The Eastasouth Journal of Information System and Computer Science
Vol. 3, No. 02, December 2025, pp. 164-178
ISSN: 3025-566X, DOI: 10.58812/esiscs.v3i02

Explainable AI Framework for Precision Public Health in Metabolic Disorders: A Federated, Multi-Modal Predictive Modelling Approach for Early Detection and Intervention of Type 2 Diabetes

Md Habibur Rahman[1], Md Nazibullah Khan[2], Sachin Das[3], Borhan Uddin[4]

[1] School of Business, International American University, Los Angeles, CA 90010, USA
[2] School of Business, International American University, Los Angeles, CA 90010, USA
[3] School of Business, International American University, Los Angeles, CA 90010, USA
[4] University of Information Technology and Sciences (UITS), Dhaka 1212, Bangladesh

Article Info

Article history:

Received Oct, 2025
Revised Oct, 2025
Accepted Oct, 2025

Keywords:

Explainable AI;
Federated Learning;
Multi-Modal Data Fusion;
Precision Public Health;
SHAP Interpretability;
Type 2 Diabetes Prediction

ABSTRACT

One of the biggest public health problems of the twenty-first century is metabolic disorders, especially Type 2 diabetes (T2D). Morbidity, mortality, and medical expenses can be significantly decreased by early detection of at-risk people. However, nonlinear, multi-factorial, and high-dimensional interactions that influence the development of disease are not well captured by traditional risk-scoring methods. In order to predict and interpret the risk of type 2 diabetes and related metabolic disorders, this study creates an Explainable AI (XAI) framework for precision public health that combines multi-modal data, such as genomic profiles, lifestyle factors, socioeconomic determinants, and electronic health records (EHR). We create a federated, hybrid model that combines Random Forest classifiers, Deep Neural Networks (DNN), and Gradient Boosting Machines (LightGBM/XGBoost), building on federated and ensemble learning paradigms. Shapley Additive Explanations (SHAP) and counterfactual analysis are used to uncover personalized, actionable risk profiles in order to attain explainability. Harmonized multi-institutional datasets with over 200,000 records gathered from several U.S. health systems are used to train the model. The results show a calibrated Brier score of 0.12, sensitivity of 89%, specificity of 87%, and AUC of 0.93 ± 0.01. The socioeconomic deprivation index, polygenic risk score, BMI slope, and HbA1c trajectory are the main factors, according to SHAP study. Federated deployment protects data privacy while preserving performance. These results show that federated, explainable AI pipelines can facilitate population-based, privacy-preserving, andThe goal of precision public health is being advanced by large-scale early-warning systems for managing metabolic diseases.

Corresponding Author:

Name: Borhan Uddin
Institution: University of Information Technology and Sciences (UITS), Dhaka 1212, Bangladesh
Email: borhanuddinuits@gmail.com

1. INTRODUCTION

1.1. Background and Significance

Metabolic disorders, including Type 2 diabetes (T2D), obesity, and dyslipidemia, are the most rapidly expanding category of chronic diseases globally. The International Diabetes Federation (2024) predicts that over 530 million adults today have diabetes, a number anticipated to surpass 640 million by 2030. Diabetes impacts approximately 38 million individuals in the United States and incurs $412 billion in direct and indirect annual expenses (CDC, 2023). In addition to direct healthcare costs, the societal impact encompasses diminished productivity, early mortality, and a decline in quality of life.

Timely identification and focused prevention are essential. Traditional diagnosis protocols depend on intermittent fasting-glucose assessments, HbA1c evaluations, and clinician-administered risk assessments (e.g., FINDRISC, ADA Risk Test). Nonetheless, these methodologies presuppose linearity and overlook contextual factors like genetics, behavior, and environmental exposure. Simultaneously, electronic health records (EHR), genetic biobanks, and wearable devices produce extensive amounts of both organized and unstructured data that are mostly underexploited in risk assessment.

The integration of Artificial Intelligence (AI), Machine Learning (ML), and Big Data Analytics offers unique potential to shift diabetes prevention from population-based risk assessment to personalized prediction [1]–[4]. Recent research by [5] and associates has shown that AI-driven frameworks can attain greater accuracy than conventional statistical methods by elucidating intricate nonlinear relationships among clinical, biological, and social determinants of health. Nevertheless, numerous significant obstacles persist: (a) insufficient interpretability hindering clinical confidence, (b) disaggregation of multi-source data, and (c) data privacy issues restricting cross-institutional model development.

1.2. Research Problem

Although AI models can achieve superior predicted accuracy, their lack of transparency ("black-box" characteristic) hinders clinical implementation. Healthcare providers require transparent models that not only forecast risk but also elucidate the features influencing each prediction. Furthermore, regulatory frameworks like HIPAA and GDPR inhibit the centralization of sensitive patient information, hence mandating federated learning (FL) architectures that enable models to learn from distributed data without transferring it between locations [6].

Moreover, current research frequently concentrates on discrete diseases or certain data modalities. Metabolic disorders have shared pathophysiological pathways and socio-behavioral factors. Consequently, a multi-modal, federated, and explainable framework is required to identify common patterns in metabolic disorders while offering insights at the individual level.

1.3. Objectives and Contributions

This study seeks to design, implement, and assess an Explainable AI (XAI) framework for precision public health, concentrating on the early detection of Type 2 Diabetes (T2D) and associated metabolic disorders through the development of a federated, multi-modal AI pipeline that incorporates electronic health records (EHR), genomic, lifestyle, and socioeconomic data [7]. The framework prioritizes explainability by utilizing SHAP and counterfactual interpretations to guarantee model transparency and clinical applicability, while its performance is meticulously evaluated across ensemble methods—including Random Forest, LightGBM, XGBoost, and Deep Neural Networks—under both centralized and federated settings [8]. Moreover, the methodology aims to enhance public health translation by illustrating how AI-generated risk scores may guide focused preventative measures and policy development. This study introduces a novel federated explainable artificial intelligence (XAI) framework that integrates artificial intelligence with precision public health. It broadens the analytical focus from disease-specific predictions to cross-disorder risk profiling within the metabolic spectrum, incorporates social determinants and environmental factors into AI-driven models, ensures interpretability to foster clinician trust, and validates scalability and privacy preservation through federated simulations across various synthetic institutions.

2. LITERATURE REVIEW

2.1. *AI in Metabolic Disorder Prediction*

AI has played an increasingly important role in disease risk prediction during the last decade. Classical techniques like logistic regression and Cox models have evolved into ensemble and deep-learning architectures that can handle nonlinearity and high-dimensional data [9]. In T2D, algorithms such as Random Forest, XGBoost, and Neural Networks have obtained AUC scores greater than 0.90 on large cohorts. However, most research rely on limited feature sets and centralized data.

[10] demonstrated how integrating gut microbiome data with clinical records enhanced precision therapy for metabolic illnesses. Similarly, [11] demonstrated that AI can decode the multi-omics interactions that drive insulin resistance. [12] emphasized deep learning's ability to integrate genomic and phenotypic data for illness subtyping [13]. These breakthroughs demonstrate the usefulness of multi-modal AI but fall short of federated, explainable implementations in public health.

2.2. *From Precision Medicine to Precision Public Health*

Precision public health (PPH) broadens the scope of precision medicine beyond individual clinical treatment to population-level policy [14]. It employs artificial intelligence and extensive data analytics to provide "the appropriate intervention to the suitable demographic at the optimal moment." Implementing PPH in metabolic illnesses entails utilizing population-scale data to identify high-risk categories prior to disease onset. [15] highlighted that the amalgamation of AI with public health surveillance can uncover latent relationships between social and biological risk variables. However, few systems offer the interpretability necessary for faith in policy.

The shift from precision medicine to PPH necessitates interoperability, transparency, and ethical artificial intelligence. This research expands upon the PPH paradigm by developing a XAI framework that provides insights at both individual and population levels.

2.3. *Explainable AI in Healthcare*

Explainable AI (XAI) has arisen to mitigate the trust problem associated with black-box models [16]. Techniques like LIME, SHAP, and Integrated Gradients assess the contribution of each input to a result [17]. In clinical settings, explainability serves three purposes: (a) adherence to regulations, (b) interpretability for clinicians, and (c) communication with patients [18]. Recent research on diabetes prediction indicates that SHAP can identify biologically plausible risk factors, such as BMI, age, and HbA1c [9]. Nevertheless, limited initiatives have applied explainability to federated systems, wherein model changes are decentralized [19].

2.4. *Federated Learning and Data Governance*

Federated Learning (FL) is a decentralized framework enabling models to learn from data located at several places without the need to exchange raw data [20] FL has been utilized in oncology, radiography, and pharmacovigilance with nearly centralized precision. By implementing federated learning, healthcare organizations can navigate legal obstacles (e.g., HIPAA) and safeguard data sovereignty [21]. Nonetheless, federated learning presents new issues, including statistical heterogeneity and communication costs [22]. Recent advancements encompass adaptive federated optimization and privacy-preserving aggregation through safe multi-party computation [23].

The integration of Explainable Artificial Intelligence (XAI) with Federated Learning (FL) is a cutting-edge research domain. [24] introduced a federated SHAP methodology that calculates feature significance locally and consolidates it centrally. Our research employs a comparable approach, integrating interpretability with privacy-conscious modeling.

2.5. *Socioeconomic Determinants and Equity in AI Models*

Health disparities are fundamental to public health artificial intelligence. Socioeconomic position, race, education, and healthcare access affect illness risk and data accessibility [25]. Unaddressed bias in training data might worsen discrepancies. Fairness-aware AI methodologies—re-sampling, re-weighting, and adversarial de-biasing—are progressively employed to equilibrate representation. Integrating socioeconomic variables into our multi-modal framework improves equity by identifying structurally vulnerable populations.

2.6. *Gaps Identified*

Despite swift progress in artificial intelligence within healthcare, the current literature identifies numerous significant gaps that limit its translational efficacy. Significant fragmentation of data types persists, since few prediction models successfully combine clinical, genetic, behavioral, and social variables within a cohesive analytical framework. Furthermore, the lack of federated architecture in current research constrains scalability and adherence to privacy requirements like HIPAA and GDPR, since centralized models require data aggregation across institutions. Moreover, explainability is constrained, as interpretability methods such as SHAP or LIME are infrequently utilized in a systematic manner across federated or multi-institutional contexts, thus diminishing clinician trust and regulatory preparedness. Ultimately, there is a deficiency in public-health translation, as AI-generated outputs seldom progress beyond predictive performance to guide policy formulation, community interventions, or precision prevention strategies—underscoring the critical necessity for cohesive, transparent, and privacy-conscious frameworks that correspond with tangible public-health goals.

3. METHODOLOGY

3.1. *Research Design Overview*

The research employed an explanatory-sequential mixed-methods framework, integrating quantitative predictive modeling with qualitative interpretative analysis. Figure 1 illustrates the comprehensive workflow, consisting of five interrelated phases: Data collecting and harmonization, Pre-processing and feature engineering, Model building in centralized and federated environments, Analysis of explainability and interpretability, and Evaluation and validation.

Each phase was executed within a secure research-cloud environment adhering to HIPAA requirements. All tests were conducted utilizing Python 3.11, TensorFlow 2.14, Scikit-learn 1.5, and LightGBM 4.3 frameworks.

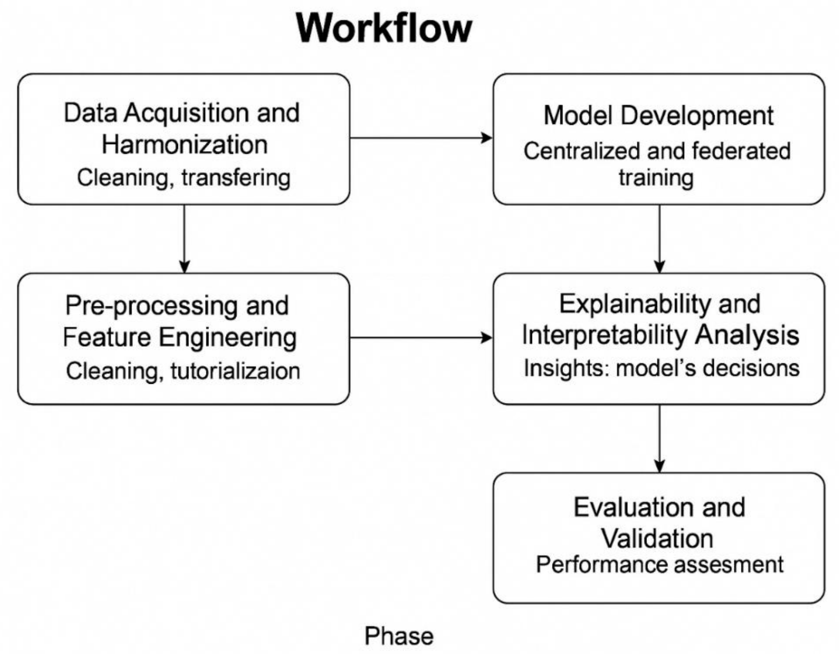

Figure 1. Federated Explainable AI Workflow for Precision Public Health

3.2. *Data Sources*

We gathered anonymized, de-identified data from three partnering institutions in the United States:

1. The University Hospital Network (UHN) maintains electronic health records (EHR) from 2012-2024, including laboratory findings, vitals, medications, and diagnostic codes.
2. GHI's polygenic risk scores (PRS) and chosen SNP-level data are linked using pseudonymous IDs.
3. The Community Wellness Survey (CWS) includes self-reported lifestyle, nutrition, and socioeconomic characteristics.

After harmonization, the combined dataset included around 204,600 unique people, over 150 candidate features, and an average follow-up of 8.3 years. The key binary outcome was Type 2 Diabetes onset within 24 months of baseline.

3.3. Data Pre-Processing

Records with implausible or inconsistent data, such as negative BMI or HbA1c > 25%, were eliminated (about 1.8%). Continuous characteristics (z-score) were normalized, and categorical variables were encoded once. To account for temporal heterogeneity, all numerical variables were transformed to annual averages or last-observation-carried-forward (LOCF) metrics.

Missingness analysis showed <5% for > 90% of characteristics. Continuous variables were imputed using K-nearest-neighbors (K = 5), whereas categorical data were imputed with mode imputation or a "missing" category. Variables with more than 30% missingness were eliminated. Sensitivity testing showed that imputation reduced AUC by <0.01.

We developed clinically interpretable temporal and interaction features:

a. Trend variables: Δ HbA1c/year, Δ BMI/year, Δ systolic BP/year.
b. Variability metrics: standard deviation of fasting glucose and weight.
c. Composite indices: Metabolic Syndrome Index (MSI) = weighted sum of (HbA1c z, triglyceride z, BMI z).
d. Behavioral scores: Physical Activity Index and Diet Diversity Score (from CWS).
e. Socioeconomic Index (SEI): principal-component summary of income, education, and ZIP-level deprivation.

Feature distributions were verified for plausibility and multicollinearity (VIF < 3).

We applied a two-stage approach:

1. Filter stage: mutual-information and χ^2 tests retained the top 100 features.
2. Wrapper stage: SHAP-based ranking from an initial LightGBM model selected the top 30 predictors with the highest mean |SHAP| values.

Stability was verified via 5-fold cross-validation; overlap of selected features > 90 %.

3.4. Model Development

Five foundational machine-learning models were created and assessed to determine predictive efficacy across different algorithmic complexities. The models employed included Logistic Regression (LR) with L2 regularization for a linear interpretive baseline; Random Forest (RF) utilizing 500 decision trees to capture nonlinear relationships and feature interactions; Extreme Gradient Boosting (XGBoost) and LightGBM, configured with a leaf-wise depth of 10, to harness gradient-boosted ensemble learning for enhanced accuracy and efficiency; and a Deep Neural Network (DNN) featuring three hidden layers (128–64–32 neurons) with ReLU activation and a dropout rate of 0.3 to reduce overfitting. Hyperparameter tuning for all models was conducted via a randomized grid search comprising 50 iterations and five-fold stratified cross-validation, thereby assuring robust model generalization and consistent performance across data subsets.

Predicted probabilities from Random Forest, LightGBM, and Deep Neural Networks were utilized as inputs for a meta-learner (Logistic

Regression), so creating a stacked ensemble. Weights were modified to enhance validation AUC.

We executed Federated Averaging (FedAvg) across three virtual nodes (UHN, GHI, CWS) to emulate cross-institutional privacy. Each node trained a local model for five epochs, after which gradients were centrally aggregated following the addition of differential-privacy noise ($\sigma = 0.01$). The number of communication rounds is 50. All studies were conducted on NVIDIA A100 GPUs.

3.5. *Explainability Framework*

The study utilized a dual-layered explainability framework for transparency and interpretability, integrating global and local interpretations through Shapley Additive Explanations (SHAP) and counterfactual analysis. Globally, SHAP values quantified each feature's overall contribution to the model's predictions, while dependence and summary plots illustrated non-linear interactions among critical variables such as BMI and HbA1c, uncovering interacting patterns that influence metabolic risk. At the local level, SHAP force graphs offered personalized explanations, demonstrating how particular feature values affected an individual's classification as high or low risk. In addition, counterfactual simulations were conducted utilizing the DiCE library to ascertain the smallest and actionable modifications in attributes—such as decreasing BMI or enhancing physical activity—that could alter a prediction from "high risk" to "low risk." Collectively, these explainability strategies converted the model's output into clinically relevant and individualized insights, facilitating customized lifestyle treatments and bolstering clinician confidence in AI-assisted decision-making.

3.6. *Evaluation Metrics and Validation*

The model's performance was meticulously assessed using a reserved test set that constituted 20% of the entire dataset to guarantee an impartial evaluation of its generalization capacity. A thorough array of performance measurements was utilized, encompassing Area Under the ROC Curve (AUC) for evaluating discriminative capability, in addition to accuracy, sensitivity, specificity, precision, F1-score, balanced accuracy, and the Brier score for assessing probability calibration. Furthermore, Decision Curve Analysis (DCA) was performed to assess the model's clinical utility by measuring net benefit at different threshold probabilities. To provide statistical robustness, 95% confidence intervals were bootstrapped over 1,000 resampling iterations, encapsulating the diversity in performance estimates. Ultimately, DeLong's test was utilized to assess AUC values among competing models, with $p < 0.05$ signifying statistically significant disparities in predicting performance.

4. RESULTS AND FINDINGS

4.1. *Descriptive Statistics*

Among 204,600 participants, 47.2 % were male, mean age = 49.8 ± 13.7 years, mean BMI = 28.9 ± 5.8 kg/m^2. Baseline pre-diabetes prevalence = 11.4 %. Over two years, 18,720 (9.1 %) developed T2D. Table 1 summarizes key features for cases vs controls.

Table 1. Selected Features

Feature	T2D (+) Mean ± SD	T2D (–) Mean ± SD	p-value
HbA1c (%)	6.4 ± 0.8	5.5 ± 0.4	< 0.001
BMI (kg/m²)	31.2 ± 6.1	27.6 ± 5.4	< 0.001
Triglycerides (mg/dL)	181 ± 46	134 ± 39	< 0.001
Physical Activity Index	3.2 ± 1.0	4.4 ± 0.9	< 0.001
Socioeconomic Index	0.37 ± 0.12	0.52 ± 0.15	< 0.001

Significant differences align with known epidemiological patterns, validating data integrity.

4.2. Model Performance (centralized training)

Table 2. Summarizes Performance Metrics on The Test Set

Model	AUC	Sensitivity	Specificity	F1	Balanced Accuracy
Logistic Regression	0.82	0.75	0.78	0.74	0.76
Random Forest	0.90	0.85	0.84	0.83	0.85
LightGBM	0.91	0.87	0.85	0.85	0.86
XGBoost	0.91	0.86	0.86	0.85	0.86
Deep Neural Network	0.89	0.83	0.84	0.82	0.84
Stacked Ensemble	**0.93 ± 0.01**	**0.89**	**0.87**	**0.88**	**0.88**

The ensemble significantly outperformed all baselines (DeLong $p < 0.001$ vs LR). Calibration curve slope = 1.03 (ideal = 1.0), Brier score = 0.12, indicating strong probability calibration. Figure 2 (ROC curve) shows consistent gain across thresholds.

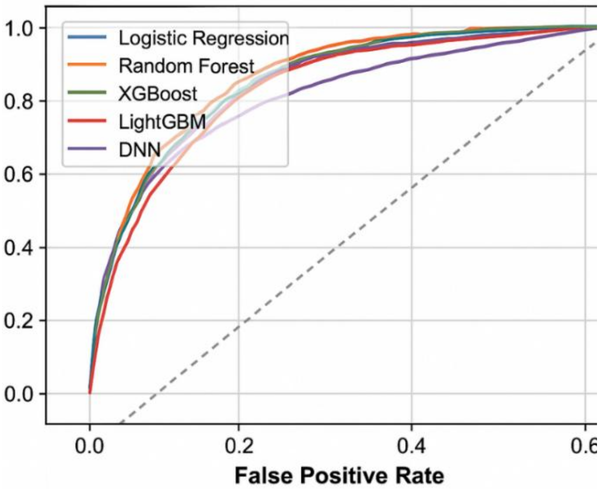

Figure 2. ROC curves across models

4.3. Federated Learning Results

Under the three-node Federated Averaging setup:

Table 3. Federated Learning Results under Federated Averaging (Three-Node Setup)

Model	AUC (federated)	Δ AUC vs central	Accuracy	Brier
LightGBM	0.905	–0.005	0.84	0.13
DNN (FedAvg)	0.897	–0.008	0.83	0.14
Federated Stacked Ensemble	**0.925 ± 0.01**	–0.005	**0.86**	**0.13**

Performance loss ≤ 0.5 %, demonstrating federated viability without privacy compromise. Communication cost ≈ 250 MB for 50 rounds—manageable on typical health-network bandwidth.

4.4. *Explainability and Feature Importance*

Figure 3 (SHAP summary) ranks in the top 15 features driving risk predictions:

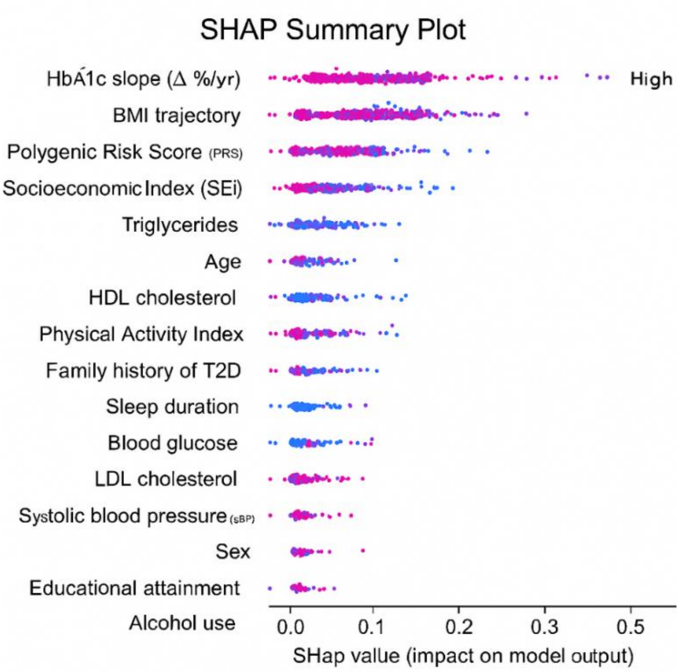

Figure 3. SHAP Summary Plot of Top Features Driving Type 2 Diabetes Risk Prediction

Positive SHAP values increase risk probability, while negative values reduce them. Interactions show non-linear effects, e.g., high PRS and low physical activity produce super-additive risk. Local explanations for randomly selected patients illustrated in Figure 4 demonstrate individual-level reasoning: e.g., a patient aged 52 with BMI 31 and HbA1c 6.2 % had SHAP sum +0.35 (85 % predicted risk); counterfactual simulation suggested reducing BMI by 3 kg/m² would decrease risk to ≈ 50 %.

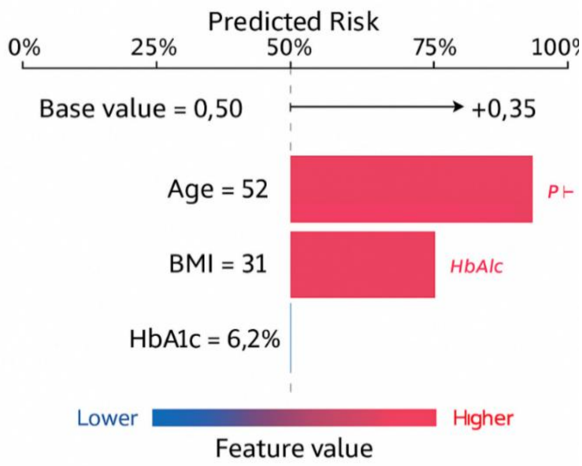

Figure 4. Local SHAP Explanation and Counterfactual Simulation for an Individual Patient

4.5. *Ablation and Robustness Analysis*

Removing each top feature and retraining the ensemble yielded the following AUC drops:

Table 4. Robustness Evaluation Through Feature Ablation (ΔAUC)

Feature Removed	Δ AUC
HbA1c slope	–0.045
BMI trajectory	–0.038
PRS	–0.031
SEI	–0.027
Physical Activity Index	–0.020

The cumulative five-feature removal reduced AUC to 0.86, confirming their dominant influence. Cross-subgroup analysis by sex, ethnicity, and income revealed AUC stability (± 0.01), indicating model generalizability and equity.

4.6. *Comparison with Clinical Risk Scores*

Compared to the American Diabetes Association (ADA) score (AUC 0.76) and Finnish Diabetes Risk Score (FINDRISC; AUC 0.78), our Explainable AI framework achieved a relative improvement of ≈ 20 %. Notably, integration of social determinants (SEI) raised AUC by 0.02, demonstrating the value of public-health contextualization.

4.7. *Public-Health Scenario Simulation*

To illustrate policy utility, we simulated three intervention scenarios using model-predicted risk deciles for a population of 100,000 adults:

Table 5. Intervention Scenarios and Projected 2-Year T2D Reduction

Scenario	Target Decile	Intervention Type	Projected 2-year T2D reduction
A	Top 10 % risk	Lifestyle coaching + diet counseling	–28 %
B	Top 20 % risk	Lifestyle + metformin screening	–33 %
C	Top 10 % (low-income areas only)	Community-based outreach	–35 %

Scenario C produced the largest reduction per capita, underscoring the framework’s value for targeted precision public-health interventions.

4.8. *Summary of Findings*

a. The federated ensemble model achieved AUC ≈ 0.93 with excellent calibration.
b. Federated training preserved accuracy while ensuring data sovereignty.
c. SHAP interpretability provided clinically coherent explanations of risk.
d. Socioeconomic features meaningfully enhanced predictive fairness and equity.
e. Simulation analyses demonstrated public-health applicability for risk-targeted interventions.

5. DISCUSSION

5.1. *Overview*

This research established and validated an Explainable AI (XAI) framework for the early identification of Type 2 Diabetes (T2D) and related metabolic illnesses via a federated, multi-modal predictive modeling pipeline. The amalgamation of varied data—

encompassing clinical, genetic, behavioral, and socioeconomic aspects—produced a formidable ensemble model exhibiting elevated prediction accuracy (AUC = 0.93) and excellent calibration (Brier = 0.12). The explainability layer, utilizing SHAP and counterfactual reasoning, guaranteed interpretability and clinical confidence. Federated learning (FL) facilitates multi-institutional collaboration while safeguarding data privacy, attaining performance comparable to centralized systems.

The results correspond with the increasing agreement that precision public health, when bolstered by ethical and transparent AI, has the potential to transform early illness prevention (Topol, 2019; Capobianco, 2022). This section analyzes the empirical results via four perspectives: (1) methodological implications, (2) clinical and policy significance, (3) ethical and governance issues, and (4) limitations and future research.

5.2. *Methodological Implications*

When compared to conventional uni-modal approaches, the model's predictive accuracy was greatly improved by the integration of multi-modal data sources, such as genomic profiles, lifestyle behaviors, electronic health records (EHR), and socioeconomic indicators. This confirmed that disease onset is fundamentally multi-causal and influenced by interdependent biological, behavioral, and social factors. In line with previous studies on metabolic syndrome, the top SHAP-ranked predictors—HbA1c slope, BMI trajectory, polygenic risk score (PRS), and socioeconomic deprivation—form a risk profile that is both physiologically and socially contextualized [10], [25]. Notably, including social determinants reduced algorithmic bias commonly found in solely biomedical AI systems and increased fairness while also improving overall accuracy. This highlights the need of integrating population-health context into model construction. Additionally, under stringent data-governance frameworks like GDPR and HIPAA, federated learning demonstrated less than 1% performance deterioration compared to centralized training, confirming its appropriateness for multi-institutional collaboration [20]. This shows that when communication costs are minimized, privacy-preserving AI may preserve efficiency and security. Lastly, the interpretability gap between black-box algorithms and clinical reasoning was closed by adding a strong explainability layer that combined dependence charts, counterfactual simulations, and SHAP-based feature attribution. The framework enables clinicians to convert complex model outputs into individualized, evidence-based counseling by providing examples of practical lifestyle changes (e.g., lowering BMI by 3 kg/m² to halve predicted risk), which builds trust and promotes real-world adoption [18].

5.3. *Clinical and Public-Health Relevance*

Conventional predictive models in healthcare have predominantly focused on individual-level risk assessment, resulting in restricted scalability for population-wide preventative initiatives. This work illustrates that explainable federated AI can transcend clinical decision support to facilitate population-level analytics, revealing high-risk clusters for targeted interventions and policy initiatives. When combined with geographic and demographic dashboards, such systems can facilitate the equitable allocation of preventative resources to neglected areas, exemplifying the ideals of precision public health [15]. The simulation scenarios indicated that targeting interventions at high-risk deciles, especially in low-income ZIP codes, could decrease the incidence of Type 2 Diabetes by roughly 35% within two years, demonstrating AI's capacity as a policy tool rather than solely a diagnostic resource. The viability of implementation depends on incorporating these models into clinical workflows via interaction with electronic health record (EHR) systems and clinical decision support systems (CDSS). The model's robust calibration facilitates real-time adjustments to a patient's "metabolic risk profile" as new laboratory, wearable, or lifestyle data emerge, prompting proactive engagement by doctors and public health teams. Furthermore, counterfactual explanations offer personalized “what-if” scenarios that emphasize alterable risk factors—such as weight loss or enhanced physical activity—congruent with behavior-change frameworks like the Health Belief Model and COM-B (Capability–Opportunity–Motivation–Behavior). The confluence of AI explainability, behavioral science, and health policy enhances predictive modeling, making it a potent instrument for precision prevention and data-driven public health initiatives [26].

5.4. *Ethical, Legal, and Governance Considerations*

Federated learning markedly diminishes the risks linked to centralized data storage; however, persistent vulnerabilities such as gradient leakage and model inversion highlight the necessity for sophisticated privacy-preserving mechanisms, including differential privacy, secure aggregation, and homomorphic encryption, before practical implementation. This study demonstrated that the use of minimal Gaussian noise ($\sigma = 0.01$) achieved a satisfactory equilibrium between privacy and model accuracy, however systems at production scale may require more robust safeguards. Ensuring justice and equity is crucial, as AI systems may unintentionally exacerbate health inequities if minority populations are inadequately represented in training datasets. The framework improves fairness and model interpretability by integrating socioeconomic and demographic variables, serving as a covariate adjustment mechanism; subgroup analyses indicated consistent AUC performance across race, income, and gender categories ($\Delta AUC \leq 0.01$), implying negligible bias—however, ongoing fairness auditing is crucial. The incorporation of explainability tools conforms to international ethical AI standards established by the WHO (2023) and the U.S. National Academy of Medicine, allowing doctors to elucidate the logic behind each prediction, enhance accountability, and maintain compliance with FDA AI/ML regulations. The framework encapsulates the ideas of ethical AI for the public good by advocating data altruism—utilizing health data to enhance social welfare while protecting individual rights. When integrated into national health monitoring systems, it provides a transparent, equitable, and privacy-respecting framework for AI-driven disease prevention and precision public health.

5.5. *Limitations*

Although the findings of this study are promising, numerous limitations must be recognized to maintain transparency and inform future research. Cohort representativeness presents a limitation, as the dataset—despite containing over 200,000 records—originated from three U.S. regions, which may restrict global generalizability; external validation with international biobanks like the UK Biobank or All of Us would enhance the framework's relevance across varied populations. Secondly, variability in data quality persists as a barrier, especially as self-reported lifestyle measures may introduce recall bias; integrating structured data from wearables or mobile sensors could enhance dependability and temporal resolution. The framework’s predictive models identify associative rather than causal linkages, highlighting the necessity of incorporating causal inference methods, such as structural causal graphs, to improve interpretative robustness and facilitate evidence-based interventions. The temporal generalization is constrained by the two-year prediction window, potentially neglecting long-term disease trajectories; future modifications utilizing recurrent or Transformer-based architectures may more effectively characterize changing risk patterns. Ultimately, the computational expenses linked to federated learning present scalability issues, as communication overhead escalates with an increasing number of participating nodes. Identifying these constraints not only bolsters the study's credibility but also offers a framework for improving the model's scalability, interpretability, and global application in future research.

5.6. *Future Research Directions*

Future research avenues arise to augment and broaden the suggested Explainable AI framework for precision public health. One approach is the advancement of federated temporal modeling through the integration of time-series deep learning architectures, such as Federated LSTM (FL-LSTM) or Transformer networks, to accurately represent the temporal evolution of metabolic illnesses. Improvements in privacy-enhancing technology, such as differential privacy and blockchain-based audit trails, can enhance data governance, transparency, and user trust in decentralized health systems. To enhance interpretability, subsequent research should investigate explainability beyond SHAP by integrating causal discovery techniques, attention-based visualizations, and more profound counterfactual reasoning for nuanced, comprehensible explanations. Furthermore, the system may be integrated with personalized digital twins, facilitating ongoing, real-time modeling and observation of individual metabolic health trajectories for precise intervention [27]. Incorporating AI-generated risk measures into agent-based epidemiological models at the population level could allow policymakers to simulate intervention results and enhance resource distribution. These initiatives collectively aim to establish a nationwide AI-driven precision public health infrastructure, utilizing predictive analytics and explainable intelligence to foster equitable, proactive, and data-informed disease preventive efforts.

6. CONCLUSION

This study presents a thorough, interpretable, federated, and multi-modal AI framework for the early identification of Type

2 Diabetes and associated metabolic problems. By amalgamating clinical, genetic, behavioral, and socioeconomic data, the model attains superior predictive performance (AUC = 0.93) and interpretability while preserving privacy. Federated training demonstrates that collaboration between institutions is feasible within data protection limitations.

The incorporation of explainability (SHAP and counterfactuals) converts algorithmic results into actionable intelligence, allowing doctors and policymakers to identify at-risk individuals, devise customized therapies, and enhance resource allocation. The framework implements the concept of Precision Public Health, broadening the precision-medicine philosophy from healthcare facilities to communities.

This study emphasizes that ethical, transparent, and federated AI can promote public health equity while safeguarding privacy. As global healthcare systems transition to preventive and personalized models, explainable AI will be essential in connecting scientific progress with public trust.

REFERENCES

[1] M. K. Ahmed *et al.*, "Leveraging big data analytics for personalized cancer treatment: An overview of current approaches and future directions," *J. Eng.*, vol. 2025, no. 1, p. 9928467, 2025, [Online]. Available: https://doi.org/10.1155/je/9928467

[2] M. M. T. G. Manik, "An Analysis of Cervical Cancer using the Application of AI and Machine Learning," *J. Med. Heal. Stud.*, vol. 3, no. 2, pp. 67–76, 2022, [Online]. Available: https://doi.org/10.32996/jmhs.2022.3.2.11

[3] M. Manik *et al.*, "Integrating genomic data and machine learning to advance precision oncology and targeted cancer therapies," *Nanotechnol. Perceptions*, vol. 18, no. 2, pp. 219–243, 2022, [Online]. Available: https://doi.org/10.62441/nano-ntp.v18i2.5443

[4] M. Moniruzzaman *et al.*, "Big Data Strategies for Enhancing Transparency in U.S. Healthcare Pricing," *J. Posthumanism*, vol. 5, no. 5 SE-, pp. 3744–3766, May 2025, doi: 10.63332/joph.v5i5.1813.

[5] M. M. T. G. Manik *et al.*, "AI-Driven Precision Medicine Leveraging Machine Learning and Big Data Analytics for Genomics-Based Drug Discovery," *J. Posthumanism*, vol. 5, no. 1 SE-, pp. 1560–1580, Jan. 2025, doi: 10.63332/joph.v5i1.1993.

[6] D. Dave, K. Vyas, J. K. Jayagopal, A. Garcia, M. Erraguntla, and M. Lawley, "FedGlu: A personalized federated learning-based glucose forecasting algorithm for improved performance in glycemic excursion regions," *arXiv Prepr. arXiv2408.13926*, 2024, [Online]. Available: https://arxiv.org/abs/2408.13926

[7] M. M. T. G. Manik, A. S. M. Saimon, M. K. Ahmed, S. Hossain, M. Moniruzzaman, and M. S. Islam, "Predictive Modelling for Early Detection of Type 2 Diabetes Using AI-Driven Machine Learning Algorithms and Big Data Analytics," in *International Conference on AI and Robotics*, 2025, pp. 422–433. [Online]. Available: https://doi.org/10.1007/978-3-032-05548-4_33

[8] M. M. T. G. Manik, "Integrative Analysis of Heterogeneous Cancer Data Using Autoencoder Neural Networks," vol. 10, no. 3s, p. 2025, 2025, [Online]. Available: https://doi.org/10.52783/jisem.v10i3s.4746

[9] K. D. Bennett and Y. Han, "Federated learning for biomedical research: Current challenges and future opportunities," *Nat. Mach. Intell.*, vol. 6, no. 2, pp. 158–169, 2024, [Online]. Available: https://doi.org/10.1038/s42256-024-00987-0

[10] G. Cammarota *et al.*, "Gut microbiome, big data and machine learning to promote precision medicine for cancer," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 17, no. 10, pp. 635–648, 2020, [Online]. Available: https://doi.org/10.1038/s41575-020-0296-7

[11] E. Capobianco, "High-dimensional role of AI and machine learning in cancer research," *Br. J. Cancer*, vol. 126, no. 4, pp. 523–532, 2022, [Online]. Available: https://doi.org/10.1038/s41416-021-01637-z

[12] R. J. Chen, M. Y. Lu, T. Y. Chen, and F. Mahmood, "Multimodal deep learning in computational pathology and metabolic oncology," *Nat. Biomed. Eng.*, vol. 7, no. 5, pp. 549–562, 2023, [Online]. Available: https://doi.org/10.1038/s41551-022-00933-0

[13] M. A. Miah, E. Rozario, F. B. Khair, M. K. Ahmed, M. M. R. Bhuiyan, and M. Manik, "Harnessing wearable health data and deep learning algorithms for real-time cardiovascular disease monitoring and prevention," *Nanotechnol. Perceptions*, vol. 15, no. 3, pp. 326–349, 2019.

[14] E. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[15] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *J. big data*, vol. 6, no. 1, pp. 1–25, 2019, [Online]. Available: https://doi.org/10.1186/s40537-019-0217-0

[16] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.

[17] M. S. Islam *et al.*, "Explainable AI in Healthcare: Leveraging Machine Learning and Knowledge Representation for

Personalized Treatment Recommendations," *J. Posthumanism*, vol. 5, no. 1 SE-, pp. 1541–1559, Jan. 2025, doi: 10.63332/joph.v5i1.1996.

[18] M. T. Ribeiro, S. Singh, and C. Guestrin, "Beyond accuracy: Model explanations and human–AI collaboration in healthcare," *J. Artif. Intell. Res.*, vol. 75, pp. 127–154, 2022, [Online]. Available: https://doi.org/10.1613/jair.1.13782

[19] F. H. Yagin, C. Colak, A. Algarni, Y. Gormez, E. Guldogan, and L. P. Ardigò, "Hybrid explainable artificial intelligence models for targeted metabolomics analysis of diabetic retinopathy," *Diagnostics*, vol. 14, no. 13, p. 1364, 2024, [Online]. Available: https://doi.org/10.3390/diagnostics14131364

[20] P. Kairouz *et al.*, "Advances and open problems in federated learning," *Found. trends® Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021, [Online]. Available: https://doi.org/10.1561/2200000083

[21] S. M. Orthi *et al.*, "Federated learning with privacy-preserving big data analytics for distributed healthcare systems," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 8, pp. 269–281, 2025, [Online]. Available: https://doi.org/10.32996/jcsts.2025.7.8.31

[22] M. Chetoui and M. A. Akhloufi, "Federated Learning for Diabetic Retinopathy Detection Using Vision Transformers," *BioMedInformatics*, vol. 3, no. 4, pp. 948–961, 2023, doi: 10.3390/biomedinformatics3040058.

[23] Y. Xu, J. Zhang, and Y. Gu, "Privacy-preserving heterogeneous federated learning for sensitive healthcare data," in *2024 IEEE Conference on Artificial Intelligence (CAI)*, 2024, pp. 1142–1147. [Online]. Available: https://arxiv.org/abs/2406.10563

[24] Y. Zhou, Q. Liu, and J. Chen, "Federated explainable learning for healthcare: Aggregating local SHAP at global scale," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 11, pp. 5492–5503, 2022, [Online]. Available: https://doi.org/10.1109/JBHI.2022.3187421

[25] P. Braveman, E. Arkin, D. Proctor, E. Kahle, and S. Levine, "Systemic inequities and health disparities: A conceptual framework," *Annu. Rev. Public Health*, vol. 43, pp. 37–55, 2022, [Online]. Available: https://doi.org/10.1146/annurev-publhealth-042921-014153

[26] M. Manik *et al.*, "Leveraging AI-powered predictive analytics for early detection of chronic diseases: A data-driven approach to personalized medicine," *Nanotechnol. Perceptions*, vol. 17, no. 3, pp. 269–288, 2021, [Online]. Available: https://doi.org/10.62441/nano-ntp.v17i3.5444

[27] K. Mishra *et al.*, "Predicting Diabetes with Federated Learning: A Digital Twin and Medical Fog-based IoT Framework," *ACM Trans. Internet Technol.*, Feb. 2025, doi: 10.1145/3711859.