**Body Fat Prediction**

Advanced Computer Science Masters Project

Abdul Jaleel Mohammed

22090668

Maria Psarrou

**Title Page**  
Please use the template provided in the assignment.

**Abstract**

The abstract should be a statement up to half a page in length describing the  
subject matter of the project report and the main findings and conclusions  
presented in the report. A reader should be able to decide what the report is  
about by reading this alone, so be clear and concise.

**Acknowledgements (if any)**

I am very pleased to present this report. This period of my student life has been truly rewarding. Several people were of immense help to me during the course of my research throughout this period of research.

I would like to thank my guide, **Dr. Maris Pssarou, Lecturer, Department of Computer Science, University of Hertfordshire, Hatfield,** for her invaluable guidance and support as my supervisor. Her insights and regular guidance were invaluable to me throughout my research.

I would like to express my deep sense of gratitude to **Mike Watkins, Lecturer, Department of Computer Science, University of Hertfordshire**, **Hatfield,** for his useful lectures, which gave me a clear understanding of the project.

**MSc Final Project Declaration**  
This declaration is about whether you had any human participants or not.  
Please use the template provided in the assignment.

**Table of Contents**

1. **Introduction of the Project**1.1 Background  
   1.2 Research Question  
   1.3 Objectives  
   1.4 Significance of the Study
2. **Literature Review**2.1 Overview of Machine Learning Models  
   2.2 Related Studies on Body Fat Prediction  
   2.3 Hybrid Model Approaches  
   2.4 Gaps in the Literature
3. **Methodology**3.1 Data Collection and Preprocessing  
   3.2 Feature Selection and Engineering  
   3.3 Model Design (Individual and Hybrid Models)  
   3.4 Evaluation Metrics
4. **Results and Analysis**4.1 Individual Model Performance  
   4.2 Hybrid Model Performance  
   4.3 Comparative Analysis  
   4.4 Statistical Validation
5. **Discussion**5.1 Interpretation of Results  
   5.2 Strengths and Limitations of the Hybrid Model  
   5.3 Implications for Health Informatics
6. **Conclusion and Recommendations**6.1 Summary of Findings  
   6.2 Recommendations for Future Research  
   6.3 Practical Applications
7. **References**
8. **Appendices**
   * A: Python Code and Scripts
   * B: Data Description
   * C: Additional Figures and Tables

**Introduction to the Project**

Obesity is defined as having an excessive amount of body fat. In the UK, it is estimated that approximately one in four adults and one in five children aged 10 to 11 are affected by obesity. The Body Mass Index (BMI) is the most commonly used tool for identifying obesity, with a BMI range of 30 to 39.9 indicating obesity. However, for individuals of Asian, Chinese, Middle Eastern, Black African, or African-Caribbean descent, a BMI of 27.5 or higher is used to classify overweight and obesity. Obesity can significantly increase the risk of developing various serious health conditions, including high blood pressure, asthma, and cancer (NHS, 2023).

The human body comprises extracellular fluid, bone, fat, and muscle cells. A harmonious balance in the proportion of these components ensures an ideal body composition. Fat tissue is classified into two types: white and brown. Brown fat tissue, primarily found in newborns, plays a crucial role in generating heat. In contrast, white fat tissue acts as the body's largest energy reserve, storing surplus energy for future use. In addition to adipocytes, which are fat cells, fat tissue also includes other cell types such as fibroblasts, preadipocytes, and macrophages (Uçar and Uçar, 2020; Sümer, 2014).

White fat tissue constitutes 15–20% of the body. It is found in two different places, under the skin and around the internal organs. Brown fat tissue is about 4% of newborn [body weight](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/body-weight) (Polat, 2017). Fat tissue is considered an endocrine organ with known heat and energy regulation functions as well as secretory feature (Uçar and Uçar, 2020).

In healthy individuals, body fat percentages should fall between 25-30% for women and 18-23% for men. Women with more than 30% body fat and men with more than 25% are classified as obese (Penn Medicine, no date).

Feature extraction has been widely used in the medical area to map redundant, relevant, and irrelevant features into a smaller set of features from the original data (Dara *et al*., 2018). For example, Das, Naik and Behera (2020) applied feature extraction methods to extract significant features from the raw data before using an Artificial Neural Network (ANN) model for medical disease classification. Their results showed that feature extraction methods could increase the accuracy of diagnosis.

This project implements a unique approach to determining obesity by creating machine-learning models that estimate body fat percentage using physical measurements such as weight, abdomen, chest, hip circumferences, and density determined from underwater weighing.

**1.2 Research Question:**

1.) "Can machine learning models accurately predict body fat percentage in humans with multiple anthropometric measurements and density determined from underwater weighing to represent they are obese?"

2.) Does the Hybrid Machine Learning model obtain higher accuracy rates compared to individual models for predicting the body fat percentage?

* 1. **Objectives:**
* Gather anthropometric measurements (e.g., waist, neck, abdomen, chest, and hip circumferences) and body density data.
* Clean, normalize, and prepare the dataset to ensure accuracy and reliability for analysis.
* Train machine learning models such as Decision Trees, Random Forests, and MLP to predict body fat percentage.
* Perform hyperparameter tuning to optimize model performance and improve accuracy.
* Develop a hybrid model by combining predictions from individual models to improve accuracy.
* Evaluate and compare model performance using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R² score, etc.
* Use predicted body fat percentages to identify individuals who are obese based on standard thresholds and evaluate the models’ accuracy in obesity classification.
* Validate the models' reliability and generalization using techniques like cross-validation and statistical testing.
* Assess whether the hybrid model significantly outperforms individual models.
* Explore the potential of the models to serve as tools for early obesity detection and health monitoring in real-world applications.

**1.4 Significance of the Study:**

**Chapter 2: Literature Review: =>**

**(Supervised Machine learning concepts, reference about different ML models, used, with formulas)**

**Chapter 3 : Methodology =>**

1. **Data collection => description.(Tables and charts about data.-statistical info)**
2. **Spearman’s correlation as well.**
3. **Standard scalar technique.**
4. **Feature selection algorithm(proper referencing). Write Each and every feature sub-section.(tables here as well)**
5. **Machine learning models.(MLP, SVR, RandomForest, Regression about it)**

**Chapter 4: Evaluation & Analysis:**

* RMSE, MAE, RMSE, MAD, SH, R squared, MSE.

**Chapter 5: Discussion:**

**Write about the bodyfat estimation is useful, talk about without vs. all anthropometric measurements were used. Write the different values.**

**Purpose of these values.**

**Chapter 6: Conclusions**

**About the methods already available limitations.**

**System based on BPF that can calculate BFP with anthropometric measurements.**

**Bibliography**

* Dara S, Tumma P, Eluri NR, Kancharla GR. (2018) “Feature Extraction In Medical Images by Using Deep Learning Approach,” *International Journal of Pure and Applied Mathematics*, 120(6), pp. 305–312.
* Das, H., Naik, B. and Behera, H.S. (2020) “Medical disease analysis using Neuro-Fuzzy with Feature Extraction Model for classification,” *Informatics in medicine unlocked*, 18(100288). Available at: <https://doi.org/10.1016/j.imu.20100288>.
* Polat, I.G., 2017. *Effect of Er stress and Sik2 Reciprocal relationship on human precursor fat cell (LiSa-2) differentiation* (Doctoral dissertation, Doktora Tezi, Gebze Teknik Üniversitesi, Kocaeli, Türkiye).
* Sümer, A., 2014. In vitro investigation of the effect of topiramate on fat cell differentiation and some carbonic anhydrase isoenzymes gene expression.
* Uçar, M.K. and Uçar, Z., 2020. Body fat amount/percentage calculation methods: Systematic review. *J. Inst. Sci. Technol*, *10*, pp.930-943.