Contributing Factors to BMI categories @ UU

Andreas Alexandrou, Nikolas Stavrou, Sotiris Zenios, Tiago Carriço
{a.alexandrou,s.zenios,n.stavrou,t.dossantossilvapeixotocarrico}@students.uu.nl
Utrecht University
Utrecht, the Netherlands

ABSTRACT

Obesity is a significant public health challenge globally, with a rising prevalence in various countries. Understanding the underlying factors contributing to different BMI categories is crucial for developing effective interventions. This project explores various explainability methods applied to different machine learning models to identify key features influencing BMI categories in individuals.

We implemented Decision Trees, Random Forests, LightGBM, RuleFit, Neural Networks, and Partial Dependence Plots to analyze the data. Our approach included plotting feature importances, visualizing model structures, and pruning the depth of trees for the Decision Trees and Random Forests. Additionally, we generated counterfactual explanations to provide insights into how slight changes in input features could lead to different obesity outcomes.

The results indicate that the consumption of vegetables, Age, and time spent using technological devices are the most relevant features in most models. RuleFit provided a rule-based interpretation and Counterfactual explanations highlighted that changes in specific features could potentially reduce BMI, offering recommendations for individuals.

By leveraging these diverse explainability methods, we provide actionable insights into the critical factors contributing to obesity. These findings can inform targeted interventions and policies aimed at improving health outcomes in these populations.

ACM Reference Format:
Andreas Alexandrou, Nikolas Stavrou, Sotiris Zenios, Tiago Carriço. 2024. Contributing Factors to BMI categories @ UU. In *Proceedings of Utrecht University (INFOMHCML'2023)*. Utrecht University, 8 pages.

1 INTRODUCTION

The increasing rates of obesity pose a challenge to public health systems worldwide, leading to higher incidences of related diseases such as diabetes, cardiovascular conditions, and certain cancers. The complexity of obesity, influenced by a plethora of factors including genetics, lifestyle and environment requires a deep analysis of its determinants. The problem we address in this project is identifying the key factors that most significantly contribute to BMI in populations from Colombia, Peru, and Mexico. We aimed to achieve a deeper understanding by leveraging machine learning models and explainability methods, where we are going to compare and analyze the level of interpretability and accuracy those models achieve. The work, therefore aims to inform targeted and effective interventions to mitigate obesity and its associated health risks.

INFOMHCML'2023, April 2023, Utrecht, the Netherlands

ACM ISBN xxxxxxx…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 RELATED WORK

In this section, we review previous research related to our project, focusing on the importance of explainability in healthcare, prior usage of the dataset we employed, and methodologies for achieving interpretable machine learning models.

2.1 Importance of Explainability in Healthcare

Explainability in machine learning models is critical, especially in healthcare applications, where decisions can have significant impacts on patient outcomes. Cinà et al. (2022) argue that explainable AI is necessary for healthcare to ensure transparency, trust, and ethical decision-making in clinical settings [2]. Their work emphasizes that healthcare practitioners need to understand the reasoning behind AI-driven decisions to make informed choices and gain confidence in the models' outputs. This insight underscores the importance of developing and employing explainable models in our study.

2.2 Usage of Dataset in Related Research

The dataset used in our project has been previously utilized in other research, which validates its relevance and applicability. For instance, Yagin et al. (2023) employed this dataset to estimate BMI levels using a trained neural network approach optimized by Bayesian techniques[6]. Their study focused on predicting obesity based on physical activity and dietary habits, identifying key factors influencing obesity through feature selection algorithms like chi-square, F-Classify, and mutual information classification. Their results demonstrated high accuracy in obesity prediction, reinforcing the dataset's utility in health-related machine learning applications. This prior usage of the dataset supports its suitability for our analysis and provides a benchmark for evaluating our models' performance.

2.3 Explainability on Decision Trees

The quest for interpretable machine learning models has led to various approaches, including the development of decision trees with short explainable rules. Souza et al. (2022) introduced a method for creating decision trees that balance the trade-off between explainability and depth, aiming to maintain model performance while ensuring that the rules are simple and interpretable [4]. Their technique, which we consider incorporating in our project, offers a potential solution for achieving interpretable models that do not compromise on accuracy. By leveraging their function for evaluating the explainability-depth trade-off, we aim to enhance the interpretability of our models, making them more suitable for practical applications in healthcare.

2.4 Counterfactuals with DiCE library

The use of counterfactual explanations is a growing area of research in the field of explainable AI, providing insights into how slight changes in input data can alter the prediction of a model. In this project, we utilize the DiCE library [3], introduced by Mothilal et al.. Unlike earlier methods, which often produce single or similar counterfactual instances, DiCE focuses on creating multiple, distinct counterfactuals. This diversity is essential for finding varied and actionable pathways to achieve desired outcomes, making the explanations more practical and useful in real-world applications.

3 METHODS

3.1 Dataset Description

The dataset used for this project is the 'Estimation of BMI levels based on eating habits and physical condition' from the UCI Machine Learning Repository [1]. It includes data on individuals' eating habits, physical conditions, and BMI levels. Each person is classified to one of the seven different BMI levels such as Overweight Level I or Normal Weight. The dataset comprises 17 attributes such as age, gender, family history of overweight, frequency of consumption of high-caloric food, physical activity frequency, and technology usage during meals.

The dataset variable table can be found on the appendix (3).

3.2 Data Preprocessing

3.2.1 BMI as Predictor. To enhance the interpretability and robustness of our models, we decided to use Body Mass Index (BMI) as the predictor variable instead of the original obesity level categories provided in the dataset. BMI is a widely recognized and standardized measure of body fat, calculated as weight in kilograms divided by height in meters squared. By using BMI, we can maintain consistency in our comparisons across individuals, regardless of their height and weight differences. Consequently, we removed the original weight and height attributes from our input features to avoid redundancy and potential multicollinearity issues.

For our analysis, we categorized BMI into four distinct groups based on widely accepted health guidelines [5]:

- **Underweight**: BMI less than 18.5
- **Normal weight**: BMI between 18.5 and 24.9
- **Overweight**: BMI between 25 and 39.9
- **Obese**: BMI 40 and above

These categories allow us to group individuals into meaningful segments that reflect different health conditions and risk levels associated with their body mass.

3.2.2 Features and Encoding. The dataset includes several binary and categorical attributes that required appropriate encoding to ensure they could be effectively utilized in our models. Binary attributes, such as gender, family history of overweight, and others, were encoded into binary format (e.g., yes/no converted to 1/0). Categorical attributes with multiple levels, such as frequency of alcohol consumption or type of transportation, were transformed into separate binary columns for each category level.

3.3 Models Used

3.3.1 Feed-Forward Neural Network. We employed a Multilayer Perceptron (MLP) for the classification task. The MLP architecture consists of an input layer, two hidden layers, and an output layer. The details of each layer are summarized in Table 1.

Layer	Attributes
Input Layer	Receives input features
First Hidden Layer	64 neurons, ReLU activation
Second Hidden Layer	32 neurons, ReLU activation
Output Layer	4 Neurons, softmax activation

Table 1: Attributes of the Multilayer Perceptron (MLP) model.

- **Data Preparation**: Features were scaled, and target labels were encoded.
- **Training Configuration**: The model was trained for 50 epochs with a batch size of 32, using 20% of the training data for validation.
- **Compilation**: The model was compiled with the Adam optimizer and *sparse_categorical_crossentropy* loss function.
- **Training Execution**: The model was fitted to the training data, and validation accuracy was monitored to prevent overfitting.

3.3.2 LightGBM. In addition to the Multilayer Perceptron (MLP), we also employed LightGBM, a gradient boosting framework that uses tree-based learning algorithms. LightGBM is categorized as a black-box model due to its complexity and the difficulty in directly interpreting its internal workings. However, it offers several advantages that motivated our choice.

LightGBM is known for its efficiency and speed, particularly with large datasets, making it suitable for our analysis. Furthermore, it provides built-in functions to visualize feature importance based on 'split' and 'gain' metrics. The 'split' importance shows how many times a feature is used in the decision-making process, while the 'gain' importance measures the contribution of a feature to the model's accuracy.

3.3.3 Decision Trees. In our analysis, we also utilized the Decision Tree model. Decision Trees are inherently interpretable, making them an excellent choice as one of our baseline models. This model type provides a clear and straightforward way to understand the relationships between features and the target variable, which is crucial for our goal of identifying key factors influencing BMI levels.

We leveraged the built-in feature importance graphs to identify the most significant features in our dataset. Additionally, we visualized the tree structure to gain deeper insights into the decision-making process of the model. However, the initial decision tree had a high depth due to the large number of features, which made it complex and challenging to interpret.

To address this, we decided to prune the tree to a maximum depth of 5. This pruning significantly increased the interpretability of the model while still maintaining a reasonable level of accuracy. By simplifying the tree, we could better understand and communicate the key factors contributing to obesity, making the Decision Tree an invaluable tool in our analysis.

3.3.4 RuleFit. We also employed RuleFit, which combines the predictive power of ensemble methods with the interpretability of linear models. RuleFit generates a sparse linear model consisting of rules derived from decision trees, offering both robustness and clarity.

The motivation for choosing RuleFit is its ability to produce understandable rules that describe the data, providing actionable insights into factors influencing BMI levels. RuleFit handles high-dimensional data well and highlights feature interactions often missed by linear models alone.

By extracting and fitting rules into a linear model, RuleFit blends interpretability and accuracy, allowing us to visualize and understand complex relationships within the data effectively. This makes RuleFit a valuable tool for analyzing obesity-related factors.

3.3.5 EBM Classifier. Lastly, we utilized the Explainable Boosting Machine (EBM) classifier to obtain both local and global explanations for our predictions. The EBM is a type of Generalized Additive Model (GAM) that combines the interpretability of linear models with the flexibility of decision trees, making it particularly suitable for our analysis.

The motivation for choosing EBM lies in its capability to provide clear insights into model behavior. For local explanations, we examined samples showing the highest prediction errors, allowing us to understand and interpret specific misclassifications. For global explanations, we used EBM to assess overall feature importance, helping us identify the most influential factors affecting BMI levels.

4 RESULTS

4.1 Black-box models

4.1.1 LightGBM.

- The LightGBM achieved an accuracy of 0.87

Feature Importances. To understand the importance of each feature in our dataset, we utilized LightGBM's built-in functions to plot feature importances based on 'split' and 'gain' metrics.

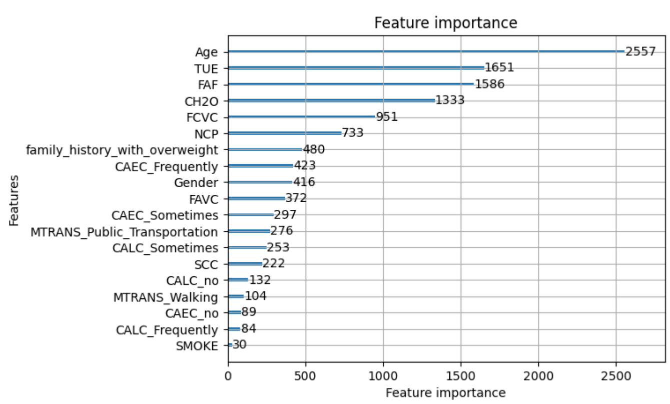

Figure 1: Feature importance based on the number of times a feature is used in the decision-making process (split importance).

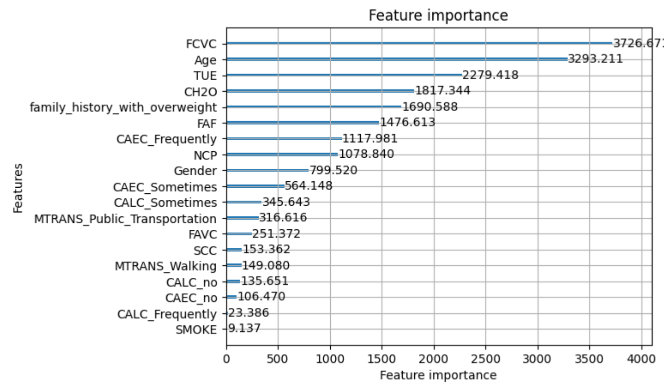

Figure 2: Feature importance based on the contribution of a feature to the model's accuracy (gain importance).

Split Importance. : The plot (Figure 1) shows that 'Age', 'TUE', and 'FAF' are among the most frequently used features in the decision-making process. This indicates that these features are the most important in the structure of the LightGBM model.

Gain Importance. : The plot (Figure 2) reveals that 'FCVC', 'Age', and 'TUE' significantly contribute to the model's accuracy. The higher gain values for these features suggest that they provide substantial information for predicting BMI category levels.

4.1.2 Random Forest.

- The Random Forest model achieved an accuracy of 0.886.

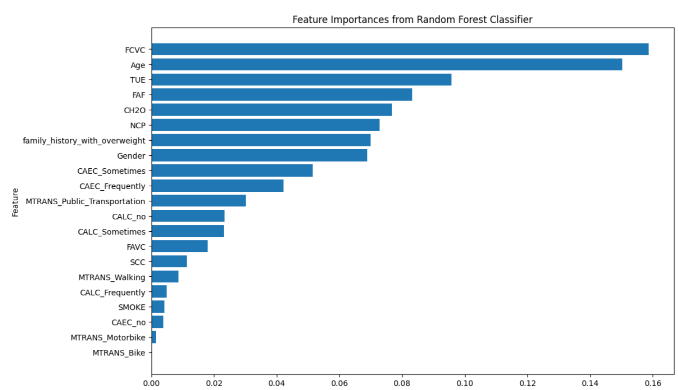

Figure 3: Feature importance from the Random Forest Classifier.

The feature importance plot (Figure 3) shows that 'FCVC', 'Age', and 'TUE' are among the most significant features in the model. This indicates that these features play a crucial role in predicting the BMI levels. Other important features include 'FAF', 'CH2O', and 'NCP', which also contribute significantly to the model's predictions. Understanding these feature importances helps us to identify key factors influencing BMI levels and can guide targeted interventions.

4.1.3 Feed-Forward Neural Network.

- The Neural Network achieved an accuracy of 0.85

The performance of the Neural Network was deemed appropriate in order to be used for the counterfactual explanations.

4.2 Intrinsically Interpretable Models

4.2.1 Decision Tree.

- The Decision Tree model achieved an accuracy of 0.813 with the max depth of 16.

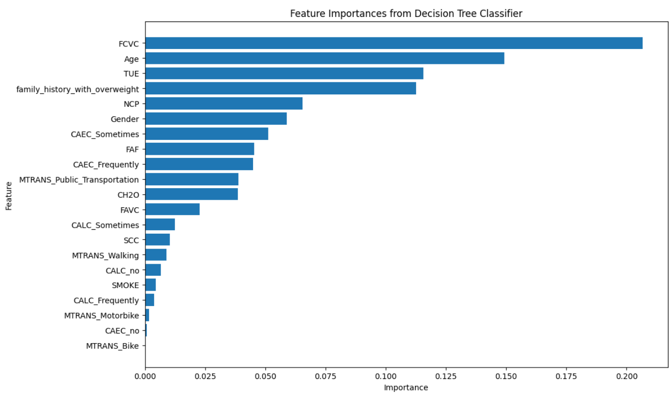

Figure 4: Feature importance from the Decision Tree Classifier.

The feature importance plot (Figure 4) shows that 'FCVC', 'Age', and 'TUE' are among the most significant features in the Decision Tree model. This indicates that these features play a crucial role in predicting BMI levels. Other important features include 'family_history_with_overweight', 'NCP', and 'Gender', which also contribute significantly to the model's predictions.

Pruning. To enhance the interpretability of the Decision Tree, we pruned the tree to a maximum depth of 5. Pruning helps in reducing the complexity of the model and makes it easier to understand the decision-making process. However, this simplification comes at a slight cost to accuracy. Given the increased interpredability the trade-off is minimal to none.

Experiment Results After Pruning:

- The pruned Decision Tree model achieved an accuracy of 0.801.

The feature importance plot after pruning (Figure 5) reveals that 'FCVC', 'family_history_with_overweight', and 'Age' remain the most significant features. However, their relative importance has been adjusted to reflect the simplified structure of the pruned tree. This pruning process makes the model more interpretable while still maintaining a reasonable level of accuracy, thereby balancing complexity and interpretability.

4.2.2 Rulefit. In addition to the neural network and other architectures, we employed the RuleFit model for classification. The RuleFit model achieved an accuracy of 70% and generated 1677 rules. This

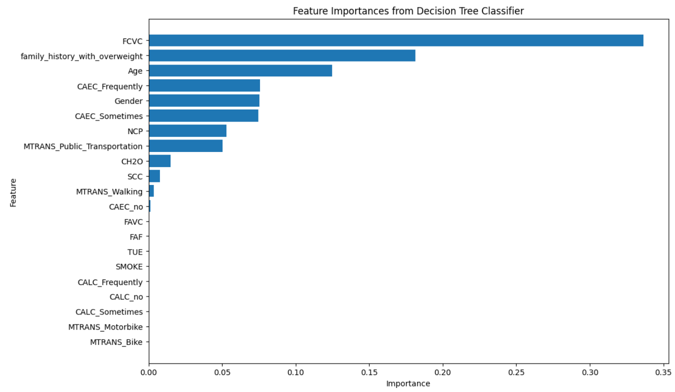

Figure 5: Feature importance from the pruned Decision Tree Classifier (max depth = 5).

performance was notably lower compared to the other models we used, all of which achieved accuracies higher than 80

The sheer number of rules produced by RuleFit significantly reduces interpretability. With 1677 rules, the model's output becomes complex and difficult to analyze, which does not utilize one of the primary benefits of rule-based models, their inherent interpretability. Attempting to limit the number of rules to enhance interpretability resulted in a further decrease in accuracy, rendering the model less useful for practical applications.

Given these outcomes, RuleFit's performance and interpretability were inadequate compared to the other models in our study.As a result, while RuleFit can be a valuable tool in certain contexts, it did not prove to be effective for interpreting the factors contributing to different obese levels in the form of rules.

4.2.3 Explainable Boosting Machine (EBM).

- The Random Explainable Boosting Machine (EBM) achieved an accuracy of 0.818.

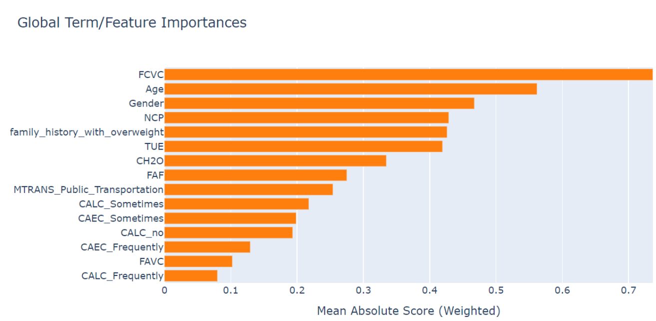

Figure 6: Random Explainable Boosting Machine (EBM)

The global feature importance plot (Figure 6) shows several key features that contribute to predicting BMI levels. The most influential feature is the frequency of vegetable consumption (FCVC), highlighting the role of healthy eating habits in managing body weight. Age and gender also emerged as important factors, reflecting the variations in obesity risk across different demographics.

The number of main meals per day (NCP) and a family history of overweight further underscore the importance of meal patterns and genetic predisposition in determining BMI levels. Additionally, lifestyle factors such as time spent using technology (TUE), water consumption (CH2O), frequency of physical activity (FAF), and mode of transportation (MTRANS_Public_Transportation) were also significant predictors, emphasizing the complex interplay of dietary, behavioral, and genetic factors in influencing obesity.

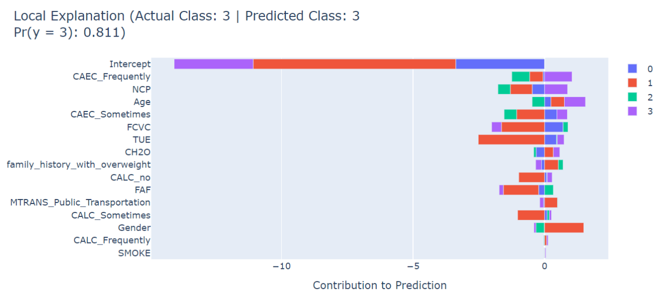

Figure 7: Local Feature importance example.

The local explanation plot (Figure 7) for a specific instance (predicted class: 3) from the EBM model provides a detailed breakdown of how various features contributed to the final prediction. In this instance, the most influential factors include frequent consumption of food between meals (CAEC-Frequently), number of main meals per day (NCP), and age, which together significantly pushed the prediction towards a higher obesity class. Other features like the frequency of vegetable consumption (FCVC) and time spent on activity (FAF) had a negative contribution. This insights helps in understanding the specific reasons behind the model's prediction for this individual, highlighting how certain behaviors and demographic factors combine to influence obesity risk.

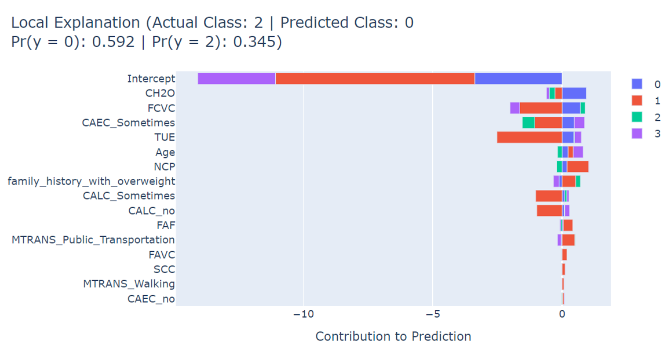

Figure 8: Local Feature importance of a misclassification example.

The local explanation for a misclassification instance where the actual class is 2, but the predicted class is 0, provides insight into the model's decision-making process. The chart shows the contribution of various features to the final prediction. In this instance, significant positive contributions from features such as CH2O (water consumption), FCVC (frequency of vegetable consumption) have pushed the prediction towards class 0. The positive contribution of these features seems to trick the system into predicting class 0 because usually the consumption of vegetables and water contributes positively to the underweight class. This misclassification highlights the complex interplay of dietary and lifestyle factors and underscores the importance of understanding individual feature impacts to improve model accuracy.

4.3 Counterfactuals

In this study, we utilized the DiCE [3] (Diverse Counterfactual Explanations) library to generate counterfactual explanations for our trained TensorFlow Neural Network model described in Section 3.3.1. By integrating our pre-trained model with DiCE, we were able to produce meaningful counterfactual explanations that highlight actionable changes in input features.

DiCE was configured to generate counterfactual explanations by varying selected features while keeping certain attributes constant. Specifically, 'Age', 'family history with overweight', and 'Gender' were kept unchanged to ensure the generated counterfactuals remained realistic and actionable.

DiCE works by perturbing the input features and using the TensorFlow model to evaluate these new instances. The library iteratively adjusts the features to find a set of counterfactuals that meet the desired outcome, which in this case is the Normal Weight category.

DiCE was also chosen as it generates multiple counterfactuals, providing a diverse set of explanations. This diversity helps us to understand different ways to achieve the desired outcome.

Feature	Original	Counterfactual 1	Counterfactual 2
Gender	Male	Male	Male
Age	24	24	24
Smoker	No	Yes	No
FCVC	2	2.90	2.80
CH2O	2.74	2.74	1.90
TUE	0.79	0.79	0.10

Table 2: Example of Produced Counterfactual Explanations

The example at Table 2 illustrates an instance of an individual who belongs to the Overweight class.The first counterfactual explanation suggests the user to increase his consumption of vegetables in his meals.The second explanation suggests the increase of vegetable consumption again, but in addition to a decrease of water consumption and time spent on using electronic devices.Even though these explanations are possible and realistic, they should never be taken as a recommendation without an expert's opinion.

4.4 Discussion/Conclusion

Our study evaluated multiple machine learning models to classify BMI levels, considering both accuracy and interpretability. The results highlighted various tradeoffs between these two critical aspects.

The Decision Tree model, when allowed to grow without depth limitations, achieved a commendable accuracy of 81%. However, the resulting complex visualization posed significant challenges for interpretation. In contrast, a pruned Decision Tree with a maximum depth of 4 maintained a slightly lower accuracy of 80% but offered high interpretability with fewer branches, making it easier to understand and analyze.

The Random Forest model and the LightGBM model both achieved higher accuracies of 87%. These models provided built-in feature importance graphs, which are beneficial for gaining a high-level understanding of the factors influencing the predictions. However, these feature importance measures fall short in delivering the detailed explanations necessary for practical applications. While useful, they do not offer the same level of interpretability as other methods, such as the visualizations from a pruned Decision Tree or the local explanations from the Explainable Boosting Model (EBM).

The EBM, which also achieved an accuracy of 82%, stands out by offering both global and local explanations alongside feature importances. This capability allows for a deeper understanding of the model's behavior. The ability to provide detailed insights at both the individual and aggregate levels ensures that EBM can effectively bridge the gap between accuracy and interpretability.

Our use of the RuleFit model underscored the importance of balancing accuracy and interpretability. Despite generating 1677 rules, RuleFit's accuracy was significantly lower than the other models at 70%. Moreover, the sheer volume of rules rendered the model impractical for interpretation, demonstrating that simply generating interpretable components is not sufficient if the overall accuracy is compromised.

Across all models, certain features consistently emerged as the most important: FCVC (consumption of vegetables), Age, and TUE (time spent using technological devices). The prominence of FCVC highlights the role of dietary habits in obesity, while Age and TUE reflect the impact of demographic factors and lifestyle choices.

In our study, we also focused on generating counterfactual explanations to provide insights into how slight changes in input features could lead to different obesity outcomes. Using the DiCE library integrated with our trained TensorFlow Neural Network model, we were able to create hypothetical scenarios where specific features were altered to observe their impact on obesity classifications. The counterfactual explanations were designed to vary only the modifiable features, such as dietary habits and lifestyle factors, while keeping immutable features like age, family history of being overweight, and gender constant. By setting realistic constraints on the variations, such as restricting age changes to within ±3 years, we ensured that the counterfactuals were both feasible and actionable. These counterfactual scenarios highlighted how small adjustments in daily routines or habits could potentially lead to significant changes in obesity outcomes, even though those changes should never be taken as valid advice as they might be encouraging other unhealthy choices in an individuals lifestyle.

By leveraging diverse explainability methods, including counterfactual explanations, we provided actionable insights into the critical factors contributing to obesity. These findings can inform targeted interventions and policies aimed at improving health outcomes in various populations. The combination of high-accuracy models with interpretable outputs ensures that our results are both accurate and practically relevant, offering valuable guidance for public health initiatives and individual lifestyle adjustments. This approach demonstrates the importance of integrating multiple explainability techniques to achieve a balanced understanding of complex health issues like obesity.

5 ETHICAL CONSIDERATIONS

In the development and implementation of our explainable AI models, ethical considerations were taken into account to ensure the integrity and ethical soundness of the project.

One of the major considerations in AI research is that of guaranteeing the privacy of the participants through anonymization of the data. However, since the chosen dataset already provides the data completely anonymized, no further steps were required for this project.

Another important ethical concern is ensuring fairness and eliminating possible biases. Since the dataset used does not include sensitive attributes and only contains attributes that are widely considered to directly affect the probability that an individual is overweight, these concerns are mitigated.

However, the study of counterfactuals throughout the project highlights the need for a thorough investigation by domain experts, given the possibility of producing results that can have a negative impact on a participant's health. For example, when studying counterfactuals, we identified situations where the model recommends the user to start smoking. While this recommendation makes sense in the scope of the project, since smoking can lead to weight loss, it should not be a viable recommendation, given how detrimental this can be to the participant's overall health.

Finally, concerns around transparency are eliminated given the nature of the project, which specifically intends to provide a transparent view of the predictions in order to identify the recommended lifestyle changes that will reduce the risk of obesity and promote healthier living. This approach ensures that users can understand the reasoning behind the AI's recommendations, fostering trust and facilitating informed decision-making.

6 MEMBER CONTRIBUTIONS

The contributions of each project member are described below:

Andreas: Andreas focused on data collection and preprocessing efforts. He was responsible for cleaning the dataset and performing initial exploratory data analysis. Additionally, Andreas implemented the baseline machine learning models and conducted feature engineering.

Sotiris: Sotiris focused on the programming tasks involved in the project. He worked alongside Andreas in implementing the baseline and explainable models. Sotiris was mostly responsible for integrating the counterfactual explainations and the black box models. He additionally conducted experiments to evaluate the trade-off between explainability and model performance, ensuring the models were both accurate and interpretable.

Tiago: Tiago was responsible for several sections of the report. He identified relevant research articles, summarized key findings, and integrated these insights into the project report. He played a crucial role in the model results and contributed to the overall project discussion and conclusions.

Nicolas: Nicolas took the lead in preparing the project report and poster. He compiled the final project report, ensuring it was comprehensive and well-organized. Nicolas also worked with Tiago on the writing of the report.

REFERENCES

[1] 2019. Estimation of Obesity Levels Based On Eating Habits and Physical Condition . UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5H31Z.
[2] Giovanni Cinà, Tabea Röber, Rob Goedhart, and Ilker Birbil. 2022. Why we do need Explainable AI for Healthcare. arXiv:2206.15363
[3] Ramaravind Kommiya Mothilal, Amit Sharma, and Chenhao Tan. 2019. Explaining Machine Learning Classifiers through Diverse Counterfactual Explanations. *CoRR* abs/1905.07697 (2019). arXiv:1905.07697 http://arxiv.org/abs/1905.07697
[4] Victor Feitosa Souza, Ferdinando Cicalese, Eduardo Laber, and Marco Molinaro. 2022. Decision Trees with Short Explainable Rules. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 12365–12379. https://proceedings.neurips.cc/paper_files/paper/2022/file/500637d931d4feb99d5cce84af1f53ba-Paper-Conference.pdf
[5] Courtney B Weir and Adam Jan. 2023. *BMI Classification Percentile And Cut Off Points*. StatPearls Publishing, Treasure Island (FL). [Updated 2023 Jun 26].
[6] Fatma Hilal Yagin, Mehmet Gülü, Yasin Gormez, Arkaitz Castañeda-Babarro, Cemil Colak, Gianpiero Greco, Francesco Fischetti, and Stefania Cataldi. 2023. Estimation of Obesity Levels with a Trained Neural Network Approach optimized by the Bayesian Technique. *Applied Sciences* 13, 6 (2023). https://doi.org/10.3390/app13063875

APPENDIX

Table 3: Dataset Variables Table

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Gender	Feature	Categorical	Gender			no
Age	Feature	Continuous	Age			no
Height	Feature	Continuous				no
Weight	Feature	Continuous				no
family_history_with_overweight	Feature	Binary		Has a family member suffered or suffers from overweight?		no
FAVC	Feature	Binary		Do you eat high caloric food frequently?		no
FCVC	Feature	Integer		Do you usually eat vegetables in your meals?		no
NCP	Feature	Continuous		How many main meals do you have daily?		no
CAEC	Feature	Categorical		Do you eat any food between meals?		no
SMOKE	Feature	Binary		Do you smoke?		no
CH2O	Feature	Continuous		How much water do you drink daily?		no
SCC	Feature	Binary		Do you monitor the calories you eat daily?		no
FAF	Feature	Continuous		How often do you have physical activity?		no
TUE	Feature	Integer		How much time do you use technological devices such as cell phone, videogames, television, computer and others?		no
CALC	Feature	Categorical		How often do you drink alcohol?		no
MTRANS	Feature	Categorical		Which transportation do you usually use?		no
NObeyesdad	Target	Categorical		Obesity level		no