AURUM JOURNAL OF HEALTH SCIENCES
A. J. Health Sci
Volume 8, No 2 | Summer 2026

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN HEALTHCARE: A REGION FIXED EFFECTS APPROACH TO THE GLOBAL HEALTH IMPACT OF OBESITY AND ITS NEAR TERM FORECASTING

*Onur Gürcan ERGEN[1], Mehmet GÖRGÜLÜ[2], İnci OKTAY[3]
[1]Health Management Department, Altınbaş University, Istanbul, Turkey, onurgurcanergen@gmail.com
(https://orcid.org/0000-0001-5756-2168)
[2]Health Management Department, Altınbaş University, Istanbul, Turkey, mehmet.gorgulu@altinbas.edu.tr
(https://orcid.org/0000-0002-3616-6294)
[3]Health Management Department, Altınbaş University, Istanbul, Turkey, inci.oktay@altinbas.edu.tr
(https://orcid.org/4349-0617-0002-0000)

Received: 29.04.2026 Research Article
pted: 25.05.2026 pp. 19-30
Published: 31.07.2026
*Corresponding author DOI: 10.XXXX/ajhs.2023.XXXXX

Abstract
The purpose of this article examines the global health impact of obesity at the World Health Organization (WHO) regional level and generates near-term forecasts of adult obesity prevalence for 2020–2022 using a region fixed-effects multiple linear regression model on a regional year panel. Methods: Region year observations ($n = 60$) for the six WHO regions were retrieved automatically from the WHO Global Health Observatory OData API between 2010–2019 via a Python 3.11 script. A two-stage strategy was applied: four predictors (overweight, underweight, physical inactivity, raised blood pressure) were projected to 2020–2022 by simple linear regression; obesity prevalence was then modelled with region fixed effects multiple linear regression. Out of sample accuracy was assessed through a time ordered back test (training: 2010–2017; test: 2018–2019). The model achieved $R^2 = 0.9989$ and a back test mean absolute percentage error of 4.02%. Overweight ($\beta = +0.7385$), underweight ($\beta = +0.8094$), and physical inactivity ($\beta = +0.2232$) showed significant positive associations with obesity (all $p < 0.01$). Forecast 2022 prevalences were 32.96% (Americas), 27.94% (Eastern Mediterranean), 22.13% (Europe), 11.23% (Africa), 9.66% (Western Pacific), and 7.53% (South-East Asia). The study concludes that the region fixed effects panel regression delivers high accuracy and interpretability for near term obesity forecasting and yields region specific insights translatable into policy under the 6P framework.
Keywords: Obesity, artificial intelligence, panel data analysis, region fixed-effects model, near-term forecasting, 6P framework

1. Introduction

Obesity, defined as excessive body fat accumulation associated with adverse health outcomes, has become one of the most prevalent and complex health problems in modern populations. The World Obesity Federation describes obesity as a chronic, progressive, and relapsing disease process (Bray, Kim, & Wilding, 2017). Pooled analyses from the Non-Communicable Disease Risk Factor Collaboration (NCD-RisC) indicate that, between 1990 and 2022, age-standardized obesity prevalence increased from 8.8% to 18.5% in women and from 4.8% to 14.0% in men, with approximately 880 million adults affected worldwide in 2022 (NCD-RisC, 2024).

The global health impact of obesity is multidimensional. It encompasses prevalence, mortality, quality of life, healthcare access, and the broader burden on health systems. Obesity is causally linked to type 2 diabetes, hypertension, dyslipidaemia, cardiovascular disease, certain cancers, osteoarthritis, and obstructive sleep apnea, generating substantial direct and indirect disease burden (Berberoğlu & Hocaoğlu, 2021). The Global Burden of

Disease (GBD) 2021 Risk Factors analysis demonstrated that high body mass index (BMI) contributes increasingly to disability adjusted life years (DALYs), particularly among older age groups (GBD 2021 Risk Factors Collaborators, 2024). Chen et al. (2024) further projected that age-standardized mortality and DALY rates attributable to high BMI will continue rising through 2035.

Beyond epidemiology, obesity imposes high economic and social costs. Systematic reviews indicate that direct and indirect obesity related costs increase the share of healthcare expenditure in gross domestic product (GDP) and threaten the long-term sustainability of health systems. Socioeconomic gradients also vary across country contexts: in low- and middle-income countries, obesity is often inversely associated with income, whereas the relationship may reverse in high-income settings (Dinsa, Goryakin, Fumagalli, & Suhrcke, 2012). These heterogeneous patterns underscore the inadequacy of one-size-fits-all global policies and emphasize the need for region-specific strategies.

Forecasting health indicators has become a critical input for capacity planning and policy design. Grøntved et al. (2024), in a scoping review of healthcare capacity forecasting, observed that many published studies lack temporal validation and rigorous performance assessment. Statistical and machine learning models can complement traditional epidemiological approaches, but their accuracy depends on data quality, leakage prevention, and explicit uncertainty quantification (Sterne et al., 2009; Starcke, Spadafora, Spadafora, & Toma, 2025). For obesity specifically, scoping reviews show that artificial intelligence (AI) has been applied to measurement, classification, risk prediction, and intervention planning (An, Shen, & Xiao, 2022; Haghighathoseini, Lin, Song, An, & Xue, 2025), with decision tree-based methods often achieving high accuracy in classification tasks (Ölçer, 2024). Nevertheless, interpretable linear panel models remain attractive for policy applications because their coefficients have direct substantive meaning.

This study examines the global health impact of obesity at the WHO regional level and produces near-term forecasts of adult obesity prevalence for the 2020–2022 period using a region-fixed-effects multiple linear regression model built on a balanced regional-year panel with MATLAB. The contribution is threefold. First, the data acquisition pipeline is fully automated through a Python script that queries the WHO Global Health Observatory (GHO) OData API, ensuring transparency and reproducibility. Second, the modelling strategy combines simple trend extrapolation for the predictors with a region fixed-effects regression for the outcome, enabling control for time-invariant regional heterogeneity. Third, a time-ordered back test protocol (training: 2010–2017; test: 2018–2019) is applied to evaluate out of sample accuracy under realistic forecasting conditions. The findings are interpreted within the 6P framework (Predictive, Preventive, Personalized, Participatory, Precision-oriented, Public Health) to translate quantitative results into actionable, regionally prioritized policy recommendations.

2. Methods

2.1. Study Design

This study is a secondary data analysis based on aggregated regional indicators. The analytical unit is the region-year observation rather than individuals or countries, and the design is therefore ecological. Findings reflect population-level trends and forecasts and should not be interpreted as individual level causal inferences (Cohen, 1990). The study uses publicly available, fully anonymized aggregated data and requires no ethics approval.

2.2. Data Sources and Acquisition

All indicators were retrieved from the WHO Global Health Observatory (GHO) OData API (https://ghoapi.azureedge.net/api/). Data extraction was implemented in Python 3.11 through a custom script that first queries the GHO indicator catalog, then requests records for each selected indicator filtered by region and year, and finally produces a single comma separated values (CSV) file for analysis. The full script is available in the corresponding doctoral thesis appendix.

The following GHO indicator codes were used: NCD_BMI_30A (adult obesity prevalence, BMI ≥ 30, age-standardized, %), NCD_BMI_25A (adult overweight prevalence, BMI ≥ 25, age-standardized, %), NCD_BMI_18A (adult underweight prevalence, BMI < 18.5, age-standardized, %), NCD_PAA (insufficient physical activity prevalence, age 18+, age-standardized, %), and BP_04 (raised blood pressure prevalence, age 30–79, %). All series were filtered to the SEX_BTSX (both sexes) category. Regional coverage included all six WHO regions: African Region (AFR), Region of the Americas (AMR), Eastern Mediterranean Region (EMR), European Region (EUR), South-East Asia Region (SEAR), and Western Pacific Region (WPR).

Three indicators initially considered for inclusion (per capita alcohol consumption, mean total cholesterol, and the cost of a healthy diet) were excluded because of insufficient temporal coverage in the WHO GHO API or incompatible regional aggregation. The diabetes indicator (NCD_GLUC_04) was retained as a sensitivity variable but not included in the core model since regional aggregates were available only through 2014.

2.3. Population, Period, and Sample

The target population consisted of region/year observations across the six WHO regions for the period 2010–2019, with 2020–2022 reserved as the forecasting horizon. The 2010–2019 window was selected because it represents the maximum overlapping period for which all five core indicators were jointly available at the regional level. A complete enumeration approach was adopted, yielding a balanced panel of $n = 60$ region-year observations (6 regions × 10 years). With three continuous predictors and five regional dummy variables (Africa as reference), the model includes eight estimated parameters and 51 degrees of freedom, satisfying minimum sample size requirements for multivariable prediction models (Riley et al., 2019).

2.4. Variables

The dependent variable was adult obesity prevalence (BMI ≥ 30, age-standardized, %). The independent variables in the core model were adult overweight prevalence, adult underweight prevalence, and insufficient physical activity prevalence. Raised blood pressure prevalence was retained for stage-one trend modeling but excluded from the second-stage regression after preliminary multicollinearity diagnostics. All variables were expressed as percentages, with all observed values verified to lie within the logical [0, 100] interval.

2.5. Modeling Strategy

A two-stage modeling approach was applied. In the first stage, for each WHO region and each predictor, a simple linear regression of the indicator on the year (Y ~ Year) was estimated to project the predictor values to 2020, 2021, and 2022. This yielded a total of 24 trend models (six regions × four predictors).

The first-stage equation is:

$$X_{b,t} = \alpha_b + \beta_b \cdot t + \varepsilon_{b,t} \qquad (1)$$

In the second stage, adult obesity prevalence was modeled using region fixed effects multiple linear regression. The fixed-effects specification controls for time-invariant regional heterogeneity, an approach commonly recommended in panel data analysis to address structural differences across units (Wooldridge, 2010). The second-stage equation is:

$$Y_{b,t} = \beta_0 + \beta_1 X_{1,b,t} + \beta_2 X_{2,b,t} + \beta_3 X_{3,b,t} + \Sigma\, \gamma_k D_{k,b} + \varepsilon_{b,t} \qquad (2)$$

In Equation (2), $Y_{b,t}$ is the obesity prevalence in region b in year t; X_1, X_2, and X_3 denote overweight, underweight, and physical inactivity prevalence, respectively; $D_{k,b}$ are dummy variables for the five non-reference regions, with Africa as the reference category; γ_k are the regional fixed effects coefficients; and $\varepsilon_{b,t}$ is the error term. All parameters were estimated by ordinary least squares using the fitlm function in MATLAB R2023a (Statistics and Machine Learning Toolbox).

Multicollinearity was assessed using variance inflation factors (VIF), with VIF < 10 set as the acceptability threshold. As a sensitivity analysis, a pooled multiple linear regression model omitting the regional fixed-effects was also estimated and compared with the main specification to evaluate the impact of unobserved regional heterogeneity.

2.6. Validation Through Time-Ordered Backtesting

To assess out of sample predictive accuracy, a time ordered hold out back test was performed (Bergmeir, Hyndman, & Koo, 2018). The model was retrained using only 2010–2017 observations (n = 48), and predictions were generated for the held-out 2018–2019 test set (n = 12). Performance was evaluated through root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). This protocol simulates a realistic forecasting scenario in which the model must extrapolate to a previously unseen time period (Starcke et al., 2025).

3. Results

3.1.1. Descriptive Statistics

During 2010–2019, mean adult obesity prevalence across the six WHO regions was 15.83% (range: 4.08%–31.80%), while mean overweight prevalence was 40.50%. The Region of the Americas exhibited the highest obesity prevalence throughout the period, followed by the Eastern Mediterranean and European regions. South-East Asia and Africa consistently showed the lowest prevalence levels but exhibited some of the steepest upward trends.

Underweight prevalence remained substantially higher in South-East Asia and Africa than in other regions, consistent with the persistence of the double burden of malnutrition (NCD-RisC, 2024).

3.2. Stage 1: Trend Estimates for Predictors

Across the 24 region-based predictor trend models, R^2 values exceeded 0.82, and all slopes were statistically significant at $p < 0.001$, indicating that regional time trends in the four predictors are well approximated by linear functions over the 2010–2019 period. The fastest increases in overweight prevalence were observed in South-East Asia (+0.95 percentage points/year) and the Eastern Mediterranean (+0.86 percentage points/year), while Europe exhibited the slowest increase (+0.22 percentage points/year). Insufficient physical activity rose most sharply in South-East Asia (+1.07 percentage points/year) and the Eastern Mediterranean (+0.52 percentage points/year), declined in Africa and Europe, and increased modestly in the Americas and Western Pacific. Underweight prevalence declined in all regions except Europe, where a small but statistically significant increase was detected. Raised blood pressure prevalence declined across all regions, most rapidly in Europe (–0.50 percentage points/year).

3.3. Stage 2: Region Fixed-Effects Regression Model

Table 1 presents the parameter estimates for the region fixed effects, multiple linear regression model. The model achieved $R^2 = 0.9989$ (adjusted $R^2 = 0.9987$), with the overall regression highly significant ($F(8, 51) = 5{,}554.35$; $p < 0.0001$) and an RMSE of 0.341 percentage points.

Table 1: Parameter Estimates of the Region Fixed Effects Multiple Linear Regression Model for Adult Obesity Prevalence (n = 60)

Variable	Coefficient (β)	Std. Error	t-statistic	p-value
Intercept	−24.681	2.523	−9.781	< 0.0001
Overweight	+0.7385	0.0386	19.141	< 0.0001
Underweight	+0.8094	0.1029	7.865	< 0.0001
Inactivity	+0.2232	0.0760	2.937	0.0050
Region [Americas]	−2.064	0.867	−2.380	0.0211
Region [E. Mediterranean]	−5.302	0.855	−6.201	< 0.0001
Region [Europe]	−2.196	0.935	−2.347	0.0228
Region [S.-E. Asia]	−10.278	1.861	−5.522	< 0.0001
Region [W. Pacific]	−2.712	0.344	−7.881	< 0.0001

Africa was used as the reference region. The dummy coefficients indicate the average difference (in percentage points) of obesity prevalence in each region relative to Africa, after controlling for the three continuous predictors.

All three continuous predictors were positively and significantly associated with obesity prevalence. A one percentage point increase in overweight prevalence was associated with a 0.74-percentage-point increase in obesity within the same region; the corresponding effects for underweight and inactivity were 0.81 and 0.22 percentage points, respectively. The positive coefficient for underweight is consistent with the double burden of malnutrition documented in low- and middle-income contexts (NCD-RisC, 2024; Dinsa et al., 2012). VIF values were all below the conventional threshold of 10 (overweight: 7.84; underweight: 6.80; inactivity: 2.29), supporting the absence of severe multicollinearity.

3.4. Backtest Validation

The time-ordered backtest produced an overall MAPE of 4.02% on the held-out 2018–2019 observations (Table 2). Region-level errors were below 3% in five of the six WHO regions, with the Western Pacific Region exhibiting systematic overprediction (+14.75% in 2018 and +16.79% in 2019).

Table 2: Backtest Performance Metrics (Train: 2010–2017; Test: 2018–2019)

Metric	Value
Test set observations (n)	12
Root Mean Squared Error (RMSE)	0.5534
Mean Absolute Error (MAE)	0.4465
Mean Absolute Percentage Error (MAPE)	4.02%

The systematic deviation in the Western Pacific Region likely reflects (i) regional heterogeneity, since the WPR includes both low prevalence high income East Asian countries (e.g., Japan, the Republic of Korea) and Pacific island states with the highest obesity prevalence worldwide (NCD-RisC, 2024); (ii) BMI–obesity relationships that differ in East Asian populations, prompting WHO to recommend lower BMI thresholds for the region; and (iii) the assumption of common slopes across regions in the fixed effects specification, which may not fully capture region specific dynamics.

3.5. Near-Term Forecasts: 2020–2022

Combining the projected predictor values from Stage 1 with the second-stage fixed-effects coefficients yielded 18 region-year forecasts for 2020–2022 (Table 3). All six regions exhibited positive expected trajectories. The Americas remained the highest-prevalence region (32.96% in 2022), followed by the Eastern Mediterranean (27.94%), Europe (22.13%), Africa (11.23%), the Western Pacific (9.66%), and South-East Asia (7.53%).

Table 3: Forecast Adult Obesity Prevalence by WHO Region, 2020–2022 (%)

Region	2020	95% CI	2021	95% CI	2022	
Africa	10.84	[10.40–11.28]	11.04	[10.53–11.54]	11.23	
Americas	31.92	[31.58–32.27]	32.44	[32.06–32.82]	32.96	
E. Mediterranean	26.76	[26.43–27.10]	27.35	[26.98–27.72]	27.94	
Europe	21.81	[21.56–22.05]	21.97	[21.72–22.22]	22.13	
S.-E. Asia	6.95	[6.49–7.41]	7.24	[6.71–7.77]	7.53	
W. Pacific	8.60	[8.29–8.91]	9.13	[8.79–9.48]	9.66	

Note: 95% CI = 95% confidence interval. CIs for 2022 follow the same pattern and are reported in the corresponding doctoral thesis.

Forecast increments between 2020 and 2022 differed substantially across regions: the Eastern Mediterranean exhibited the largest absolute increase (+1.18 percentage points), followed by the Western Pacific (+1.06), the Americas (+1.04), South-East Asia (+0.58), Africa (+0.39), and Europe (+0.33). The relatively flat trajectory in Europe is consistent with the plateauing trend documented in high income regions (Mihalache et al., reported by NCD-RisC, 2024).

3.6. Sensitivity Analysis: Pooled Versus Fixed-Effects Specification

Compared with the region fixed-effects model, the pooled multiple linear regression specification yielded a substantially lower R^2 and produced a counterintuitive negative coefficient for physical inactivity ($\beta = -0.1963$), reversing the sign obtained in the fixed effects model. This sign flip is consistent with Simpson's paradox: regional heterogeneity in baseline obesity levels confounds the within-region relationship between inactivity and obesity unless properly accounted for. The comparison thus supports the use of a fixed effects specification for this panel.

4. Discussion

4.1. Comparison With Existing Literature

The regional rankings produced by the fixed effects model align closely with independent global estimates. In particular, the 2022 forecasts for the six WHO regions fall within the corresponding regional ranges reported by NCD-RisC (2024), with absolute differences typically below 1–2 percentage points in five of the six regions. The Americas (32.96%) sits within the 30–33% range reported by NCD-RisC; Eastern Mediterranean (27.94%) and Europe (22.13%) fall in their respective expected bands; and Africa (11.23%) lies near the center of the 10–13% range. South-East Asia (7.53%) is at the upper end of the NCD-RisC 5–8% range. These convergences across very different methodological approaches (a parsimonious panel regression here versus the Bayesian hierarchical model used by NCD-RisC) suggest that the structural patterns captured by the present model are robust.

Methodologically, the most comprehensive forecasting exercise covering 2020–2022 is the GBD 2021 Adult BMI Collaborators' analysis, which used a Generalized Ensemble Modeling (GenEM) approach combining 12 sub-models (GBD 2021 Risk Factors Collaborators, 2024). Although the present study uses a single, parsimonious specification, both approaches share the core epistemological principle of using past data to validate and forecast future trajectories. The regional rankings produced here match those reported by GBD 2021, including the relative ordering of high-prevalence (Americas, Eastern Mediterranean) and lower-prevalence (South-East Asia) regions. The convergence between a high-complexity ensemble model and a low-complexity panel regression demonstrates that interpretable models can yield policy-relevant estimates when applied to well-structured ecological data.

The positive and significant coefficient for underweight ($\beta = +0.8094$) initially appears counterintuitive but is consistent with the double burden of malnutrition. NCD-RisC (2024) shows that several South Asian and Sub Saharan African settings exhibit simultaneously elevated underweight and rising obesity, particularly stratified by socioeconomic position (Dinsa et al., 2012). The observed coefficient is therefore best interpreted as an ecological reflection of this coexistence rather than a direct causal pathway.

4.2. The Western Pacific Deviation

The systematic overprediction in the Western Pacific Region merits particular attention. Three explanations are mutually compatible. First, the WPR aggregates extremely heterogeneous populations: high-income East Asian countries with relatively low obesity prevalence coexist with Pacific island nations exhibiting the world's highest prevalence rates. Second, BMI cardiometabolic risk relationships differ in East Asian populations at the same BMI level, leading WHO to recommend lower BMI thresholds for the region (NCD-RisC, 2024). Third, the fixed-effects specification assumes common slopes across regions; only the intercepts vary. Mixed effects models with region-specific slopes are a natural extension and are recommended for future work in this area.

4.3. The 6P Framework: From Findings to Policy

The 6P framework (Predictive, Preventive, Personalized, Participatory, Precision-oriented, Public Health) provides a structured way to translate quantitative findings into actionable policy. The Predictive dimension is directly served by the 2020–2022 forecasts and their 95% confidence intervals, which can be operationalized into regional dashboards and early warning thresholds (Grøntved et al., 2024; Jakobsen et al., 2021). The Preventive dimension is informed by the magnitude of the model coefficients: a one percentage point reduction in overweight prevalence is associated with a 0.74-percentage-point reduction in obesity prevalence, while a corresponding reduction in inactivity yields a 0.22-percentage-point reduction. These elasticities support targeting overweight reduction as the highest-leverage near-term intervention point.

The Personalized and Precision-oriented dimensions emphasize that interventions should be calibrated to the regional context. The fastest-rising regions (Eastern Mediterranean, Western Pacific, Americas) require different policy mixes than the relatively plateaued European Region. The Participatory dimension calls for engagement of national ministries, civil society, and communities in priority-setting; while the Public Health dimension underscores the structural determinants—nutrition transition, urbanization, food environments—that no single intervention can address (Swinburn et al., reported in NCD-RisC, 2024).

4.4. Strengths and Limitations

The study has three main strengths. First, the data acquisition pipeline is fully automated and reproducible: the Python script provides a transparent, version-controllable record of how the analytical dataset was assembled. Second, the region fixed-effects specification controls for time-invariant regional heterogeneity, mitigating Simpson's paradox risks demonstrated in the sensitivity analysis. Third, the time-ordered backtest provides a realistic out-of-sample assessment, addressing a methodological gap noted by Grøntved et al. (2024) in the healthcare forecasting literature.

Several limitations should be acknowledged. The ecological design precludes individual-level causal inference; the coefficients describe regional trends, not within-person mechanisms (Cohen, 1990). The sample size (n = 60) is modest and constrains model complexity, although it satisfies the minimum requirements for the eight estimated parameters (Riley et al., 2019). Three indicators of theoretical interest (alcohol consumption, total cholesterol, cost of a healthy diet) had to be excluded due to data gaps in the WHO GHO API; future work could integrate alternative sources (OECD, World Bank, FAO). The Western Pacific deviation represents a region-specific limitation that mixed effects models could address. Finally, reliance on a single data source ensures internal consistency but limits the scope for cross-source validation.

5. Conclusion

Region fixed effects multiple linear regression, applied to an automatically curated WHO GHO panel of 60 region-year observations, yields highly accurate (R^2 = 0.9989) and interpretable near-term forecasts of adult obesity prevalence (backtest MAPE = 4.02%). All three core predictors (overweight, underweight, and physical inactivity) showed statistically significant positive associations with obesity prevalence, and forecast 2020–2022 trajectories were upward in every WHO region. The forecasts converge with independent estimates from NCD-RisC (2024) and

GBD 2021 across five of the six regions, supporting the validity of the parsimonious approach. The Western Pacific Region exhibited systematic over-prediction, indicating a need for region specific slope models in future work.

Translated into the 6P framework, these findings suggest that overweight reduction is the highest leverage near-term intervention target, that the Eastern Mediterranean and Western Pacific regions require accelerated action, and that a plateauing trajectory in Europe should not be mistaken for resolution. The combination of automated data acquisition, time ordered validation, and fixed effects specification provides a methodological template that can be replicated, extended, and integrated into routine regional monitoring and policy prioritization processes. By demonstrating that interpretable models can produce policy-relevant accuracy at the global scale, the study contributes to a broader research agenda on how artificial intelligence and machine learning can support equitable, evidence based public health planning.

ACKNOWLEDGEMENTS

The authors thank the World Health Organization for maintaining the Global Health Observatory and providing public access to its OData API. No external funding was received for this study.

CONFLICT OF INTEREST

The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

AUTHOR STATEMENT

This study used only publicly available, fully aggregated regional data from the WHO Global Health Observatory and did not involve human or animal subjects; therefore, no ethical approval or informed consent was required. The Python data acquisition script is available on request from the corresponding author. The study is based on the doctoral thesis of the first author.

References

An, R., Shen, J., & Xiao, Y. (2022). Applications of artificial intelligence to obesity research: Scoping review of methodologies. Journal of Medical Internet Research, 24(12), e40589. https://doi.org/10.2196/40589

Berberoğlu, Z., & Hocaoğlu, Ç. (2021). Küresel sağlık sorunu "obezite": Güncel bir gözden geçirme. Celal Bayar University Journal of Sciences and Health, 8(3), 543–552. https://doi.org/10.34087/cbusbed.886473

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Bray, G. A., Kim, K. K., & Wilding, J. P. H. (2017). Obesity: A chronic relapsing progressive disease process. A position statement of the World Obesity Federation. Obesity Reviews, 18(7), 715–723. https://doi.org/10.1111/obr.12551

Chen, Y., Ma, L., Han, Z., & Xiong, P. (2024). The global burden of disease attributable to high body mass index in 204 countries and territories: Findings from 1990 to 2019 and predictions to 2035. Diabetes, Obesity and Metabolism, 26(9), 3998–4010. https://doi.org/10.1111/dom.15748

Choong, C., Brnabic, A., Chinthammit, C., et al. (2024). Applying machine learning approaches for predicting obesity risk using US health administrative claims database. BMJ Open Diabetes Research & Care, 12(5), e004193. https://doi.org/10.1136/bmjdrc-2024-004193

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45(12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304

Dinsa, G. D., Goryakin, Y., Fumagalli, E., & Suhrcke, M. (2012). Obesity and socioeconomic status in developing countries: A systematic review. Obesity Reviews, 13(11), 1067–1079. https://doi.org/10.1111/j.1467-789X.2012.01017.x

GBD 2021 Risk Factors Collaborators. (2024). Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. The Lancet, 403(10440), 2162–2203. https://doi.org/10.1016/S0140-6736(24)00933-4

Grøntved, S., Kirkeby, M. J., Johnsen, S. P., et al. (2024). Towards reliable forecasting of healthcare capacity needs: A scoping review and evidence mapping. International Journal of Medical Informatics, 189, 105527. https://doi.org/10.1016/j.ijmedinf.2024.105527

Haghighathoseini, A., Lin, S.-Y., Song, G., An, R., & Xue, H. (2025). A systematic review on applications of artificial intelligence for obesity prevention. Obesity Reviews, e70062. https://doi.org/10.1111/obr.70062

Jakobsen, E., Olsen, K. E., Bliddal, M., et al. (2021). Forecasting lung cancer incidence, mortality, and prevalence to year 2030. BMC Cancer, 21, 985. https://doi.org/10.1186/s12885-021-08696-6

NCD Risk Factor Collaboration (NCD-RisC). (2024). Worldwide trends in underweight and obesity from 1990 to 2022: A pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. The Lancet, 403(10431), 1027–1050. https://doi.org/10.1016/S0140-6736(23)02750-2

Ölçer, E. (2024). Makine öğrenmesi temelli obezite durum tahminlemesi. Bilgisayar Bilimleri ve Mühendisliği Dergisi, 17(2), 156–164. https://doi.org/10.54525/bbmd.1469701

Riley, R. D., Ensor, J., Snell, K. I. E., et al. (2019). Calculating the sample size required for developing a clinical prediction model. BMJ, 368, m441. https://doi.org/10.1136/bmj.m441

Starcke, J., Spadafora, J., Spadafora, P., & Toma, M. (2025). The effect of data leakage and feature selection on machine learning performance for early Parkinson's disease detection. Bioengineering, 12(8), 845. https://doi.org/10.3390/bioengineering12080845

Sterne, J. A. C., White, I. R., Carlin, J. B., et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. BMJ, 338, b2393. https://doi.org/10.1136/bmj.b2393

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). Cambridge, MA: MIT Press.

World Health Organization. (n.d.). Global Health Observatory (GHO) OData API. Retrieved from https://ghoapi.azureedge.net/api/ [REPORT]