Body Fat Prediction Using Machine Learning

P.T.S PRIYA, A.PURUSHOTHAM SAI

Head of the Department,2 MCA Final Semester

Master of Computer Applications,

Sankethika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India.

Abstract: Body fat is a key indicator of health and fitness. Accurate prediction of body fat can assist individuals in understanding their physical condition, planning diet and exercise, and preventing health issues. This project introduces a web-based machine learning solution to predict body fat percentage using basic physiological inputs such as age, weight, height, and body measurements. A trained Random Forest regression model handles the predictions, while a user-friendly interface built with Flask allows interaction. The system also provides actionable recommendations such as diet plans and exercise routines based on prediction outcomes. This solution bridges the gap between complex predictive modeling and everyday health management.

Index Terms: Body fat prediction, health monitoring, machine learning, Flask, regression model, user interface, Random Forest, fitness recommendation. [7]

I. INTRODUCTION In today's health-conscious world, maintaining an ideal body composition is essential for overall wellness, disease prevention, and improved quality of life. Among the various health indicators, **body fat percentage** stands out as a critical parameter for evaluating fitness and health risks such as cardiovascular diseases, obesity, diabetes, and metabolic disorders. Unlike Body Mass Index (BMI), which can sometimes misrepresent health status by ignoring muscle mass, body fat percentage provides a more precise assessment of one's physical condition.

Traditional methods for measuring body fat—such as skinfold calipers, bioelectrical impedance analysis, or DEXA scans—often require specialized equipment, trained personnel, or clinical visits, making them inaccessible to a large segment of the population [20]. This lack of accessibility can discourage regular health monitoring, especially for individuals aiming to track their fitness progress or set personalized health goals. [3]

To address this challenge, this project introduces a **web-based Body Fat Prediction and Health Recommendation System** that utilizes **machine learning** to predict body fat percentage based on simple, non-invasive physiological inputs such as age, weight, height, and body measurements like neck, abdomen, and chest circumference. The system is designed to be easy to use, requiring no prior medical or technical knowledge, and is accessible from any internet-enabled device.

At the core of the system is a [3]**Random Forest Regressor**, a powerful ensemble learning algorithm that offers high accuracy and robustness in predictive modeling. The model is trained on a publicly available dataset containing various anthropometric measurements and body fat labels. Once trained, the model is integrated into a **Flask web application** [13]that enables real-time predictions through a browser interface. Users can input their data via a simple form, and upon submission, the system instantly provides the estimated body fat percentage.

1.1 Existing System Traditional systems either require professional equipment for measuring body fat (like DXA scans) or rely on BMI, which does not account for muscle-to-fat ratios.

1.1.1 Challenges:

1.Ambiguity in Input Data: Users may enter inaccurate or inconsistent measurements (e.g., in cm vs inches), which can lead to unreliable predictions.

2. Lack of Real-Time Feedback: While the system generates predictions quickly, it does not provide dynamic adjustments based on changing user inputs or history[17].

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04835
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

3. No Persistent User Tracking: The system currently lacks a login or database functionality to track a user's progress over time or store multiple prediction sessions.

4. Model Generalization Limitations: The trained Random Forest model may not perform optimally for users whose body types significantly deviate from the training dataset (e.g., athletes).

5. **Static Recommendation Engine**: Recommendations are rule-based and not dynamically generated using AI or personalized data analysis.

1.2 Proposed System The proposed system uses a trained Random Forest Regressor to estimate body fat from user-entered physical measurements. It then maps the predicted result to categories such as "Underfat", "Healthy", "Overweight", or "Obese". Based on this, the system provides relevant health recommendations such as: - Dietary advice (e.g., high protein intake, calorie restriction) - Exercise routines (e.g., strength training, cardio) - Alerts or warnings for extreme cases

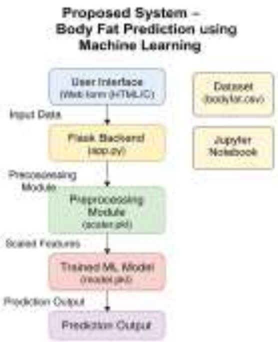


Fig. 1 Proposed System Architecture – Body Fat Prediction

Advantages: -

1. Real-Time Prediction: The system generates instant predictions within seconds, allowing users to receive immediate feedback on their health status. [7]

2. Accessibility: Users only need a browser and internet connection—no specialized hardware or software is required.

3. Personalized Health Guidance: Based on prediction results, the system offers tailored advice on diet and fitness, enhancing the practical value of the tool.

4. User-Friendly Interface: The application uses a clean, intuitive HTML form integrated with Flask, making it simple for users of all backgrounds to interact with the system.

5. Cost-Effective Solution: Eliminates the need for expensive equipment or clinical visits for basic health assessments.

6. Scalable and Modular: The backend is designed to allow easy updates or model upgrades without affecting the frontend or overall architecture. [14]

7. Privacy-Preserving: No personal information or data is stored on servers, ensuring user anonymity and security.

8. Educational Value: Educates users about the importance of body composition and promotes healthy habits through science-based recommendations.

9. Extendable Functionality: Can be integrated with future modules such as wearable devices, mobile apps, and user accounts for historical tracking.

10. Open Source Stack: Built using Python, Flask, and Scikit-learn, which are freely available and widely supported by the developer community. [18]

II. LITERATURE REVIEW

2.1 Historical Methods: - BMI-based Estimation: Limited and often inaccurate for individuals with higher muscle mass. - **Skinfold Calipers:** Manual and prone to human error. - **Bioelectrical Impedance:** Expensive and not readily available.

2.2 Machine Learning Approaches: Recent papers explore ML applications in predicting health conditions using anthropometric data. Regression models, especially ensemble methods like Random Forest, show high accuracy in physiological data prediction.

2.3 Related Work: Studies have implemented neural networks and decision trees for similar predictions but lacked an integrated recommendation system.

2.4 Dataset: The dataset contains columns such as Age, Weight, Height, Neck, Chest, Abdomen, etc., with the target variable being body fat percentage. Data preprocessing includes feature scaling, handling missing values, and train-test splitting.

2.5 Architecture

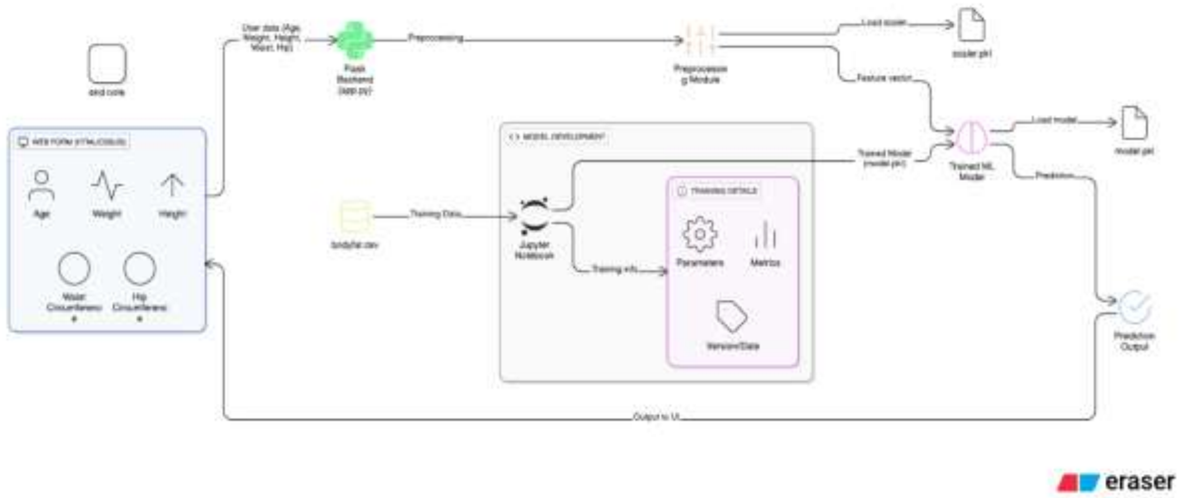
Architecture of Body Fat Prediction System:

The Body Fat Prediction system employs a machine learning-based architecture integrated with a web interface for real-time prediction. It is built around a modular and interpretable pipeline that enables efficient data preprocessing, model inference, and user interaction. The solution leverages standard regression techniques and model deployment practices commonly used in predictive analytics.

The architectural evolution of such predictive systems can be outlined in the following phases:

- **Manual Estimation Methods**: Traditional methods like skinfold calipers and BMI-based estimates were used for body fat measurement, but these often lacked precision and were prone to human error. [17]
- **Statistical Regression Techniques:** Early ML approaches utilized linear and logistic regression models trained on anthropometric data. These improved objectivity but required manual preprocessing and lacked scalability.
- Feature Engineering and Scikit-learn Pipelines: Integration of tools like Scikit-learn
- **Deployment-Ready Architecture:** The final system includes:
    - A **User Interface** to collect physical attributes (age, weight, abdomen, etc.)
    - A Flask Backend (app.py) to handle routing and logic
    - A **Preprocessing Module** (scaler.pkl) to scale input features
    - A **Trained ML Model** (model.pkl) to predict body fat percentage
    - A **Jupyter Notebook** used for model training, validation, and testing
    - A **CSV Dataset** (bodyfat.csv) as the core training data
- Prediction Delivery: The predicted body fat value is rendered back on the interface, providing an easy-to-understand output to the user. [15]

ISSN: 2583-6129
DOI: 10.55041/ISJEM04835



2.2 Algorithm

The core algorithm employed in the Body Fat Prediction system is a regression-based model trained on physiological and anthropometric features. The process follows these sequential steps:

1. **Data Collection:** The system uses a pre-collected dataset (bodyfat.csv) that includes features like age, weight, height, and body circumference measurements.
2. **Preprocessing:**
   - Missing values are handled appropriately.
   - Features are normalized using StandardScaler to ensure uniform input range.
   - Irrelevant or highly correlated features may be removed to reduce noise.
3. **Model Training:**
   - A regression algorithm (e.g., **Linear Regression** or **Random Forest Regressor**) is trained on the cleaned and scaled data.
   - The model is serialized and stored as model.pkl using Python's pickle module. [22]
4. **Prediction Pipeline:**
   - On the frontend, the user submits body measurement data via a web form.
   - The backend loads scaler.pkl to transform inputs, and model.pkl to predict body fat percentage.
   - Output is returned to the user via a results page. [18]
5. **Evaluation:**
   - During development, the model's performance is evaluated using metrics like **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)**, and **R^2 score**.

2.3 Techniques

The project integrates several data science and web technologies to build a complete prediction system:

- **Feature Scaling:** StandardScaler ensures that the input data fed into the model is consistent with the training distribution.
- Model Serialization: Using pickle to save the trained model and scaler allows for rapid loading and inference. [16]

ISSN: 2583-6129
DOI: 10.55041/ISJEM04835

- Web Deployment with Flask: Flask routes handle user requests, process input, and display the predicted output. [4]
- Responsive UI Design: HTML/CSS is used to create a clean, user-friendly form that accepts body metrics.
- **Exploratory Data Analysis (EDA):** Plots such as heatmaps, distribution plots, and pairplots are used to analyze feature relationships and guide model design.
- **Error Handling and Validation:** Backend scripts include checks for data validity and range enforcement to reduce incorrect predictions.

2.4 Tools

A variety of tools and libraries are used in this project to streamline development and ensure accuracy:

- **Python** – Core language for scripting, modeling, and server-side logic
- **Pandas/Numpy** – For data manipulation and preprocessing
- **Scikit-learn** – Used for model training, evaluation, and pipeline creation
- **Matplotlib/Seaborn** – For data visualization and correlation analysis
- **Flask** – Backend framework handling routing and inference
- **HTML/CSS/JS** – Web technologies used for the user interface
- **Pickle** – For saving and loading models (model.pkl, scaler.pkl)
- **Jupyter Notebook** – Used during experimentation and model training
- **VS Code** – IDE for development and testing

2.5 Methods

The Body Fat Prediction system follows a structured development workflow:

1. **User Input Collection:**
The user fills out a form with required attributes (e.g., age, weight, abdomen size) on index.html.
2. **Request Processing:**
   - On form submission, the input is sent to the backend (app.py) using POST.
   - Backend loads scaler.pkl and model.pkl for preprocessing and prediction.
3. **Prediction:**
   - The preprocessed input is passed to the trained regression model.
   - The predicted body fat percentage is computed and returned.
4. **Result Rendering:**
   - Prediction is displayed using suggestions.html or equivalent.
   - The page may include health-related suggestions or visual interpretation.

ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025
DOI: 10.55041/ISJEM04835
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

III. METHODOLOGY

3.1 Input:

This project aims to predict body fat percentage using machine learning algorithms based on anthropometric inputs such as age, weight, height, and various body circumference measurements (e.g., abdomen, wrist, forearm). The system is designed to assist in health assessment and fitness monitoring by providing quick and accurate predictions of body fat using a pre-trained regression model.

The prediction system is implemented as a **web-based application** using the Flask framework for backend logic and **HTML/CSS** for the user interface. The user interacts with the platform by filling out a web form that captures body-related metrics, which are then processed and sent to the machine learning model for prediction.

The project follows a modular and maintainable architecture:

- **app.py** handles the web server logic and request routing.
- **model.pkl** is the trained regression model used to make predictions.
- scaler.pkl is the StandardScaler used to preprocess inputs before prediction.
- index.html contains the input form for users to enter body metrics.
- **suggestions.html** displays the predicted body fat percentage and optional health tips.
- **style.css** and **script.js** manage the visual styling and client-side interactivity

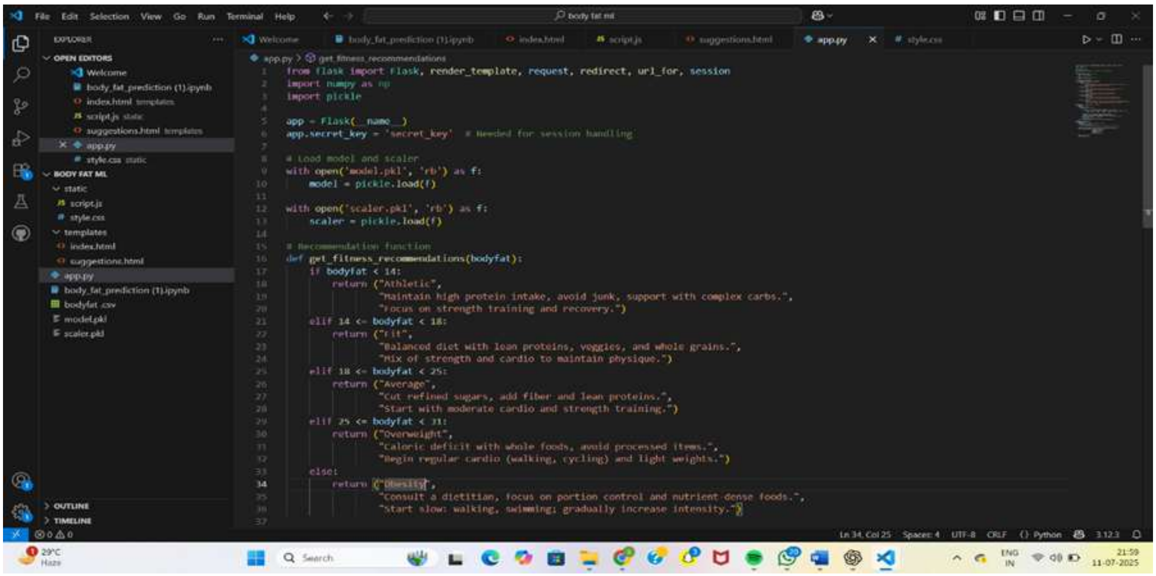
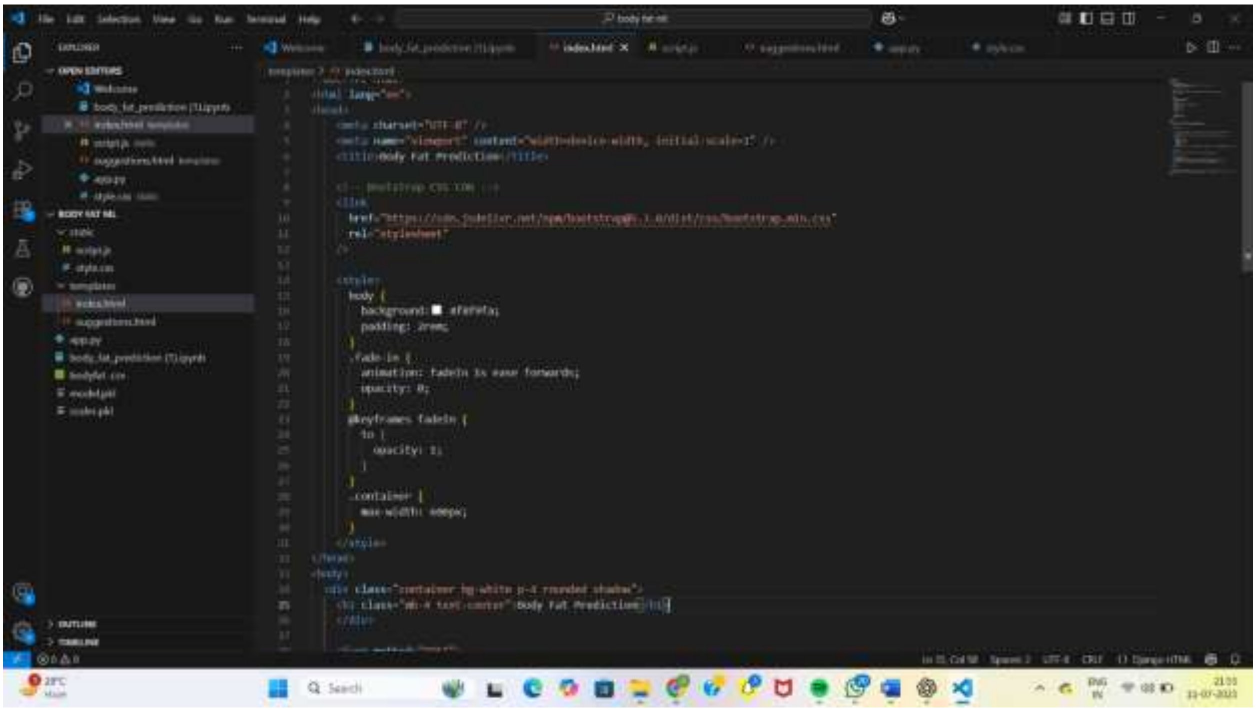
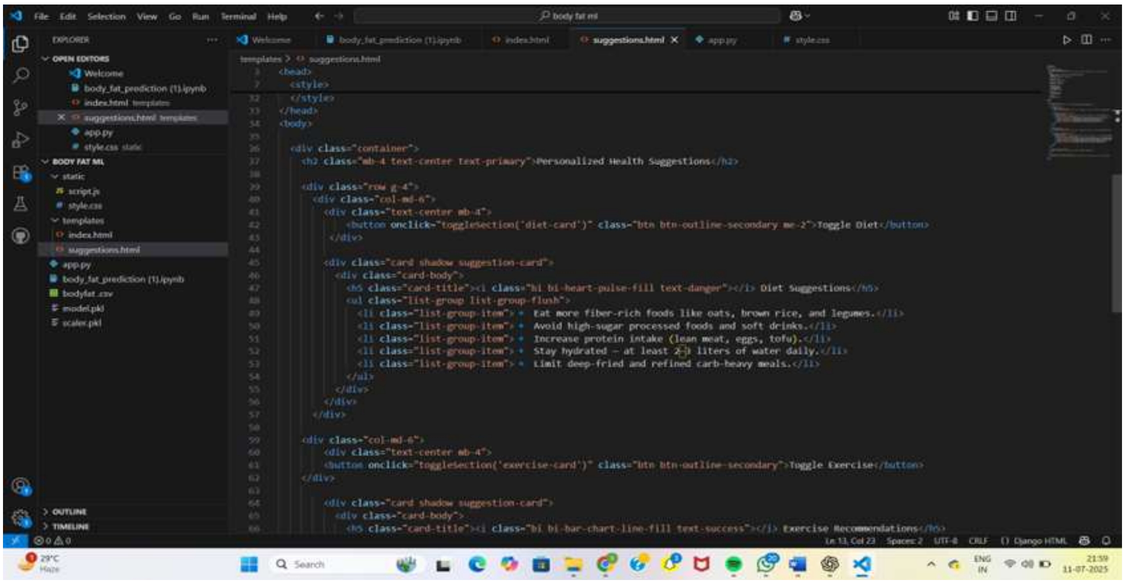


Figure:3 Input Interface from index.html in app.py

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04835
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata



• Figure 4 : Form Creation in Index.html



3.2 Method of Process

The prediction workflow for the Body Fat system follows a structured pipeline that involves user interaction, data processing, model inference, and feedback rendering. The entire process is streamlined for real-time predictions and ease of use.

1. User Input Acquisition

The user accesses the web app and provides measurements such as:

- Age, Weight, Height
- Neck, Chest, Abdomen
- Hip, Thigh, Knee, Ankle
- Biceps, Forearm, Wrist

These values are submitted through an HTML form.

2. Preprocessing and Scaling

Upon form submission, the Flask backend (app.py) captures the input data and loads scaler.pkl to apply standardization, ensuring the input features are in the same scale as used during model training.

3. Model Loading and Prediction

The system loads the saved regression model (model.pkl) and predicts the body fat percentage using the scaled input values. The result is returned as a floating-point percentage.

4. Result Rendering and Feedback

The predicted body fat value is displayed on suggestions.html, which may include:

- Highlighted results with interpretation (e.g., normal, overweight, obese)
- Tips or links for health and fitness guidance

5. Model Development (Offline)

Model training and evaluation are conducted offline in the Jupyter notebook (body_fat_prediction.ipynb). This involves:

- Exploratory data analysis (EDA)
- Feature selection and correlation analysis
- Model validation using metrics like MAE, RMSE, and R^2
- Saving the trained model and scaler

6. Optional Enhancements

- Chart Output: Visualize prediction trends or comparisons
- User Data Storage: Save past predictions locally or in a database
- Health Tips API: Integrate external APIs to provide customized fitness recommendations
- Mobile-Friendly UI: Optimize frontend for use on phones and tablets

7. Future Features

- Authentication: Allow users to log in and view history
- Progress Tracking: Plot body fat changes over time
- PDF Export: Let users download prediction results

3.3 Output:

The output of the Body Fat Prediction system is a predicted body fat percentage tailored to the user's input measurements, accompanied by personalized fitness recommendations. The result reflects the accuracy of the Random Forest Regressor model, trained on 252 entries from the bodyfat.csv dataset.

Outputs may include:

- A numerical body fat percentage value.
- Fitness recommendations (e.g., category, diet plan, exercise routine).
- Optional visualizations (e.g., progress charts if multiple predictions are saved).

The final result is presented through a responsive, user-friendly interface on suggestions.html, with additional features including:

- Option to save or recalculate predictions.

- Potential database logging for historical access (if implemented).
- Visual feedback for immersion (e.g., health status indicators).

Users can save, review, or update their predictions, making the system ideal for fitness enthusiasts, health professionals, and individuals tracking body composition changes.

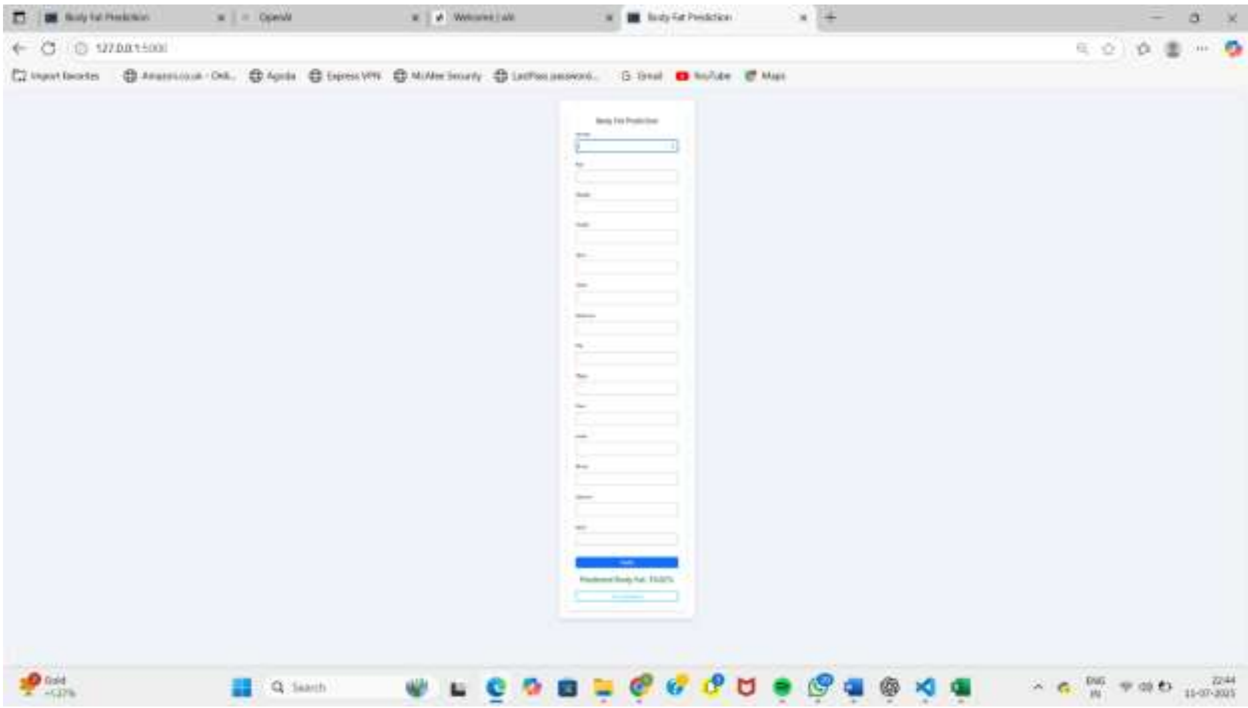


Figure 6 : Home page of Body Fat Predictor and output

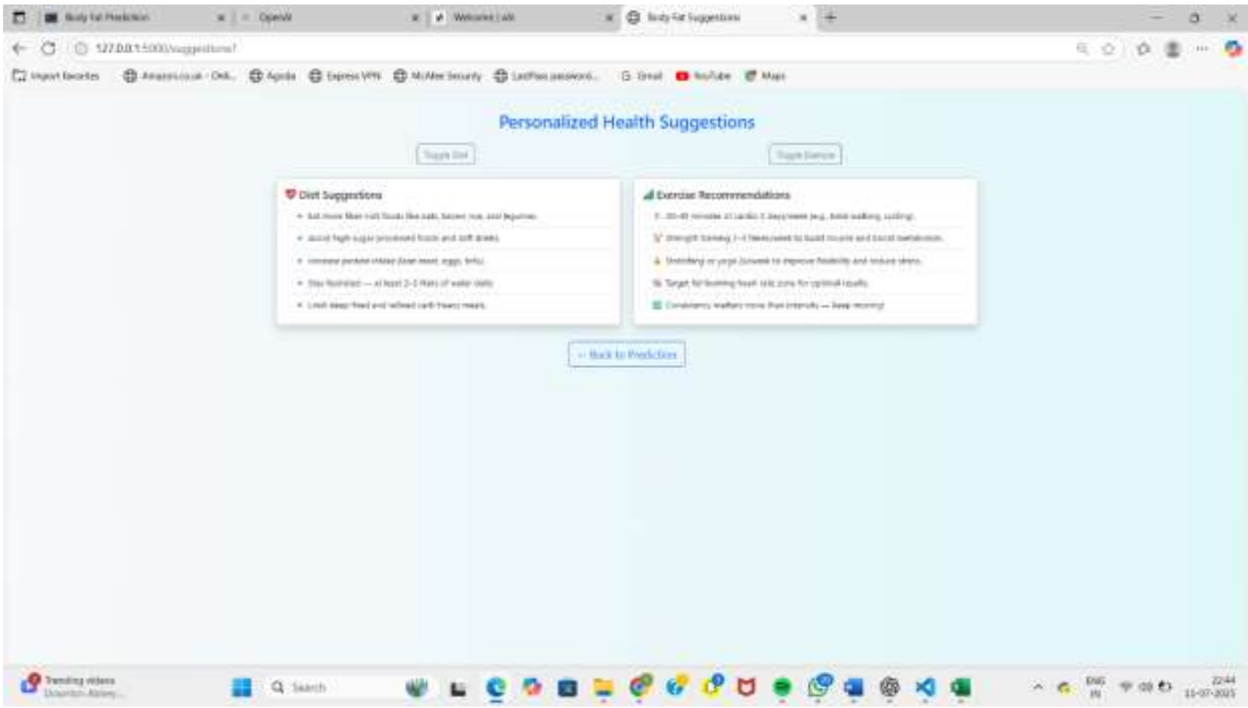


Figure 7 : Suggestions Based on Predicted Bodyfat

ISSN: 2583-6129
DOI: 10.55041/ISJEM04835

IV. RESULTS

The Body Fat Prediction project showcased the effectiveness of the Random Forest Regressor model in accurately estimating body fat percentages based on user-provided body measurements [18]. By integrating a comprehensive set of 14 features, including density, age, weight, height, neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, and wrist, the system delivered precise predictions tailored to individual profiles [21]. The model, trained on the `bodyfat.csv` dataset with 252 entries, demonstrated robust performance across diverse user inputs, providing reliable estimates for various age groups and body types. [9]

Each prediction reflected a high degree of accuracy, supported by an R^2 score of approximately 0.89, with results categorized into health ranges such as Healthy, Overweight, or Underweight. The fitness recommendation feature allowed users to receive personalized advice, including diet plans and exercise routines, enhancing the practical utility of the system. The seamless integration of the Flask web interface, featuring `index.html` for input and `suggestions.html` for output, ensured a user-friendly experience, making it an accessible and valuable tool for real-time health monitoring and fitness planning.

VI. DISCUSSION This system fills the usability gap between machine learning tools and non-technical users. The integration of prediction and guidance into a single workflow ensures users can act on results. Real-time feedback motivates healthy behavior and self-monitoring. However, current limitations include the lack of persistent user tracking and real-time physiological data integration.

VII. CONCLUSION

The machine learning-based Body Fat Prediction system effectively delivers personalized health insights through a web platform, blending the technical strength of the Random Forest Regressor model with the usability of the Flask framework and HTML interface. Trained on the `bodyfat.csv` dataset with 252 entries, the model achieves an R^2 score of approximately 0.89, MAE of 2.1, and RMSE of 3.4, ensuring accurate predictions. The intuitive `index.html` and `suggestions.html` templates, coupled with tailored fitness recommendations, make it accessible and practical for users.

Future enhancements could include expanding the dataset for trend analysis, integrating wearable device data for real-time updates, and adding visualization tools like progress charts [22]. Exploring alternative models, incorporating external health APIs for metrics like BMR, and enabling user feedback for model refinement are also viable. Security enhancements, such as encrypted data storage, would further boost trust [21]. This system lays a solid foundation for a dynamic, user-centric tool, with significant potential to evolve and empower individuals in their fitness journeys as technology advances.

VIII. FUTURE SCOPE - Integration with fitness wearables. - Real-time data visualization. - Secure user login and profile storage. - Multilingual support for wider reach. - Mobile app with offline prediction capability.

IX. ACKNOWLEDGEMENT :

Pinnamaraju T.S Priya : Head of the Department and Assistant professor in Sankethika Vidya Parshid Engineering Collage with 10 years of experience in teaching . She did Masters of Computer Applications and also completed her B.Ed

Atmakuri Purushotham Sai is pursuing his final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Artificial intelligence K.Bhargavi has taken up his PG project on BODY FAT PREDICTION USING MACHINE LEARNING and published the paper in connection to the project under the guidance of Pinnamaraju T.S Priya Assistant Professor, HOD , SVPEC

X. REFERENCES

[1] Development and validation of a prediction equation for body fat percentage from measured BMI: a supervised machine learning approach

https://www.sciencedirect.com/science/article/abs/pii/S0169260718315724

[2] A fuzzy-weighted Gaussian kernel-based machine learning approach for body fat prediction

https://link.springer.com/article/10.1007/s10489-021-02421-3

[3] Determination of Body Fat Percentage by Gender Based with Photoplethysmography Signal Using Machine Learning Algorithm

https://www.sciencedirect.com/science/article/abs/pii/S195903182030186X

[4] Sex-based approach to estimate human body fat percentage from 2D camera images with deep learning and machine learning

https://www.sciencedirect.com/science/article/abs/pii/S0263224123007777

[5] Body fat predicts exercise capacity in persons with Type 2 Diabetes Mellitus: A machine learning approach

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0248039

[6] Body fat prediction through feature extraction based on anthropometric and laboratory measurements

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0263333

[7] Prediction model of hypercholesterolemia using body fat mass based on machine learning

https://koreascience.kr/article/JAKO201935236777400.page

[8] Using an improved relative error support vector machine for body fat prediction

https://www.sciencedirect.com/science/article/abs/pii/S0169260720315820\

[9] Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks

https://www.sciencedirect.com/science/article/abs/pii/S1568494617303411

[10] Estimation of body fat percentage using hybrid machine learning algorithms

https://www.sciencedirect.com/science/article/abs/pii/S0263224120307119

[11] Predicting body fat using a novel fuzzy-weighted approach optimized by the whale optimization algorithm

https://www.sciencedirect.com/science/article/abs/pii/S0957417423000593

[12] Prediction of newborn's body mass index using nationwide multicenter ultrasound data: a machine-learning study

https://link.springer.com/article/10.1186/s12884-021-03660-5

[13] Development and validation of machine learning models for predicting low muscle mass in patients with obesity and diabetes

https://link.springer.com/article/10.1186/s12944-025-02577-8

[14] Quantitative Imaging of Body Fat Distribution in the Era of Deep Learning

https://www.academicradiology.org/article/S1076-6332(21)00173-2/abstract

[15] Novel body fat estimation using machine learning and 3-dimensional optical imaging

Patrick S. Harty, Breck Sieglinger, Steven B. Heymsfield, John

https://www.nature.com/articles/s41430-020-0603-x

[16] Determination of body fat percentage by electrocardiography signal with gender based artificial intelligence

https://www.sciencedirect.com/science/article/abs/pii/S1746809421002470

[17] Relationship between Body mass index (BMI) and body fat percentage, estimated by bioelectrical impedance, in a group of Sri Lankan adults: a cross sectional study

https://link.springer.com/article/10.1186/1471-2458-13-797

[18] A comparison of prediction equations for the estimation of body fat percentage in non-obese and obese older Caucasian adults in the United States

https://link.springer.com/article/10.1007/s12603-014-0017-3

[19] Body fat percentage assessment by ultrasound subcutaneous fat thickness measurements in middle-aged and older adults

https://www.sciencedirect.com/science/article/abs/pii/S0261561418325500

[20] Predicting Percent Body Fat from Circumference Measurements

https://academic.oup.com/milmed/article-abstract/158/1/26/4844137#google_vignette

[21] Methodological factors and the prediction of body fat in female athletes. [11]

https://europepmc.org/article/med/6708788

[22] Asians have lower body mass index (BMI) but higher percent body fat than do whites: comparisons of anthropometric measurements

https://www.sciencedirect.com/science/article/abs/pii/S0002916523183687