CLINICAL DECISION SUPPORT FOR EARLY IDENTIFICATION OF OBESITY-RELATED COMPLICATION RISK

Group 5 – Frederick Damptey & Benjamin Odoom Asomaning

SAT 5141 – Clinical Decision Support & AI Modeling

Background & Motivation

•Obesity is a major global public-health challenge driving T2DM, hypertension, and cardiovascular disease [1].

•Early identification of high-risk individuals can reduce long-term morbidity and mortality [1], enable targeted prevention and lifestyle interventions.

•Clinicians often have limited time and may under-recognize obesity risk in busy primary-care settings.

•Clinical Decision Support Systems (CDSS) assist clinicians to make complex, data-rich decisions [2] using multiple patient information.

AI & CDSS for Obesity

- CDSS can integrate multiple risk factors and provide consistent, evidence-based recommendations [2].
- AI-driven CDSSs can augment human judgement in diagnosis, risk prediction and treatment planning [2,8].
- Recent work shows strong performance of ML for obesity risk prediction and monitoring [3,7].
- However, many models are “black boxes”, limiting clinician trust and adoption [4,5,7].

Population and Justification

Target population:

- Adults at high risk of obesity-related complications

Justification:

- High disease burden and modifiable risk factors.
- Opportunity for early intervention during routine visits.
- Supports over-loaded clinicians by highlighting patients needing closer

Prior Obesity ML Studies - Literature Review

Machine-learning models have shown strong performance in predicting obesity risk and related outcomes.

Helforoush & Sayyad (2024) used a hybrid metaheuristic ML approach, improving precision of obesity risk prediction [3].

Shen et al. (2024) developed a visualization-based obesity prediction system, emphasizing interpretability for clinicians [4].

Nguyen et al. (2023) systematic review: AI methods consistently outperform traditional statistical models in obesity prediction tasks [7].

Literature Review Cont’d

Shortliffe & Sepúlveda (2018) highlight how CDSS, when combined with AI, can enhance but not replace clinician judgment [2].

Esteva et al. (2019) showed deep learning’s potential in healthcare but stress transparency and clinical validation [8].

Lee et al. (2025) used ensemble ML models on survey data to predict obesity in T2DM patients, achieving high performance and demonstrating real-world utility [5].

Gap: Many AI models are not packaged into usable CDSS tools with clinician-in-the-loop workflows.

Aim and Objectives

Aim:
Build an AI-enabled CDSS that predicts obesity-related complication risk and supports clinician decision-making.

Objectives:

- ❑ Train and validate ML models to classify patients into 7 obesity-risk categories.
- ❑ Evaluate performance using accuracy, AUC, F1-score, sensitivity, and specificity.
- ❑ Use SHAP explainability to identify key predictors and improve clinical trust.

- ❑ Prototype a clinician-in-the-loop Streamlit interface for real-time use.

Dataset & Population

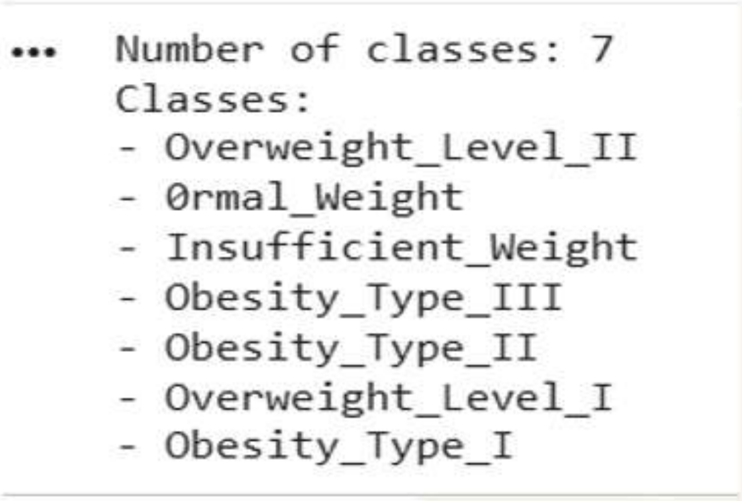
Data source: Obesity Risk Dataset from Kaggle (20,758 samples, 18 features) [6].

Variables: age, gender, height, weight, physical activity level, caloric intake, family history, and lifestyle behaviors.

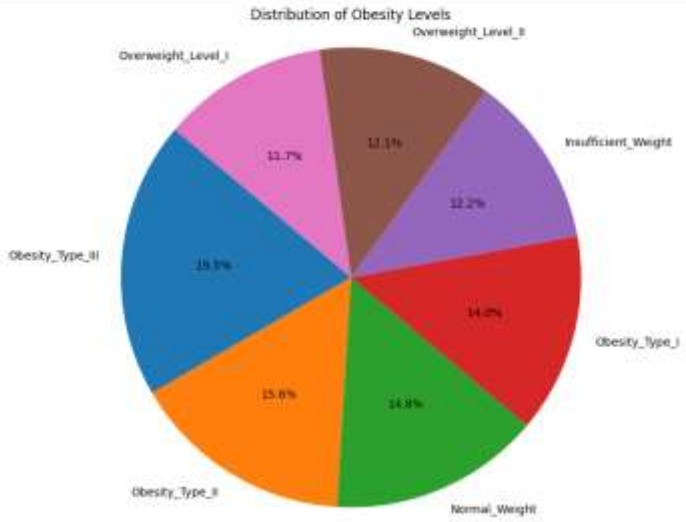
Target variable: Obesity_Risk, 7 obesity weight_based classes.

✓ Insufficient_Weight, Normal_Weight, Overweight_Level I & II, Obesity_Type I–III.

De-identified, no missing values, suitable for secondary analysis in research.



```
Number of classes: 7
Classes:
- Overweight_Level_II
- 0rmal_Weight
- Insufficient_Weight
- Obesity_Type_III
- Obesity_Type_II
- Overweight_Level_I
- Obesity_Type_I
```



Data Quality & Validation

Data exploration:

- Completeness (no missing values), feature relevance, spelling.

Used tratified k-fold cross-validation to preserve class distribution and improve generalizability.

Modeling Pipeline Overview

1. **Data Ingestion and exploration:** Loaded and inspected dataset structure

2. **Preprocessing:**

✓ Cleaned data
✓ Defined features and target.
✓ Train_test_split (80/20)

3. AutoGluon Tabular: Automated model selection and stacking.

4. Evaluation: Accuracy, AUC, F1, confusion matrix.

5. Explainability: SHAP, local XAI and contrastive ecplanable AI to rank features and visualize impact.

6. Clinician-in-the-Loop Interface: Streamlit tool for real-time use.

Preprocessing

- Dropped the column ‘id’.
- Defined features and label
- Corrected misspelt class.
- Train-test split (80/20)
- 3-fold CV in training phase.
- Recreated train and test dataframes for AutoGluon

Data splitting

```python
# Import train_test_split module from sklearn
from sklearn.model_selection import train_test_split

# Split the data into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

# Print X and y shape
print("Shape of training features (X_train):", X_train.shape)
print("Shape of testing features (X_test):", X_test.shape)
print("Shape of training labels (y_train):", y_train.shape)
print("Shape of testing labels (y_test):", y_test.shape)
```

```
Shape of training features (X_train): (16606, 16)
Shape of testing features (X_test): (4152, 16)
Shape of training labels (y_train): (16606,)
Shape of testing labels (y_test): (4152,)
```

AutoGluon & Model Details

AutoGluon explored multiple algorithms:

Derivatives of Random Forest, XGBoost, CatBoost, LightGBM, neural networks, etc.

Metric prioritized: Balanced accuracy to handle class imbalance in training.

Best model: AutoGluon using top-performing base learners.

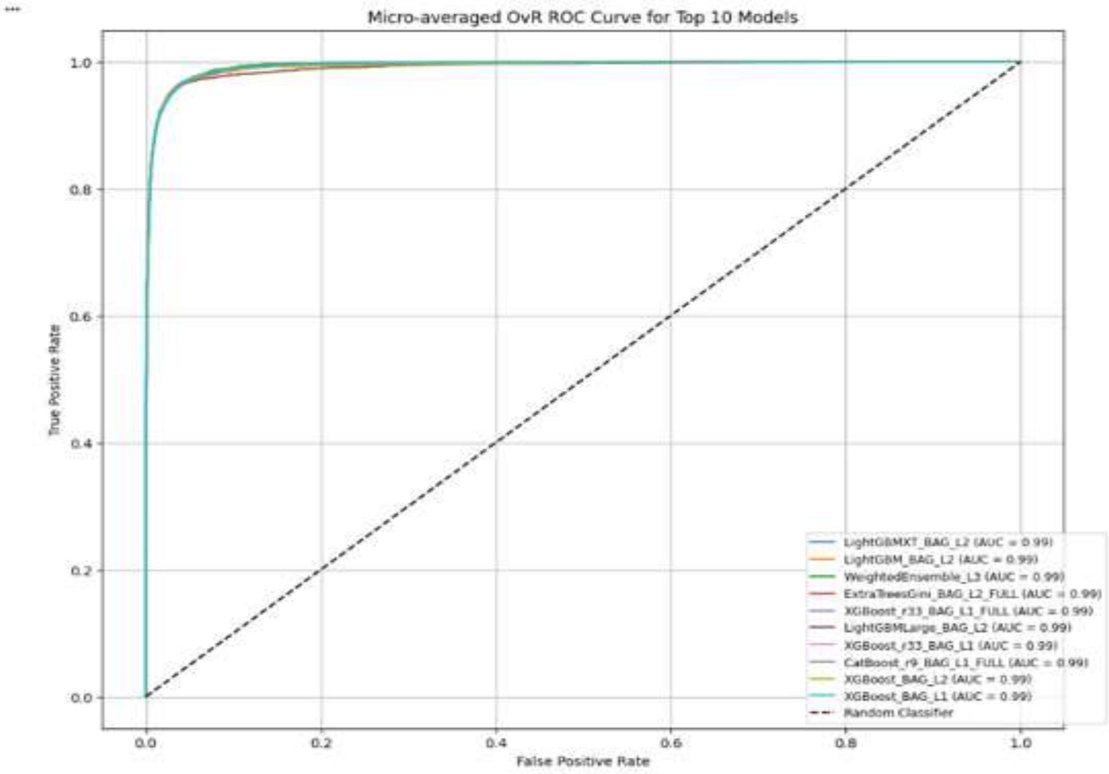
Top Performing Model Performance

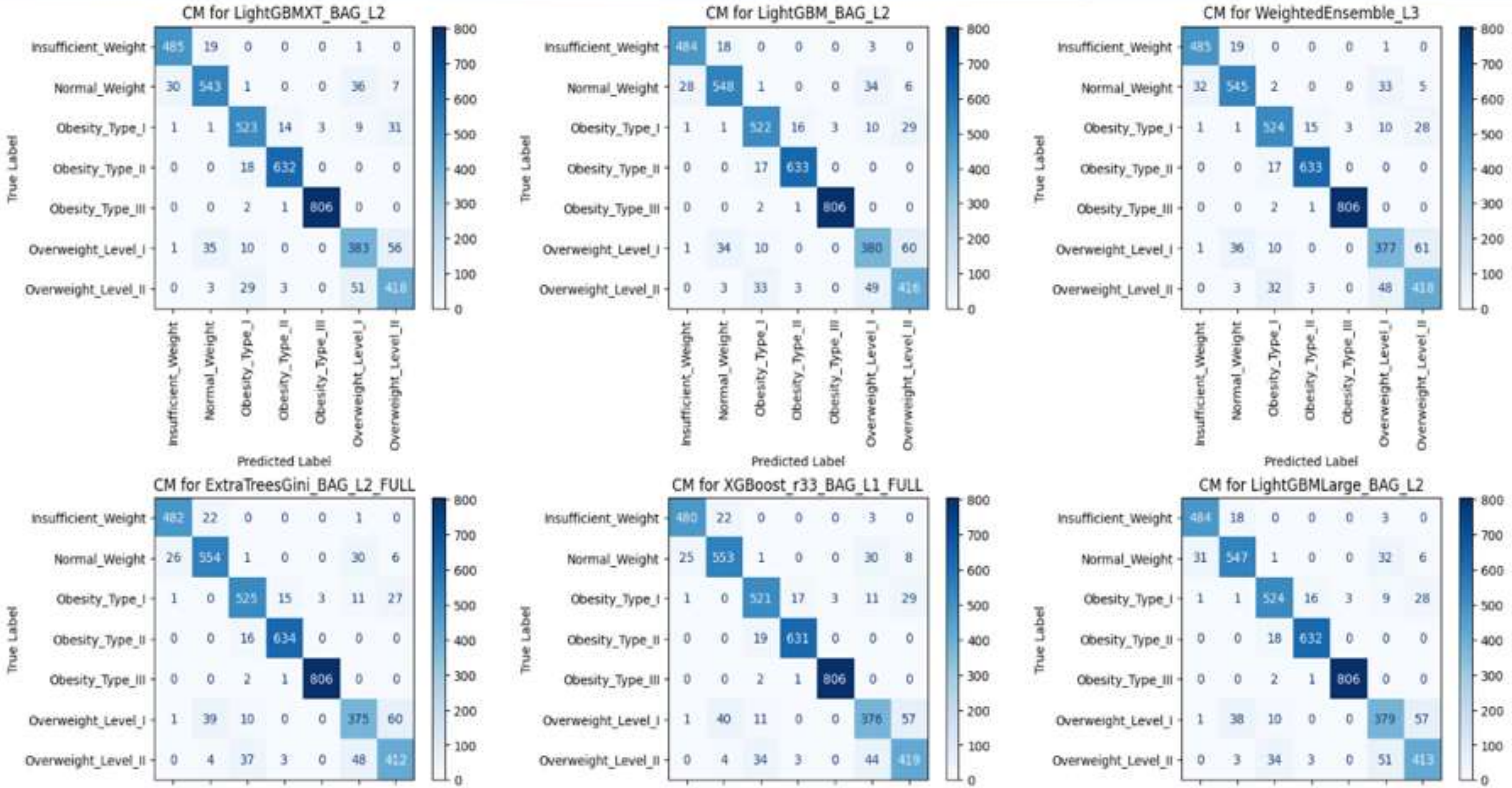
Leaderboard Model Performance Metrics

model	test_accuracy	AUC
LightGBMXT_BAG_L2	0.903820	0.991364
LightGBM_BAG_L2	0.903219	0.990861
WeightedEnsemble_L3	0.902981	0.992028
ExtraTreesGini_BAG_L2_FULL	0.902392	0.990488
XGBoost_r33_BAG_L1_FULL	0.902232	0.991461
LightGBMLarge_BAG_L2	0.902113	0.989349
XGBoost_r33_BAG_L1	0.901999	0.991504
CatBoost_r9_BAG_L1_FULL	0.901854	0.991991
XGBoost_BAG_L2	0.901757	0.990716
XGBoost_BAG_L1	0.901374	0.991621

ROC Curve

- AUC ≈ 0.99 for all top 10 models.
- Indicates excellent discriminative performance for all 7 classes.
- Model generalizes well across classes.




CM for LightGBMXT_BAG_L2
CM for LightGBM_BAG_L2
CM for WeightedEnsemble_L3
CM for ExtraTreesGini_BAG_L2_FULL
CM for XGBoost_r33_BAG_L1_FULL
CM for LightGBMLarge_BAG_L2
True Label
Predicted Label
Insufficient_Weight
Normal_Weight
Obesity_Type_I
Obesity_Type_II
Obesity_Type_III
Overweight_Level_I
Overweight_Level_II

Classification reports

Classification Report for Model: LightGBMXT_BAG_L2

	precision	recall	f1-score	support
Insufficient_Weight	0.94	0.96	0.95	505
Normal_Weight	0.90	0.88	0.89	617
Obesity_Type_I	0.90	0.90	0.90	582
Obesity_Type_II	0.97	0.97	0.97	650
Obesity_Type_III	1.00	1.00	1.00	809
Overweight_Level_I	0.80	0.79	0.79	485
Overweight_Level_II	0.82	0.83	0.82	504
accuracy			0.91	4152
macro avg	0.90	0.90	0.90	4152
weighted avg	0.91	0.91	0.91	4152

Classification Report for Model: LightGBM_BAG_L2

	precision	recall	f1-score	support
Insufficient_Weight	0.94	0.96	0.95	505
Normal_Weight	0.91	0.89	0.90	617
Obesity_Type_I	0.89	0.90	0.89	582
Obesity_Type_II	0.97	0.97	0.97	650
Obesity_Type_III	1.00	1.00	1.00	809
Overweight_Level_I	0.80	0.78	0.79	485
Overweight_Level_II	0.81	0.83	0.82	504
accuracy			0.91	4152
macro avg	0.90	0.90	0.90	4152
weighted avg	0.91	0.91	0.91	4152

Classification Report for Model: WeightedEnsemble_L3

	precision	recall	f1-score	support
Insufficient_Weight	0.93	0.96	0.95	505

Classification Report for Model: WeightedEnsemble_L3

	precision	recall	f1-score	support
Insufficient_Weight	0.93	0.96	0.95	505
Normal_Weight	0.90	0.88	0.89	617
Obesity_Type_I	0.89	0.90	0.90	582
Obesity_Type_II	0.97	0.97	0.97	650
Obesity_Type_III	1.00	1.00	1.00	809
Overweight_Level_I	0.80	0.78	0.79	485
Overweight_Level_II	0.82	0.83	0.82	504
accuracy			0.91	4152
macro avg	0.90	0.90	0.90	4152
weighted avg	0.91	0.91	0.91	4152

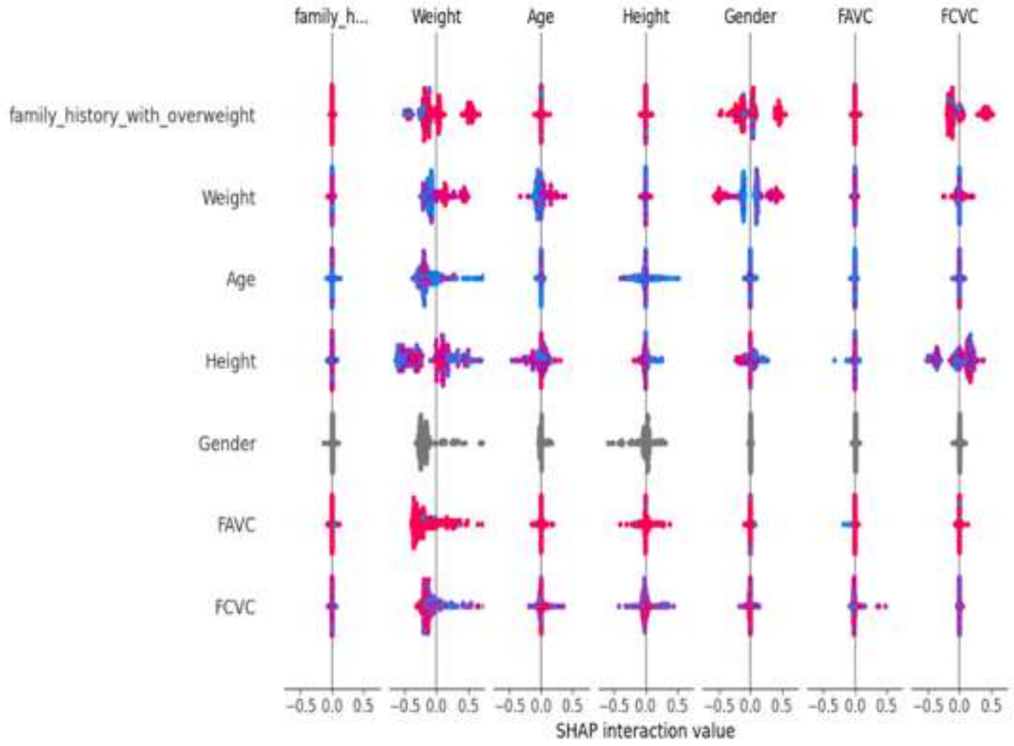
Classification Report for Model: ExtraTreesGini_BAG_L2_FULL

	precision	recall	f1-score	support
Insufficient_Weight	0.95	0.95	0.95	505
Normal_Weight	0.89	0.90	0.90	617
Obesity_Type_I	0.89	0.90	0.90	582
Obesity_Type_II	0.97	0.98	0.97	650
Obesity_Type_III	1.00	1.00	1.00	809
Overweight_Level_I	0.81	0.77	0.79	485
Overweight_Level_II	0.82	0.82	0.82	504
accuracy			0.91	4152
macro avg	0.90	0.90	0.90	4152
weighted avg	0.91	0.91	0.91	4152

Classification Report for Model: XGBoost_r33_BAG_L1_FULL

SHAP Explainable AI

- The SHAP figure below shows the top 6 factors which are highly predictive of obesity risk among all features.
  - Weight, Age, Height, FH_Obesity, FAVC, FCVC
  - Gender is the least predictive



CDSS Prototype & Clinical Workflow

Implemented a **Streamlit clinician-in-the-loop interface**:
•Input - case.

•Output - CDSS returns predicted obesity-risk class + XAI explanation.

•Feedback - Clinician can **accept, reject or override** the suggestion and assign reason if necessary.

•Justification - Justification text and timestamps logged for auditability.

CDSS Prototype & Clinical Workflow

•Data are de-identified; no direct patient identifiers [6].

Aligns with key principles from HIPAA deidentification stipulations and healthcare AI guidelines.

•Transparency**:** SHAP explanations and clinician-override logs.

•CDSS is designed to **support**, not replace, clinician judgment [2].

Prototype & Clinical Workflow

Controls

Case selection

Index of test-set case

34 − +

Show debug info

Obesity Risk Predictor – Explainable Clinical Viewer

This interface allows you to:

- Review **model performance** on the held-out test set.
- Inspect **individual cases** (features, ground truth, prediction, probabilities).
- Provide **expert feedback** on the model's predictions.
- View **Explainable AI** (XAI):
  - **Local what-if explanation** – which features most change this case's prediction.
  - **Contrastive explanation** – closest case with a *different* BMI category.

Model performance on test set

Accuracy

0.910

Balanced accuracy

0.900

Balanced accuracy accounts for all BMI classes equally, which is impo
some categories (e.g., severe obesity or underweight) are less freque

---

Case 34 – Review & Explain

Case & Prediction | Local XAI | Contrastive XAI

Contrastive explanation (nearest different-class case)

Streamlit app

Case 34 – Review & Explain

- Case & Prediction   Local XAI   Contrastive XAI

Contrastive explanation (nearest different-class case)

Nearest different-class case has label: `Overweight_Level_II`

	Current case	Contrastive case
SCC	0	0
FAF	1.453042	2.0
TUE	0.969085	1.0
CALC	0	Frequently
MTRANS	Public_Transportation	Public_Transportation
Age_Gro	<21	<21
FH_Obe	Yes	Yes
High_Ca	Yes	Yes
Activity_	Low	Moderate
y_pred	None	Overweight_Level_I

Differences between these columns highlight features that may be clinically important for distinguishing between BMI categories in similar individuals.

If the contrastive case seems too far or not clinically plausible, you may:

- Restrict the distance search (e.g., only within a certain age or BMI window).

Expert feedback

How do you judge this prediction?

- Accept (model prediction is correct)
- Reject (model prediction is incorrect)
- Override (provide a different class)

Select the correct class:

Insufficient_Weight

Reason / comments (optional but recommended):

e.g. recent weight change, comorbidities, measurement error, etc.

Submit feedback

Case 34 – Review & Explain

- Case & Prediction   Local XAI   Contrastive XAI

Local what-if explanation

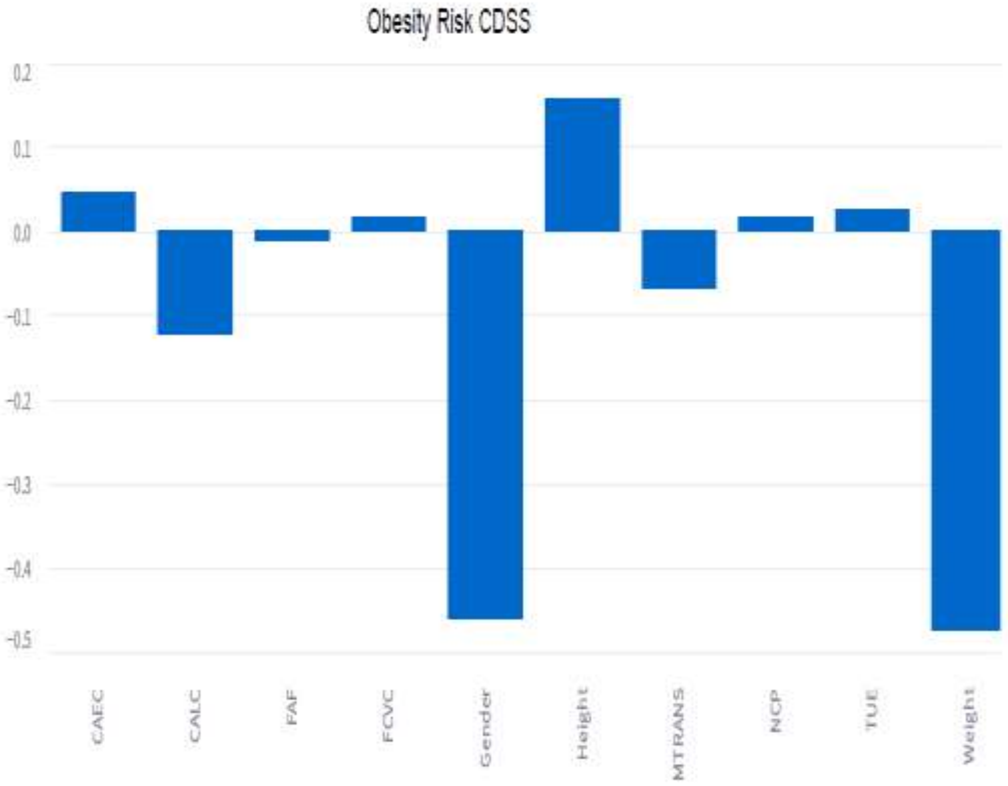
Class to explain

Overweight_Level_II

Base probability for `Overweight_Level_II` : `0.5513807535171509`

Local XAI table (top features by change in probability):

	feature	delta_prob	abs_delta
3	Weight	-0.4737	
0	Gender	-0.4598	
2	Height	0.158	
14	CALC	-0.123	
15	MTRANS	-0.0668	
8	CAEC	0.0459	
13	TUE	0.0249	
6	FCVC	0.0172	
7	NCP	0.0159	
12	FAF	-0.0115	



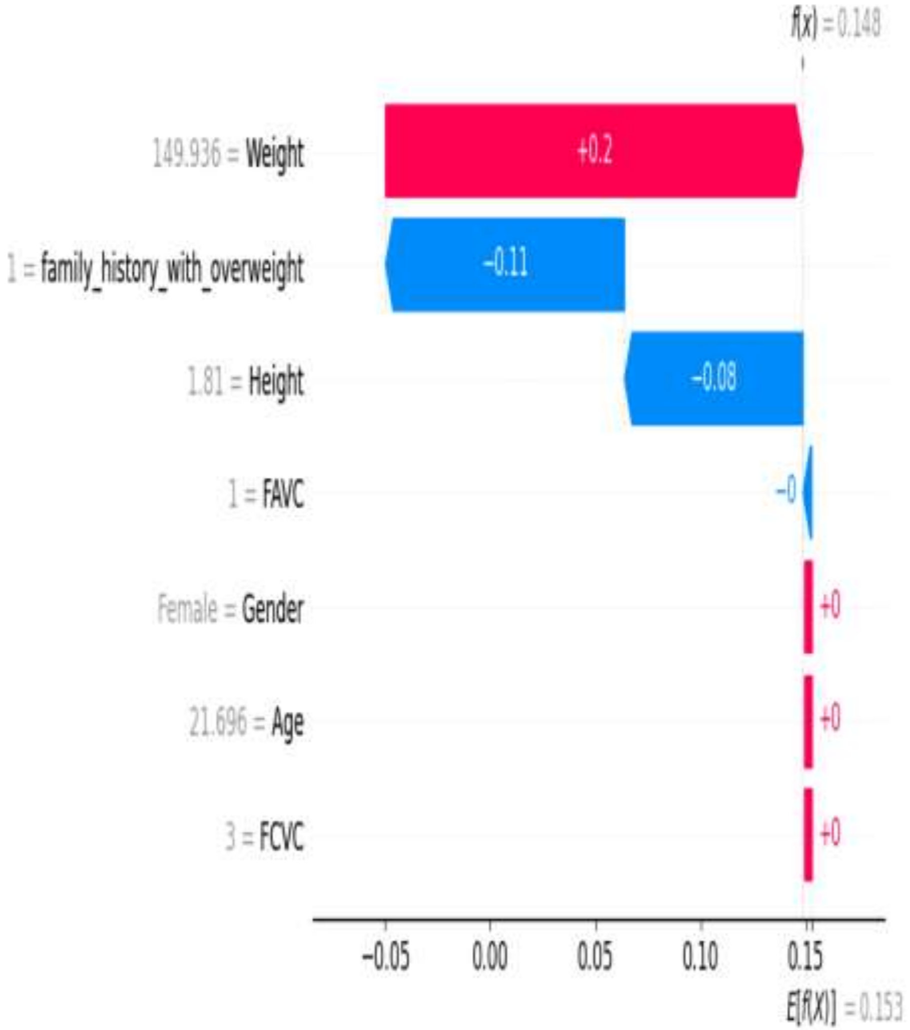
- Bars **above zero**: increasing/changing that feature in the tested way would **increase** the model's probability for `Overweight_Level_II`.
- Bars **below zero**: would **decrease** that probability.

If the plot looks strange or flat, corrective actions might include:

- Checking feature scaling and distributions.
- Verifying that the model actually uses those features (e.g., feature importance).

Explaining this test case:

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS
9603	Female	21.695892	1.809871	149.935848	1	1	3.0	3.0	Sometimes	0	2.36651	0	1.995582	0.890527	Sometimes	Public_Transportation



Challenges

- Limitations with Autogluon model.
- Streamlit app development was a challenge.
- The 7 multiclass task posed challenges with computing class performances and outputting results.
- The raw data, including categorical values fed into Autogluon made SHAP computation and plot very challenging.
- Integrated LIME and perturbation analysis plots XAI in streamlit may not be ideal for clinical deployment.
- Time constraints for robust development.

Conclusions & Impact

Developed a good-performing, explainable CDSS for obesity-risk prdiction.

Models achieved accuracy >0.90 and 0.99 AUC across seven risk classes.

SHAP explanations and clinician-in-the-loop design enhanced transparency and reduced automation bias.

Potential impact on targeted population:

- Earlier identification of high-risk patients.
- More personalized counseling and resource allocation.

Support for population-level prevention strategies.

References

1. H. C. Whitlock, L. M. Williams, and D. M. Colditz, “Obesity as a global public health challenge: Epidemiology, risk factors, and prevention,” *The Lancet*, vol. 387, no. 10026, pp. 231–243, 2016.

2. E. H. Shortliffe and M. J. Sepúlveda, “Clinical decision support in the era of artificial intelligence,” *JAMA*, vol. 320, no. 21, pp. 2199–2200, 2018.
3. Z. Helforoush and H. Sayyad, “Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach,” *Frontiers in Big Data*, 2024.
4. J. Shen et al., “Visualization obesity risk prediction system based on machine learning techniques,” *Scientific Reports*, 2024.
5. C. Lee et al., “A machine learning model for predicting obesity risk in patients with T2DM using health survey data,” *Diabetes & Metabolism Journal*, 2025.
6. J. P. Kochar, “Obesity Risk Dataset,” Kaggle, 2024.
7. Q. T. Nguyen, H. T. Vo, and T. M. Le, “Artificial intelligence applications in obesity prediction: A systematic review,” *Healthcare Analytics*, vol. 5, no. 2, pp. 101–118, 2023.
8. A. Esteva et al., “A guide to deep learning in healthcare,” *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.
9. https://www.kaggle.com/datasets/jpkochar/obesity-risk-dataset

Thank You