Er. Madeeha Laiq, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

AWTG-PD: A Trajectory-First Personalized Framework for Early Warning of Near-Term Weight Gain

Er. Madeeha Laiq[1], Er. Animesh Kushwaha[2], Er. Ajaz Ahmad[3]
Department of CSE,
MVIET, Kaushambi, India[1,3,]

Department of CSE
Jawaharlal Nehru College of Technology, Rewa[2]

Abstract- Early identification of emerging weight gain may enable preventive intervention before a sustained increase becomes established. Conventional weight-gain prediction often uses static characteristics or long prediction horizons, while short-horizon forecasting primarily estimates future numerical weight. This study proposes AWTG-PD (Adaptive Weight Trajectory Guidance with Personal Deviation), a longitudinal machine-learning framework combining multi-window weight trajectories with deviation from an individual-specific recent baseline. The study used the publicly available FitLife360 health and fitness tracking dataset obtained from Kaggle. Weight features included 7-days, 14-days, and 30-day averages and changes, trajectory slope, personal-baseline deviation, and deviation z-score. A near-term gain event was defined as a future 14-day average weight increase of at least 0.10 kg. Chronological training, validation, and test periods were used, followed by ablation, threshold, early-warning, risk-stratification, bootstrap, and calibration analyses. On the held-out test set, AWTG-PD achieved ROC-AUC 0.7018 (95% CI: 0.6745–0.7263) and PR-AUC 0.1818 (95% CI: 0.1537–0.2144). At threshold 0.65, precision was 0.1914, recall 0.3281, F1 0.2418, and alert rate 16.44%. The model generated early warnings for 124 users with a mean lead time of 9.008 days. At the same alert rate, a 30-day baseline generated 119 early-warning users with a mean lead time of 8.807 days. Risk stratification identified a high-trajectory/high-deviation group with a gain-event rate of 19.76% and relative risk of 2.459. A lifestyle-only model achieved ROC-AUC 0.5050. Calibration indicated substantial overestimation of absolute risk, so model outputs should currently be interpreted as relative alert scores rather than calibrated probabilities. The findings support trajectory-based early warning as a methodological approach, while external validation in real-world longitudinal cohorts remains necessary.

Keywords— Weight gain prediction, longitudinal machine learning, trajectory modeling, personalized risk, early warning, adaptive intervention, digital health.

I. INTRODUCTION

Overweight and obesity are major public-health challenges. The World Health Organization reported that 2.5 billion adults were overweight in 2022, including more than 890 million adults living with obesity [1]. Because weight gain can accumulate gradually, identifying an unfavorable trajectory before a substantial increase occurs is a relevant preventive-computing problem.

Machine-learning methods have increasingly been applied to obesity and weight-related risk prediction. A systematic review and meta-analysis found that many machine-learning studies reported discrimination above 0.70, while also noting substantial heterogeneity across datasets and methods [2]. Jawara et al. developed elastic-net and XGBoost models using the All of Us Research Program to predict at least 10% total body-weight gain within two years; XGBoost achieved an AUC of 0.706 [3]. An electronic-medical-record study predicting at least a 10% BMI increase within 12 months reported an AUC of 0.75 for XGBoost [4]. These studies demonstrate that future weight gain can be modeled, but their prediction horizons are substantially longer than the short-term changes that may be useful for an early-warning system. Longitudinal measurements provide additional information about direction, persistence, and recent deviation. Rolling averages, changes, and trajectory slopes can therefore represent an evolving state rather than a single static measurement.

Short-horizon weight forecasting has also emerged as a research direction. Cheng et al. developed a sequence-based framework for seven-day body-weight forecasting using behavioral and physiological sequences, with FitLife360 as the primary benchmark and PMData as a supplementary real-world feasibility dataset [5]. That work focused on numerical future-weight forecasting. In contrast, the present study formulates near-term gain as an event-oriented early-warning problem and explicitly evaluates alert burden and lead time.

Personalization is a second challenge. An absolute change may have different significance for different individuals because normal weight variability is not identical across users. A rolling personal baseline can provide an interpretable reference against which current weight is compared. Recent just-in-time adaptive intervention research similarly emphasizes real-time monitoring, personalization, and delivery of support at appropriate moments [6], [7].

Accordingly, this study proposes AWTG-PD, a trajectory-first framework that links longitudinal trajectory, personal deviation, risk scoring, early warning, risk stratification, and adaptive intervention. The study asks whether trajectory information is more informative than current weight, whether personal deviation improves risk stratification, whether warnings can precede a qualifying event, and whether risk states can be converted into interpretable intervention levels.

II. LITERATURE REVIEW

Prior weight-gain prediction studies have largely targeted outcomes over months or years. The All of Us study reported modest discrimination for two-year weight gain and found that survey-based behavioral information did not materially improve XGBoost performance [3]. The EMR study by Choong et al. similarly demonstrated moderate discrimination for 12-month BMI increase [4].

Longitudinal forecasting provides a different perspective. The recent FitLife360 study demonstrated the feasibility of short-horizon sequence forecasting and explicitly characterized FitLife360 as a longitudinal dataset [5]. AWTG-PD is complementary rather than a direct replacement: it uses longitudinal observations to detect a future event and to determine whether an alert occurs sufficiently early.

Adaptive intervention research indicates growing interest in just-in-time support. A 2025 scoping review of weight-management JITAIs included 35 studies and reported that 31.4% used machine learning for decision-making [6]. A broader scoping review found that reporting of JITAIs remains inconsistent, motivating explicit separation of prediction, decision rules, and intervention components [7].

The research gap addressed here is the integration of short-horizon event detection with personal-baseline context and operational early-warning evaluation. The contribution is therefore a unified framework rather than a claim of universal predictive superiority over simpler baselines.

III. MATERIALS AND METHODS

1. Dataset

The study used the FitLife360 Health & Fitness Tracking Dataset obtained from Kaggle [8]. The dataset card describes FitLife360 as a longitudinal dataset simulating health and fitness tracking records for 3,000 participants over approximately one year. For this study, the processed analytical data contained 1,000 longitudinal users and 137,000 observations across the training, validation, and test periods.

Partition	Days	Observations	Users
Training	30–110	81,000	1,000
Validation	111–140	30,000	1,000
Test	141–166	26,000	1,000
Total	30–166	137,000	1,000

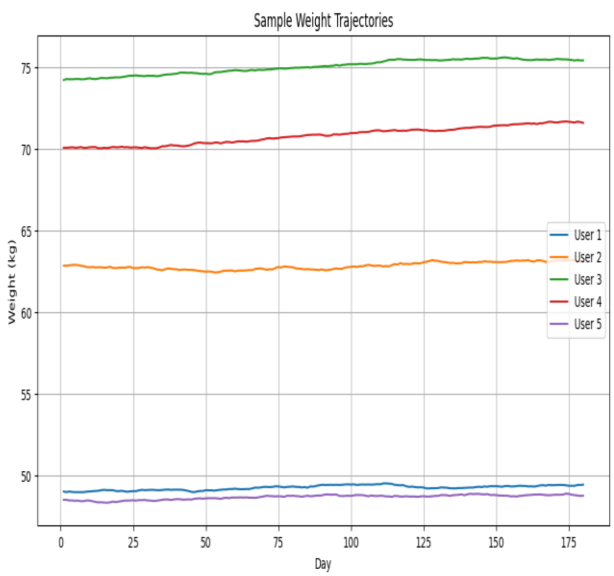

Fig. 1. Representative longitudinal weight trajectories across five users, illustrating between-user heterogeneity in absolute weight level and temporal evolution.

2. Temporal Study Design

Chronological partitioning was used to preserve the direction of time: training observations ended at day 110, validation covered days 111–140, and testing began at day 141. Because the intended application is longitudinal monitoring of previously observed users, users occur across temporal partitions; therefore, this is a within-user temporal generalization design rather than an unseen-user generalization study.

3. Feature Engineering

For user i at day t, current weight is denoted by W(i,t). Rolling averages were calculated over 7, 14, and 30 days. Weight change over a window k was calculated as $\Delta W(i,t,k) = W(i,t) - W(i,t-k)$, where $k \in \{7,14,30\}$. Trajectory slope was calculated over the corresponding historical window. The final trajectory representation contained current weight, three rolling averages, three rolling changes, and 7-day slope.

The 30-day personal baseline was calculated as the mean of the user's recent historical weights. Personal deviation was defined as current weight minus this baseline. A deviation z-score was calculated by dividing deviation by the user's recent 30-day standard deviation. The final 12-feature AWTG-PD representation comprised the eight trajectory features plus Personal_Baseline_30d, Weight_Deviation, Deviation_7d, and Deviation_Z.

4. Outcome Definition

Future 14-day average weight was calculated from the subsequent 14 observations. Future change was defined as future 14-day average weight minus current weight. A gain event was assigned when future 14-day average change was at least 0.10 kg. The test-set event prevalence was approximately 9.59%.

5. Proposed AWTG-PD Framework

The framework contains five conceptual stages: longitudinal feature construction, personal-baseline estimation, machine-learning risk scoring, threshold-based early warning, and explainable intervention assignment.

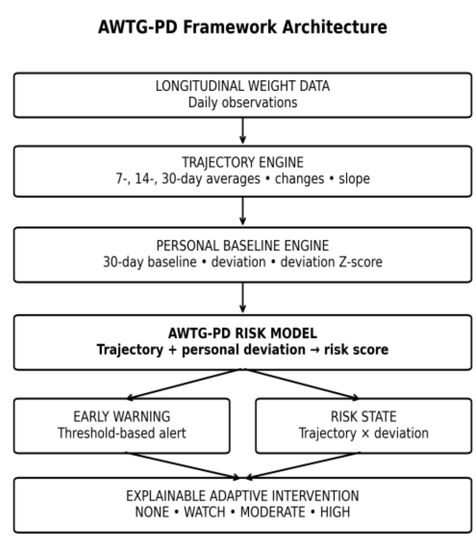

Fig. 2. Architecture of the proposed AWTG-PD framework.

6. Risk Stratification and Intervention

High trajectory was defined as 30-day weight change ≥0.24 kg, while high personal deviation was defined as absolute deviation ≥0.10 kg. These rules produce four states: NONE (low trajectory/low deviation), WATCH (low trajectory/high deviation), MODERATE (high trajectory/low deviation), and HIGH (high trajectory/high deviation).

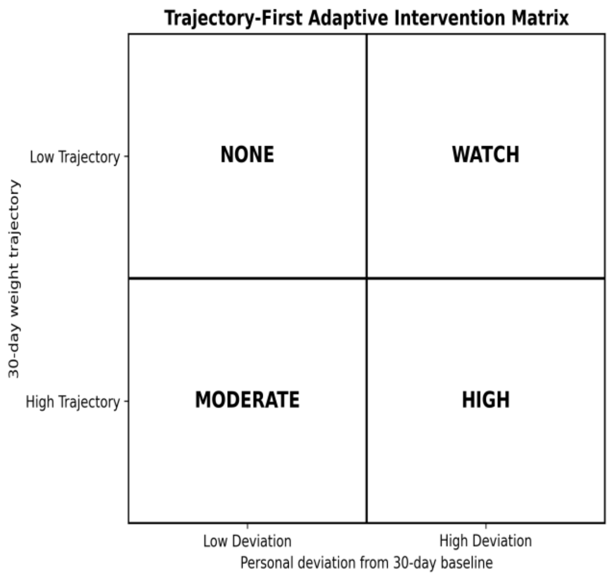

Fig. 3. Trajectory-first adaptive intervention matrix used to translate trajectory and personal deviation into interpretable intervention levels.

7. Early-Warning Algorithm

Algorithm 1 — AWTG-PD Early-Warning Procedure

- Sort observations chronologically.
- Calculate 7-, 14-, and 30-day weight features and trajectory slope.
- Estimate the 30-day personal baseline and deviation measures.
- Generate the AWTG-PD risk score.
- Apply the selected threshold and generate an alert when exceeded.
- Identify the first qualifying future gain event and calculate lead time.
- Assign the trajectory/deviation risk state.
- Map the state to NONE, WATCH, MODERATE, or HIGH intervention.
- Generate a numerical explanation.

8. Model Evaluation

Performance was evaluated using ROC-AUC, PR-AUC, precision, recall, F1 score, accuracy, false-positive rate, false-negative rate, alert rate, and early-warning lead time. Because event prevalence was approximately 9.59%, PR-AUC was treated as an important complementary measure. User-level bootstrap resampling was used to estimate 95% confidence intervals for ROC-AUC and PR-AUC. Calibration was evaluated using the Brier score and decile-based calibration values.

IV. RESULTS

1. Component Ablation

Model	Features	ROC-AUC	PR-AUC
Current Weight	1	0.4877	0.0981
7-day Change	1	0.6218	0.1375
30-day Change	1	0.7116	0.1812
Trajectory Only	8	0.7120	0.1894
Trajectory + Personal Deviation	11	0.7118	0.1926
Trajectory + Z-score	9	0.7113	0.1906
Full AWTG	12	0.7114	0.1916
Lifestyle-only	8	0.5050	0.0999

The ablation results show that longitudinal trajectory information was substantially more informative than current weight alone. The trajectory-only configuration achieved the highest ROC-AUC (0.7120), while trajectory plus personal deviation achieved the highest PR-AUC (0.1926). The lifestyle-only model was close to random discrimination.

2. Final Test Performance

Metric	AWTG-PD
Threshold	0.65
Precision	0.1914
Recall	0.3281
F1	0.2418
Accuracy	0.8027
Alert rate	16.44%
TN	20,051
FP	3,456

FN	1,675
TP	818
False-positive rate	0.147
False-negative rate	0.6719

At threshold 0.65, the model achieved moderate discrimination with a 16.44% alert rate. The confusion matrix contained 20,051 true negatives, 3,456 false positives, 1,675 false negatives, and 818 true positives.

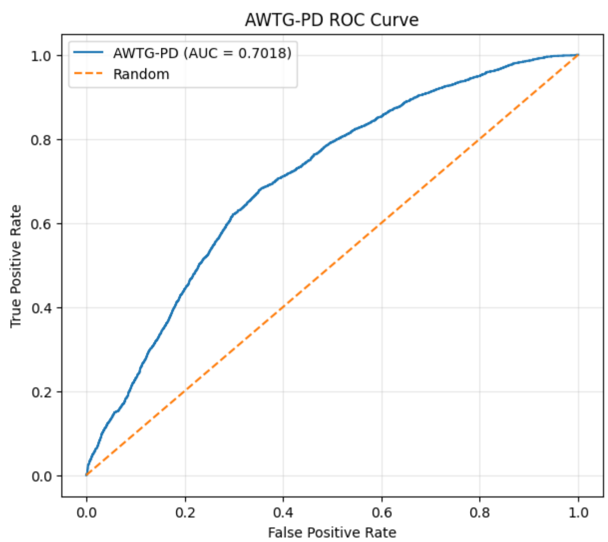

Fig. 4. Receiver operating characteristic curve of the final AWTG-PD model; ROC-AUC = 0.7018.

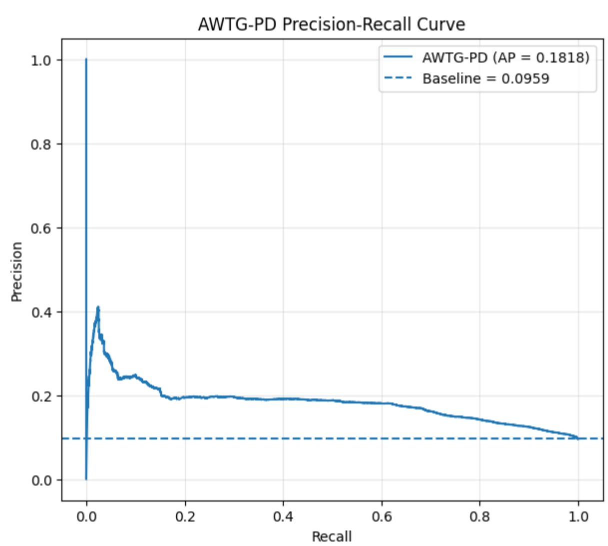

Fig. 5. Precision-recall curve of the final AWTG-PD model; PR-AUC = 0.1818 and prevalence baseline ≈0.0959.

3. Bootstrap Discrimination and Calibration

Metric	Estimate	95% CI
ROC-AUC	0.7018	0.6745–0.7263
PR-AUC	0.1818	0.1537–0.2144

User-level bootstrap analysis produced stable uncertainty intervals around both discrimination measures. The Brier score was 0.2104. The calibration curve showed that predicted risk was systematically higher than observed event rates across risk deciles, indicating overestimation of absolute probability.

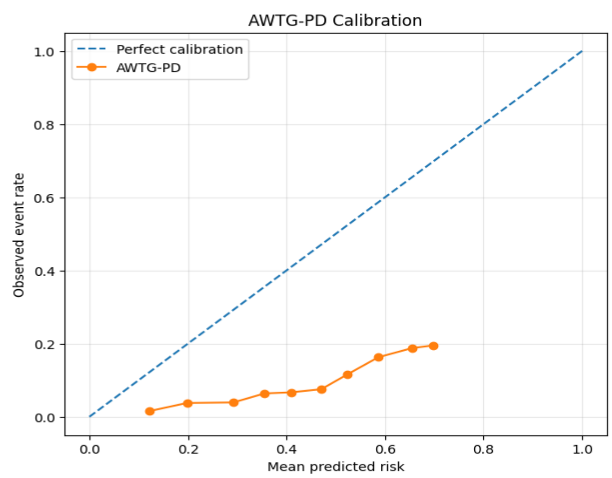

Fig. 6. Calibration curve of AWTG-PD. The deviation from the perfect-calibration line indicates overestimation of absolute event probabilities.

4. Early-Warning Performance

Threshold	Alert Rate	Early-Warning Users	Mean Lead	Median Lead
0.55	29.75%	182	9.577 d	9 d
0.60	23.31%	153	9.288 d	8 d
0.65	16.44%	124	9.008 d	8 d
0.70	3.14%	32	7.156 d	5 d

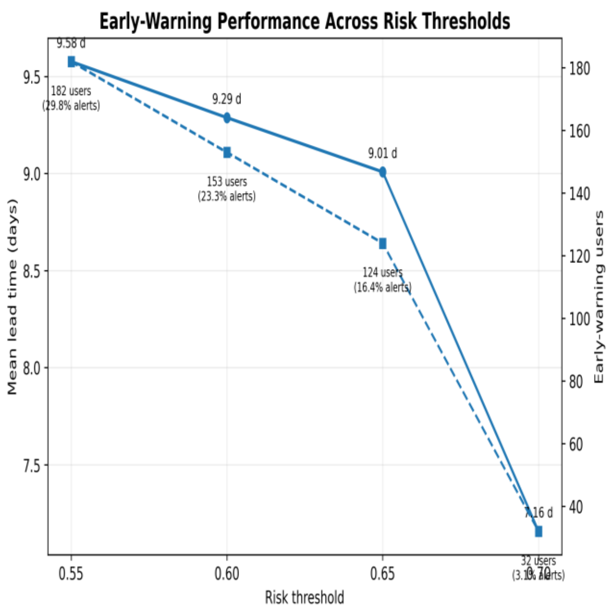

Fig. 7. Early-warning performance across thresholds, showing the trade-off between alert burden, warning population, and mean lead time.

5. Fair Alert-Budget Comparison

Model	Precision	Recall	F1	Alert Rate	Warned Users	Mean Lead
AWTG-PD	0.1914	0.3281	0.2418	16.44%	124	9.008d
30-day Baseline	0.1965	0.3369	0.2482	16.44%	119	8.807d

At the same 16.44% alert budget, the 30-day baseline slightly exceeded AWTG-PD on precision, recall, and F1. AWTG-PD nevertheless identified five additional early-warning users and showed a slightly longer mean lead time. Accordingly, the proposed framework is not claimed to be universally superior to the simpler baseline.

6. Personalized Risk Stratification

Risk Group	Observations	Gain Events	Gain Rate	Relative Risk
Low trajectory + Low deviation	14,620	1,175	8.04%	1.000
High trajectory + Low deviation	374	69	18.45%	2.296
Low trajectory + High deviation	7,105	478	6.73%	0.837
High trajectory + High deviation	3,901	771	19.76%	2.459

The association between risk group and gain-event status was statistically significant ($\chi^2 = 607.4868$, $p = 2.40 \times 10^{-131}$, $df = 3$). The high-trajectory/high-deviation group had the highest event rate and relative risk.

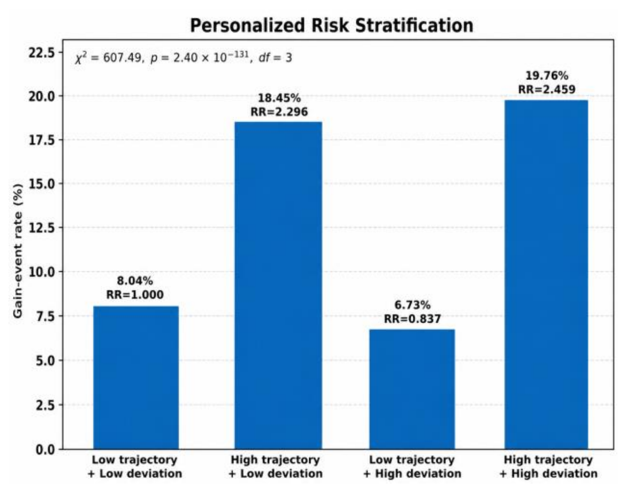

Fig. 8. Gain-event rates across trajectory and personal-deviation risk groups. The high-trajectory/high-deviation group shows the highest observed event rate.

7. Lifestyle Analysis

Feature	Future-Change Corr.	Event Corr.
Steps	0.0053	0.0078
Sleep Hours	−0.0016	0.0047
Work Hours	−0.0041	−0.0021
Sitting Hours	−0.0085	−0.0067
Dinner Hour	−0.0032	−0.0021

The lifestyle-only model achieved ROC-AUC 0.5050 and PR-AUC 0.0999. Lifestyle variables therefore showed limited standalone predictive discrimination for the defined near-term event and are better treated as contextual inputs for future recommendation logic.

V. DISCUSSION

The main finding is that longitudinal trajectory carries substantially more predictive information than current weight alone. Current weight produced ROC-AUC 0.4877, whereas 30-day change reached 0.7116 and the multi-window trajectory model reached 0.7120. This pattern supports the trajectory-first hypothesis that direction and persistence of recent change are more informative for the defined near-term event than an isolated absolute measurement.

Personal deviation provided additional value primarily in precision-recall behavior and risk stratification. The trajectory-plus-personal-deviation model achieved the highest PR-AUC among the tested AWTG configurations, and the high-trajectory/high-deviation group showed a 19.76% event rate with relative risk 2.459. However, isolated high deviation without an elevated trajectory did not increase risk in this dataset, indicating that deviation should be interpreted in trajectory context.

The early-warning analysis demonstrates the value of explicitly evaluating warning timing. At threshold 0.65, AWTG-PD produced a mean lead time of 9.008 days. Nevertheless, the matched-alert comparison showed that the 30-day baseline had slightly better precision, recall, and F1. This result prevents the proposed framework from being characterized as a superior classifier solely on conventional metrics.

The contribution of AWTG-PD is instead its integration of prediction and action-oriented interpretation. The model produces a risk score, while the trajectory/deviation layer translates that score into a transparent state and intervention level. This design is consistent with current work on personalized and just-in-time digital interventions [6], [7].

The lifestyle analysis produced near-random standalone discrimination. This does not imply that lifestyle is unimportant for weight management; rather, the measured lifestyle variables did not provide strong standalone discrimination for this particular short-horizon endpoint. Their more defensible role is contextualization of future recommendations.

The calibration findings are a critical limitation. Predicted scores were substantially higher than observed event rates, so the current model should not be used to communicate absolute probability without recalibration. Future work should apply calibration using an independent calibration set and evaluate clinical utility prospectively.

Limitations

First, FitLife360 is documented by its Kaggle repository [8], so the findings should be interpreted as methodological evidence rather than clinical effectiveness evidence. Second, the temporal design contains the same users across training, validation, and test periods; this evaluates longitudinal prediction for previously observed users but not generalization to unseen users. Third, calibration was inadequate for absolute probability interpretation. Fourth, recall at threshold 0.65 was 0.3281, leaving many qualifying events undetected. Fifth, the simpler 30-day baseline slightly outperformed AWTG-PD on conventional classification metrics at the matched alert rate. Finally, the adaptive intervention engine has not been prospectively tested for behavioral or clinical benefit.

VI. CONCLUSION

AWTG-PD provides a trajectory-first framework for near-term weight-gain early warning that combines multi-window longitudinal features with individual baseline deviation. On the held-out test set, the model achieved ROC-AUC 0.7018 and PR-AUC 0.1818, while threshold 0.65 produced a 16.44% alert rate and a mean early-warning lead time of 9.008 days. Risk stratification identified a high-trajectory/high-deviation group with a 19.76% event rate and relative risk 2.459. At a matched alert budget, however, the 30-day baseline retained slightly better precision, recall, and F1. Thus, AWTG-PD is best characterized as an integrated personalized early-warning and intervention framework rather than a universally superior classifier. External validation, unseen-user evaluation, probability calibration, and prospective intervention studies are required before real-world deployment.

REFERENCES

1. World Health Organization, “Obesity and overweight,” WHO Fact Sheet, Dec. 8, 2025.
2. F. M. Delpino, Â. K. Costa, M. C. do Nascimento, et al., “Does machine learning have a high performance to predict obesity among adults and older adults? A systematic review and meta-analysis,” Nutrition, Metabolism and Cardiovascular Diseases, vol. 34, no. 9, pp. 2034–2045, 2024, doi: 10.1016/j.numecd.2024.05.020.
3. D. Jawara, K. V. Lauer, M. Venkatesh, et al., “Using machine learning to predict weight gain in adults: An observational analysis from the All of Us Research Program,” Journal of Surgical Research, vol. 306, pp. 43–53, 2025, doi: 10.1016/j.jss.2024.11.042.
4. “Identifying individuals at risk for weight gain using machine learning in electronic medical records from the United States,” 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12046438/
5. L. Cheng, L. Wang, G. Wang, B. Lu, B. Wu, Y. Xiao, and Q. Huang, “Domain-informed weight forecasting: leveraging behavioral and physiological sequences from wearables,” Frontiers in Public Health, vol. 14, 2026, Art. no. 1810381, doi: 10.3389/fpubh.2026.1810381.
6. J. Y. J. Koh, S. X. Tan, X. M. Tan, et al., “Just-in-time adaptive interventions for weight management among adults with excess body weight: Scoping review,” Journal of Medical Internet Research, vol. 27, 2025, Art. no. e76625, doi: 10.2196/76625.
7. T.-C. C. Hsu, P. Whelan, J. Gandrup, et al., “Personalized interventions for behaviour change: A scoping review of just-in-time adaptive interventions,” British Journal of Health Psychology, vol. 30, no. 1, 2025, Art. no. e12766, doi: 10.1111/bjhp.12766.
8. J. Taheri, “FitLife: Health & Fitness Tracking Dataset,” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/jijagallery/fitlife-health-and-fitness-tracking-dataset
9. G. S. Collins, K. G. M. Moons, P. Dhiman, et al., “TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods,” BMJ, vol. 385, 2024, Art. no. e078378, doi: 10.1136/bmj-2023-078378.