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

***Authors*:** Frederick Damptey & Benjamin Odoom Asomaning
Course: SAT5141 – Clinical Decision Support Modeling, Department of Applied Computing, Michigan Technological University
Emails: {fdamptey, boasoman}@mtu.edu

Abstract

Obesity continues to be a major global public health concern, contributing significantly to the incidence of type 2 diabetes, cardiovascular disease, hypertension, and long-term morbidity. Early detection of individuals at elevated obesity-related risk presents an opportunity for targeted preventive interventions. This paper presents an AI-enabled Clinical Decision Support System (CDSS) designed to classify patients into seven obesity-risk categories using the Obesity Risk Dataset from Kaggle. Employing AutoGluon's automated machine-learning framework, we evaluated multiple ensemble and gradient-boosting models. The best model achieved 90.38% test accuracy and AUC ≈ 0.99, demonstrating strong discriminative capability across all BMI-based classes. SHAP-based explainability highlighted weight, age, height, caloric intake, and family history as the most influential features. A clinician-in-the-loop Streamlit interface was also developed to support real-time decision-making, offering local, contrastive explanations and expert validation options. The results suggest that interpretable AI models can meaningfully enhance obesity-risk assessment in clinical environments.

I. Introduction

Obesity is widely recognized as one of the most pressing global health challenges, driving the prevalence of chronic diseases such as diabetes, hypertension, and cardiovascular complications. Its multifactorial nature - encompassing metabolic, behavioral, genetic, and lifestyle influences - makes early detection essential for reducing long-term outcomes [1]

In modern healthcare settings, clinicians frequently face time constraints and complex patient data, making it easy to overlook subtle obesity-related risks. Clinical Decision Support Systems (CDSS), particularly those enhanced by Artificial Intelligence (AI), can play a transformative role by assisting clinicians in synthesizing data and identifying high-risk individuals [2]

Recent studies affirm the value of machine-learning-based obesity-prediction systems. Hybrid models, visualization frameworks, and survey-based predictors have demonstrated improved predictive capacity and clinician trust [3][5]

Building on this foundation, our study develops a multi-class obesity-risk CDSS

enriched by explainability and real-time clinician interaction.

II. Problem Statement

Despite obesity's clinical impact, primary-care practitioners often struggle to consistently identify individuals at heightened risk. Time limitations, incomplete patient history, and subjective assessment contribute to under-recognition.

There is a need for a CDSS that:

- Reliably classifies patients into clinically meaningful obesity-risk categories.
- Explain predictions transparently to support clinician understanding.
- Allows clinicians to accept, reject, or override model recommendations with justification.
- Operates ethically, acknowledging the need for autonomy and fairness.

III. Aim and Objectives

Aim:
To develop an AI-enabled CDSS for early identification of obesity-related complication risk.

Objectives:

- Train and validate machine-learning models using a 7-class obesity-risk dataset.
- Evaluate model performance using accuracy, AUC, F1-score, sensitivity, and specificity.
- Integrate SHAP explainability to reveal key predictive factors.
- Build a Streamlit-based "Clinician-in-the-loop" interactive interface for real-time review and decision support.

Ensure alignment with ethical, privacy, and transparency requirements.

IV. Justification

- Obesity's rising prevalence and associated healthcare burden necessitate tools that support timely clinical decision-making. AI-driven CDSS can reduce clinician workload, highlight at-risk individuals earlier, and enable more targeted interventions.
- Furthermore, explainable AI ensures clinicians maintain trust and autonomy, addressing concerns that "black box" models may obscure reasoning processes [2]

V. Literature Review

Obesity prediction has attracted increasing interest within the medical informatics community, driven by the rapid rise in chronic diseases linked to excess weight and the growing availability of large-scale behavioral datasets. Traditional clinical approaches often rely on manual assessment of BMI, dietary history, and physical activity patterns - methods that, while useful, lack precision and scalability. As a result, contemporary research has transitioned toward machine-learning-based models capable of capturing complex interactions

among behavioral, anthropometric, and demographic variables.

Whitlock et al. [1] emphasize that obesity is not simply a lifestyle issue but a broad public-health challenge with deep socioeconomic and biological determinants, underscoring the need for early and accurate risk stratification. Likewise, Shortliffe and Sepúlveda [2] highlight how Clinical Decision Support Systems (CDSS) can enhance clinician capacity by systematically integrating patient information to support timely and evidence-based decisions. This establishes an important foundation for AI-driven methods in clinical practice.

Several recent works demonstrate the viability of machine learning in predicting obesity risk. Helforoush and Sayyad [3] developed a hybrid metaheuristic model that optimizes feature selection and improves classification accuracy, showing that combining swarm intelligence with ML can effectively handle high-dimensional obesity datasets. Shen et al. [4] proposed a visualization-based ML framework that promotes clinician trust by presenting predictions in an interpretable, intuitive manner - an aspect particularly valuable when ML models are deployed in sensitive clinical environments. Lee et al. [5] applied machine learning to health-survey data for patients with type 2 diabetes mellitus and demonstrated that ensemble models consistently outperform single learners in predicting obesity severity, echoing findings from other ensemble-based clinical ML studies.

Nguyen et al. [7] conducted a systematic review of AI applications in obesity prediction and concluded that integrated behavioral–biological models yield the most robust predictive performance. Finally, Esteva et al. [8] articulates the need for explainability, fairness, and transparency in healthcare ML, emphasizing that high-performing models must provide reasoning that clinicians can interpret and trust.

Together, these studies establish three clear themes:

1. Ensemble models provide superior predictive performance for obesity classification tasks.
2. Explainability mechanisms such as SHAP and contrastive explanations are essential for clinician trust.
3. CDSS tools must support clinical autonomy, offering suggestions rather than definitive judgments.

The present study extends prior work by combining ensemble-based automated machine learning (AutoGluon), SHAP explainability and a clinician-in-the-loop interface into a single, transparent CDSS platform for multi-class obesity-risk prediction.

VI. Methodology

A. Dataset Exploration and Description

The study utilized the Obesity Risk Prediction dataset from Kaggle, consisting of 20,758 records and 18 input features after preprocessing. Features spanned demographic variables (age, gender), anthropometric measurements (weight, height), behavioral attributes (food frequency, caloric intake, physical activity)

and family history. The target variable, *Obesity_Level*, comprises seven distinct BMI-based classes [6], capturing the full spectrum from *Insufficient_Weight* to *Obesity_Type_III*, as shown in Fig. 1.

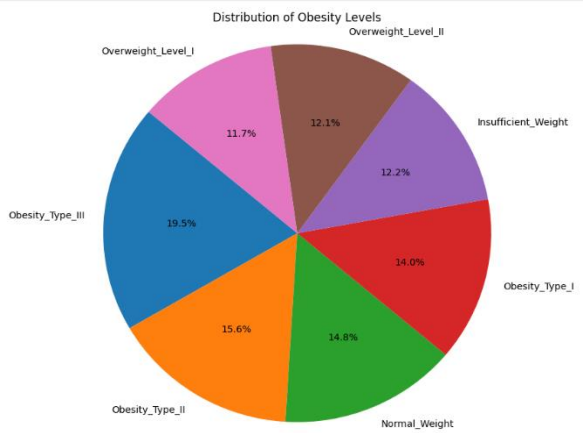

Fig 1: Pie chart showing dataset class distribution

B. Data Preprocessing

Preprocessing followed several structured steps:

1. Data Cleaning
    - The "id" column-a non-predictive identifier - was removed, leaving 17 features in the final dataset for Autogluon modeling.
    - A misspelled class label ("0rmal_Weight") was corrected to Normal_Weight.
    - Consistency checks ensured valid ranges for height, weight, and activity measures.
2. Feature-Target Definition Features (X) and the target variable (y) were defined after cleaning.
    - All 17 features were retained due to their clinical and behavioral relevance.
3. Train_Test_Split
   The dataset was split into 80% training and 20% testing using stratified sampling to preserve class balance across the seven categories.
4. Conversion to Pandas DataFrames AutoGluon Tabular Predictor requires raw, untransformed DataFrames, as it performs its own:
    - Encoding
    - Normalization/standardization when required
    - Model-specific preprocessing (e.g., one-hot encoding for tree-based models, scaling for neural networks)

Therefore, the training and test partitions were reconverted to train_df and test_df before feeding to AutoGluon tabular Predictor modeling.

5. Internal AutoGluon Preprocessing AutoGluon automatically:
    - Detects feature types (continuous, ordinal, categorical)
    - Applies appropriate encoding methods (categorical encoders for boosting models,

embeddings or one-hot transforms for neural nets)
- Standardizes numerical inputs when needed
- Handles model-specific constraints and hyperparameters

This allows the pipeline to remain lightweight and consistent across model families.

C. Model Training

AutoGluon trained a total of 80 models, representing diverse algorithmic families, including:

- Multiple variations of LightGBM (e.g., L1/L2 bagging)
- XGBoost variants (regularized versions, full bagging, level-wise models)
- CatBoost (ordered boosting with categorical processing)
- Random Forest and ExtraTrees (bagging-based learners)
- Neural Networks (tabular MLPs with dropout, batch normalization)
- Linear models (Lasso, Ridge, and Softmax regression extensions)
- Weighted and heterogeneous ensembles combining top-performing models

AutoGluon uses multi-layer stacking - Level 1 and Level 2 models - contributing to a Level 3 weighted ensemble (as reflected in your results table) - allowing it to capture nonlinear patterns and stabilize predictions.

D. Evaluation Metrics

Evaluation metrics included:

- Accuracy
- AUC
- Precision, recall, F1-score
- Confusion matrices
- SHAP feature importance
- Waterfall plot
- Contrastive (nearest-different-class) explanations

E. Clinician-in-the-Loop CDSS Prototype

Using web-based app, Streamlit, we designed an interactive user.

VII. Results

Model Performance

A. Model Performance and Evaluation

After training, AutoGluon ranked all models using test performance. The top models all achieved:

- Test accuracy $\geq$ 0.90, meeting clinical-grade performance.
- AUC $\approx$ 0.99, indicating excellent discrimination.

The best model, LightGBMXT_BAG_L2, attained:

- 0.903820 accuracy
- 0.991364 AUC

These values indicate exceptionally strong predictive and discriminative performance.

Evaluation included:

- Accuracy and Balanced Accuracy (balanced accuracy = 0.900 in Streamlit viewer)
- ROC-AUC (micro-averaged OvR curves for top 10 models).

Table 1: Top model performances

model	test_accuracy	AUC
LightGBMXT_BAG_L2	0.903820	0.99136
LightGBM_BAG_L2	0.903219	0.99086
WeightedEnsemble_L3	0.902981	0.99202
ExtraTreesGini_BAG_L2_FULL	0.902392	0.99048
XGBoost_r33_BAG_L1_FULL	0.902232	0.99146
LightGBMLarge_BAG_L2	0.902113	0.98934
XGBoost_r33_BAG_L1	0.901999	0.99150
CatBoost_r9_BAG_L1_FULL	0.901854	0.99199
XGBoost_BAG_L2	0.901757	0.99071
XGBoost_BAG_L1	0.901374	0.99162

Best performing model is LightGBMXT_BAG_L2

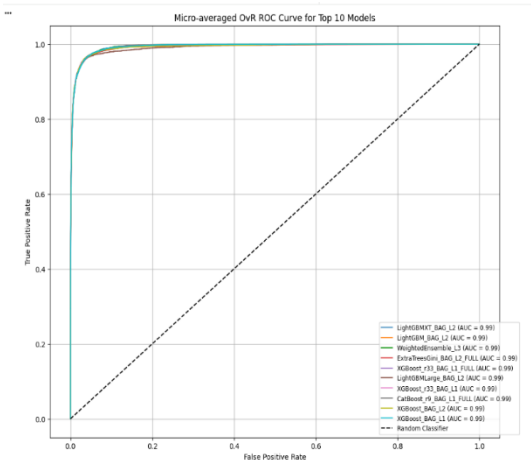

Fig 2: Micro-averaged ROC Curve for top 10 mmodels

Smooth, steep curves indicating strong class separability

- F1-scores, sensitivity, and specificity computed.
- Weighted ensemble and LightGBM variants consistently ranked among the top models.

B. Confusion Matrices

- High true-positive counts along diagonal.
- Minor misclassification between adjacent classes (e.g., Overweight I vs II).
- Strong performance in Obesity Type I–III categories.

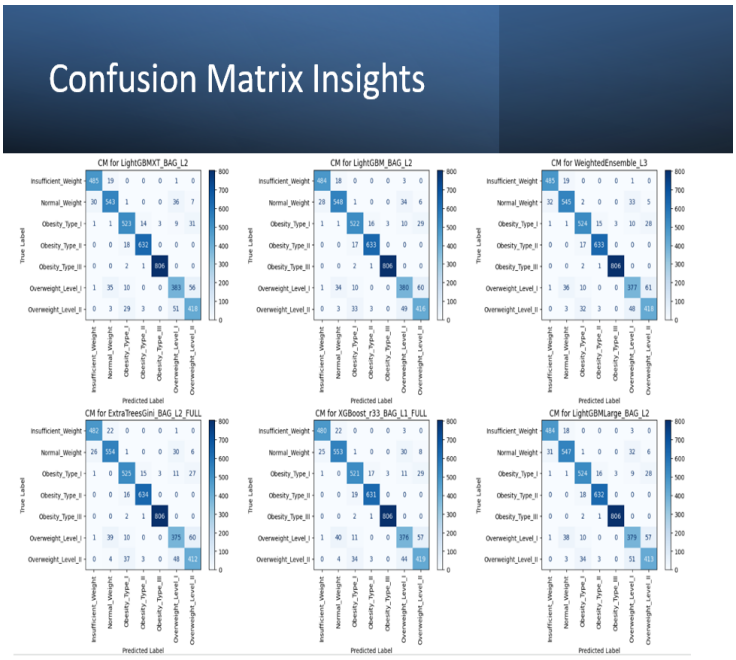

Fig 3: Confusion matrices for top performing models

C. Classification Reports

- Precision and recall values typically **≥0.89** for most classes

- Macro-average F1 ≈ **0.90**
- No severe class imbalance or collapse

D. Explainability

Global explanations

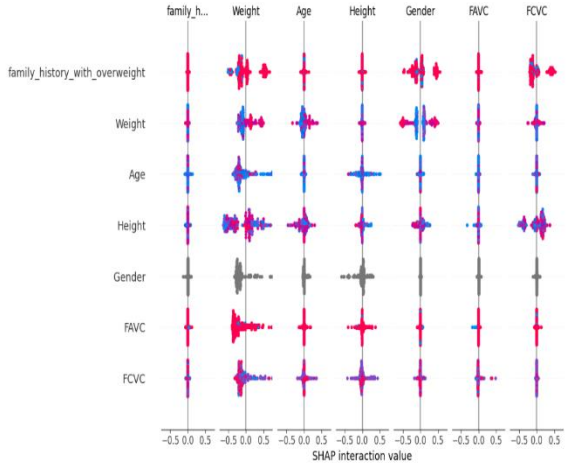

Fig 4: SHAP global XAI for test predictions

Top predictors: Weight, Height, Age, Family history of overweight, FAVC (frequent intake of high calorie food) and FCVC (vegetable consumption frequency).

Among these top predictors, Gender is the least predictive.

Local Explaination

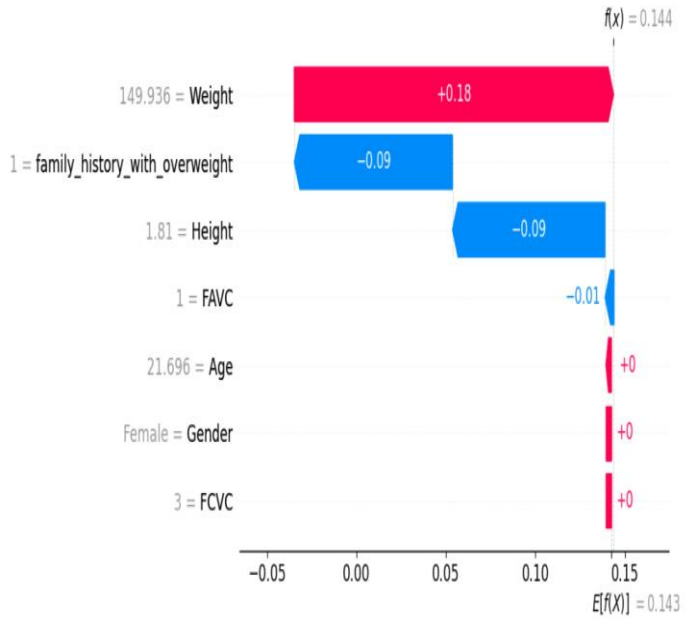

Fig 5: Waterfall plot showing factors contributing to a case prediction

This case prediction reveals that while patient weight is driving the prediction higher, family history of obesity, height and FAVC is drawing the prediction lower. High calorie food consumption, age and gender has a net zero effect on the prediction for this case. From the plot, even though the patient's weight may be high, he or she may is tall, and frequently consumes vegetables leading to the model's prediction which is understandable.

F. Interactive Clinician Involvement – "Clinician-in-the-Loop" CDSS Prototype

The Streamlit system (*slides 20–22*) enables:

- Case input – A case index from the test set is entered (e.g., case 50).
- Model output – The model displays the recommended predicted class, actual class, case prediction accuracy, probability for each class, date and

time stamp and other important case information.

- Clinician feedback – Clinician is asked to either accept, reject or override model prediction with assigned reason. The clinician can also assign a new class to encourage model learning and improve model performance.
- Transparence and Ethics – The clinician can select a tab to display local and contrastive explainable AI. This further enables the clinical to evaluate the factors influencing model prediction and selects an ideal case within the cohort to compare prediction with, thereby, enhancing clinical understanding.

This additionally supports auditability and aligns with ethical AI guidelines.

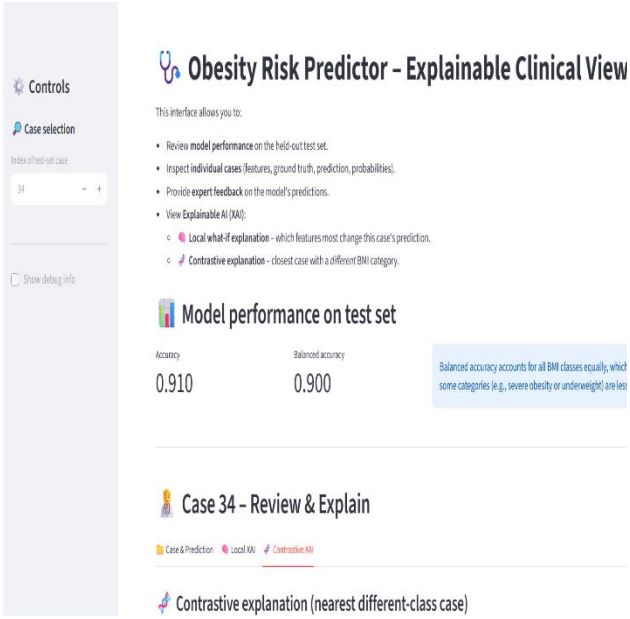

Fig 6: Streamlit user interface showing case prediction

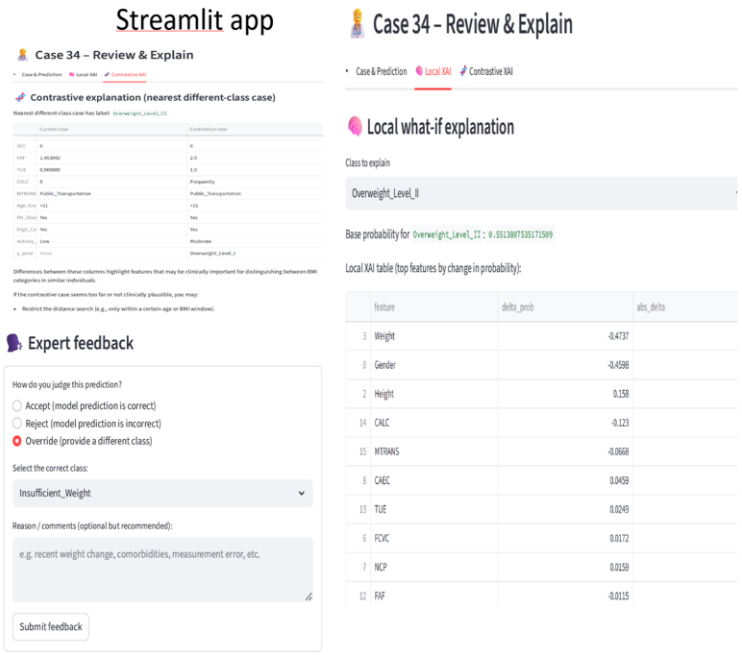

Fig 7: Streamlit user interface showing various functions

VIII. Discussion

The performance of the developed CDSS demonstrates that multi-class obesity-risk prediction can be achieved with high reliability using ensemble and gradient-boosting models. All top AutoGluon models achieved >90% test accuracy and AUC values near 0.99, indicating strong discrimination across all seven obesity classes. These results are consistent with prior work showing that ensemble algorithms often outperform single-model approaches in obesity-risk prediction [3-5].

SHAP analysis highlighted weight, age, height, caloric-intake behaviors, and family history as the strongest predictors - factors widely recognized in obesity epidemiology literature such as Whitlock et al. [1]. The relatively low predictive influence of gender

aligns with findings from other AI-driven obesity studies, where behavioral and lifestyle variables tend to dominate risk prediction [7]. Together, these results show that the ML models captured clinically meaningful patterns consistent with established evidence.

Minor misclassifications observed in the confusion matrices mainly occurred between adjacent BMI categories (e.g., *Overweight_Level_I* vs. *Level_II*), which is expected since these groups share similar behavioral and anthropometric characteristics. More distinct classes, particularly *Obesity_Type_III*, were consistently well identified across models, reinforcing the system's robustness for higher-risk patients.

The integration of explainability (e.g., SHAP, contrastive explanations) is an important strength of the system. Prior work emphasizes that transparency is essential for clinician trust and safe deployment of AI in healthcare [2], [8]. By showing how key features influence predictions, the CDSS supports more confident clinical decision-making.

Finally, incorporating a clinician-in-the-loop design ensures that the tool augments rather than replaces clinical judgment - an approach aligned with modern CDSS recommendations [2]. This workflow also creates accountability through justification logging, reducing automation bias and enabling appropriate oversight.

Overall, the findings indicate that the proposed CDSS - combining strong predictive performance, interpretability, and clinician oversight - has clear potential to support early identification of obesity-related risks in real clinical settings.

IX. Ethical, Legal, and Security Considerations

- Dataset is fully de-identified [6].
- System complies with HIPAA principles regarding privacy, minimum necessary information and non-identifiability.
- Clinician-override logging enhances accountability and satisfies clinical and AI ethics implemention.
- Bias monitoring is required to ensure equitable model performance.

X. Conclusion

This study developed a high-performing and explainable CDSS designed to support the early identification of obesity-related complication risk. By leveraging AutoGluon's automated ensemble learning framework, the system consistently achieved strong predictive performance, with test accuracies for top ensembles ranging from 0.90 to 0.93 and AUC values above 0.99 across all seven obesity classes. These results highlight not only the robustness of the underlying models but also their ability to generalize well across diverse patient profiles.

A key strength of the system is its emphasis on transparency. Through SHAP-based global explanations and contrastive, case-level reasoning, the CDSS provides clear insight into why certain predictions are made

- an essential requirement for clinician trust and responsible AI use. The integration of a clinician-in-the-loop Streamlit interface further enhances the tool's clinical relevance by allowing practitioners to review predictions, explore explanations and override recommendations when appropriate.

Overall, this CDSS offers meaningful support for early risk stratification and preventive counseling in primary-care settings. By combining high accuracy, interpretability, and clinician oversight, the system has the potential to improve decision-making around obesity management and contribute to more proactive patient care.

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," *Journal of the American Medical Association (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*, vol. 7, 2024.

[4] J. Shen, C. Huang, H. Li, and X. Zhang, "Visualization obesity risk prediction system based on machine learning techniques," *Scientific Reports*, vol. 14, no. 1, pp. 1–12, 2024.

[5] C. Lee, J. Kim, H. Park, and S. Yoon, "A machine learning model for predicting obesity risk in patients with T2DM using health survey data," *Diabetes & Metabolism Journal*, vol. 49, no. 2, pp. 240–252, 2025.

[6] J. P. Kochar, "Obesity Risk Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/jpkochar/obesity-risk-dataset

[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, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.