IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF HONE CALORIE EXPENDITURE USING ADVANCED ML

[1]Utsab Ray, [2]Arka Rajak, [2]Souvick Kumar Guha, [2]Aniket Roy, [3]Karabi Ganguly

[1]B.Tech – Department of Biomedical Engineering, [2] B.Tech – Department of Electronics and Communication Engineering,
[3]Associate Professor – Department of Biomedical Engineering
[1]Department of Biomedical Engineering,
[1]JIS College of Engineering, Kalyani, West Bengal, Nadia, India

Abstract: This research paper presents a novel approach to accurately predict calories burnt by individuals using advanced Machine Learning techniques. In today's fast-paced lifestyle, where obesity is a common concern, monitoring and managing calorie intake and expenditure is crucial. Existing methods rely on limited devices and manual calculations using MET charts and formulas, posing challenges. To address this, an XGBoost regression model is trained on a comprehensive dataset of many data points, achieving an impressive mean absolute error of 2.7 as compared to other ML algorithms. By continuously feeding the model with more data, its predictive capabilities are expected to improve further. The proposed system empowers individuals to make informed decisions about their diet and exercise routines, aiding in the fight against obesity.

***Keywords* - Machine Learning, XGBoost regression, calories burnt prediction, ML Algorithms.**

I. INTRODUCTION

Calories are often associated solely with food and weight reduction, but in reality, they represent a measure of heat energy. One calorie is defined as the amount of energy required to raise the temperature of 1 gram of water by 1°C. While commonly used to quantify energy in the context of the human body, calories also apply to various energy-releasing systems unrelated to humans. Different foods contain varying amounts of energy, indicated by their calorie counts. During exercise or intense workouts, the body temperature and heart rate increase. Carbohydrates present in the body are broken down into glucose, which is further converted into energy through the utilization of oxygen. To predict the amount of energy expended during exercise, several factors come into play. These include the duration of exercise, average heart rate per minute, temperature, height, weight, gender, and age of the individual.

During exercise, carbohydrates are metabolized into glucose, which is then converted into energy through the process of aerobic respiration, utilizing oxygen. This increased demand for energy in the muscles necessitates a higher oxygen intake. Consequently, the heart rate rises significantly to pump more oxygen-rich blood to the working muscles. The enhanced blood flow supplies the required oxygen for the breakdown of glucose molecules, facilitating energy production. However, only a portion of the energy derived from glucose is utilized for muscle contraction, while the remainder is dissipated as heat. As a result, the body temperature elevates, triggering the body's natural cooling mechanism through sweating. To accurately predict calorie expenditure, key parameters such as exercise duration, average heart rate per minute, body temperature, height, weight, and gender are considered. By integrating these factors into the predictive model, a comprehensive understanding of energy utilization during exercise can be achieved, enabling individuals to better track and manage their fitness goals.

In this study, a machine learning XGBoost regressor algorithm is employed to estimate the number of calories burned based on the aforementioned parameters. By inputting data such as exercise duration, heart rate, temperature, height, weight, and age, the algorithm can provide predictions on energy expenditure. This approach offers valuable insights for individuals seeking to monitor their calorie burn during physical activities, enabling them to make informed decisions regarding their fitness routines and overall health management. Understanding the relationship between various factors and calorie expenditure through machine learning techniques contributes to a better comprehension of human physiology and helps individuals adopt healthier lifestyles. By accurately estimating calories burned during exercise, individuals can track their progress and align their fitness goals accordingly.

II. LITERATURE SURVEY

Nipas and others [1] in their study utilized a regression model, specifically Random Forest regression, to accurately predict calories burned by individuals based on data analysis and model testing, achieving an impressive accuracy rate of 95.77%.
Yash and others [2] proposed a real-time machine learning system using a wearable accelerometer to determine the eating habits of individuals, allowing for calorie consumption and burnt estimation, and predicting health habits with a cloud-based logistic regression model.
Ludwig DS and others [3] reviewed and explored the controversies surrounding the role of carbohydrates in optimal health, longevity, and the prevention of obesity, diabetes, cardiovascular disease, cancer, and early death.
Brian claggett and others [4] investigated the association between carbohydrate intake and mortality, specifically examining the impact of low carbohydrate diets and the replacement of carbohydrates with plant-based or animal-based fat and protein in their study.
Joanne Slavin and others [5] classified carbohydrates into sugars, starches, and fibers, with sugars being intrinsic in fruits and milk products, while added sugars enhance flavor and palatability but offer minimal nutritional value. The Nutrition Facts label provides information on total sugars per serving without differentiating between naturally occurring and added sugars.
Cara B Ebbeling and others [6] discussed on the carbohydrate-insulin model suggests that increased consumption of processed, high-glycemic-load carbohydrates contributes to obesity by promoting calorie deposition, increasing hunger, and reducing energy expenditure, providing a new perspective on dietary interventions for weight loss beyond calorie restriction and dietary fat reduction.
Hilary Green and others [7] aims to identify and promote healthful carbohydrates and carbohydrate sources while limiting those with negative impacts, evaluate the impact of food processing on carbohydrate quality, and address the challenges of developing healthier food products in relation to regulations, science, technology, and consumer education.
Dale A Schoeller and others [8] reviewed the examination of the potential thermodynamic mechanisms behind increased weight loss rates in individuals consuming high-protein and/or low-carbohydrate diets, highlighting the observed greater weight loss compared to low-fat diets despite no significant differences in energy availability or expenditure, emphasizing the need for further research on weight loss composition and satiety effects.
Dena M Bravata and others [9] analyzed 107 articles and data from 94 dietary interventions to examine the effects of low-carbohydrate diets on adult participants, revealing that a subset of 663 participants consumed diets with 60 g/d or less of carbohydrates, and only 71 participants consumed diets with 20 g/d or less of carbohydrates.
R A Khanferyan and others [10] examined the frequency of consumption of sweet carbonated drinks containing carbohydrates in the Russian population and their contribution to overall caloric intake, revealing a relatively low consumption frequency and a limited impact on the caloric and carbohydrate intake of the diet.
.

III. METHODOLOGY

The work outline is forecasted in figure I.

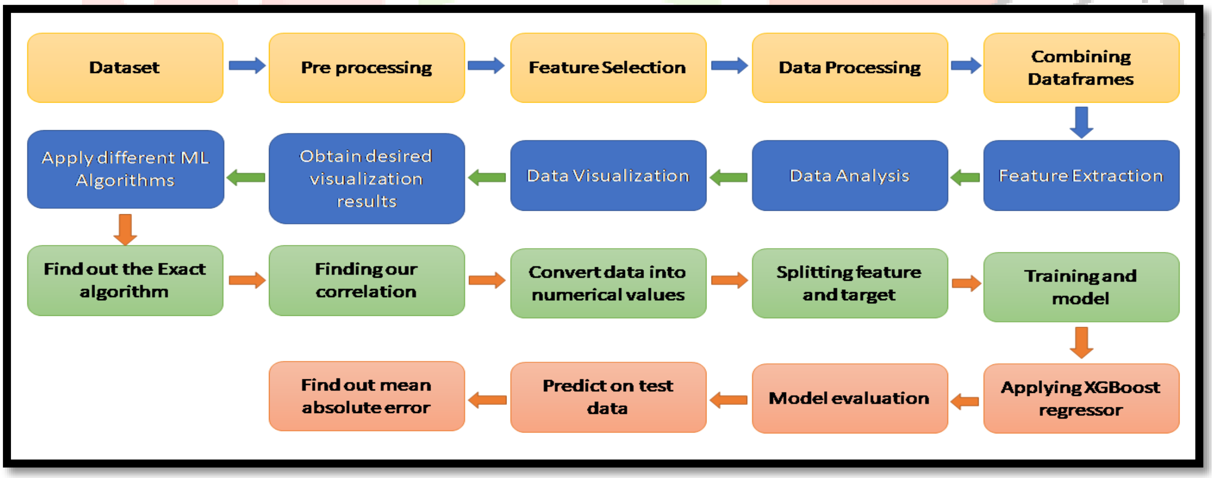


Figure 1: Methodology

In the current endeavor, the main goal is to collect the correct information set to train our artificial intelligence models, which will help us estimate the amount of calories an individual will burn. Prior to the statistics feeding operation, the records must be pre-processed. Following the completion of data processing, the data is organised as plots or graphs using a variety of visualization approaches. Here, we compare these models using the XG Boost regressor for a machine learning (ML) model, and then we evaluate these models.
Two parameters make up the XGBoost regressor. The adjustment of the decision tree is influenced by the regulator parameter lambda (), which stands for regulation parameter, and the threshold parameter gamma().
If Gain = Parallels Weight(left choice tree) + Comparison Weight(right decision tree) and Similarity Weight(SM) = (Residue)^2/(No of Residues)
Weight for Similarity (root)- The xgboost approach is more effective than others because if (Gain >) then decision tree bifurcation occurs for more levels else not.
In the field of data analysis and machine learning, several key steps play a crucial role in deriving meaningful insights from datasets. The dataset serves as the foundation, containing valuable information that requires preprocessing to ensure usability and accuracy. Preprocessing involves cleaning the data, handling missing values, and removing anomalies to create a reliable dataset.

Feature selection is a critical step that aids in identifying the most relevant features for analysis, eliminating unnecessary ones, and optimizing model performance. This process reduces training time and improves the model's ability to learn from the data. It allows us to focus on the most influential factors and extract meaningful patterns.

Data processing involves manipulating and transforming the dataset to extract valuable information. This includes combining dataframes, merging relevant data sources, and structuring the data for further analysis. Feature extraction is a technique used to transform raw data into meaningful features, such as reducing pixel-based images to condensed attribute collections, allowing for efficient analysis and modeling.

Data analysis involves applying statistical methods, algorithms, and techniques to uncover patterns, relationships, and insights within the dataset. Data visualization plays a significant role in presenting the analyzed data in a visually appealing and informative manner, enabling better understanding and interpretation of the results.

Machine learning algorithms are applied to the data to build models that can make predictions or classify new instances based on patterns learned from the dataset. Various algorithms, such as XGBoost regressor, are applied and evaluated to identify the most suitable model for the specific task at hand.

Correlation analysis helps determine the strength and direction of relationships between variables, providing insights into how different features affect the target variable. Converting data into numerical values allows for effective modeling and analysis, as many algorithms operate on numerical data.

Splitting the dataset into features and target variables is essential for training the model. The model is then trained using the training data to learn patterns and make predictions. Evaluation of the model's performance is conducted using metrics such as mean absolute error to assess its accuracy and effectiveness.

Based on the time of the workout, as well as variables like age, gender, body temperature, and heart rate at different moments during the activity, this dataset was analyzed to create predictions about the number of calories burned. Using these machine learning techniques, we are looking for a machine learning model that has a smaller mean absolute error and yields more accurate results. Ultimately, these processes enable researchers and analysts to derive meaningful insights from the data, uncover hidden patterns, make accurate predictions, and make informed decisions based on the results. The significance of each step lies in its contribution to refining and transforming the data, extracting relevant features, and building models that provide valuable insights and predictions.

.

IV. Results and Discussions

Importing the Dependencies

```
[ ] import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    from sklearn.model_selection import train_test_split
    from xgboost import XGBRegressor
    from sklearn import metrics
```

Figure 1: Importing dependencies

Data Collection & Processing

```
[ ] # loading the data from csv file to a Pandas DataFrame
    calories = pd.read_csv('/content/calories.csv')
```

```
# print the first 5 rows of the dataframe
calories.head()
```

	User_ID	Calories
0	14733363	231.0
1	14861698	66.0
2	11179863	26.0
3	16180408	71.0
4	17771927	35.0

```
[ ] exercise_data = pd.read_csv('/content/exercise.csv')
```

```
[ ] exercise_data.head()
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp
0	14733363	male	68	190.0	94.0	29.0	105.0	40.8
1	14861698	female	20	166.0	60.0	14.0	94.0	40.3
2	11179863	male	69	179.0	79.0	5.0	88.0	38.7
3	16180408	female	34	179.0	71.0	13.0	100.0	40.5
4	17771927	female	27	154.0	58.0	10.0	81.0	39.8

Figure 2: Data Collection and Processing

Combining the two Dataframes

```
[ ] calories_data = pd.concat([exercise_data, calories['Calories']], axis=1)
```

```
[ ] calories_data.head()
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	male	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	female	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	male	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	female	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	female	27	154.0	58.0	10.0	81.0	39.8	35.0

```
[ ] # checking the number of rows and columns
    calories_data.shape

(15000, 9)
```

Figure 3: Dataframe Combination

```
# getting some informations about the data
calories_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   User_ID     15000 non-null  int64
 1   Gender      15000 non-null  object
 2   Age         15000 non-null  int64
 3   Height      15000 non-null  float64
 4   Weight      15000 non-null  float64
 5   Duration    15000 non-null  float64
 6   Heart_Rate  15000 non-null  float64
 7   Body_Temp   15000 non-null  float64
 8   Calories    15000 non-null  float64
dtypes: float64(6), int64(2), object(1)
memory usage: 1.0+ MB
```

```
# checking for missing values
calories_data.isnull().sum()
```

```
User_ID       0
Gender        0
Age           0
Height        0
Weight        0
Duration      0
Heart_Rate    0
Body_Temp     0
Calories      0
dtype: int64
```

Figure 4: Data Information

Data Analysis

```
# get some statistical measures about the data
calories_data.describe()
```

	User_ID	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
count	1.500000e+04	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000	15000.000000
mean	1.497736e+07	42.789800	174.465133	74.966867	15.530600	95.518533	40.025453	89.539533
std	2.872851e+06	16.980264	14.258114	15.035657	8.319203	9.583328	0.779230	62.456978
min	1.000116e+07	20.000000	123.000000	36.000000	1.000000	67.000000	37.100000	1.000000
25%	1.247419e+07	28.000000	164.000000	63.000000	8.000000	88.000000	39.600000	35.000000
50%	1.499728e+07	39.000000	175.000000	74.000000	16.000000	96.000000	40.200000	79.000000
75%	1.744928e+07	56.000000	185.000000	87.000000	23.000000	103.000000	40.600000	138.000000
max	1.999965e+07	79.000000	222.000000	132.000000	30.000000	128.000000	41.500000	314.000000

Figure 5: Data Analysis

Data Visualization

```
sns.set()
```

```
# plotting the gender column in count plot
sns.countplot(calories_data['Gender'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: Fu
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fcbbd756110>
```

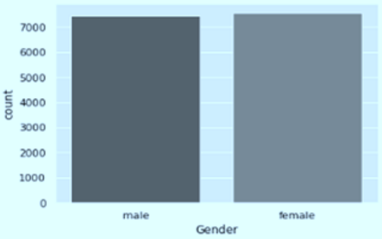


Figure 6: Data Visualization for gender

```
# finding the distribution of "Age" column
sns.distplot(calories_data['Age'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fcbbd200550>
```

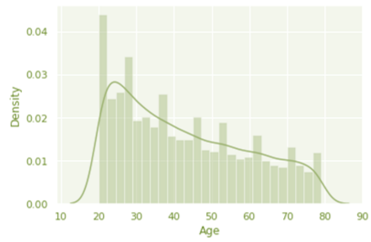


Figure 7: Data Visualization for age

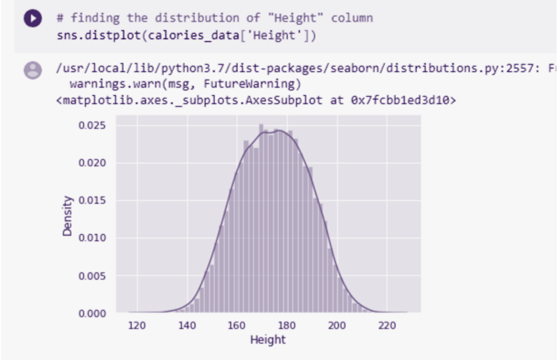


Figure 7: Data Visualization for height

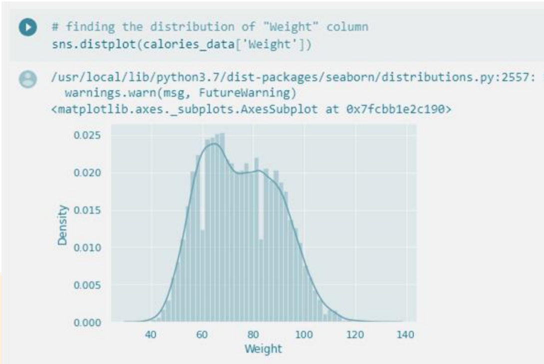


Figure 7: Data Visualization for weight

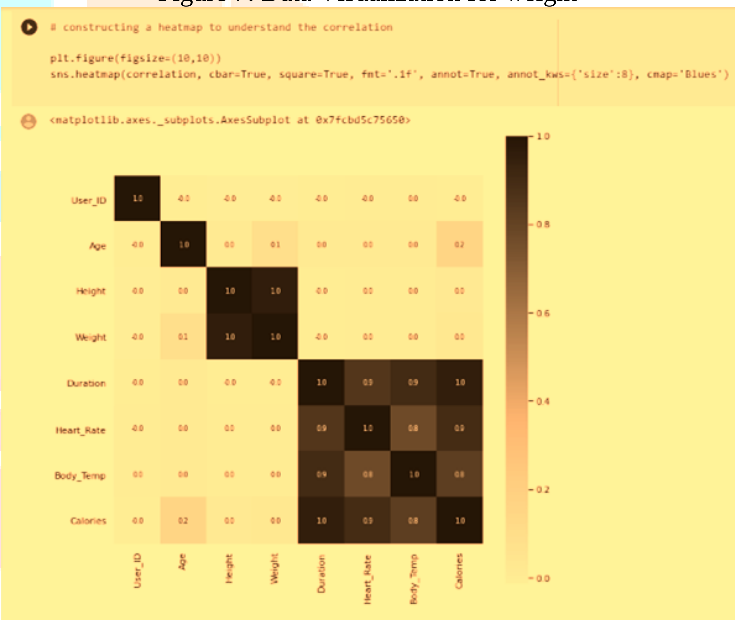
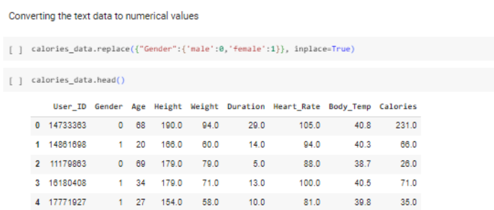


Figure 8: Creating a heatmap to understand correlation



Converting the text data to numerical values

```
[ ] calories_data.replace({"Gender":{'male':0,'female':1}}, inplace=True)
```

```
[ ] calories_data.head()
```

	User_ID	Gender	Age	Height	Weight	Duration	Heart_Rate	Body_Temp	Calories
0	14733363	0	68	190.0	94.0	29.0	105.0	40.8	231.0
1	14861698	1	20	166.0	60.0	14.0	94.0	40.3	66.0
2	11179863	0	69	179.0	79.0	5.0	88.0	38.7	26.0
3	16180408	1	34	179.0	71.0	13.0	100.0	40.5	71.0
4	17771927	1	27	154.0	58.0	10.0	81.0	39.8	35.0

Figure 9: Converting data into numerical values

Separating features and Target

```
[ ] X = calories_data.drop(columns=['User_ID','Calories'], axis=1)
    Y = calories_data['Calories']
```

```
print(X)
```

```
       Gender  Age  Height  Weight  Duration  Heart_Rate  Body_Temp
0           0   68   190.0    94.0      29.0       105.0       40.8
1           1   20   166.0    60.0      14.0        94.0       40.3
2           0   69   179.0    79.0       5.0        88.0       38.7
3           1   34   179.0    71.0      13.0       100.0       40.5
4           1   27   154.0    58.0      10.0        81.0       39.8
...       ...  ...     ...     ...       ...         ...        ...
14995       1   20   193.0    86.0      11.0        92.0       40.4
14996       1   27   165.0    65.0       6.0        85.0       39.2
14997       1   43   159.0    58.0      16.0        90.0       40.1
14998       0   78   193.0    97.0       2.0        84.0       38.3
14999       0   63   173.0    79.0      18.0        92.0       40.5

[15000 rows x 7 columns]
```

```
print(Y)
```

```
0        231.0
1         66.0
2         26.0
3         71.0
4         35.0
         ...
14995     45.0
14996     23.0
14997     75.0
14998     11.0
14999     98.0
Name: Calories, Length: 15000, dtype: float64
```

Figure 10: Separating features and target

Splitting the data into training data and Test data

```
[ ] X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

```
print(X.shape, X_train.shape, X_test.shape)
```

```
(15000, 7) (12000, 7) (3000, 7)
```

Figure 11: Splitting data into training and test data

XGBoost Regressor

```
[ ] # loading the model
    model = XGBRegressor()
```

```
# training the model with X_train
model.fit(X_train, Y_train)
```

```
[10:06:32] WARNING: /workspace/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
             colsample_bynode=1, colsample_bytree=1, gamma=0,
             importance_type='gain', learning_rate=0.1, max_delta_step=0,
             max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
             n_jobs=1, nthread=None, objective='reg:linear', random_state=0,
             reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
             silent=None, subsample=1, verbosity=1)
```

Figure 12: Applying XGBoost Regressor

Mean Absolute Error

```
[ ] mae = metrics.mean_absolute_error(Y_test, test_data_prediction)
```

```
print("Mean Absolute Error = ", mae)
```

```
Mean Absolute Error =  2.7159012502233186
```

Figure 13: Computing the mean absolute error

Evaluation

Prediction on Test Data

```
[ ] test_data_prediction = model.predict(X_test)
```

```
[ ] print(test_data_prediction)
```

```
[129.06204  223.79721   39.181965 ... 145.59767   22.53474   92.29064 ]
```

Figure 14: Final prediction and Evaluation

V. Conclusion

Through the utilization of multiple instance parameters and various factors, we have successfully enhanced the innovation and effectiveness of calorie prediction systems. Accuracy, a crucial component in prediction-based systems, has been significantly improved, resulting in more expressive regression outputs that are both understandable and bounded with a certain level of accuracy. Notably, the XGB Regressor has emerged as a standout performer, exhibiting high accuracy in its findings. The mean absolute error, a measure of the discrepancy between observed and predicted values, stands at an impressive 2.71 for the XGB Regressor, indicating low error rates. Thus, we can confidently assert that the XG Boost Regressor serves as the optimal model for predicting calorie burn. Furthermore, the proposed approach demonstrates flexibility, which can be further enhanced through variations and adaptations.

In this study, we have focused on the seven primary factors that influence calorie burn; however, it is important to recognize that other factors also play a significant role. To maintain overall health and fitness, it is crucial to not only understand the number of calories burned but also monitor calorie consumption. Machine learning (ML) techniques can be employed to construct a user interface (UI) where individuals can input their values and receive comprehensive results showcasing their calorie burn. By integrating these features with our recommended diet and exercise regimen, we can develop a fully functional application that empowers users to track their calorie burn effectively and make informed decisions regarding their health and fitness journey.

VI. Acknowledgment

We would like to thank Department of Biomedical Engineering and Department of Electronics and Communication Engineering, JIS College of Engineering for their support.

References

[1] M. Nipas, A. G. Acoba, J. N. Mindoro, M. A. F. Malbog, J. A. B. Susa and J. S. Gulmatico, "Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm," 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 2022, pp. 1-4, doi: 10.1109/ICPC2T53885.2022.9776710.

[2] Y. Jain, D. Chowdhury and M. Chattopadhyay, "Machine Learning Based Fitness Tracker Platform Using MEMS Accelerometer," 2017 International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, India, 2017, pp. 1-5, doi: 10.1109/ICCECE.2017.8526202.

[3] Ludwig DS, Hu FB, Tappy L, Brand-Miller J. Dietary carbohydrates: role of quality and quantity in chronic disease. BMJ. 2018 Jun 13;361:k2340. doi: 10.1136/bmj.k2340. PMID: 29898880; PMCID: PMC5996878.

[4] Seidelmann SB, Claggett B, Cheng S, Henglin M, Shah A, Steffen LM, Folsom AR, Rimm EB, Willett WC, Solomon SD. Dietary carbohydrate intake and mortality: a prospective cohort study and meta-analysis. Lancet Public Health. 2018 Sep;3(9):e419-e428. doi: 10.1016/S2468-2667(18)30135-X. Epub 2018 Aug 17. PMID: 30122560; PMCID: PMC6339822.

[5] Slavin J, Carlson J. Carbohydrates. Adv Nutr. 2014 Nov 14;5(6):760-1. doi: 10.3945/an.114.006163. PMID: 25398736; PMCID: PMC4224210.

[6] Ludwig DS, Ebbeling CB. The Carbohydrate-Insulin Model of Obesity: Beyond "Calories In, Calories Out". JAMA Intern Med. 2018 Aug 1;178(8):1098-1103. doi: 10.1001/jamainternmed.2018.2933. PMID: 29971406; PMCID: PMC6082688.

[7] Lamothe LM, Lê KA, Samra RA, Roger O, Green H, Macé K. The scientific basis for healthful carbohydrate profile. Crit Rev Food Sci Nutr. 2019;59(7):1058-1070. doi: 10.1080/10408398.2017.1392287. Epub 2017 Nov 30. PMID: 29190114.

[8] Buchholz AC, Schoeller DA. Is a calorie a calorie? Am J Clin Nutr. 2004 May;79(5):899S-906S. doi: 10.1093/ajcn/79.5.899S. PMID: 15113737.

[9] Bravata DM, Sanders L, Huang J, Krumholz HM, Olkin I, Gardner CD, Bravata DM. Efficacy and safety of low-carbohydrate diets: a systematic review. JAMA. 2003 Apr 9;289(14):1837-50. doi: 10.1001/jama.289.14.1837. PMID: 12684364.

[10] Khanferyan RA, Radzhabkadiev RM, Evstratova VS, Galstyan AG, Khurshudyan SA, Semin VB, Vrzhesinskaya OA, Akimov MY. [Consumption of carbohydratecontaining beverages and their contribution to the total calorie content of the diet]. Vopr Pitan. 2018;87(2):39-43. Russian. doi: 10.24411/0042-8833-2018-10017. Epub 2018 Feb 26. PMID: 30592867.