© April 2025| IJIRT | Volume 11 Issue 11 | ISSN: 2349-6002

Fat Stat: Body Fat Percentage Estimation Using Machine Learning and Computer Vision

P Spandana Valli[1], Sappa Anudeep[2], Dommeti Charmi Sri Tulasi[3], Gattu Nikhilesh[4], Datti Bhavishya Sree[5], Valeri Adinarayana[6]

[1,2,3,4,5,6]*Department of CSE ANITS, Visakhapatnam, India*

***Abstract*—Accurate body fat assessment is vital for individualized health and exercise planning. Traditional methods relying solely on static data such as height, weight, and circumferences of- ten overlook individual variations in body shape and fat distribution, leading to imprecise estimates. To address these limitations, we propose a novel approach that combines machine learning with real-time body image analysis. Our system captures user inputs—including height, weight, age, gender, and specific body part measurements—and employs Random Forest Regression to predict body fat percentage from dimensions extracted via a calibrated camera. This integration of dynamic image processing and model-based predictions yields a robust performance with an accuracy of 93%, significantly outperforming conventional techniques. Moreover, the system provides personalized dietary and exercise recommendations, ensuring high adaptability to diverse body types. Overall, this cost- effective and scalable hybrid approach offers a user-friendly alternative for more precise body fat estimation.**

***Index Terms*—Body Fat Estimation, Machine Learning, Computer Vision, OpenCV, Random Forest**

I. INTRODUCTION

Body fat percentage estimation is a fundamental aspect of clinical assessment, fitness tracking, and health management. Obesity is a pressing public health issue, associated with increased risks of diabetes, heart disease, and other chronic illnesses. Accurate body fat measurement is vital for early detection and management of obesity-related health risks. Traditional approaches such as BMI calculations and linear regression models, although widely used, lack precision as they do not account for muscle mass variations and unique fat distribution. Advanced methods like Dual-energy X-ray Absorptiometry (DEXA) and hydrostatic weighing provide accurate measurements but are impractical due to their cost and complexity.

Recent advancements in machine learning have introduced more accurate and scalable alternatives for body fat prediction. This study employs Random Forest Regression combined with real-time image processing to address the limitations of conventional techniques. By integrating user-specific at- tributes and image-based measurements, the system enhances accuracy and provides dynamic predictions. The objective of this research is to develop a model that not only predicts body fat percentage but also offers personalized dietary and exercise recommendations. This approach is suitable for fitness enthusiasts, healthcare professionals, and researchers seeking an efficient, adaptable, and affordable solution.

Machine learning models, particularly Random Forest Regression, have demonstrated superior predictive capabilities by capturing complex, non-linear relationships between variables. This model uses multiple decision trees to improve accuracy and reduce overfitting. Real-time body image analysis further enhances the system's capability to adapt to various body types. By combining these techniques, our research aims to create a user-friendly, scalable system that provides fast and reliable body fat estimates.

Additionally, this system has significant applications in clinical and fitness environments. Medical practitioners can use it for early detection of obesity-related risks, while individuals can monitor their health and receive personalized lifestyle recommendations. This approach addresses the limitations of static models and offers an innovative solution for real-time body fat estimation. Our research bridges the gap between advanced machine learning techniques and practical health applications, providing an accurate and adaptable method for body fat prediction.

II. RELATED WORKS

Several studies have explored machine learning techniques for body fat prediction. Fan et al. (2022) proposed a feature extraction framework using anthropometric and laboratory measurements, applying Linear Regression and Random For- est models to predict body fat percentage. Their findings emphasized the potential of machine learning in enhancing accuracy through feature engineering and model optimization.

Alves et al. (2021) introduced a gender-based machine learning approach for body fat estimation using a dataset of 163 individuals. Their model, incorporating Random Forest and Extreme Gradient Boosting, achieved reliable predictions, highlighting the significance of gender-specific considerations in body fat assessment.

Metshein et al. (2024) utilized bioelectrical impedance analysis to establish correlations between body mass index and electrical properties of body tissues. This study demonstrated the potential of alternative methodologies in improving the accuracy of non-invasive body fat estimation.

Mahesh et al. (2023) compared various regression models for body fat prediction using a dataset of 252 participants. Random Forest Regression emerged as the most accurate model, achieving an RMSE of 0.276. This study reinforced the effectiveness of ensemble learning techniques in predicting body fat.

Carletti et al. (2018) proposed an innovative framework using depth images and deep learning techniques for body fat estimation. Their model, based on ResNet-50, demonstrated high accuracy, especially when trained on a diverse dataset of body scans.

Building upon these prior works, our research combines Random Forest Regression with real-time image analysis and camera calibration to deliver a robust, adaptable, and precise solution for body fat prediction and personalized health recommendations.

III. SYSTEM ARCHITECTURE

The proposed Fat Stat system is designed as an integrated framework that combines advanced machine learning with computer vision to accurately estimate body fat percentage in real-time. The architecture is organized into several interlinked modules:

A. Data Acquisition Module

This module collects both anthropometric and image-based data. Users input key measurements such as height, weight, age, and gender through a web interface, while a calibrated camera captures a series of images. A reference image is first acquired to determine the pixel-to-centimeter conversion factor, followed by multiple images (e.g., front, back, left, and right) to cover the targeted body region comprehensively.

B. Image Processing Module

Captured images are preprocessed to enhance measurement accuracy. The key steps include:

- Camera Calibration: Utilizing reference images, intrin- sic and distortion parameters are computed to undistort subsequent images.
- Noise Reduction: Gaussian blur is applied to reduce image noise.
- Edge Detection: Canny edge detection highlights prominent edges.
- Contour Extraction: Contour detection algorithms de- lineate the boundaries of specific body parts, enabling the extraction of accurate dimensional features.

The pixel-to-centimeter ratio derived from the reference image is then used to convert the extracted measurements into real- world units.

C. Feature Integration and Prediction Module

Image-based features (e.g., body part width and circumferences) are combined with the user's numerical inputs to form a comprehensive feature vector. This vector is fed into a pre-trained Random Forest Regression model, which predicts the body fat percentage based on learned patterns from a labeled dataset. Image-based features (e.g., body part width and circumferences) are combined with the user's numerical inputs to form a comprehensive feature vector. This vector is then fed into a trained Random Forest Regression model with 100 trees and a maximum depth of 2, which predicts body fat percentage based on learned patterns from a labeled dataset. This approach ensures robust and accurate predictions.

D. System Integration and User Interface

The backend, implemented using Flask, manages data flow between the user interface, the image processing pipeline, and the prediction module. MySQL handles data storage, while Python libraries such as OpenCV, NumPy, Pandas, and Scikit-learn enable efficient computation. Real-time processing capabilities ensure that users receive immediate feedback, with visualization overlays (e.g., detected contours) displayed via the interface.

IV. METHODOLOGY

The methodology for Fat Stat comprises several stages, ensuring both precision and real-time responsiveness in body fat estimation.

A. Data Collection and Preprocessing

User data is acquired in two primary forms:

1) **Anthropometric Measurements:** Users provide height, weight, age, and gender.
2) **Image Capture:** A calibrated camera is used to capture a reference image and additional images from multiple angles covering the target body part.

The reference image establishes the pixel-to-centimeter scaling factor, while subsequent images undergo a preprocessing pipeline:

- **Camera Calibration:** Calibration images are used to generate camera matrices and distortion coefficients.
- **Gaussian Blurring:** Smoothing filters reduce high-frequency noise.
- **Edge Detection:** The Canny algorithm extracts significant edges, preparing the image for contour analysis.

B. **Contour Detection:** Contours corresponding to the boundaries of the target body region are identified, and dimensional measurements are extracted.

C. Feature Extraction and Integration

The system computes the physical dimensions (e.g., body part width) by applying the pixel-to-centimeter ratio. These image-derived features are then normalized and combined with the user's numerical data to form an input feature vector. Feature scaling, such as Min-Max normalization, is performed to standardize the values across different units.

D. Machine Learning Model Training and Prediction

The integrated dataset is used to train a Random Forest Regression model. The training phase involves:

- **Hyperparameter Tuning:** A grid search is employed to determine the optimal number of trees, maximum depth, and minimum samples per leaf.
- **Cross-Validation:** The dataset is partitioned into training and testing sets (e.g., 80% training and 20% testing) to ensure the model generalizes well.
- **Performance Evaluation:** The model is evaluated using Mean Squared Error (MSE) and R-squared (R^2) metrics.

Once trained, the model is deployed to predict body fat percentage in real time by processing the combined feature vector.

E. Real-Time Processing and Output Generation

During operation, the system continuously captures and processes images. The user can adjust their posture prior to image capture, ensuring optimal data quality. The processed features are immediately fed into the prediction model, and the output is compared with established health thresholds to generate personalized dietary and exercise recommendations. Visualization modules provide real-time feedback by overlaying detected contours and measurement data on the user's image.

F. Implementation Environment

The system is developed in Python, utilizing:

- **Flask 3.1.0** for backend web services.
- **MySQL-Connector 2.2.9** for database management.
- **OpenCV 4.11.0.86** for image processing.
- **NumPy 2.0.2** and **Pandas 2.2.3** for numerical computation and data handling.
- **Joblib 1.4.2** for efficient model serialization.
- **Scikit-learn 1.6.0** for implementing and tuning the Random Forest model.

This integrated environment ensures efficient real-time processing and robust body fat prediction.

V. EXPERIMENTAL SETUP

1) Hardware: The system captures images using a standard camera setup for accurate body measurement. Proper camera calibration is performed to correct distortions and ensure precise measurements. The processing unit comprises a computer with sufficient computational power to handle real-time image processing and feature extraction.

2) *Software:* The system is developed in Python and utilizes essential libraries including OpenCV, NumPy, Pandas, Joblib, and Scikit-learn. OpenCV is responsible for image processing, edge detection, and contour extraction. Joblib facilitates efficient loading of the machine learning model, while Pandas is used for handling input and output data.

3) *Camera Calibration:* Camera calibration is initiated using a set of reference images to generate camera matrices (cmtx.npy and dmtx.npy). These matrices are stored and subsequently used to undistort images, ensuring accurate representation of body measurements.
4) *Body Measurement Estimation:* The system processes input images by applying edge detection techniques to extract contours representing different body parts. The detected contours are analyzed to compute body part circumferences, which are later used as input features for the trained model.
5) *Machine Learning Model:* A pre-trained model (new-trainedmodel.pkl) is loaded using Joblib. This model predicts body fat percentage based on extracted features such as height, weight, and body part circumferences. The model is trained on a dataset containing key anthropometric measurements.
6) *Real-Time Processing:* The system captures real-time
images and allows users to adjust their posture before the image is processed. The user's height in pixels is determined by manually selecting reference points, and a pixel-to-cm ratio is calculated to ensure accurate estimations. Edge detection and segmentation techniques are applied to extract relevant contours.
7) *Performance Metrics and Visualization:* Performance evaluation includes: - Frame Processing Speed: Measuring the frames per second (FPS) for system efficiency. - Circumference Calculation: Systematically analyzing extracted contours to estimate body part circumferences. - Model Accuracy: Validating predictions against ground-truth values to ensure reliability.
Visualization tools such as OpenCV overlays display detected contours and real-time feedback for user guidance.
8) *Testing and Evaluation:* The system undergoes rigorous testing with various input images under different lighting conditions and postures. Performance is assessed based on prediction accuracy, processing speed, and usability. Manual validation is performed by comparing predicted values with known reference measurements.
9) *System Termination:* The system ensures controlled shutdown using appropriate termination functions, preventing abrupt interruptions and ensuring data integrity. The os.kill method is utilized for stability and reliability.

VI. FUTURE SCOPE

This study successfully demonstrates the use of Random Forest Regression to predict fat percentage for specific body parts based on user inputs and body part measurements. The model's high accuracy and interpretability make it suitable for practical applications in health and fitness. Future work will focus on expanding the dataset to include more diverse samples and improving measurement techniques for body part width.

A. *Integration with Wearable Technology*

Future advancements can include integration with smart- watches and fitness trackers to provide real-time body fat analysis.

Continuous monitoring will allow users to track changes over time and receive personalized health recommendations.

B. *AI-Powered Predictive Analysis*

Machine learning algorithms can be used to predict body fat trends based on user habits, including diet, exercise, and lifestyle.

AI-powered suggestions can help users maintain a healthy body composition and reduce the risk of obesity-related dis- eases.

C. *Mobile and Cloud-Based Applications*

Developing a cloud-based platform will allow users to access and store their body fat percentage data securely.

A mobile application with an intuitive user interface can enhance user experience and make health tracking more accessible.

D. *Integration with Healthcare Systems*

Future versions of the system can be integrated with hospital databases and Electronic Health Records (EHRs) to assist medical professionals.

This will help in diagnosing and treating obesity-related conditions more effectively.

VII. RESULTS

A. *Predicted Body Fat Percentage for Specific Body Parts*

The FatStat model predicts body fat percentage for specific body parts using a Random Forest Regressor. The model analyzes body measurements—such as height, weight, and circumference of targeted body regions—to estimate fat distribution in localized areas.

A higher predicted fat percentage in a body part suggests greater fat accumulation in that region, while a lower value indicates a leaner composition.

B. *Model Performance for Body Part-Specific Fat Estimation*

The Random Forest Regressor was evaluated using Mean Squared Error (MSE) and R^2 score for both training and testing datasets:

- Training Set Performance:
- MSE: 3.42
- R^2 Score: 0.935
- Testing Set Performance:
- MSE: 2.25
- R^2 Score: 0.956

These results indicate that the model effectively learns patterns for predicting fat percentage in specific body parts, achieving high accuracy with minimal error.

C. *Visualization and Interpretation*

Predicted vs. Actual Body Fat Percentage for Specific Body Parts The scatter plot below compares actual and predicted localized fat percentage values. The red dashed line represents a perfect 1:1 correlation, while the blue points depict model predictions. The close clustering of points around the line suggests strong predictive accuracy.

REFERENCES

[1] N. Mahesh, P.B. Pati, K. Deepa, and S. Yanan, ”Body Fat Prediction using Various Regression Techniques,” in *Proc. 2nd IEEE Int. Conf. on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2023.

[2] R. Araujo, et al., ”Gender-based approach to estimate the human body fat percentage using Machine Learning,” in *2021 IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, 2021.

[3] M. Fujihara, et al., ”A novel prediction method for body fat by using Choquet integral regression model,” in *2009 IEEE Int. Conf. on Fuzzy Systems*, 2009.

[4] S. Hussain, et al., ”A Fuzzy Adaptive Binary Global Learning Colonization-MLP for Body Fat Prediction,” in *2019 IEEE Congr. on Evolutionary Computation (CEC)*, 2019.

[5] A. Singh and H. Tawfik, ”Obesity Risk Prediction Using Machine Learning Approach,” in *2023 IEEE Int. Conf. on Artificial Intelligence and Computer Applications (ICAICA)*, 2023.

[6] Y. Xu, et al., ”3D Shape-based Body Composition Prediction Model Using Machine Learning,” in *2018 IEEE Int. Conf. on Big Data (Big Data)*, 2018.

[7] Z. Fan, et al., ”Comparison Of Different Machine Learning Methods Applied to Obesity Level Estimation,” in *2022 IEEE Int. Conf. on Big Data (Big Data)*, 2022.

[8] L. Shao, et al., ”Estimate Height Weight and Body Mass Index From Face Image Using Machine Learning,” in *2023 IEEE Int. Conf. on Image Processing (ICIP)*, 2023.

[9] A. Kupusinac, et al., ”Obesity Level Estimation based on Machine Learning Methods and Artificial Neural Networks,” in *2021 IEEE Int. Conf. on Electrical, Computer and Energy Technologies (ICECET)*, 2021.

[10] F. Ferdowsy and H.R. Pourghasemi, ”Fetal Weight Estimation via Ultrasound Using Machine Learning,” in *2019 IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, 2019.

[11] N.M. Nascimento, et al., ”New Advances in Body Composition As- sessment with ShapedNet: A Single Image Deep Regression Approach,” arXiv preprint arXiv:2310.09709, 2023.

[12] T.T. Mueller, et al., ”Body Fat Estimation from

Surface Meshes using Graph Neural Networks," arXiv preprint arXiv:2308.02493, 2023.
[13] S. Chhatkuli, et al., "Body Composition Estimation Based on Multimodal Multi-task Deep Neural Network," arXiv preprint arXiv:2205.11031, 2022.
[14] F. Keivanian, et al., "A fuzzy adaptive evolutionary-based feature selection and machine learning framework for single and multi-objective body fat prediction," arXiv preprint arXiv:2303.11949, 2023.
[15] OpenCV documentation, https://opencv.org/
[16] Scikit-learn Random Forest, https://scikit-learn.org/
[17] DEXA Scan Research, Journal of Body Composition Studies, 2023.