Childhood Obesity Prediction – Prototype ML Results

This report summarizes the prototype machine learning model results using a public obesity dataset, as a stand-in for Lachesis project data.

Dataset Overview

• Records: 2,111
• Features: Age, Gender, Height, Weight, Lifestyle factors (e.g., family history, eating habits, activity, transport)
• Target: NObeyesdad (7 obesity categories)

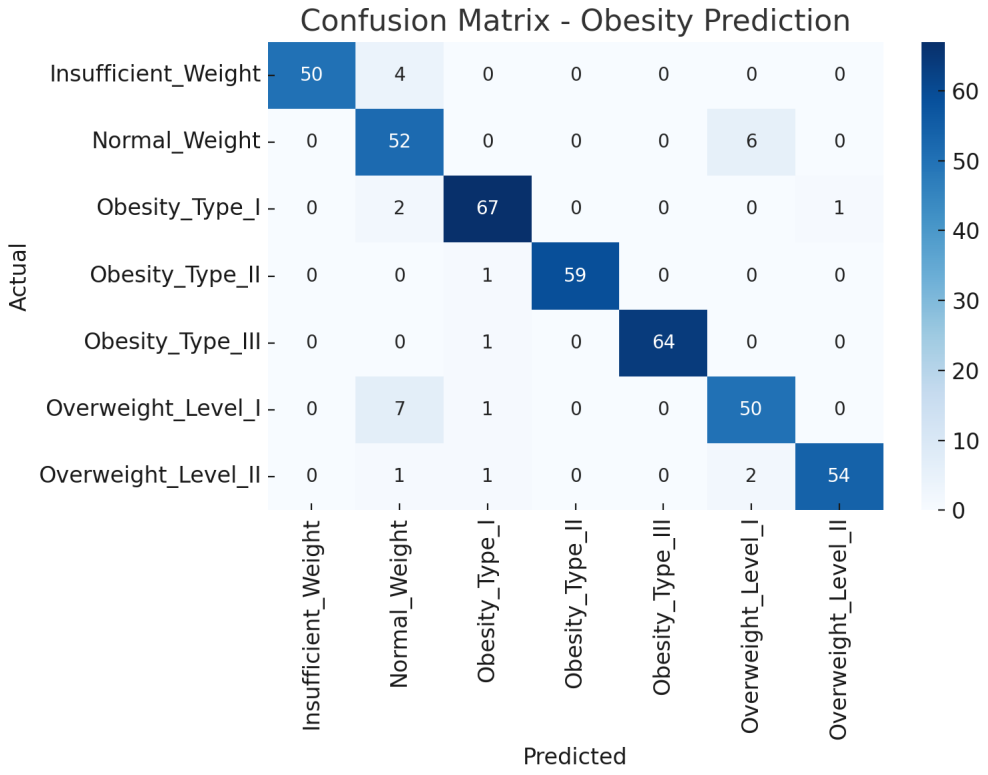
Methodology

Random Forest Classifier with 400 estimators, 80/20 train-test split, and one-hot encoding for categorical features.

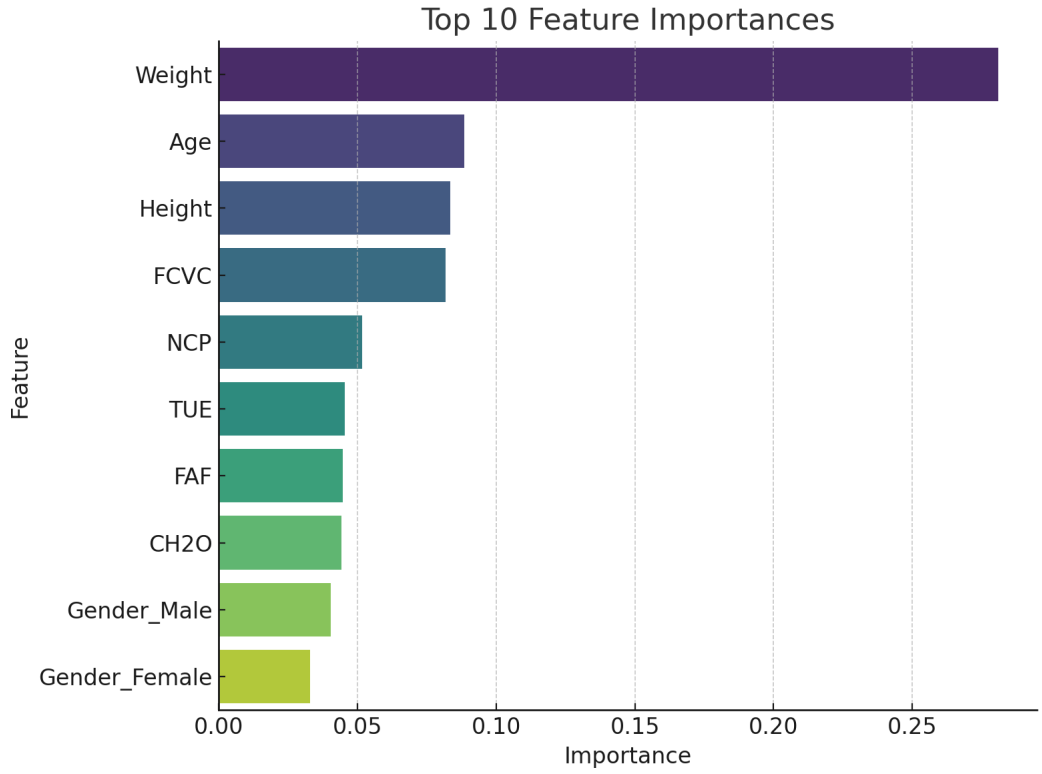
Results

• Accuracy: 93.62%
• Macro F1 Score: 93.60%
• Weighted F1 Score: 93.74%

Confusion Matrix



Top 10 Feature Importances



Next Steps

• Expand feature engineering (correlations, interaction terms)
• Apply the pipeline to Lachesis-specific datasets
• Conduct further evaluation (ROC curves, SHAP values)