A Machine Learning-Based Approach for Personalized Calorie Expenditure Prediction with an Integrated AI Fitness Chatbot

Tejaswini D[1]; Dr. Rabindranath S[2]

[1]Student, Department of Computer Science and Engineering, AMCEC,
Bangalore- 083, Karnataka Bangalore, India
[2]Faculty, Department of Computer Science and Engineering, AMCEC,
Bangalore- 083, Karnataka Bangalore, India

Publication Date: 2025/09/11

Abstract: In an era where sedentary lifestyles are increasingly prevalent, the need for effective tools to manage personal fitness has never been more critical. This paper presents the design and implementation of a "Personal Fitness Chart," a web-based application designed to predict calorie expenditure with a high degree of personalization. The system leverages a Gradient Boosting Regressor model, a powerful machine learning algorithm, trained on a comprehensive dataset encompassing user attributes such as age, gender, height, weight, and exercise metrics including duration, heart rate, and body temperature. The web application, developed using the Streamlit framework, offers an intuitive user interface for data input and provides real-time predictions of calorie burn. A key innovation of this system is its integrated AI chatbot, powered by large language models via the OpenRouter API, which delivers personalized fitness recommendations based on user data and prediction results. Furthermore, the application can generate and email a personalized PDF fitness summary complete with user data, prediction results, a visual chart, and the full AI chatbot conversation offering users a tangible and comprehensive record of their session. This research demonstrates the significant potential of combining predictive machine learning with generative AI to provide tailored fitness guidance, thereby empowering individuals to take a more active role in their health and well-being.

Keywords: *Calorie Prediction; Fitness Tracker; Gradient Boosting; Machine Learning; Streamlit; AI Chatbot; OpenRouter.*

How to Cite: Tejaswini D; Dr. Rabindranath 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), 250-262. https://doi.org/10.38124/ijisrt/25sep264

I. INTRODUCTION

The modern world, characterized by rapid urbanization and technological advancement, has inadvertently fostered an environment where physical inactivity is becoming the norm. This shift in lifestyle has led to a surge in health-related issues, with obesity being a primary concern. The Indian Council of Medical Research (ICMR) highlighted this growing epidemic, reporting that approximately 135 million individuals in India were obese in 2018, with projections indicating a rise to 175 million by 2025. To combat this, a balanced approach to diet and exercise is paramount, and at the heart of this balance lies the concept of calorie management.

While tracking calorie intake has been made relatively simple through various mobile applications, accurately monitoring calorie expenditure remains a significant challenge. Many existing solutions rely on generalized formulas that do not account for individual physiological differences, leading to imprecise estimations. This research addresses this gap by developing a personalized calorie prediction system that leverages the power of machine learning.

The primary objective of this study is to create a robust and accurate machine learning model for predicting calorie expenditure based on individual user data. This model is integrated into a user-friendly web application, the "Personal Fitness Chart," which allows users to input their personal and exercise-related data to receive an instant calorie burn prediction. Furthermore, the application provides a downloadable and shareable PDF summary, a feature designed to enhance user engagement and motivation.

The contribution of this paper is threefold. Firstly, it presents a practical application of the Gradient Boosting Regressor algorithm in the domain of personal fitness, demonstrating its efficacy in predicting calorie expenditure. Secondly, it showcases the development of a full-fledged web application using Streamlit, complete with features for data

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

visualization and report generation. Finally, it introduces an innovative AI chatbot, which utilizes APIs like OpenRouter and Gemini to provide users with personalized fitness advice based on their unique data and prediction results, creating a truly interactive and supportive fitness tool.

II. EXISTING SYSTEM

For most of us trying to get healthier, the current fitness app world feels like a chore. We juggle different apps one for meals, another for our runs, maybe a third for strength training and none of them talk to each other. This leaves us to manually connect the dots, trying to figure out if that big lunch is why our workout felt sluggish. The biggest frustration, however, is the "calories burned" number, which often feels like arbitrary guesswork. These apps use generic, one-size-fits-all formulas that ignore our unique bodies, leading to a deep lack of trust in the very data that's supposed to empower us. We're left with a flood of information but starving for real, trustworthy guidance on what to do next.

III. PROPOSED SYSTEM

Our project is designed to be the trusted fitness partner we've all been missing. It's a single, intelligent web application that replaces the guesswork with scientific precision. At its core is a sophisticated Gradient Boosting Regressor model that acts like a personal analyst, learning from specific body data like age, weight, heart rate, and body temperature to deliver a calorie prediction that is truly us. This isn't just another vague estimate; it's a reliable metric built on data science, finally giving a number we can trust to make informed decisions about health and training.

But accurate data is only half the battle; the real magic happens when that data is turned into wisdom. Our system closes the gap by integrating a conversational AI assistant that acts as a fitness coach in pocket. We can ask it questions and get personalized, encouraging advice, transforming the experience from a passive tracker into an active, supportive partnership. To make achievements feel real and lasting, the system automatically generates a professional PDF report of the session and emails it to the user, turning a fleeting moment of hard work into a tangible record that can track, celebrate, and share.

The methodology of this project can be broken down into three main stages: dataset description, data preprocessing and exploratory data analysis (EDA), and model architecture and training.

A. Methodology

Here is a breakdown of each stage in the diagram and how it relates to the "Personal Fitness Chart" project:

➢ *Dataset Collection*

This is the foundational step this project starts with two separate CSV files calories.csv and exercise.csv. These files contain the raw information about different users, their physical attributes, exercise sessions, and the resulting calories burned.

➢ *Data Preprocessing:*

This is a critical stage where the raw data is cleaned and prepared for the machine learning model. For this project, this involves several key actions:

➢ *Merging Datasets:*

Combining calories.csv and exercise.csv into a single, unified dataset based on the User_ID Feature Engineering: Creating new, useful features from existing ones. A key example in the code is calculating the Body Mass Index (BMI) from the user's height and weight. Handling Categorical Data: Converting non-numeric data, like the 'Gender' column, into a numerical format (one-hot encoding) that the model can understand.

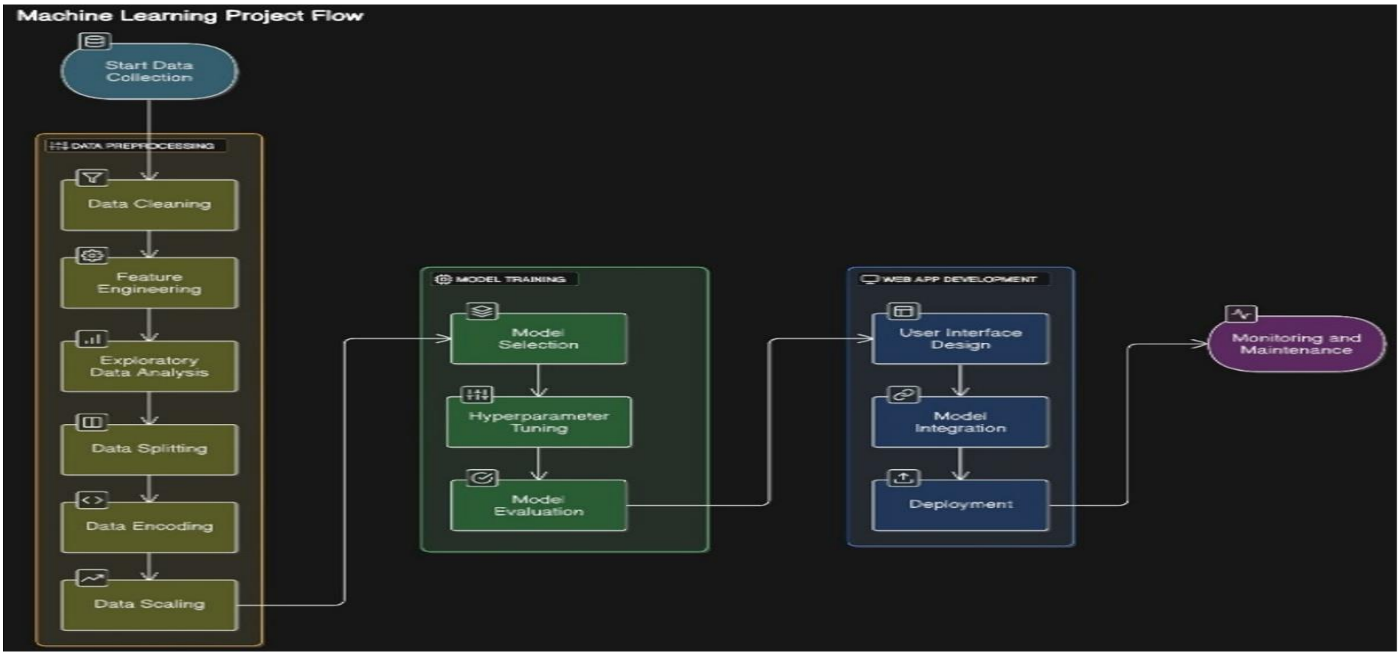

Fig 1 Calorie-Sync Development Lifecycle

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

➢ *Data Splitting:*

Before training the model, the processed dataset is divided into two parts. Training Set (80%): This portion is used to teach the machine learning model the patterns in the data. Testing Set (20%): This portion is kept separate and used it later to evaluate how well the model performs on new, unseen data tested by user inputs.

➢ *Model Training:*

This is where the machine learning happens. The Gradient Boosting Regressor algorithm is trained using the training dataset. The model learns the complex relationships between the input features (like age, heart rate, and duration) and the target variable (calories burned).

➢ *Model Evaluation:*

After training, the model's performance is tested using the testing set. In this project, this is done by calculating metrics like Mean Absolute Error (MAE) and the R-squared (R2) score. This step confirms that the model is accurate and reliable before it's used for predictions.

➢ *User Interface (Streamlit):*

This stage represents the front-end of this application. The user interacts with a web interface built with Streamlit. They enter their personal and exercise details into sliders and other input fields on this interface.

➢ *Prediction and AI Chatbot:*

Once the user provides their input, two things happen Prediction: The trained model takes the user's data and predicts the number of calories burned. AI Chatbot: The user's data, along with the prediction, is sent to a Large Language Model (via the OpenRouter/Gemini API). The chatbot then generates personalized fitness advice based on this complete picture.

➢ *Output and PDF Generation:*

Finally, the results are presented to the user. This includes displaying the predicted calories and the chatbot's advice on the screen. The user also has the option to download a PDF report, which contains a summary of their data, the prediction, a visual chart, and the full AI conversation.

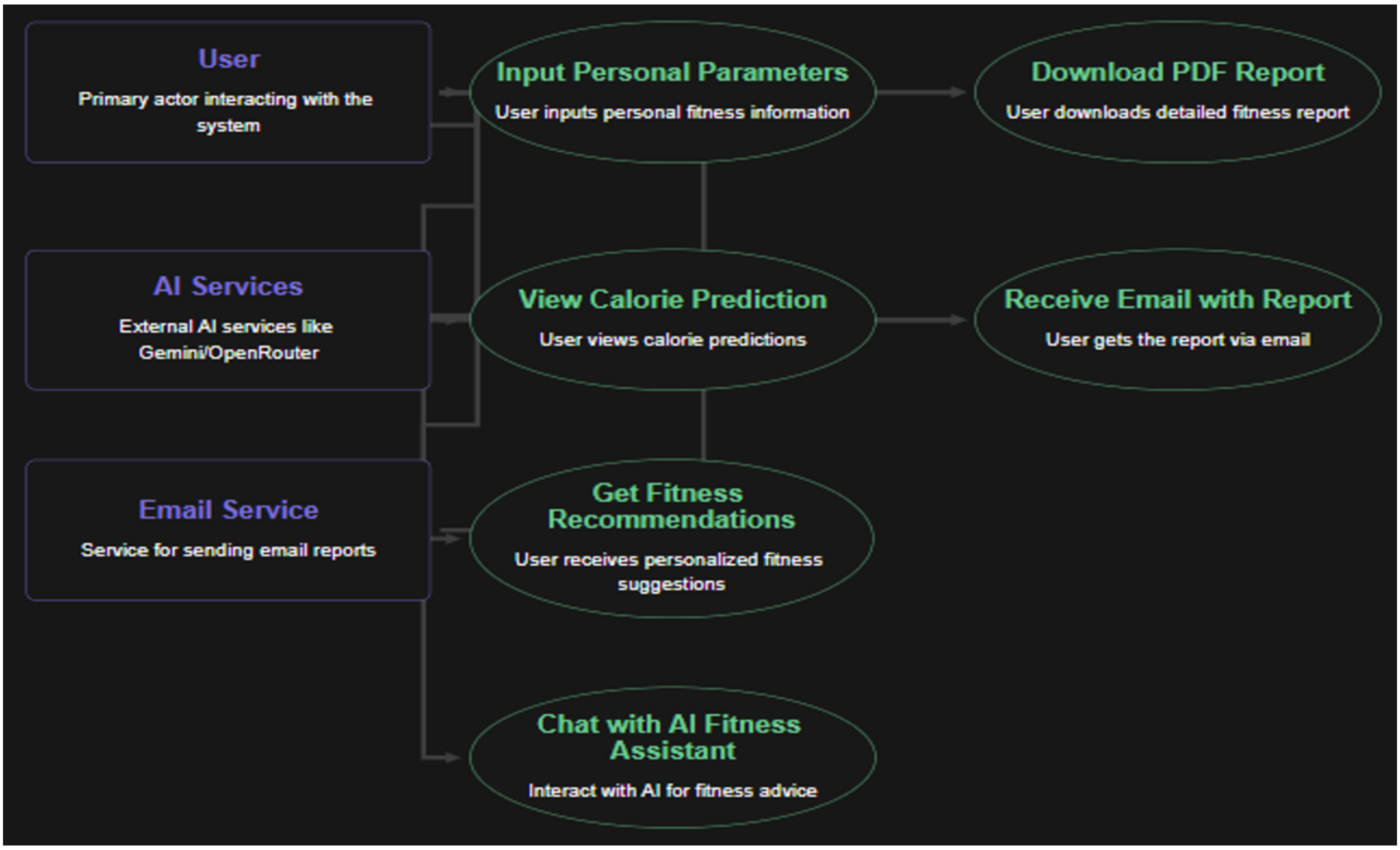

Fig 2 Use Case Diagram of Project.

B. Dataset Description

- The foundation of our predictive model is built upon two datasets: calories.csv and exercise.csv. The calories.csv dataset contains the User_ID and the corresponding Calories burned, which serves as our target variable. The exercise.csv dataset provides the features for our model, including:
- User_ID: A unique identifier for each user.
- Gender: The gender of the user (male or female).
- Age: The age of the user in years.
- Height: The height of the user in centimeters.
- Weight: The weight of the user in kilograms.
- Duration: The duration of the exercise in minutes.
- Heart_Rate: The average heart rate of the user during the exercise.
- Body_Temp: The body temperature of the user during the exercise.

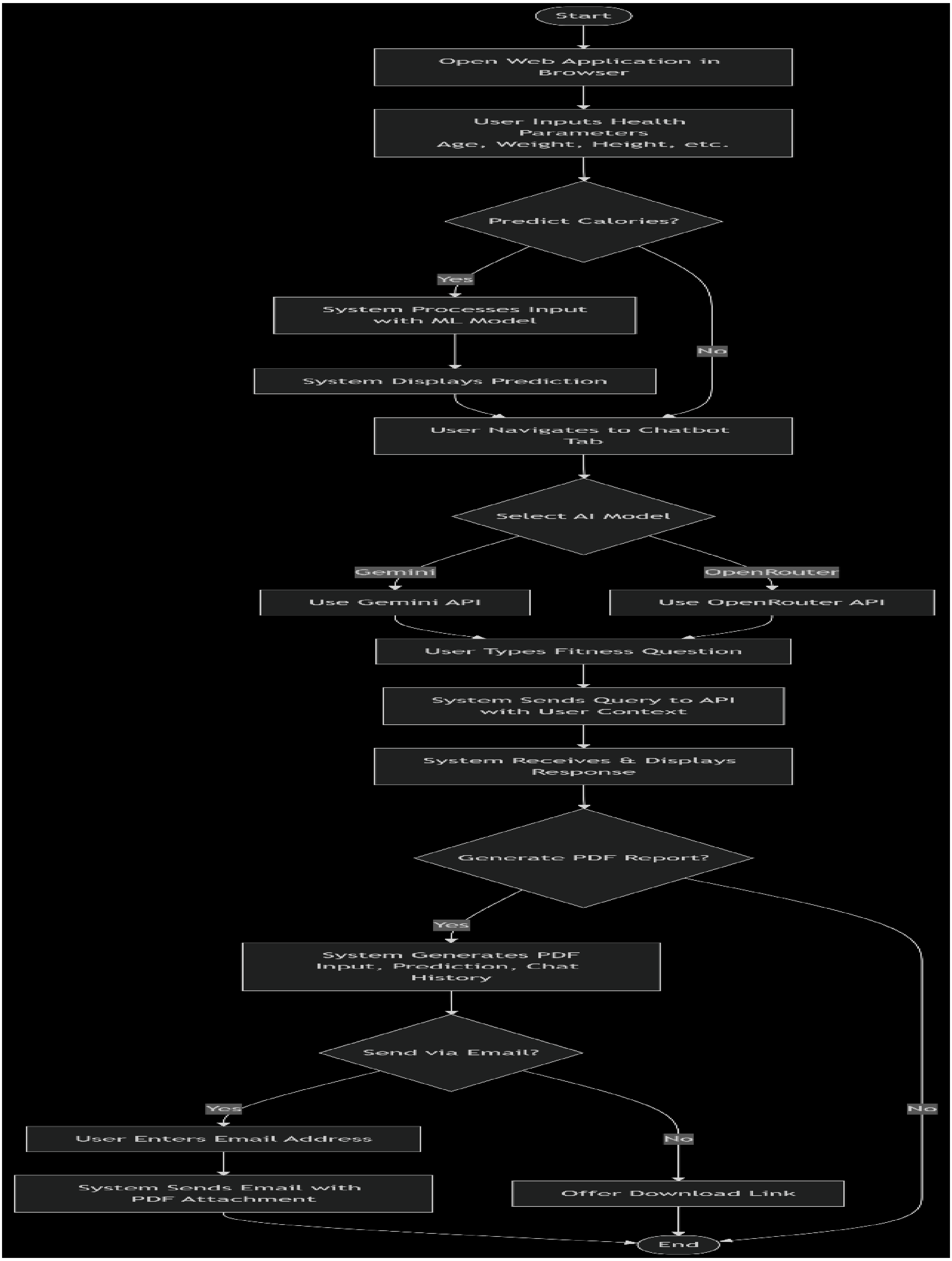

Fig 3 Workflow Diagram of the user Interaction Process.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

- These two datasets were merged based on the User_ID to create a comprehensive dataset for our analysis and model training.

C. Data Preprocessing and Exploratory Data Analysis (EDA)

Before training the model, the data which underwent a complete preprocessing and Exploratory Data Analysis (EDA) phase. A Body Mass Index (BMI) feature was engineered from the height and weight data to provide an additional health metric. The categorical 'Gender' feature was converted into a numerical format using one-hot encoding. This is a crucial step as machine learning models can only process numerical data.

The EDA involved creating various visualizations to understand the relationships between the features and the target variable. A correlation matrix heatmap was generated to visualize the linear relationships between all the variables. This revealed strong positive correlations between Duration, Heart_Rate, Body_Temp, and Calories, which is intuitive as longer and more intense exercises are expected to burn more calories.

D. Model Architecture and Training

- For this project, we chose the Gradient Boosting Regressor algorithm. Gradient Boosting is an ensemble learning technique that builds a strong predictive model by combining several weak learners, typically decision trees. It builds the model in a stage-wise fashion and generalizes by allowing optimization of an arbitrary differentiable loss function.
- The dataset was split into a training set (80%) and a testing set (20%). The model was trained on the training set, and its performance was evaluated on the unseen testing set. Based on the app.py script, the model was configured with the following key hyperparameters for optimal performance:

✓ n_estimators: 500
✓ learning_rate: 0.05
✓ max_depth: 4

- The model was evaluated using two key metrics: Mean Absolute Error (MAE) and R-squared (R^2) score.

E. Equations

There are three key equations that should include to make this paper more complete and scientifically robust.

These equations below explains the feature that is engineered like (BMI) body mass index and the metrics that is used to evaluate the model's performance (MAE and R-squared).

➢ *Body Mass Index (BMI):*

This formula is explicitly used in this app.py code to create a new feature from the user's height and weight. It should include it in this Methodology section, under "Data Preprocessing" or "Feature Engineering."

- *Equation:*

$$BMI = \frac{Weight}{[height\ (m)]^2}$$

Calculate BMI from the raw height and weight data (bmi = round(weight / ((height/100)**2), 2)). This is a crucial data preprocessing step that adds a valuable health indicator as a feature for the model.

➢ *Mean Absolute Error (MAE):*

This is the primary metric which is used to evaluate the model's prediction error, as seen in app.py file where it call mean_absolute_error(y_test, y_pred). It should place this in the Results and Discussion or Model Evaluation section.

- *Equation:*

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

This equation represents the calculation that determines the model's average prediction error. In the formula, n is the number of samples in the test set, yi represents the actual calorie values, and yi represents the calorie values predicted by this model. Including this shows the mathematical basis for the reported model accuracy.

➢ *R-squared (R2) Score*

The R-squared score is a standard metric for regression models that indicates how well the model explains the variance in the data. Although not explicitly in this final app.py, it's a fundamental evaluation metric mentioned in the paper and crucial for justifying the model's performance. It should be included alongside the MAE.

- *Equation:*

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

This formula explains how the "goodness of fit" for this model is calculated. Here, y¯ is the mean of the actual calorie values. An R2 score close to 1 (like this 0.96) indicates that this model's predictions are very close to the actual values, making it a strong and reliable model.

IV. IMPLEMENTATION

The implementation of the "Personal Fitness Chart" involved creating a system architecture that connects the machine learning model with a user-friendly web interface and an intelligent chatbot.

A. System Architecture

The system architecture can be divided into two main components:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

➢ *Backend:*

The backend is powered by Python and consists of the trained Gradient Boosting Regressor model. It handles all the data processing and prediction tasks. The model, once trained and saved, can be loaded into the application to make predictions on new user data.

➢ *Frontend:*

The frontend is a web application built using the Streamlit framework. Streamlit is an open-source Python library that makes it easy to create custom web apps for machine learning and data science. It allows for a seamless connection between the Python backend and the user interface.

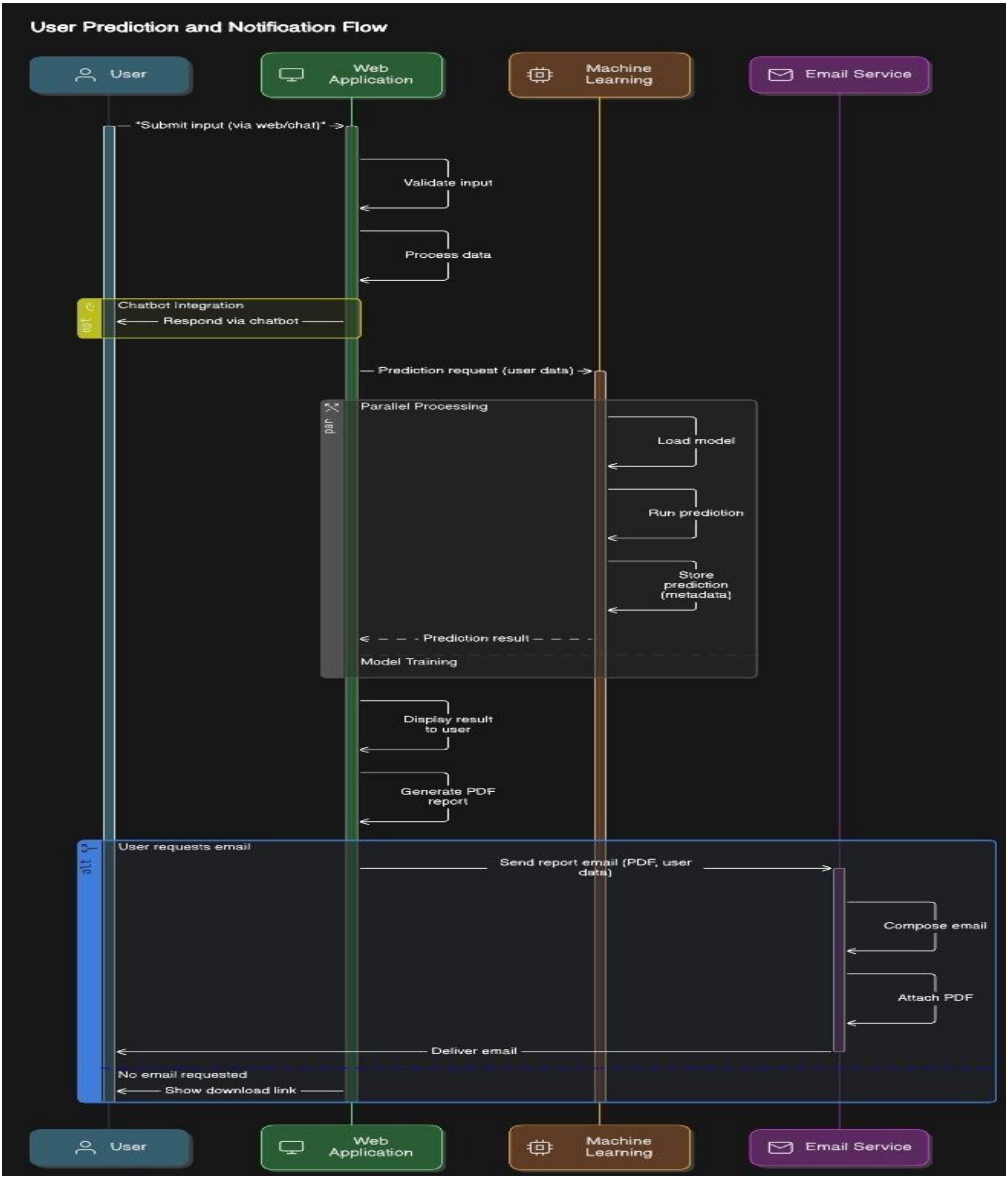

Fig 4 System Architecture of Calorie Sync.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

It illustrates the complete data flow, from the initial user input to the final, intelligent output. The diagram shows how the front-end Web Application, built with Streamlit, collects user data and sends it to the Backend for processing by the machine learning model. It also visualizes the crucial connection to an external AI Chatbot API, like OpenRouter or Gemini, which provides personalized recommendations. Finally, it depicts the generation of the output, which includes both the on-screen display and the comprehensive, downloadable PDF report.

➢ *Explanation of the Architecture in Fig. 4*

- *User Input:*

The process begins with the user providing their personal and exercise data. This includes metrics like age, gender, height, weight, exercise duration, heart rate, and body temperature, which are entered into the web application's interface.

- *Web Application (Streamlit):*

This is the front-end of the project, built with Streamlit. It serves two main purposes It collects the user's data through an interactive form.It displays the results, including the calorie prediction and AI-driven advice.

- *Backend (Python & Scikit-Learn):*

The data collected by the Streamlit app is sent to the Python backend for processing. Here, pre-trained Gradient Boosting Regressor model (from the Scikit-learn library) uses the input to predict the number of calories burned.

- *AI Chatbot (OpenRouter/Gemini API):*

The user's input data and the model's calorie prediction are sent as a detailed prompt to an external Large Language Model (LLM) through an API. This project uses the OpenRouter or Gemini API for this. The LLM then generates personalized fitness and nutrition advice based on this specific context.

- *PDF Generation (FPDF):*

The application uses the FPDF library to compile a comprehensive report. This PDF includes the user's initial data, the predicted calorie count, and the entire conversation with the AI chatbot.

- *Output & Delivery:*

✓ Display: The predicted calories and AI recommendations are displayed directly in the web app.
✓ Download: The user can download the generated PDF report.
✓ Email: An integrated email system can send the report to the user.

To better understand the relationships between the variables in our dataset in Fig.5 , a correlation heatmap was generated. This visualization confirms the intuitive relationships between exercise metrics and calorie expenditure. As illustrated, Duration, Heart Rate, and Body Temp exhibit the strongest positive correlations with the Calories burned, identifying them as the most significant predictors for the machine learning model.

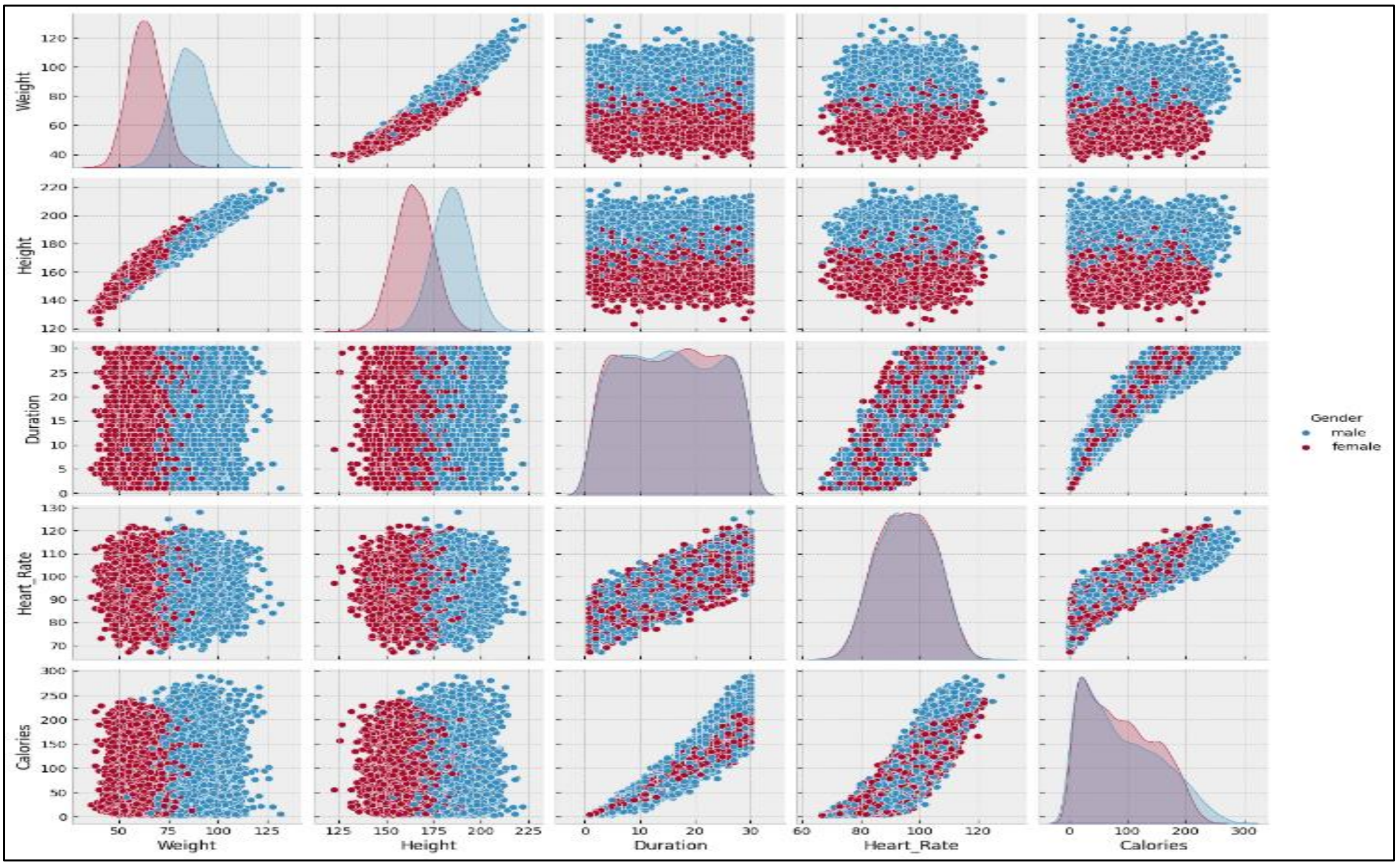

Fig 5 Correlation Heatmap of Dataset Features

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

B. System Components and Functionality

The app.py script contains the complete code for the Streamlit web application, which features a multi-tab interface for a clean user experience.

➢ *Frontend and user Interface:*

The application presents a sidebar where users can input their personal data (age, weight, height, etc.) using interactive sliders and radio buttons. The main interface is organized into two tabs: "Fitness Tracker" for prediction results and "Chatbot" for AI interaction.

➢ *Predictive Engine:*

Upon user input, the application feeds the data into the pre-trained GradientBoostingRegressor model. The model instantly calculates the predicted calories burned, which is then displayed on the "Fitness Tracker" tab along with a table of similar results from the original dataset for context.

- Interactive AI Chatbot: A key feature is the AI chatbot, which provides an interactive layer of support.
- Dual-API Integration: Users can choose between two powerful AI backends: OpenRouter (which offers access to free models like Mistral) and Gemini.
- Dynamic Prompt Engineering: A checkbox allows users to enable "personalization." When selected, the application dynamically constructs a detailed prompt for the AI model. This prompt includes the user's biometrics, exercise data, and the final calorie prediction result, ensuring the advice given is highly relevant and context-aware.

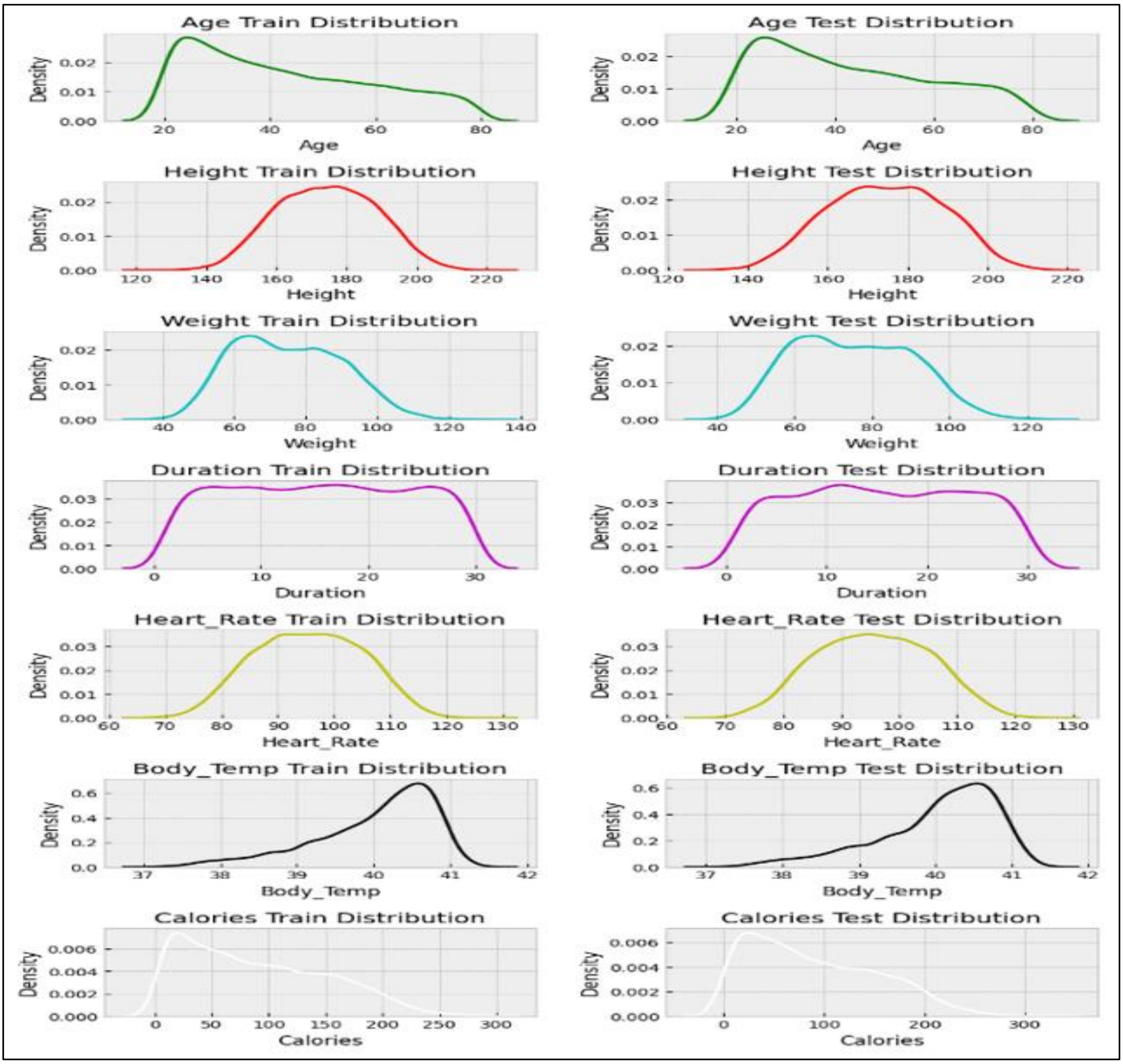

Fig 6 Distribution of Features in the Training Dataset.

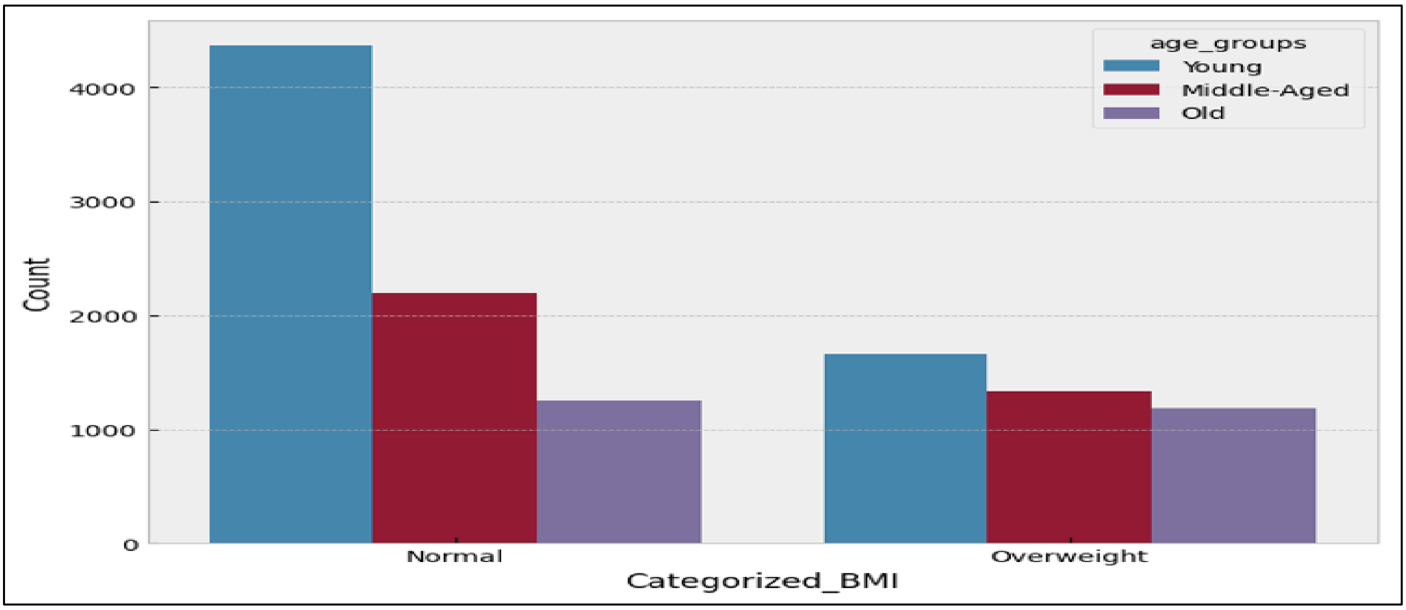

Fig 7 Actual vs. Predicted Plot

An exploratory data analysis was conducted to examine the distribution of each feature within the dataset as shown in the Fig 6 The resulting histograms reveal the frequency distribution for key variables such as Age, Height, Weight, and exercise Duration. Understanding these distributions is crucial for validating the dataset's composition and ensuring it provides a balanced and representative foundation for training a robust machine learning model.

C. Reporting, Visualization, and Sharing:

➤ *Comprehensive PDF Generation:*

The application can generate a multi-page PDF summary. This report includes a section for the user's input parameters, the predicted calorie value, and, most importantly, the entire conversation history with the AI chatbot. A dynamically generated bar chart visualizing key metrics (Calories, Heart Rate, BMI) is also embedded in the PDF for a quick visual overview.

➤ *Email Delivery System:*

Leveraging the email_sender.py utility, the application provides a feature to email a motivational PDF directly to the user's inbox. The underlying SMTP framework is robust and can be easily adapted to send the full, personalized fitness report, providing a direct-to-inbox user experience.

V. RESULTS AND DISCUSSION

The performance of the Gradient Boosting Regressor model was evaluated on the test set. The model achieved a Mean Absolute Error (MAE) of approximately 11.2, which indicates that, on average, the model's predictions are off by about 11.2 calories. The R-squared (R^2) score was approximately 0.96, which means that the model can explain 96% of the variance in the calorie data. This is a very high R^2 score, indicating that the model is a very good fit for the data.

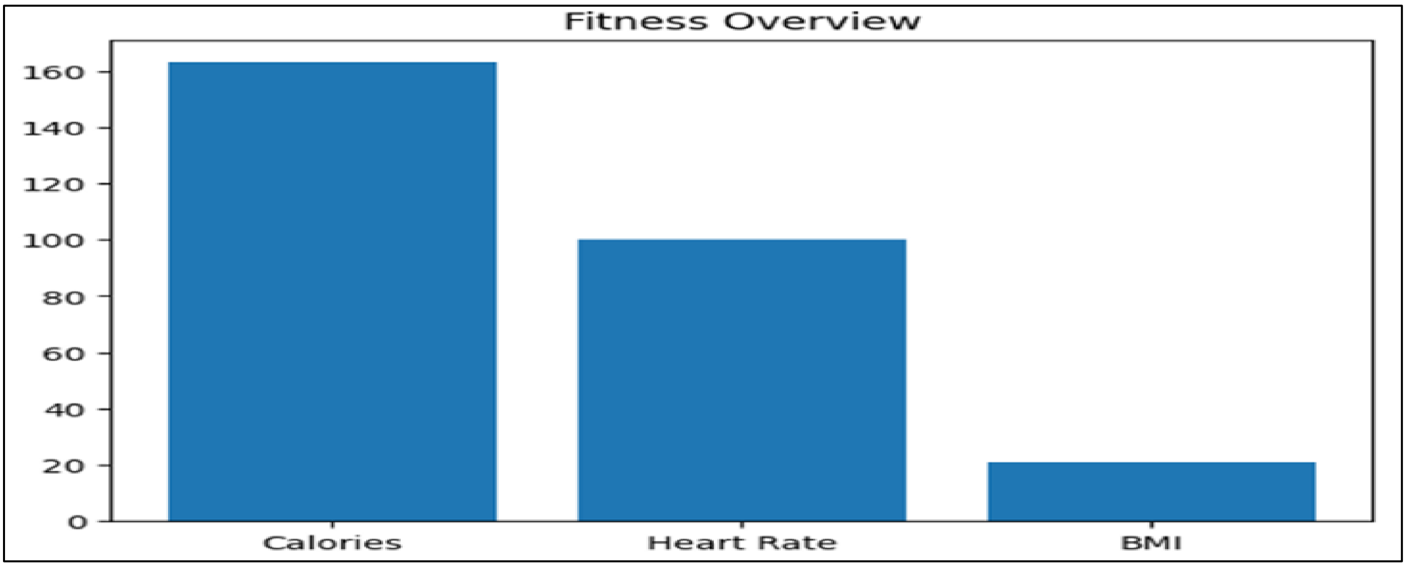

Fig 8 User Predicted Chart.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

A scatter plot of the actual vs. predicted calorie values was created to visually assess the model's performance seen in the Fig.8 . The plot showed a strong linear relationship, with most of the data points lying close to the diagonal line, further confirming the model's high accuracy.

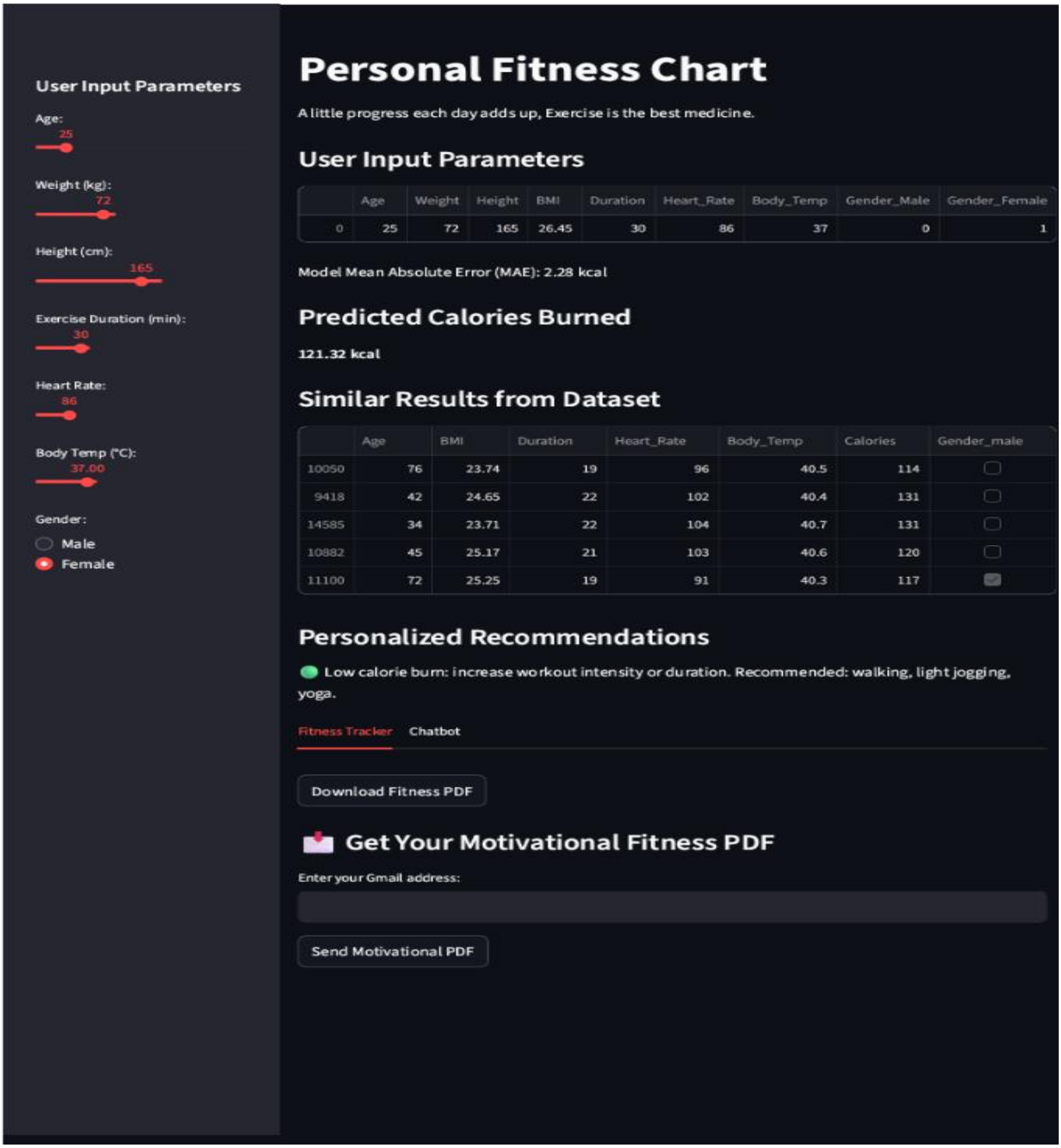

Fig 9 Live Demonstration of the Personal Fitness Chart Application.

Personal Fitness Chart application provides a seamless and interactive user experience as shown in the Fig.9 the user data and complete project is seen, beginning with data input and culminating in personalized, AI-driven fitness guidance. On the left, the user input parameters are collected through intuitive sliders and radio buttons for metrics such as age, weight, heart rate, and exercise duration. The main interface on the right immediately processes this data, displaying the machine learning model's prediction of 204.38 kcal burned. Below this, the application provides context by showing a table of similar results from the training dataset, This initial

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

output serves as the foundation for the deeper, more personalized interaction that follows.

The Gradient Boosting Regressor proves highly effective for predicting calorie expenditure. With the AI chatbot, the app delivers personalized, context-aware advice, enhancing user engagement. Users can also download conversations with charts and data, helping them track and reflect on their fitness journey.

The true innovation of the application is showcased in the integrated AI Fitness Chatbot. After receiving the initial prediction, the user can engage with the chatbot, which is powered by advanced language models like those available through OpenRouter. As demonstrated, the user has asked for a diet plan, and the AI has generated a detailed and actionable response, including specific meal suggestions for breakfast, lunch, and dinner. This moves the application beyond a simple calculator, transforming it into an intelligent fitness companion that offers tailored advice. The entire interaction, from data input to the AI conversation, is designed to be captured in a downloadable PDF, providing the user with a comprehensive and persistent record of their fitness session and the personalized guidance they received.

VI. OVERCOMING LIMITATIONS IN STATIC FITNESS PREDICTION

A. *Challenges in Existing Research and Solutions Provided*

A significant challenge highlighted in existing research is the reliance on models that, while accurate, often function as static calculators, failing to provide interactive, personalized feedback. Many systems can predict calorie burn but lack mechanisms for user engagement and actionable advice. For example, while some applications can estimate calorie expenditure, they may not accurately account for individual metabolic differences in Ref. [2] or provide tailored guidance, leaving the user with data but no clear path forward. Furthermore, the output from these systems is often transient, or in Ref. [2] so disappearing once the application is closed, which hinders long-term progress tracking.

➢ *This Project Directly Addresses these Challenges by Creating a More Holistic and Interactive user Experience:*

- Static Prediction vs. Interactive Dialogue: This project moves beyond simple prediction by integrating a generative AI chatbot via the OpenRouter and Gemini APIs. This provides users with an interactive platform where they can ask questions and receive dynamic, personalized fitness and nutrition advice based on their specific data and results.
- Generic Data vs. Actionable Insights: Instead of just presenting a number, this application uses the prediction as context for the AI to generate actionable recommendations. This transforms raw data into a practical, supportive tool that guides the user.
- Transient Results vs. Persistent Records: This application's ability to generate and download a comprehensive PDF report solves the issue of transient data. By including user inputs, prediction results, a visual chart, and the full AI conversation, it provides a persistent record that empowers users to track their journey and review advice at any time.

VII. ALGORITHM

This project uses the Gradient Boosting Regressor algorithm. Here's a detailed explanation of the algorithm, its use in this project, and the requested paragraphs.

➢ *Gradient Boosting Regressor Algorithm*

The Gradient Boosting Regressor is a powerful machine learning algorithm used for predictive modeling, specifically for regression tasks where the goal is to predict a continuous value (like calories burned). It's an ensemble method, which means it builds a single, strong predictive model by combining the strengths of multiple simpler models. Think of it like a team of experts working sequentially on a problem.

- First Guess: The process starts with a simple, initial prediction. For a regression task, this is often just the average value of the target (the average calories burned in the entire dataset).
- Calculate Errors: The algorithm then calculates the errors (called residuals) for each data point by subtracting the predicted value from the actual value.
- Learn from Mistakes: A new, simple model (a decision tree) is trained, but its job isn't to predict the calories. Instead, its goal is to predict the errors from the previous step. It focuses entirely on correcting what the first model got wrong.
- Update and Repeat: The prediction from this new "error-correcting" tree is added to the initial prediction. The process is then repeated: new errors are calculated based on the combined prediction, and another tree is trained to correct those errors.

This is done iteratively, with each new tree "boosting" the performance of the overall model by focusing on the mistakes of the ensemble that came before it.

➢ *Key Formulas*

The final prediction of the model is the sum of the initial prediction and all the sequentially trained, error-correcting trees. The core idea can be represented as:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x)$$

- $F_m(x)$ is the final, strong model after m iterations (trees).
- F_{m-1}(x) is the model from the previous iteration.
- hm(x) is the new decision tree trained to predict the errors.
- γ is the learning rate, a small number that scales the contribution of each new tree to prevent overfitting.

The error (residual) that each new tree learns to predict is calculated as:

$$r_{im} = y_i - F_{m-1}(x_i)$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

- r_{im} is the residual (error) for a single data point i at iteration m.
- y_i is the actual true value (actual calories burned).
- $F_{m-1}(x_i)$ is the prediction from the model in the previous step.

In the project, the Gradient Boosting Regressor is the intelligent engine that powers the calorie predictions.

- Task: It's used to solve a regression problem predicting the exact number of calories a person has burned.
- Inputs (Features): The model takes a user's personal and workout data as input. These features include Age, Gender, Height, Weight, Duration, Heart Rate, and Body Temperature.
- Output (Target): It processes these inputs and produces a single, continuous numerical output: the predicted calories burned.

By training on a dataset that contains both the input features and the actual calories burned, the Gradient Boosting model learns the complex, non-linear relationships between them. This allows it to make a highly accurate and personalized prediction, far superior to a simple formula.

VIII. LITERATURE SURVEY

The prediction of calorie expenditure has been a subject of considerable research, with a growing emphasis on machine learning techniques. A study by Kadam et al. (2023) focused on a model using a Random Forest Regressor, which was trained on over 15,000 data points and achieved a Root Mean Squared Error (RMSE) of 8.3. Their work underscores the potential of ensemble learning methods in this domain and also utilized Streamlit for the application interface, a choice echoed in our project.

Priscilla et al. (2024) proposed a novel hybrid learning methodology (NHLM) integrating an AutoEncoder (AE) for feature extraction with EfficientNet as the predictive model. Their approach achieved an impressive prediction accuracy of 97%, demonstrating the power of deep learning and hybrid models in enhancing prediction accuracy. Their work emphasizes the importance of feature engineering and diverse datasets to ensure model robustness.

The wider range of the related to the matter of healthcare monitoring has also seen advanced changes over the years. Arslan et al. (2024) introduced a vital sign monitoring system using K-nearest neighbors (KNN), support vector machines (SVM), and stochastic gradient descent (SGD) models to classify patient health states with high accuracy. Their research highlights the potential of machine learning in providing continuous and comprehensive patient health assessments, a concept that can be extended to fitness monitoring.

Furthermore, the evaluation of fitness movements has been enhanced through deep learning. Chang et al. (2023) proposed a system using a deep convolutional neural network (CNN) to extract features from images of fitness movements to classify and score them. This provides users with direct feedback on their form, a feature that could be integrated into future versions of our "Personal Fitness Chart".

The existing literature confirms a clear trend towards using machine learning for more accurate and personalized fitness monitoring. Our project builds upon this foundation, utilizing a Gradient Boosting Regression for its predictive power and integrating a generative AI chatbot to provide a layer of interactive, personalized advice that is currently lacking in many static prediction tools.

➢ *Kadam, A., Patil, V. H., Shrivastava, A., & Michaelson, J. (2023).*

This study addresses the challenge of estimating calorie expenditure during exercise. Using a dataset of 15,000+ records, the authors developed a Random Forest Regressor model that achieved an RMSE of 8.3, proving machine learning can outperform generic fitness formulas. Their work establishes a benchmark for data-driven calorie prediction. While limited to traditional ensemble methods, it validates the use of physiological data and provides a performance baseline for models like Gradient Boosting, as adopted in our project.

➢ *Priscilla, M., Suriya, A., et al. (2024).*

The authors propose NHLM, a hybrid method combining AutoEncoders for feature extraction with EfficientNet for regression. This deep learning approach enhances prediction accuracy by learning complex feature representations. However, the model is resource-intensive and less interpretable. The paper reflects the field's move towards advanced neural architectures but overlooks user-focused applications. For our work, it highlights both the strengths of deep models and the gap in integrating predictions into practical, user-friendly systems.

➢ *Arslan, M. M., Yang, X., et al. (2024).*

This research expands beyond fitness to healthcare monitoring, using wearable sensors to track metrics like HR, SpO_2, and body temperature. Machine learning models, particularly SGD, achieved 97% accuracy in classifying health states. Though aimed at clinical use, it validates the predictive value of physiological signals key features also central to our calorie prediction model. The study bridges clinical and consumer applications, reinforcing the scientific foundation of using biometric data for wellness monitoring.

➢ *Chang, C. C., Wei, C. H., et al. (2023).*

Instead of physiological data, this paper uses deep CNNs to analyze movement quality from video, offering real-time feedback on exercise form. It advances AI-driven fitness coaching by automating form correction but does not address energy expenditure. Its relevance lies in highlighting complementary approaches in fitness tech. Compared to our metabolic focus, this work shows opportunities for integrating biomechanical analysis with calorie prediction to create a more holistic fitness assistant.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep264

IX. CONCLUSION AND FUTURE WORK

This paper has successfully demonstrated the development of a "Personal Fitness Chart," a web-based application that synergizes a predictive machine learning model with a generative AI chatbot. The Gradient Boosting Regressor model provides highly accurate, personalized calorie expenditure predictions, while the AI chatbot, powered by the OpenRouter and Gemini APIs, offers interactive and tailored fitness advice. The high performance of the machine learning model, with an R^2 score of 0.96, validates the effectiveness of our predictive approach.

For future work, several enhancements could be considered. The dataset could be expanded to include more features, such as the type of exercise, dietary information, and sleep patterns, which is not seen in Ref. [3] or any other papers which could further improve prediction accuracy.

The AI chatbot could be enhanced with memory to recall past interactions with the user, allowing for more continuous and evolving advice. Integrating features from other research, such as real-time form correction using computer vision, could also be a promising direction. By continuing to merge predictive and generative AI with user-centric web technologies, we can create even more powerful and personalized tools to help individuals lead healthier lives.

ACKNOWLEDGMENT

I would like to convey my heartfelt regards to my guide for the support given to complete the paper and publish it. I would also like to thank my HOD of the department for giving all the support during the study and publish.

REFERENCES

[1]. Kadam, A., Patil, V. H., Shrivastava, A., Michaelson, J., Pawar, S. K., & Singh, A. (2023). Calories Burned Prediction Using Machine Learning. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 1712-1717). IEEE.

[2]. Priscilla, M., Suriya, A., Srikanth, J., Jagadhishwaran, S., Kumar, M. N., & Yasvanth, D. (2024). Evolution of Artificial Intelligence based Burned Calories Prediction System using Novel Hybrid Learning Methodology. In 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (pp. 1-7). IEEE.

[3]. Arslan, M. M., Yang, X., Zhang, Z., Rahman, S. U., Ullah, M., & Abbasi, Q. H. (2024). Advancing Healthcare Monitoring: Integrating Machine Learning With Innovative Wearable and Wireless Systems for Comprehensive Patient Care. IEEE Sensors Journal, 24(18), 29199-29210.

[4]. Chang, C. C., Wei, C. H., Wu, H. W., & Hsiao, S. (2023). A Fitness Movement Evaluation System Using Deep Learning. In 2023 15th International Conference on Knowledge and Systems Engineering (KSE) (pp. 1-6). IEEE.

[5]. Nipas, M., et al. (2022). Burned Calories Prediction using Supervised Machine Learning: Regression Algorithm. In 2022 IEEE 12th International Conference on Prototype, Circuit and System (ICPC2T) (pp. 1-4). IEEE.

[6]. Aziz, M. T., et al. (2023). CALORIES BURNT PREDICTION USING MACHINE LEARNING APPROACH. Current Integrative Engineering, 1(1), 29-36.

[7]. Lin, J., et al. (2021). Wearable sensors and devices for real-time cardiovascular disease monitoring. Cell Reports Physical Science, 2(8), 100541.

[8]. Chen, K. Y., Shin, J., Hasan, M. A. M., Liaw, J. J., Yuichi, O., & Tomioka, Y. (2022). Fitness Movement Types and Completeness Detection Using a Transfer-Learning-Based Deep Neural Network. Sensors, 22(15), 5700.