also used target-derived quantities. Such variables expose the target to the model and make the resulting test metrics invalid. The revised experiment therefore discards all engineered variables and uses only the eight original predictors specified in Table 3. It also excludes Rest_Hours because the notebook generated it through a random rule rather than an observed recovery measurement. Fatigue score and recovery time are not created

Copyright © 2026 IJSECS

as learning targets; their previous values were deterministic or heuristic post-hoc calculations and cannot be validated against the supplied data.

3.6 Model Development

Linear Regression: Linear regression was used as a baseline because it is computationally efficient and provides an interpretable estimate of the linear association between the selected predictors and Calories_Burned. Ridge Regression: Ridge regression applies L2 regularization to constrain coefficient magnitudes. The regularization parameter was set to $\alpha = 1.0$in the revised experiment. Lasso Regression: Lasso regression applies L1 regularization, which can shrink less informative coefficients toward zero. The regularization parameter was set to $\alpha = 1.0$. Random Forest Regression: Random forest regression combines decision trees and can represent nonlinear relationships. The model was trained with 200 trees and random_state = 42. Support Vector Regression: Support vector regression with a radial basis function kernel was included to evaluate a nonlinear margin-based regression approach. The settings were $C = 1.0$, $\epsilon = 0.1$, and $\gamma =$ scale. Long Short-Term Memory (LSTM) Exclusion: Long short-term memory networks were not included in the revised comparison. The supplied data contain independent tabular records and no participant-level time ordering or repeated time points. The original notebook reshaped each row to an input tensor with one timestep and eight features (samples, 1, 8). A sequence length of one does not provide temporal dependencies for the recurrent layer to learn; it therefore does not constitute a meaningful time-series experiment. Excluding long short-term memory is more methodologically appropriate than reporting its earlier result as a comparable benchmark.

3.7 Performance Evaluation

R-Squared Score (R^2): The coefficient of determination measures the proportion of variation in the observed target explained by model predictions on the test set. An R^2value of 1 indicates a perfect fit, 0 indicates performance equivalent to predicting the test-set mean, and a negative value indicates performance worse than that baseline.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

Root Mean Square Error (RMSE): The root mean square error is used to measure the average magnitude of prediction errors by calculating the square root of energy differences between actual and predicted values.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

Mean Absolute Error (MAE): The mean absolute error measures the average magnitude of prediction errors by calculating the mean of absolute differences between actual and predicted values.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (4)$$

Mean Absolute Percentage Error (MAPE): The mean absolute percentage error measures the average percentage difference between actual and predicted values.

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (5)$$

4. Results and Discussion

4.1 Results

4.1.1 Dataset Preparation and Model Implementation

The revised experiment uses age, gender, weight, height, resting BPM, average BPM, maximum BPM, and session duration as predictors. Calories_Burned is the sole target variable. Table 4 summarizes the definitions and units of these variables.

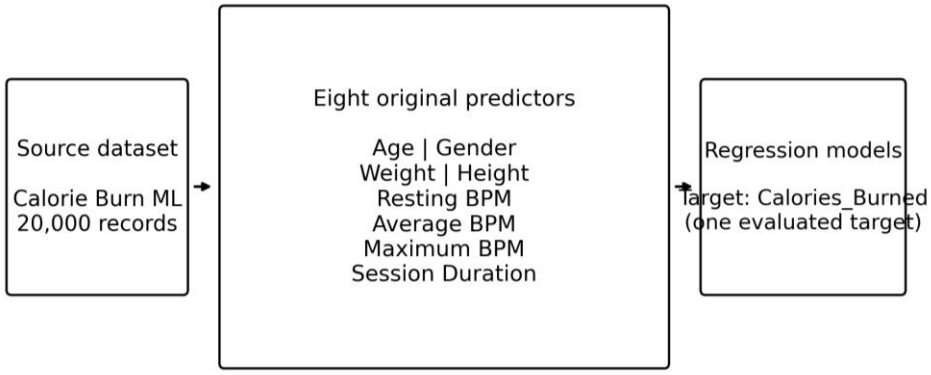
Table 4. Research Variables

Feature	Data Type	Description

Copyright © 2026 IJSECS

Age	Numerical	Age of the participant (years)
Gender	Categorical	Gender encoded as Male = 1 and Female = 0
Weight (kg)	Numerical	Body weight of the participant in kilograms
Height (m)	Numerical	Body height of the participant in meters
Resting BPM	Numerical	Heart rate before exercise (beats per minute)
Average BPM	Numerical	Average heart rate during exercise (beats per minute)
Max BPM	Numerical	Maximum heart rate during exercise (beats per minute)
Duration (hours)	Numerical	Duration of the exercise session in hours
Calories Burned	Numerical (target)	Dataset-provided calorie expenditure label

The selected columns contained no missing values or duplicate records. Gender was encoded using the binary mapping described in Section 3. The dataset was split into 16,000 training records and 4,000 testing records using an 80:20 ratio and random_state = 42. StandardScaler was fitted on the training predictors only and then applied to the test predictors. No Min-Max scaling was applied.



Fatigue Score and Recovery Time are excluded because independent observed labels are unavailable.

Figure 2. Predictor and Target Scope of the Revised Study

A leakage audit was performed before modeling. The variables shown in Table 5 were excluded because they directly incorporate the target, are generated by random or heuristic rules, or lack independent outcome measurements. The revised feature matrix therefore contains only the eight original predictors listed in Table 4.

Table 5. Variables Excluded After the Target Leakage Audit

Excluded Variable(s)	Reason for Exclusion
Calories_Per_BPM; Calorie_Efficiency	Directly calculated from Calories_Burned and retained in the original 69-feature model matrix; target leakage.
Calories_Per_Hour; Log_Calories	Direct transformations of Calories_Burned; not valid predictors for predicting that target.
Rest_Hours	Generated through a random BPM-based rule in the notebook; it is not an observed recovery measurement.
Fatigue Score	Calculated after prediction from normalized BPM values; no independent fatigue label is present for training or validation.
Recovery Time	Generated by a post-hoc heuristic using predicted calories and user inputs; no independently observed recovery-time label is available.

4.1.2 Testing Results

The dataset was partitioned into training and testing subsets using an 80:20 ratio. The training subset contained 16,000 observations and the test subset contained 4,000 observations. All five regression algorithms used the same partition and the same eight predictors. StandardScaler was fitted on the training subset only. Model-specific parameters are summarized in Table 7.

Table 6. Algorithms Used in the Study

Algorithm	Category
Linear Regression	Machine learning
Ridge Regression	Machine learning

Copyright © 2026 IJSECS

Lasso Regression	Machine learning
Random Forest Regression	Machine learning
Support Vector Regression (SVR)	Machine learning

The training and evaluation settings are summarized in Table 7. Figure 1 presents the revised experimental workflow, from predictor selection and train-test splitting to model evaluation. Long short-term memory is not part of the revised experiment because the dataset does not contain temporal sequences.

Table 7. Training Parameters

Parameter	Value
Training data	80% (16,000 records)
Testing data	20% (4,000 records)
Random state	42
Input features	8 original predictors
Feature scaling	StandardScaler; fit on training data only
Ridge regression	alpha = 1.0
Lasso regression	alpha = 1.0
Random forest	n_estimators = 200; random_state = 42
SVR	RBF kernel; C = 1.0; epsilon = 0.1; gamma = scale

4.1.3 Performance Evaluation

Model performance was evaluated on the 4,000 held-out test records using R^2, RMSE, MAE, and MAPE. The reported errors are calculated from the unrounded test predictions; values in Table 8 are shown to six decimal places. MAPE is applicable here because all target values in the supplied dataset are positive.

Table 8. Performance Comparison of Calories_Burned Prediction Models

Model	R²	RMSE	MAE	MAPE (%)
Linear Regression	0.660121	291.135777	231.217512	19.635562
Ridge Regression	0.660122	291.135212	231.215439	19.635357
Lasso Regression	0.660163	291.117649	231.111289	19.624093
Random Forest	0.613915	310.295012	244.089860	20.569086
SVR	0.485730	358.120344	276.915822	24.076665

Lasso regression produced the highest test R^2(0.660163) and the lowest error values (RMSE = 291.117649, MAE = 231.111289, and MAPE = 19.624093%). However, its advantage over linear regression ($R^2 = 0.660121$) and ridge regression ($R^2 = 0.660122$) is very small. The results show moderate predictive performance rather than a perfect fit. Random forest achieved $R^2 = 0.613915$, while support vector regression achieved $R^2 = 0.485730$.

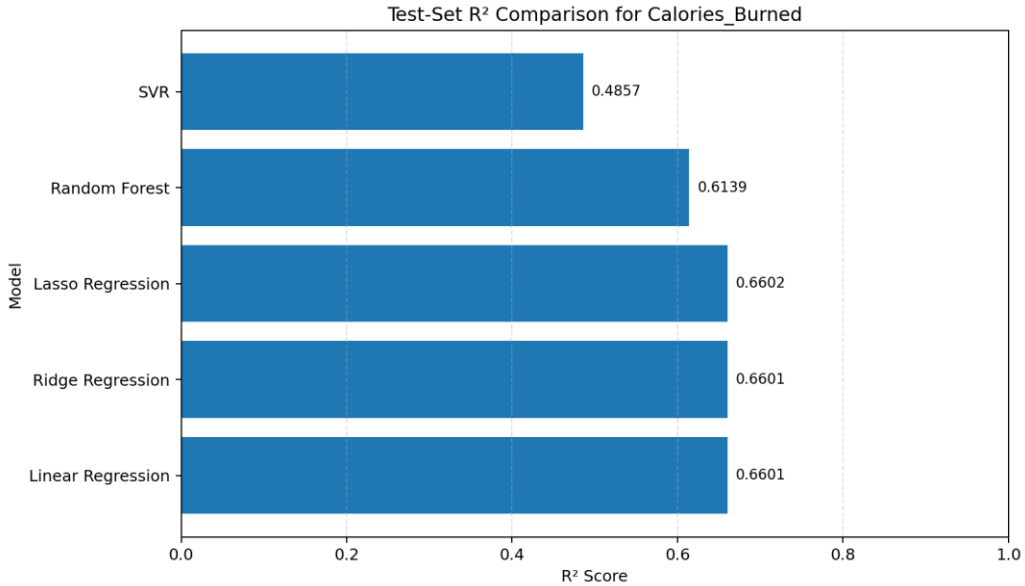


Figure 3. Test-Set R² Scores for Calories_Burned Prediction Models

Copyright © 2026 IJSECS

Figure 4 compares MAPE across the five models. The error values are materially higher than those in the original manuscript because the revised experiment removes target-derived variables and evaluates only the eight prespecified original predictors. No zero-error result is observed after this leakage-control audit.

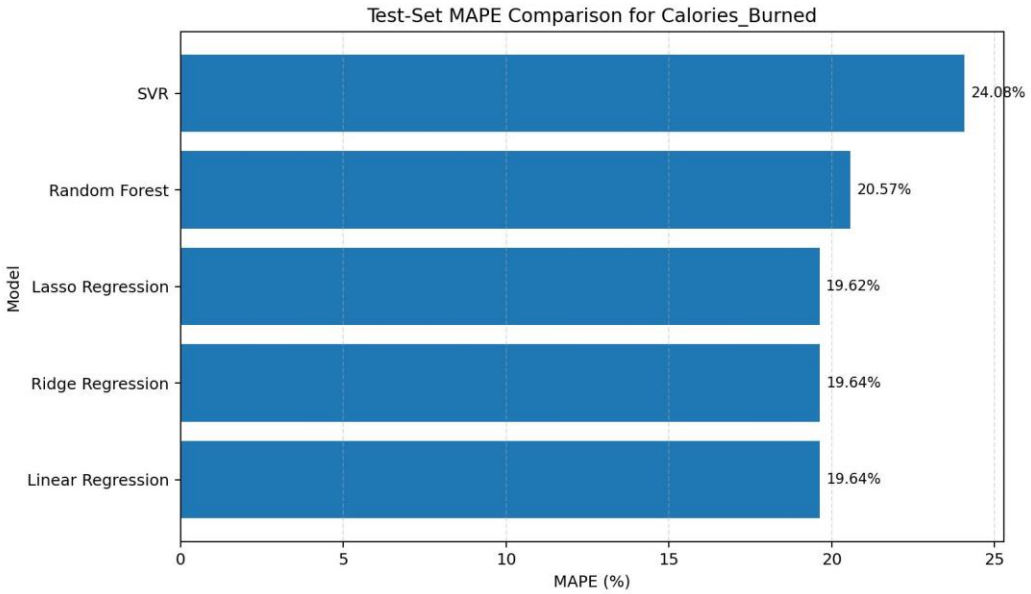


Figure 4. Test-Set MAPE for Calories_Burned Prediction Models

4.1.4 Target alvailability and Validation Scope

The source dataset provides a Calories_Burned column but does not provide independent observed labels for fatigue score or recovery time. These variables therefore cannot be trained or evaluated as regression targets using the supplied files. Table 9 summarizes the validation scope adopted in the revised study.

Table 9. Availability and Validation Status of Study Outcomes

Outcome	Availability and Status
Calories_Burned	Available as a target column in the source CSV; modeled and evaluated on a held-out test set. Measurement provenance is not documented in the supplied files.
Fatigue Score	No independent observed target label; previous values were heuristic calculations and are not reported as model predictions.
Recovery Time	No independent observed target label; previous values were post-hoc heuristic estimates and are not reported as model predictions.

Tan *et al*. (2024) compared machine learning methods for calorie prediction and provides a relevant context for the present regression task. The revised results, however, should be interpreted according to the feature and label limitations of the supplied Calorie Burn ML data. The present evaluation uses eight original predictors and does not claim validation against direct calorimetry or another independent measurement source. The five-model comparison shows that linear and regularized linear regressors perform similarly on the held-out records. Lasso has the smallest numerical errors, but the differences relative to linear regression and ridge regression are marginal. This finding does not establish that calorie expenditure has a universally linear physiological relationship; it describes the pattern learned from the supplied dataset and the selected predictors. Long short-term memory-based sports analytics studies provide context for applications that contain sequential observations. The present dataset does not include participant-specific time series or timestamps for constructing sequences. Therefore, long short-term memory was excluded rather than reshaping independent rows into one-step inputs and interpreting the resulting score as evidence of temporal learning. Research on fatigue and recovery remains relevant to exercise monitoring, as discussed in the literature review. Nevertheless, the current dataset contains no independent fatigue or recovery target measurements. The present work therefore makes no empirical claim about the accuracy of either outcome and identifies their measurement and validation as directions for future research.

4.2 Discussion

The revised experiment evaluates only the Calories_Burned label. The original notebook produced perfect or near-perfect scores using a 69-feature matrix that retained Calories_Per_BPM and Calorie_Efficiency, both calculated directly from the target. Calories_Per_Hour and Log_Calories were also generated during feature

Copyright © 2026 IJSECS

construction, and the feature-importance screening included target-derived quantities. These relationships make the original zero-error linear regression result unsuitable as evidence of generalization. Removing target-derived variables and restricting the inputs to the eight variables listed in Table 4 yields a test R^2 of approximately 0.66 and non-zero errors. The revised results indicate moderate agreement with the dataset-provided Calories_Burned labels. Lasso, linear regression, and ridge regression have almost identical performance, while random forest and support vector regression produce lower R^2values on the same held-out records. These metrics assess prediction of the provided label only. Because the supplied files do not explain how Calories_Burned was measured or generated, the results cannot be treated as validation against physiological ground truth.

Long short-term memory was excluded from the revised model comparison because the source data are cross-sectional and the original one-timestep reshape does not encode temporal dependencies. This design choice avoids comparing a recurrent model on an artificial sequence with algorithms designed for tabular observations. The revised experiment consequently contains five conventional regression models. Fatigue score and recovery time are not presented as predictions or practical recommendations. In the original notebook, fatigue score was calculated from normalized BPM values using manually selected weights and thresholds, while recovery time was generated by a post-hoc interpolation and adjustment formula based partly on predicted calories. These rules do not provide independent target observations, and the category boundary for normal fatigue was not validated. Such calculations are therefore excluded from the empirical findings. This study has several limitations. It relies on one Kaggle tabular dataset, and the provenance of the Calories_Burned label is not documented in the supplied files. The evaluation uses a single random 80:20 split rather than external validation, and it does not incorporate direct calorimetry, longitudinal wearable streams, or independently assessed fatigue and recovery outcomes. Future work should use documented biometric measurements, participant-level temporal sequences, external test cohorts, and validated fatigue and recovery labels before making broader physiological or practical claims.

5. Conclusion and Recommendations

This study re-evaluated machine learning regression for estimating Calories_Burned using the Kaggle Calorie Burn ML dataset. The revised experiment used eight original predictors: age, gender, height, weight, resting BPM, average BPM, maximum BPM, and session duration. Five algorithms were evaluated using an 80:20 train-test split, with StandardScaler fitted on the training subset only. Target-derived features and post-hoc heuristic outputs were excluded from the modeling pipeline. On the held-out test subset, lasso regression obtained $R^2 = 0.6602$, RMSE = 291.12, MAE = 231.11, and MAPE = 19.62%. Linear regression and ridge regression yielded nearly identical results, while random forest and support vector regression obtained lower R^2scores of 0.6139 and 0.4857, respectively. The earlier perfect linear regression result was not retained because the original feature matrix contained variables calculated directly from Calories_Burned, which caused target leakage. The revised contribution is a more defensible comparison of five regression algorithms after explicit feature auditing and training-only preprocessing. The results represent prediction of the supplied Calories_Burned column, not validated measurement of actual energy expenditure. Fatigue score and recovery time are not claimed as predicted outcomes because the dataset does not include independent observed labels for these variables. Long short-term memory was also excluded because the cross-sectional records do not provide a valid temporal sequence. The study is limited by the single-source tabular dataset, undocumented label provenance, and the use of one random train-test split. Future research should validate calorie estimates against documented physiological measurements, use external datasets, and collect longitudinal sensor data. Fatigue and recovery modeling should be attempted only after independent outcome labels and validated assessment protocols are available. These steps are necessary before the models can support practical fitness or recovery recommendations.

References

Arent, S. M., Walker, A. J., Sanders, D. J., & Stevenson, J. L. (2024). Monitoring fatigue and recovery in athletes: Integrating physiological and psychological measures. *Sports Medicine*, *54*(3), 455–472.

Bongiovanni, T., Genovesi, F., *et al*. (2020). Acute and delayed neuromuscular responses to high-intensity interval training (HIIT) and sprint interval training (SIT). *International Journal of Sports Physiology and Performance*, *16*(3), 1–10.

Copyright © 2026 IJSECS

Bowman, T. G., O'Connell, K., Luedke, L. E., & Rauh, M. J. (2024). Association of training load and nonfunctional overreaching in collegiate athletes. *BMJ Open Sport & Exercise Medicine*, *10*(1), e001903.

Chen, Y. (2025). Hybrid Transformer-LSTM model for athlete performance prediction in sports training management. *Physical Education and Sports Training*.

Goodall, S., Howatson, G., & Thomas, K. (2020). Intensity-dependent recovery of neuromuscular function after repeated-sprint exercise. *Medicine & Science in Sports & Exercise*, *52*(12), 2520–2529.

Hanová, P., Lehnert, M., & De Ste Croix, M. (2020). Time course of recovery from resistance exercise in male youth athletes. *Journal of Strength and Conditioning Research*, *34*(10), 2920–2926.

James, C. A., Thiebaud, R. S., *et al.* (2022). Effects of high-intensity interval training on neuromuscular fatigue and recovery. *European Journal of Applied Physiology*, *122*(6), 1315–1325.

Jia, Y., & Liang, J. (2026). Real time performance prediction in sports using machine learning algorithms. *International Journal of Data Science*, 285–297.

Minett, G. M., & Duffield, R. (2018). Is recovery from exercise influenced by prior fatigue? A systematic review. *Sports Medicine*, *48*(2), 239–258.

Ogunmiloro, B. (2026). Artificial intelligence-powered predictive modeling for calorie burns in underserved regions: AI-based calorie burn prediction. *Journal of Ideas in Health*, *9*(2).

Tan, A., Embi, Z. C., & Hashim, N. (2024). Comparison of machine learning methods for calories burn prediction. *Journal of Informatics and Web Engineering*, *3*(1), 182–191.

Tejaswini, T. D., & Rabindranath, R. S. (2025). A machine learning-based approach for personalized calorie expenditure prediction with an integrated AI fitness chatbot. *International Journal of Innovative Science and Research Technology*, *10*(9).

Vaara, J. P., Kyrolainen, H., Koivu, M., Tulppo, M., & Finni, T. (2014). The effect of 90 days of repetitive strenuous work on heart rate variability among young adult men. *Applied Ergonomics*, *45*(6), 1433–1438.

Weippert, M., Behrens, K., Rieger, A., Behrens, M., & Rudiger, H. (2022). Heart rate variability and training load in endurance athletes: A systematic review. *Frontiers in Physiology*, *13*, 868637.

Xu, Y., Zhang, N., Wang, J., Li, S., & Zhao, Y. (2024). Modeling physiological stress and recovery patterns using wearable sensor data. *Scientific Reports*, *14*, 51842.

Copyright © 2026 IJSECS