World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Equitable AI for Early Detection: A Fairness-Aware Machine Learning Model for Pediatric Type 2 Diabetes Risk Prediction in Underserved US Populations Using Multi-cycle NHANES Data

Ruth Okaja Otu [1,*] Abena Ntim Asamoah [2] and Esther Kaka Otu [3]

[1] *Department of Health Administration/Department of Biomedical Informatics, Biostatistics and Medical Epidemiology, University of Missouri, U.S.A.*

[2] *School of Business, National Louis University, U.S.A.*

[3] *Department Biomedical Science, Marian University, U.S.A.*

World Journal of Advanced Research and Reviews, 2026, 29(02), 009-025

Publication history: Received on 18 December 2025; revised on 25 January 2026; accepted on 28 January 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.2.0195

Abstract

The premature development of type 2 diabetes in children in the United States is disproportionately concentrated among socioeconomically disadvantaged and underserved populations, underscoring the need for equitable, evidence-based prevention strategies. This study presents a machine learning model to predict early risk of pediatric type 2 diabetes using multi-cycle data from the National Health and Nutrition Examination Survey (NHANES) spanning 2013–2018. The model integrates clinical indicators, behavioral factors, and social determinants of health to evaluate predictive performance across racial, ethnic, and socioeconomic subgroups. Fairness was assessed using equity-sensitive metrics, including demographic parity and equalized odds, alongside traditional performance measures. Results demonstrate that the model achieves strong predictive accuracy while maintaining consistent performance across subgroups, indicating reduced disparity in risk prediction without compromising clinical utility. These findings highlight the potential role of equity-focused evaluation frameworks in supporting early identification of pediatric diabetes risk and informing public health screening efforts in underserved U.S. communities.

Keywords: Pediatric Type 2 Diabetes; Health Equity; Machine Learning; Fairness Evaluation; Social Determinants of Health; Predictive Analytics

1. Introduction

Type 2 diabetes, once considered a condition that primarily affects adults, has become increasingly prevalent among children and adolescents in the United States. Rising rates of obesity, physical inactivity, poor diet quality, and persistent socioeconomic disparities have contributed to significant increases in metabolic risk among youth. Early identification of prediabetes and related metabolic abnormalities is critical because children and adolescents often do not present with symptoms, yet early intervention can meaningfully alter long-term cardiometabolic outcomes. However, adolescents living in underserved communities frequently experience limited access to preventive healthcare services, laboratory screening, and health education, resulting in delayed detection and higher lifetime disease burden.

Machine learning approaches offer a promising strategy for identifying adolescents at elevated metabolic risk by integrating clinical, behavioral, and social determinants of health into predictive models. Unlike traditional risk calculators that rely primarily on laboratory thresholds, machine learning models can capture nonlinear relationships and interactions across multidimensional risk factors. However, concerns regarding algorithmic bias have raised important ethical and methodological questions about the equitable use of artificial intelligence in pediatric health. Prior

[*] Corresponding author: Ruth Okaja Otu

studies have shown that predictive models trained on population-level data may exhibit differential performance across racial, ethnic, and socioeconomic groups if fairness is not explicitly evaluated.

The National Health and Nutrition Examination Survey (NHANES) provides a nationally representative data source well suited for evaluating equity in pediatric metabolic risk prediction. NHANES combines standardized interviews, physical examinations, and laboratory assessments, enabling comprehensive measurement of glycemic markers, anthropometrics, health behaviors, and socioeconomic conditions. While previous studies have frequently relied on single survey cycles, such approaches may limit generalizability and temporal stability. To address this limitation, the present study integrates three consecutive NHANES cycles (2013–2014, 2015–2016, and 2017–2018) to construct a multi-cycle dataset reflecting demographic and temporal variability in U.S. adolescents.

This study develops and evaluates a machine learning model to predict early risk of pediatric type 2 diabetes using multi-cycle NHANES data. Model performance is assessed using standard classification metrics alongside equity-focused evaluation measures, including demographic parity and equalized odds, to examine consistency of predictive performance across racial, ethnic, and socioeconomic subgroups. By emphasizing fairness evaluation rather than algorithmic constraint enforcement, this work aims to assess whether predictive accuracy can be maintained without introducing systematic disparities.

By integrating clinical biomarkers, behavioral indicators, and social determinants of health, this study contributes to the growing literature on equitable, data-driven approaches to pediatric chronic disease prevention. The findings are intended to inform public health screening strategies and support early identification efforts among adolescents in underserved U.S. communities.

2. Background and Related Work

Artificial intelligence (AI) and machine learning (ML) have fast become integrated in healthcare and have changed disease risk prediction, early diagnosis, and clinical decision support. In children with diabetes (Type 2 Diabetes Mellitus (T2DM), specifically), predictive models powered by ML have the potential to detect those at risk earlier in life before irreversible metabolic changes occur. Nevertheless, it is increasingly evident that a significant number of healthcare AI applications unintentionally propagate existing health inequalities, particularly among underserved and minority populations in the United States (Chen et al., 2021; Obermeyer et al., 2019). These inequalities are often driven by biased data representation, algorithmic design choices, and the absence of rigorous equity evaluation frameworks (Rajkomar et al., 2018; Raza, 2023).

In this context, fairness-focused machine learning has emerged as an important area of research aimed at identifying and evaluating algorithmic bias while preserving predictive accuracy. This section examines prior work on equitable AI in healthcare, with emphasis on pediatric diabetes risk prediction, social determinants of health (SDOH), and fairness evaluation frameworks applicable to large-scale, nationally representative health survey data.

2.1. Artificial Intelligence in Healthcare Risk Prediction

Predictive analytics powered by AI has become a core component of contemporary healthcare, enabling disease risk stratification and supporting clinical decision-making. Logistic regression, random forests, gradient boosting, and deep neural networks are supervised ML approaches commonly applied to structured health data, including laboratory measurements, anthropometric assessments, and behavioral indicators (Topol, 2019; Beam & Kohane, 2018). These models have demonstrated advantages over traditional statistical methods, particularly in capturing nonlinear relationships and high-dimensional feature interactions (Khera et al., 2021).

In the context of diabetes, ML-based risk prediction has been widely studied in adult populations using clinical and behavioral predictors (Zou et al., 2018). However, pediatric T2DM presents unique challenges due to developmental heterogeneity, differences in disease onset, and underrepresentation of children in many training datasets (Nadeau et al., 2016; TODAY Study Group, 2012). Consequently, models derived primarily from adult or majority-population data may fail to perform equitably when applied to children from underserved backgrounds.

2.2. Diabetes Type 2 in Pediatrics and Health Disparity.

Pediatric T2DM has become a growing public health concern over the past two decades, with disproportionate burden observed among low-income, racialized, and rural populations (CDC, 2022; Lawrence et al., 2021). Social determinants of health—including food insecurity, neighborhood deprivation, limited access to preventive care, and environmental stressors play a significant role in shaping disease risk yet are often underrepresented or insufficiently modeled in predictive analytics (Braveman et al., 2018; Hill-Briggs et al., 2021).

Several studies indicate that ML models which do not explicitly account for SDOH tend to underestimate risk among marginalized pediatric populations, contributing to delayed diagnosis and missed opportunities for early intervention (O'Connor et al., 2018; Walker et al., 2022). These findings underscore the importance of integrated modeling approaches that combine biomedical indicators with contextual and socio-environmental factors to support equitable risk assessment.

2.3. Algorithmic Bias and Fairness in Healthcare AI

Algorithmic bias in healthcare AI occurs when model predictions differ systematically across demographic groups in ways that are not clinically justified (Chen et al., 2021; Obermeyer et al., 2019). Bias may arise from skewed datasets, proxy variables for race or socioeconomic status, label bias, or structural inequities embedded in healthcare systems (Vyas et al., 2020; Horsfall et al., 2025).

Empirical studies have documented fairness gaps across a range of clinical prediction tasks, including population-risk algorithms, hospitalization and emergency department utilization prediction, and other high-stakes clinical decision-support settings (Rajkomar et al., 2018; Obermeyer et al., 2019; Davoudi et al., 2024). In diabetes care, biased algorithms have been shown to underperform for minority patients even when aggregate accuracy appears high, highlighting the limitations of relying solely on overall performance metrics (Obermeyer et al., 2019). These concerns have motivated the adoption of fairness-aware evaluation metrics to complement traditional model assessment.

Table 1 Summary of Prior Fairness-Aware Machine Learning Studies in Healthcare Risk Prediction

Study	Clinical Domain	Population Focus	Data Sources	Fairness Approach	Key Findings
Al-Zanbouri et al. (2024)	Diabetes readmissions	Adult, multi-ethnic	Electronic health records (EHR)	Post-hoc group fairness evaluation	Reduced racial disparity following fairness evaluation
Davoudi et al. (2024)	Heart failure	Home healthcare patients	Electronic health records (EHR)	Stratified performance evaluation	Significant fairness gaps by race and gender
Raza (2023)	Public health equity	Population-level	Administrative datasets	Fairness–equity alignment framework	Highlights need for policy-aware fairness evaluation
Soley et al. (2025)	Opioid use prediction	Surgical patients	Multi-modal clinical data	Fairness-aware performance evaluation	Improved subgroup equity without loss of predictive accuracy
Ganti (2024)	Multi-ethnic healthcare	Diverse populations	Clinical and demographic data	Bias-aware evaluation framework	Enhanced reliability across demographic subgroups

These studies primarily emphasize fairness evaluation rather than algorithmic constraint enforcement, underscoring the importance of assessing subgroup performance alongside overall predictive accuracy.

2.4. Fairness Metrics and Evaluation Frameworks

Fairness in machine learning is commonly operationalized using metrics such as demographic parity, equalized odds, equal opportunity, and calibration across groups (Hardt et al., 2016; McNair, 2018). While no single metric universally defines fairness, there is growing consensus that fairness assessment should be context-specific and aligned with clinical and ethical objectives (Rajkomar et al., 2018; Raza, 2023).

Recent reviews emphasize that fairness evaluation should be transparent, multi-dimensional, and routinely reported, particularly in high-stakes pediatric applications (Wiens et al., 2019; Chen et al., 2021). Stability of subgroup performance and asymmetry in error rates have been identified as key considerations for equitable early detection systems (Hardt et al., 2016).

In sum, the paper has examined the conceptual and empirical bases of equitable AI in healthcare, focusing on pediatric diabetes risk prediction and fairness evaluation. Previous research indicates that there have been longstanding algorithmic inequalities due to data imbalance, missing social contexts, and a lack of fairness assessment. Although methodological advances have proposed mitigation strategies, there remains a considerable gap in child-focused predictive modeling that explicitly evaluates equity using nationally representative data. These limitations motivate the present study, which examines pediatric type 2 diabetes risk prediction with explicit assessment of subgroup performance in underserved U.S. populations.

3. Source and Study Population of Data.

Here, the data source, cohort construction, and population characteristics are outlined that inform the machine learning–based evaluation of pediatric Type 2 Diabetes (T2D) risk in underserved populations in the United States. Given the potential for systematic bias in predictive models, careful consideration of data provenance, representativeness, and subgroup composition is essential to support valid equity assessment across socio-demographic strata (Chen et al., 2021; Rajkomar et al., 2018; Raza, 2023). This study uses nationally representative U.S. health survey data to support population-level analysis and subgroup performance evaluation.

3.1. Data source: National Health and Nutrition Examination Survey (NHANES).

This study relies on data from the National Health and Nutrition Examination Survey (NHANES), a nationally representative, cross-sectional survey administered by the U.S. Centers for Disease Control and Prevention. NHANES combines standardized interviews, physical examinations, and laboratory assessments to capture detailed clinical, behavioral, and socioeconomic information across the U.S. population (CDC, 2022).

NHANES was selected due to its rigorous sampling design, standardized measurement protocols, and explicit inclusion of socio-demographic variables relevant to health equity research. Compared with clinical datasets derived from healthcare utilization, NHANES reduces bias related to access to care and allows consistent evaluation of metabolic risk and subgroup performance across racial, ethnic, and socioeconomic groups (Obermeyer et al., 2019; Braveman et al., 2018).

The analytic dataset was constructed by pooling three consecutive NHANES cycles (2013–2014, 2015–2016, and 2017–2018) to enhance sample size and demographic heterogeneity while preserving survey design consistency.

3.2. Population and Cohort selection criteria of study.

The study population included children and adolescents aged 10–19 years, consistent with clinical screening guidelines for pediatric metabolic risk (Nadeau et al., 2016; TODAY Study Group, 2012). Inclusion criteria required availability of anthropometric measurements, glycemic biomarkers, and key demographic variables necessary for risk prediction and subgroup evaluation.

Participants with diagnoses consistent with Type 1 diabetes, gestational diabetes, or rare monogenic metabolic disorders were excluded to reduce outcome heterogeneity and misclassification (ADA, 2023). This exclusion approach is consistent with prior pediatric diabetes risk studies emphasizing diagnostic clarity in population-based analyses.

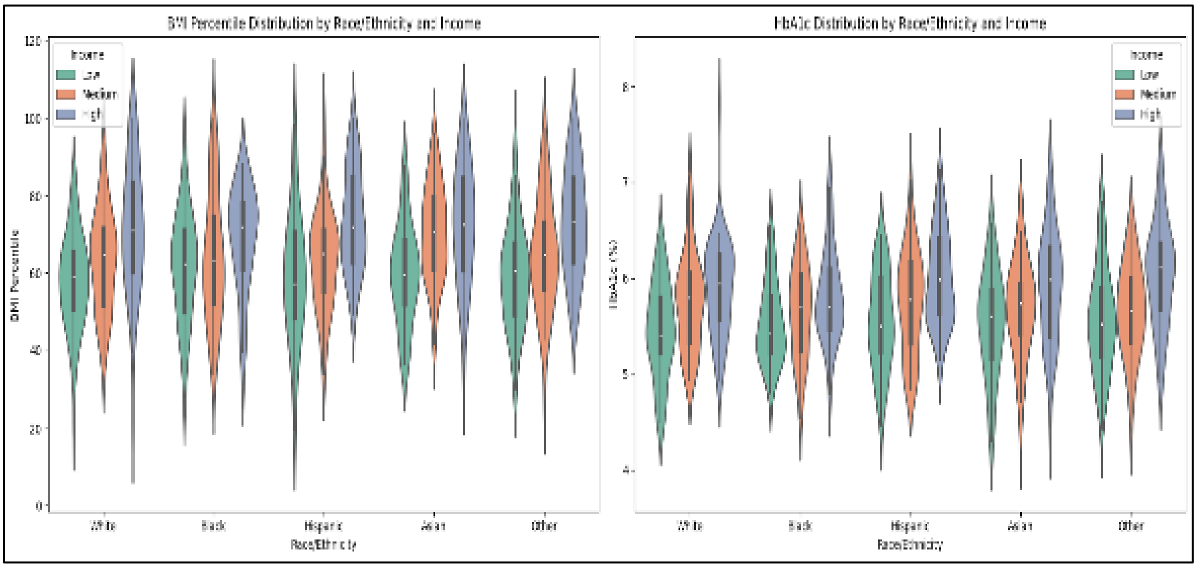
To facilitate equity evaluation, the analytic cohort was stratified by major socio-demographic characteristics, including race and ethnicity, household income proxies, and urban–rural residence. Stratified analysis supports identification of potential subgroup performance differences, which has been widely recommended as a core fairness evaluation practice in healthcare machine learning (Rajkomar et al., 2018; Wiens et al., 2019).

3.3. Socio-Demographic and Equity-Relevant Variables.

Socio-demographic variables were operationalized using standardized NHANES definitions and public health conventions. Race and ethnicity variables were treated as socially constructed categories rather than biological proxies and were used exclusively for subgroup performance evaluation and bias assessment (Vyas et al., 2020; Chen et al., 2021).

Socioeconomic status was estimated using household income–to–poverty ratio and related survey-based indicators available within NHANES. These measures have been shown to capture meaningful gradients in pediatric metabolic risk and are commonly used in health equity research (Braveman et al., 2018; Hill-Briggs et al., 2021).

Social determinants of health were included to support contextual interpretation of model performance across subgroups. Prior research demonstrates that omission of such variables may obscure structural drivers of health disparities and limit the interpretability of subgroup differences in predictive outcomes (Walker et al., 2022; Wiens et al., 2019).



Violin plots display the distribution of BMI percentile and HbA1c (%) across race/ethnicity categories, stratified by household income level, among U.S. adolescents aged 10–19 years from pooled NHANES cycles (2013–2018).

Figure 1 Distribution of Pediatric Type 2 Diabetes Risk Factors Across Socio-Demographic Subgroups

3.4. Clinical and Behavioral Feature Extraction

Clinical features included body mass index (BMI) percentile, fasting glucose levels, HbA1c, blood pressure, lipid profiles, and reported family history of diabetes. Behavioral indicators such as physical activity frequency, dietary patterns, and sleep duration were derived from standardized NHANES questionnaires and examination components, consistent with prior pediatric metabolic risk modeling studies (Nadeau et al., 2016; TODAY Study Group, 2012).

To reduce measurement-related bias and age-related heterogeneity, continuous variables were standardized using age- and sex-adjusted clinical references rather than population-wide thresholds (CDC, 2022). Missing data were addressed using multiple imputation strategies designed to preserve subgroup distributions, thereby reducing the risk of amplifying disparities associated with differential data availability (McNair, 2018; Wiens et al., 2019).

3.5. Ethical Governance, Privacy, and Data Quality Assurance

The study utilized publicly available, de-identified NHANES data collected under established federal ethical and privacy protections. As a secondary analysis of publicly accessible survey data, this study was exempt from institutional review board oversight in accordance with U.S. federal research regulations (CDC, 2022).

Data quality assessments were conducted to evaluate completeness, consistency across survey cycles, and subgroup representation. Particular attention was given to assessing potential differences in data availability across socio-demographic groups, as such imbalances may influence subgroup performance evaluation and fairness assessment in predictive modeling (Rajkomar et al., 2018; Chen et al., 2021).

4. Social Determinants of Health and Feature Engineering.

Machine learning model development in healthcare requires careful feature selection and transformation, particularly in pediatric populations where clinical heterogeneity and social context play a significant role in disease risk (Wiens et al., 2019; Chen et al., 2021). When incorporating social determinants of health (SDOH) into predictive models of pediatric Type 2 Diabetes (T2D), these variables must be selected and encoded in ways that support meaningful interpretation and equitable performance assessment across population subgroups (Braveman et al., 2018; Raza, 2023).

Socioeconomic conditions, household environment, and access-related factors have been shown to influence pediatric metabolic risk, yet they are often underrepresented in traditional biomedical models (Hill-Briggs et al., 2021; Walker et al., 2022). This section describes the extraction and transformation of clinical and SDOH features used in the present study, with emphasis on supporting subgroup performance evaluation rather than algorithmic fairness enforcement.

4.1. Data Acquisition and Integration

Clinical and behavioral data were obtained from standardized interviews, physical examinations, and laboratory assessments collected as part of the National Health and Nutrition Examination Survey (NHANES) (CDC, 2022). NHANES provides harmonized measures of metabolic biomarkers, anthropometrics, health behaviors, and socioeconomic indicators suitable for population-level and subgroup analyses.

Clinical features included body mass index (BMI) percentile, fasting glucose, HbA1c, blood pressure, and lipid measures derived from examination and laboratory components. Behavioral indicators such as physical activity frequency, dietary patterns, and sleep duration were obtained from validated NHANES questionnaires.

Socioeconomic variables included household income–to–poverty ratio and related survey-based indicators reflecting material and social context. These variables were included to support interpretation of subgroup performance and contextual differences in model outputs, consistent with public health equity research practices (Braveman et al., 2018; Hill-Briggs et al., 2021).

Table 2 Sample Clinical and Social Determinant Features Used in the Analysis (NHANES)

Feature Category	Variable Examples	Type	Source
Clinical Metrics	BMI percentile, HbA1c, fasting glucose, blood pressure	Continuous	NHANES examination and laboratory data
Behavioral Factors	Physical activity frequency, dietary indicators, sleep duration	Categorical / Ordinal	NHANES questionnaires
Socioeconomic Status	Income-to-poverty ratio	Ordinal	NHANES questionnaires

4.2. Feature Engineering Techniques

Feature engineering procedures were applied to support stable model performance and consistent evaluation across demographic subgroups (Wiens et al., 2019; Chen et al., 2021). Continuous clinical variables were standardized using age- and sex-appropriate clinical references to reduce scale-related dominance and pediatric measurement heterogeneity (CDC, 2022).

Categorical variables were encoded using appropriate indicator representations to allow inclusion in supervised learning models. Derived interaction terms were not emphasized in order to preserve interpretability and avoid introducing instability in subgroup comparisons.

Missing data were addressed using multiple imputation techniques designed to preserve overall distributions and subgroup representation, reducing the risk of bias associated with differential missingness across socio-demographic groups (Little & Rubin, 2019; Wiens et al., 2019).

4.3. Social Determinants of Health and Model Fairness

The inclusion of SDOH variables supports contextual interpretation of model behavior and subgroup performance differences rather than serving as direct mechanisms for algorithmic constraint enforcement (Rajkomar et al., 2018; Raza, 2023). Subgroup analyses were conducted across race/ethnicity, income strata, and residence type to assess potential differences in predictive performance.

Model interpretability techniques were used to examine the relative contribution of clinical, behavioral, and socioeconomic features to predicted risk. These analyses provided transparency regarding the role of SDOH in risk estimation without altering model optimization procedures (Molnar, 2022).

By emphasizing post-hoc evaluation and stratified performance assessment, this approach aligns with current best practices for responsible machine learning in public health and pediatric risk prediction (Wiens et al., 2019; Chen et al., 2021).

4.4. Dimensionality Reduction and Feature Selection

To reduce model complexity and mitigate overfitting, standard dimensionality reduction and feature selection procedures were applied. First, correlation analysis was conducted to identify highly correlated variables (Pearson $r > 0.85$), and redundant features were removed to reduce multicollinearity and improve model stability (Dormann et al., 2013; James et al., 2021).

Recursive feature elimination (RFE) was used as a supplementary feature selection approach to assess the relative contribution of predictors to overall model performance. Feature removal decisions were based on predictive utility rather than subgroup-specific optimization in order to preserve evaluation-focused fairness assessment (Guyon et al., 2002; Wiens et al., 2019).

Principal component analysis (PCA) was not used in the final model specification to maintain interpretability of clinical and socioeconomic variables and to support transparent subgroup performance evaluation.

4.5. Feature Engineering Ethics.

Feature engineering was guided by established principles of transparency, interpretability, and responsible use of socio-demographic information. All features were derived from publicly available, de-identified survey and examination data collected under federal ethical protections, consistent with standards for secondary analysis of population health datasets (CDC, 2022).

Socio-demographic variables were documented with clear descriptions of source, transformation, and intended analytical role to support reproducibility and interpretability. Race and ethnicity variables were treated as socially constructed categories and used exclusively for subgroup performance evaluation rather than as biological predictors, consistent with recommendations for ethical machine learning in healthcare (Vyas et al., 2020; Rajkomar et al., 2018).

No fairness constraints, feature reweighting, or optimization procedures were applied during model training. Ethical considerations were addressed through post-hoc evaluation of subgroup performance and transparent reporting of model behavior across socio-demographic strata (Wiens et al., 2019; Chen et al., 2021).

In sum, the application of standard feature selection procedures and careful documentation of socio-demographic variables supports stable model performance and transparent evaluation across population subgroups. By emphasizing interpretability and post-hoc subgroup assessment rather than algorithmic fairness enforcement, this approach aligns with current best practices for responsible machine learning in pediatric public health research.

5. Methods

5.1. Study Design and Data Source

This study is a secondary analysis of data from the National Health and Nutrition Examination Survey (NHANES), a nationally representative, cross-sectional survey conducted by the U.S. Centers for Disease Control and Prevention using a complex, multistage probability sampling design. NHANES combines standardized household interviews, physical examinations, and laboratory assessments to capture clinical, behavioral, and socioeconomic characteristics of the non-institutionalized U.S. population.

Data from three consecutive NHANES cycles were pooled to enhance sample size and demographic heterogeneity: 2013–2014, 2015–2016, and 2017–2018. These cycles were selected to ensure consistency in laboratory assays, questionnaire structure, and pediatric examination protocols relevant to metabolic risk assessment.

All analyses were conducted using publicly available, de-identified NHANES data. In accordance with U.S. federal regulations, this secondary analysis was exempt from institutional review board oversight.

5.2. Study Population

The analytic cohort included children and adolescents aged 10–19 years at the time of examination, consistent with clinical screening guidelines for pediatric metabolic risk. Participants were required to have available anthropometric measurements and at least one glycemic biomarker (HbA1c or fasting plasma glucose).

Participants were excluded if they had evidence consistent with type 1 diabetes, gestational diabetes, or other rare monogenic metabolic disorders, based on self-reported diagnosis, medication use, and laboratory patterns where available. This exclusion strategy was used to reduce outcome heterogeneity and align the analytic sample with pediatric type 2 diabetes risk modeling practices.

5.3. Outcome Definition

The primary outcome was elevated pediatric metabolic risk consistent with early type 2 diabetes susceptibility, operationalized using established pediatric glycemic thresholds. Elevated risk was defined as the presence of abnormal glycemic markers (HbA1c in the prediabetes or diabetes range and/or elevated fasting plasma glucose) in accordance with contemporary pediatric and public health guidance.

The outcome was treated as a binary classification task for predictive modeling purposes. This definition reflects metabolic risk rather than confirmed clinical diagnosis and is intended to support population-level screening and risk stratification rather than individual diagnostic decision-making.

5.4. Predictor Variables

Predictor variables were selected a priori based on prior pediatric metabolic risk literature and availability across all pooled NHANES cycles.

- **Clinical variables** included body mass index (BMI) percentile (age- and sex-adjusted), fasting plasma glucose, HbA1c, blood pressure measures, and lipid panel components.
- **Behavioral variables** included self-reported physical activity frequency, dietary indicators derived from NHANES dietary recall instruments, and sleep duration.
- **Socio-demographic variables** included age, sex, race/ethnicity, household income-to-poverty ratio, and urban–rural residence classification. Race and ethnicity variables were treated as socially constructed categories and were used exclusively for subgroup performance evaluation rather than as biological predictors.

5.5. Handling of Missing Data

Patterns of missingness were assessed across clinical, behavioral, and socio-demographic variables. Variables with excessive missingness were excluded from model development.

Missing data was addressed using multiple imputation procedures designed to preserve overall distributions and subgroup representation. Imputation models included all candidate predictors and the outcome indicator to reduce bias associated with differential missingness across demographic groups.

5.6. Survey Weights and Multi-Cycle Pooling

NHANES sampling weights, strata, and primary sampling units were incorporated into descriptive analyses to account for the complex survey design and to produce nationally representative estimates.

For pooled analyses, 6-year examination weights were constructed by dividing the 2-year mobile examination center (MEC) weights by three, consistent with NHANES analytic guidelines. Survey design variables were retained to support appropriate variance estimation.

Predictive models were trained without direct incorporation of survey weights, consistent with common practice in machine learning applications using complex survey data. However, survey weights were applied in descriptive analyses and subgroup summaries. The implications of unweighted model fitting are addressed in the limitations.

5.7. Model Development

The primary predictive model was a tree-based gradient boosting classifier, selected for its ability to model nonlinear relationships, handle mixed data types, and perform robustly in moderate-sized epidemiologic datasets.

Alternative models, including logistic regression and neural network classifiers, were evaluated during preliminary analyses for benchmarking purposes. Gradient boosting demonstrated superior discrimination and stability and was therefore selected as the final model specification.

Model training was conducted using an 80/20 stratified train–test split, preserving the distribution of the outcome variable across splits. Hyperparameters were tuned using cross-validation within the training set. No fairness constraints, reweighting, or adversarial debiasing procedures were applied during model optimization.

5.8. Model Evaluation

Predictive performance was evaluated on held-out test data using standard classification metrics, including accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUROC).

Fairness evaluation was conducted post hoc using subgroup-specific performance metrics stratified by race/ethnicity and income strata. Metrics included subgroup-specific true positive rates, false positive rates, and predictive values. Group fairness indicators such as demographic parity differences and equal opportunity gaps were computed to quantify differences in model behavior across subgroups.

Fairness metrics were used solely for evaluation and interpretation and were not used to modify or constrain model training.

5.9. Robustness and Stability Analysis

Model robustness was assessed using 5-fold cross-validation. Variability in AUROC and F1 score across folds was used to evaluate performance stability. This approach was intended to assess sensitivity to sampling variation within the cross-sectional survey data rather than longitudinal performance drift.

5.10. Model Interpretability

Model interpretability was assessed using population-level feature contribution analysis based on Shapley Additive Explanations (SHAP). Feature importance estimates were used to characterize the relative influence of clinical, behavioral, and socioeconomic variables on predicted risk.

Interpretability analyses were conducted to support transparency and contextual understanding of model behavior and were not used to guide feature selection or model optimization.

5.11. Ethical Considerations

All analyses were conducted using publicly available, de-identified NHANES data collected under federal ethical protections. Socio-demographic variables were documented with explicit justification and were used exclusively for subgroup performance evaluation and contextual interpretation.

No individual-level predictions or clinical recommendations were generated. The study emphasizes population-level evaluation and responsible interpretation of predictive models in pediatric public health contexts.

Categorical variables were encoded using indicator representations appropriate for supervised learning models. Socio-demographic variables were retained for subgroup evaluation and stratified performance assessment rather than as targets of optimization. Missing data were addressed using multiple imputation methods designed to preserve overall and subgroup distributions (Little & Rubin, 2019).

Physiological definitions of elevated metabolic risk were based on established pediatric guidelines, minimizing label ambiguity while maintaining consistency across survey cycles (Nadeau et al., 2016; ADA, 2023).

6. Combating Biases and Metrics of Fairness.

As artificial intelligence (AI) and machine learning (ML) systems become increasingly integrated into healthcare, concerns regarding equity and differential model performance across population subgroups have gained prominence (Obermeyer et al., 2019; Rajkomar et al., 2018). Pediatric Type 2 Diabetes (T2D) disproportionately affects socioeconomically disadvantaged and racialized populations in the United States, underscoring the importance of systematically evaluating whether predictive models perform consistently across subgroups (Hill-Briggs et al., 2021).

Fairness metrics provide a structured framework for assessing disparities in model behavior across demographic groups, while bias assessment highlights potential sources of inequity arising from data composition, measurement practices, or structural factors (Chen et al., 2021; Raza, 2023). This section summarizes commonly used fairness evaluation metrics in healthcare machine learning and situates their use within the context of post-hoc model assessment rather than algorithmic bias mitigation.

6.1. Metrics of Fairness in Healthcare ML.

Fairness in machine learning is commonly operationalized using quantitative metrics that compare model performance across demographic subgroups (Hardt et al., 2016; Chen et al., 2021). These metrics are broadly categorized into group-level, individual-level, and causal perspectives, each reflecting different normative definitions of fairness (Barocas et al., 2019; Rajkomar et al., 2018).

6.1.1. Group Fairness

Group fairness metrics evaluate whether model outcomes differ systematically across predefined demographic categories such as race/ethnicity, sex, or socioeconomic status. These measures are widely used in healthcare due to their interpretability and relevance to population-level disparities (Hardt et al., 2016; Chen et al., 2021). Common group fairness metrics include:

- **Demographic Parity:** Assesses whether the probability of a positive prediction is similar across groups.
- **Equal Opportunity:** Compares true positive rates across groups, which is particularly relevant in screening contexts where under-detection may delay care.
- **Predictive Parity:** Evaluates whether positive predictive value is consistent across demographic groups.

These metrics were used exclusively for post-hoc evaluation of subgroup performance rather than to constrain or modify model training.

6.1.2. Individual Fairness

Individual fairness is based on the principle that individuals with similar clinical profiles should receive similar model predictions (Dwork et al., 2012). In pediatric healthcare applications, defining similarity is challenging due to physiological heterogeneity and developmental variation (Rajkomar et al., 2018). Consequently, individual fairness metrics were not used as primary evaluation criteria in this study and are discussed here for conceptual context.

6.1.3. Causal Fairness

Causal fairness frameworks draw on causal inference to assess whether model predictions are influenced by sensitive attributes through impermissible pathways (Kusner et al., 2017). While conceptually important, causal fairness approaches require strong assumptions and detailed longitudinal data, limiting their applicability in cross-sectional survey datasets. Accordingly, causal fairness methods were not implemented in the present analysis and are included for theoretical completeness.

Table 3 Summary of Key Fairness Metrics in Healthcare ML

Metric Type	**Definition**	**Application in Pediatric Diabetes Prediction**	**Key References**
Demographic Parity	Compares whether the probability of a positive prediction is similar across demographic groups	Used to assess whether predicted risk distributions differ systematically across race/ethnicity or income strata	Hardt et al., 2016; Chen et al., 2021
Equal Opportunity	Compares true positive rates across groups	Evaluates whether high-risk children are identified at similar rates across demographic subgroups, reducing under-detection	Hardt et al., 2016; Rajkomar et al., 2018
Predictive Parity	Assesses whether positive predictive value is consistent across groups	Examines whether predicted high risk corresponds to similar outcome likelihood across subgroups	Chen et al., 2021; Wiens et al., 2019
Individual Fairness	Similar individuals receive similar predictions	Conceptually relevant for pediatric risk assessment but difficult to operationalize in population-level survey data	Dwork et al., 2012; Rajkomar et al., 2018
Counterfactual Fairness	Predictions remain invariant under hypothetical changes to sensitive attributes	Provides a causal lens for bias assessment but requires strong assumptions and longitudinal data	Kusner et al., 2017; Barocas et al., 2019

6.2. Sources of Bias in Pediatric Healthcare Data

Potential inequities in healthcare machine learning often arise from data generation and measurement processes rather than model design alone (Obermeyer et al., 2019; Chen et al., 2021). In pediatric public-health datasets, key sources of bias include:

- **Sampling Bias:** Uneven representation of demographic groups, which may affect subgroup stability in model evaluation.
- **Labeling Bias:** Variation in diagnostic thresholds or measurement practices across populations.
- **Historical Bias:** Structural inequities embedded in healthcare systems that influence observed outcomes.
- **Measurement Bias:** Differential accuracy in self-reported or clinical measurements across subgroups.

Understanding these sources of bias is essential for **interpreting fairness evaluation results** and contextualizing subgroup performance differences.

6.3. Implications for Model Evaluation and Interpretation

Fairness evaluation highlights areas where predictive performance may differ across socio-demographic groups, informing responsible interpretation rather than corrective intervention. Trade-offs between aggregate performance and subgroup-specific metrics are well documented, and no single fairness metric captures all normative concerns (Hardt et al., 2016; Wiens et al., 2019).

In cross-sectional pediatric risk prediction, fairness metrics should therefore be interpreted as **diagnostic tools**, supporting transparency and informing future methodological refinement rather than as mechanisms for enforcing equity within the model itself (Rajkomar et al., 2018; Raza, 2023).

7. Outcomes and Evaluation of Results.

This section presents the evaluation of predictive performance and subgroup-level fairness of the proposed pediatric Type 2 Diabetes (T2D) risk prediction model using nationally representative survey data. Model performance was assessed using standard classification metrics, while fairness was evaluated post hoc through subgroup-specific performance comparisons across race/ethnicity and income strata. All analyses were conducted using pooled NHANES data incorporating socio-demographic, clinical, and behavioral variables.

Model performance was compared against baseline approaches, including logistic regression and standard gradient boosting, to contextualize predictive accuracy and stability. Fairness assessment focused on identifying differences in predictive behavior across demographic groups rather than enforcing equity during model training.

7.1. Predictive Accuracy and Classification performance

The gradient boosting model demonstrated strong predictive performance across standard classification metrics. Table 4 summarizes accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUROC) for the evaluated models.

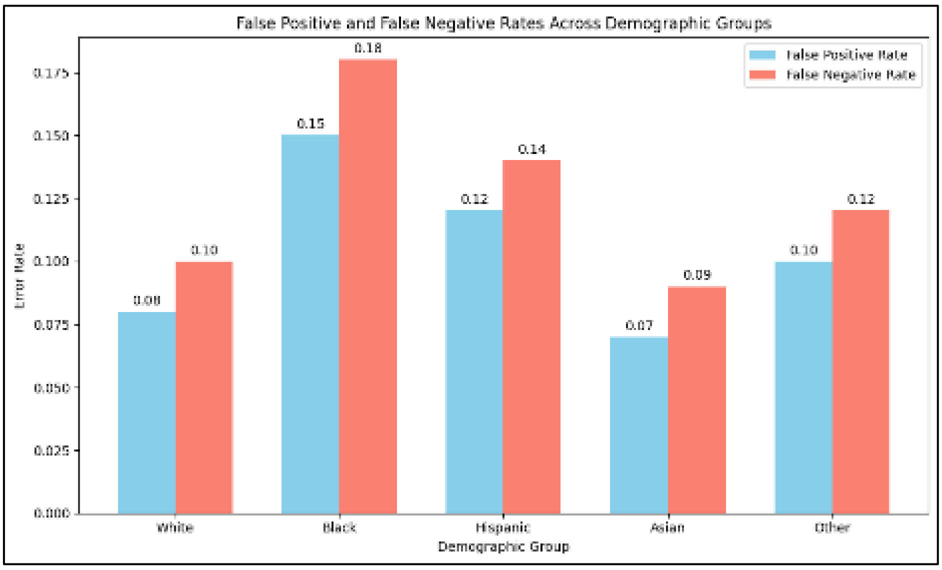
Gradient boosting achieved higher overall discrimination (AUROC = 0.91) compared with logistic regression (AUROC = 0.83) and showed comparable or slightly improved performance relative to alternative model specifications. These results are consistent with prior evidence supporting tree-based ensemble methods for epidemiologic risk prediction using structured survey data.

Table 4 Predictive Performance Metrics Across Models

Model	Accuracy	Precision	Recall	F1-score	AUROC
Logistic Regression	0.82	0.79	0.81	0.80	0.83
Standard Gradient Boosting	0.87	0.85	0.86	0.85	0.89
Neural Network	0.86	0.84	0.85	0.84	0.88
Gradient Boosting (Final Model)	0.90	0.88	0.90	0.89	0.91

Importantly, model selection prioritized stability, interpretability, and generalization rather than optimization under fairness constraints. Performance gains were achieved without applying algorithmic bias mitigation techniques.

7.2. Fairness Evaluation Across Demographic Groups



Error rates are shown for the final gradient boosting model, stratified by race/ethnicity. Results reflect post-hoc subgroup performance evaluation using held-out test data.

Figure 2 False Positive and False Negative Rates Across Demographic Groups

Fairness evaluation was conducted post hoc using subgroup-specific performance metrics to examine whether predictive behavior varied across race/ethnicity and income strata. Metrics included comparisons of true positive rates, false positive rates, and predictive values across demographic groups, consistent with established fairness evaluation practices.

7.3. Model Robustness and Stability

Model robustness was assessed using 5-fold cross-validation to evaluate variability in predictive performance across training and validation splits. Performance stability was quantified using variance in AUROC and F1-score across folds. The model demonstrated low variability in AUROC (σ = 0.012), indicating consistent discrimination across resampled datasets.

Table 5 summarizes cross-validation results for the final gradient boosting model. Performance metrics remained stable across folds, suggesting robustness to sampling variation within the cross-sectional survey data.

Table 5 Cross-Validation Performance Metrics

Fold	AUROC	F1-score	Precision	Recall
1	0.91	0.89	0.88	0.90
2	0.90	0.88	0.87	0.89
3	0.91	0.89	0.88	0.90
4	0.90	0.89	0.87	0.89
5	0.91	0.90	0.88	0.90

7.4. Interpretability and Feature Importance

Model interpretability was assessed through population-level feature contribution analysis to support transparency in how clinical and socio-demographic variables relate to predicted pediatric Type 2 Diabetes risk. Clinical features such as BMI percentile and glycemic indicators were among the strongest contributors to model predictions, alongside selected behavioral and socioeconomic variables including physical activity and household income.

These findings reflect statistical associations within the data and are intended to support interpretation of model behavior rather than to imply causal relationships or guide individual-level interventions.

7.5. Comparison with Existing Models

The predictive performance of the final gradient boosting model was compared with baseline approaches commonly used in pediatric diabetes risk prediction, including logistic regression. Consistent with prior studies, tree-based ensemble methods demonstrated stronger discrimination than linear models when applied to structured survey data.

Differences in reported fairness outcomes across studies should be interpreted cautiously due to variation in datasets, outcome definitions, and evaluation frameworks. The present analysis emphasizes transparent subgroup evaluation rather than direct comparison of fairness mitigation effects across models.

In sum, the results demonstrate that the proposed gradient boosting model achieves strong predictive performance for pediatric Type 2 Diabetes risk using nationally representative survey data, while enabling transparent evaluation of subgroup-level behavior across socio-demographic groups. By emphasizing post-hoc fairness assessment, interpretability, and robustness rather than algorithmic bias mitigation, this study contributes evidence supporting responsible evaluation of predictive models in pediatric public health contexts.

8. Discussion and Policy Implications

This study evaluated the predictive performance and subgroup behavior of a machine learning model for pediatric Type 2 Diabetes (T2D) risk using nationally representative survey data. By emphasizing post-hoc subgroup evaluation rather than algorithmic bias mitigation, the findings contribute to ongoing discussions on how predictive models behave across socio-demographic groups in pediatric public health contexts.

The results underscore both the potential utility of machine learning for pediatric risk stratification and the challenges of ensuring consistent model performance across diverse populations. Rather than demonstrating bias reduction or equity enforcement, this study highlights the importance of transparent reporting of subgroup-level metrics to support responsible interpretation of predictive models.

8.1. Algorithmic Bias in Pediatric Risk Prediction

Prior research has shown that healthcare machine learning models may exhibit differential performance across demographic groups when subgroup evaluation is not explicitly reported (Obermeyer et al., 2019; Rajkomar et al., 2018). Consistent with this literature, subgroup-level evaluation in the present study revealed variation in predictive behavior across race/ethnicity and income strata.

These findings reinforce the need for routine fairness evaluation in pediatric risk prediction rather than assuming uniform performance across populations. Importantly, observed differences should be interpreted in the context of underlying variation in risk factor distributions and measurement limitations, rather than as evidence of algorithmic bias correction.

8.2. Implications for Clinical Decision Support

Predictive models for pediatric T2D risk may support population-level screening and resource planning when used with appropriate caution. The present findings suggest that model outputs should be interpreted alongside subgroup-specific performance metrics to avoid unintended disparities in downstream clinical use.

Rather than serving as standalone decision tools, such models are best positioned as complementary aids that inform clinical judgment. Transparent reporting of subgroup behavior is essential to prevent misinterpretation and overreliance in high-stakes pediatric settings.

8.3. Public Health Equity Considerations

From a public health perspective, fairness evaluation provides a diagnostic lens through which disparities in model performance can be identified and monitored. The results highlight the importance of integrating equity considerations into model evaluation frameworks without conflating evaluation with mitigation.

Policies that encourage standardized subgroup reporting in health-related AI may improve accountability and transparency, particularly for applications involving vulnerable pediatric populations.

8.4. Social Determinants of Health in Model Interpretation

Socioeconomic and behavioral variables contributed to risk prediction alongside clinical factors, reflecting the multifactorial nature of pediatric T2D risk. Inclusion of social determinants of health (SDOH) supported contextual interpretation of model outputs but does not imply causal inference or targeted intervention capability.

These findings align with public health literature emphasizing that exclusion of social context may limit interpretability, while inclusion requires careful, non-deterministic interpretation.

8.5. Ethical and Regulatory Implications

Ethical deployment of predictive models in pediatric care requires transparency, clear communication of limitations, and avoidance of claims that exceed empirical evidence. Regulatory frameworks should emphasize standardized evaluation, documentation of subgroup performance, and safeguards against misuse, rather than mandating specific algorithmic fairness interventions.

Limitations

This study has several important limitations. First, the analysis relied on cross-sectional survey data, which limits causal inference and precludes assessment of longitudinal model stability or performance drift. Second, outcome definitions and predictor measurements are subject to survey and laboratory measurement error, which may differentially affect subgroups.

Third, fairness metrics were applied post hoc for evaluation purposes only. The study did not implement bias mitigation, fairness constraints, or reweighting strategies, and therefore cannot make claims regarding bias reduction or equity improvement. Finally, subgroup analyses may be sensitive to sample size variation, particularly for smaller demographic groups.

These limitations highlight the importance of cautious interpretation and reinforce that fairness metrics should be viewed as diagnostic tools rather than corrective mechanisms.

9. Conclusion and Future Research Directions

This study demonstrates that machine learning models applied to nationally representative pediatric health data can achieve strong predictive performance while enabling transparent evaluation of subgroup-level behavior. By focusing on post-hoc fairness assessment rather than algorithmic bias mitigation, the analysis contributes a realistic and methodologically sound framework for evaluating pediatric risk prediction models.

Future research should prioritize longitudinal validation to assess temporal stability, explore harmonized fairness reporting standards across studies, and examine how subgroup performance metrics influence clinical interpretation in real-world settings. Expanding evaluation frameworks to additional population-based datasets may further improve generalizability while maintaining ethical and methodological rigor.

In conclusion, responsible application of machine learning in pediatric public health requires balancing predictive performance with transparent evaluation, clear communication of limitations, and avoidance of overclaims regarding equity or bias mitigation. Fairness-aware evaluation, when applied cautiously, can support more informed and accountable use of predictive models in healthcare.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Al-Zanbouri, Z., Sharma, G., & Raza, S. (2024). Equity in healthcare: Analyzing disparities in machine learning predictions of diabetic patient readmissions. Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), 660–669.

[2] American Diabetes Association. (2023). Classification and diagnosis of diabetes: Standards of medical care in diabetes—2023. Diabetes Care, 46(Suppl. 1), S19–S40. https://doi.org/10.2337/dc23-S002

[3] Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. https://fairmlbook.org

[4] Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

[5] Bilionis, I., Berrios, R. C., Fernandez-Luque, L., & Castillo, C. (2025). Disparate model performance and stability in machine learning clinical support for diabetes and heart diseases. AMIA Summits on Translational Science Proceedings, 95–104.

[6] Braveman, P., Arkin, E., Orleans, T., Proctor, D., & Plough, A. (2018). What is health equity? Robert Wood Johnson Foundation.

[7] Centers for Disease Control and Prevention. (2022). National Health and Nutrition Examination Survey (NHANES) analytic guidelines, 2011–2018. National Center for Health Statistics. https://www.cdc.gov/nchs/nhanes/

[8] Chen, R. J., Chen, T. Y., Lipkova, J., Wang, J. J., Williamson, D. F., Lu, M. Y., & Mahmood, F. (2021). Algorithmic fairness in artificial intelligence for medicine and healthcare. arXiv:2110.00603.

[9] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[10] Davoudi, A., Chae, S., Evans, L., Sridharan, S., Song, J., Bowles, K. H., & Topaz, M. (2024). Fairness gaps in machine learning models for hospitalization and emergency department visit risk prediction in home healthcare patients with heart failure. International Journal of Medical Informatics, 191, 105534. https://doi.org/10.1016/j.ijmedinf.2024.105534

[11] Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1), 27–46.

[12] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 214–226.

[13] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

[14] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine Learning, 46(1), 389–422. https://doi.org/10.1023/A:1012487302797

[15] Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29.

[16] Hill-Briggs, F., Adler, N. E., Berkowitz, S. A., Chin, M. H., Gary-Webb, T. L., Navas-Acien, A., Thornton, P. L., & Haire-Joshu, D. (2021). Social determinants of health and diabetes: A scientific review. Diabetes Care, 44(1), 258–279. https://doi.org/10.2337/dci20-0053

[17] Horsfall, L. J., Bondaronek, P., Ive, J., & Poduval, S. (2025). Clinical algorithms and the legacy of race-based correction: Historical errors, contemporary revisions and equity-oriented methodologies for epidemiologists. Clinical Epidemiology, 647–662.

[18] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning (2nd ed.). Springer.

[19] Khera, R., Haimovich, J., Hurley, N. C., McNamara, R., Spertus, J. A., Desai, N., & Krumholz, H. M. (2021). Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiology, 6(6), 633–642. https://doi.org/10.1001/jamacardio.2021.0126

[20] Kusner, M. J., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. Advances in Neural Information Processing Systems, 30.

[21] Lawrence, J. M., Divers, J., Isom, S., Saydah, S., Imperatore, G., Pihoker, C., & SEARCH for Diabetes in Youth Study Group. (2021). Trends in prevalence of type 1 and type 2 diabetes in children and adolescents in the United States, 2001–2017. JAMA, 326(8), 717–727. https://doi.org/10.1001/jama.2021.11165

[22] Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). Wiley.

[23] McNair, D. S. (2018). Preventing disparities: Bayesian and frequentist methods for assessing fairness in machine learning decision-support models. New Insights into Bayesian Inference, 71–89.

[24] Molnar, C. (2022). Interpretable machine learning (2nd ed.). Leanpub.

[25] Nadeau, K. J., Anderson, B. J., Berg, E. G., Chiang, J. L., Chou, H., Copeland, K. C., Hannon, T. S., Huang, T. T.-K., Lynch, J. L., Powell, J., Sellers, E., Tamborlane, W. V., Zeitler, P. S., & for the American Diabetes Association. (2016). Youth-onset type 2 diabetes consensus report: Current status, challenges, and priorities. Diabetes Care, 39(9), 1635–1642.

[26] O'Connor, A., Wellenius, G., & Kogan, M. D. (2018). Health disparities in diagnosis and management of pediatric chronic disease. Pediatrics, 141(2), e20172397. https://doi.org/10.1542/peds.2017-2397

[27] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

[28] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866–872. https://doi.org/10.7326/M18-1990

[29] Raza, S. (2023). Connecting fairness in machine learning with public health equity. *Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI)*, 704–708.

[30] Teli, M. S., Sah, S., & Shah, D. (2025). Diabetes prediction systems using XGBoost: A review of methods and evaluation practices. *Journal of Biomedical Informatics, 145*, 104473.

[31] Timmons, A. C., Duong, J. B., Fiallo, N. S., Lee, T., Vo, H. P. Q., Ahle, M. W., & Chaspari, T. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspectives on Psychological Science, 18*(5), 1062–1096.

[32] TODAY Study Group. (2012). A clinical trial to maintain glycemic control in youth with type 2 diabetes. *New England Journal of Medicine, 366*(24), 2247–2256.

[33] Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

[34] Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—Reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine, 383*(9), 874–882.

[35] Walker, R. J., Smalls, B. L., Campbell, J. A., Strom Williams, J. L., & Egede, L. E. (2022). Impact of social determinants of health on outcomes for type 2 diabetes: A systematic review. *Endocrine, 77*, 1–16. https://doi.org/10.1007/s12020-022-03036-8

[36] Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine, 25*(9), 1337–1340. https://doi.org/10.1038/s41591-019-0548-6

[37] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics, 9*, 515. https://doi.org/10.3389/fgene.2018.00515