

# Data Visualisation Competition Submission

u7469758

October 2025

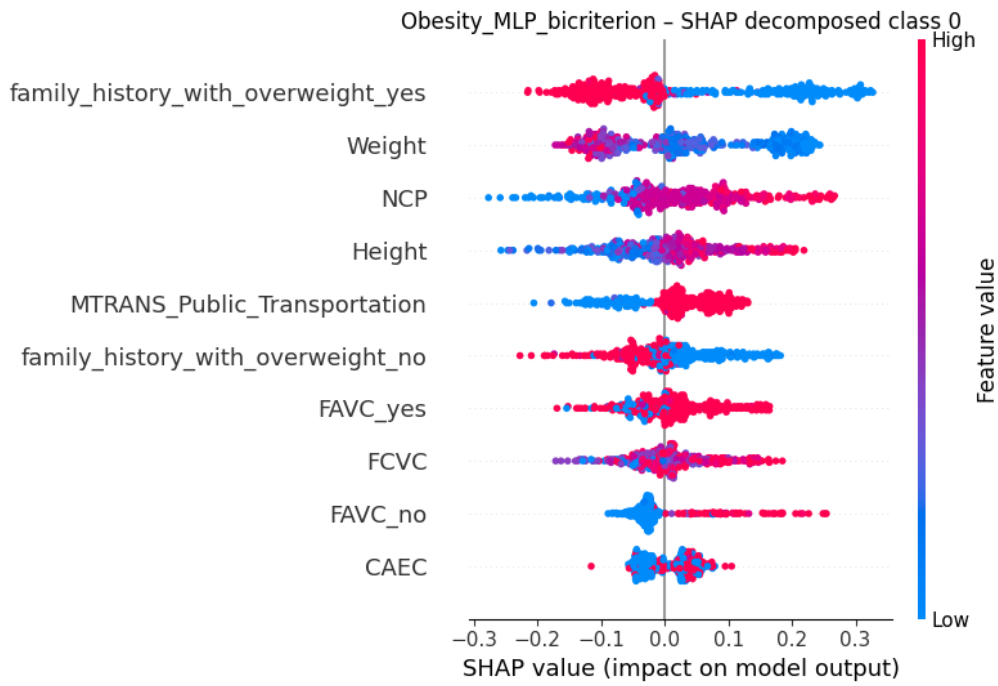
Machine learning models are increasingly applied in clinical settings, yet their complexity often leads to the propagation of bias and reduces the interpretability of the decision-making process. “SHIELD: SHapley and Information-theory based framework for **E**quitable **L**earning via **D**issimilar feature grouping” addresses this by grouping maximally dissimilar features, learning on a latent representation, and decoding attributions back to native variables. SHAP quantifies each feature’s contribution by showing how different feature values positively or negatively influence the model’s prediction.

The paired plots contrast SHAP summaries of ungrouped (raw) versus SHIELD (bicriterion grouping) for predicting *Insufficient Weight* in the Obesity dataset via MLP classifier. In the ungrouped case, attribution concentrates on anthropometrics (**Weight**, **Height**, **Age**) with many instances near zero SHAP value, meaning the participants’ records essentially did not contribute to any decision (i.e., wasted). Under SHIELD, explanatory credit is more equitably spread out. This demonstrates a more nuanced consideration of lifestyle and contextual features, such as **NCP** (# of meals per day), **FCVC** (consumption of vegetables), **CAEC** (consumption of food between meals), **MTRANS\_Public\_Transportation** and **family\_history\_with\_overweight\_{yes/no}**, as the model engages more instances and reveals environmental contributors rather than relying almost exclusively on **Height** and **Weight**. The graphic further illustrates that this equitable learning is achieved while remaining consistent with domain knowledge. For example, individuals without a family history of overweight and with low vegetable intake are less likely to belong to the insufficient weight class, whereas those with lower body weight and infrequent consumption of high-calorie foods (**FAVC\_no**) are more likely to be classified as insufficient weight.

Practical implications of this framework include broader actionable levers for clinicians (e.g., dietary patterns, snacking, transport), greater sample efficiency through better use of instances, panel rationalisation by reducing dependence on immutable traits, and improved auditability via decoder-mapped SHAP.



(a) ungrouped case



(b) grouped case via bicriterion method

Figure 1: SHAP plots for predicting Insufficient Weight class in Obesity dataset with MLP classifier under ungrouped and grouped cases