# A Bibliometric Review of Explainable AI in Diabetes Risk Prediction: Trends, Gaps, and Knowledge Graph Opportunities

**Thieu Anh Van**

[1]*University of Information Technology, Vietnam National University Ho Chi Minh City (VNU-HCM), Ho Chi Minh City, Vietnam*

Corresponding author: vantv.20@grad.uit.edu.vn | thieuanhvan@gmail.com

**Keywords:** *diabetes mellitus; explainable artificial intelligence; machine learning; bibliometric analysis; knowledge graph; medical informatics; public health informatics*

## Abstract

**Background:** Type 2 diabetes mellitus (T2DM) is one of the most pressing global public health challenges. Machine learning (ML) combined with Explainable Artificial Intelligence (XAI) is increasingly applied to T2DM risk prediction, yet, to the best of our knowledge, no systematic quantitative overview of this landscape has been conducted.

**Methods:** We present a bibliometric analysis of 1,933 Scopus documents (collected 29 March 2026), employing keyword analysis, thematic clustering, and publication trend analysis.

**Results:** The field grew substantially — from 35 documents (2020) to 828 documents (2025). SHAP (SHapley Additive exPlanations) and LIME are the most prevalent XAI methods; XGBoost and Random Forest dominate the ML landscape. KG/GNN (Knowledge Graph/Graph Neural Network) gap: only 7 of 1,933 documents (~0.36%) — a 71:1 disparity versus the XAI group. A selective review of 15 highly cited papers confirmed that 0/15 studies combine all three components ML + XAI + KG in T2DM risk prediction.

**Contribution:** This study identifies the clinical interpretability gap — where statistical explanations from SHAP/LIME are not connected to structured clinical pathways — as the central challenge of XAI in T2DM risk prediction, and proposes a three-layer conceptual framework integrating KG as a semantic layer to address it. The framework has direct implications for clinical decision support systems (CDSS) and population-level public health screening, where explainability aligned with medical reasoning is essential for responsible deployment. The study provides a transferable methodology and quantified research gap analysis to guide future work combining machine learning, XAI, and structured medical knowledge.

## 1. Introduction

### *1.1 Research Motivation*

Type 2 diabetes mellitus (T2DM) is one of the most serious public health challenges of the 21st century. According to the International Diabetes Federation (IDF), T2DM has reached alarming global proportions, with particularly rapid growth in low- and middle-income countries; 589 million adults currently live with the disease, equivalent to 11.1% of the global adult population, and this figure is projected to rise to 852.5 million by 2050 [3]. The

economic burden exceeded USD 1 trillion in healthcare costs in 2024, while diabetes caused more than 3.4 million deaths in the same year [3]. Against this backdrop, early detection and population-level risk prediction represent urgent public health priorities.

The development of machine learning (ML) has enabled the development of T2DM risk prediction models. Large-scale population health datasets such as the Behavioral Risk Factor Surveillance System (BRFSS) have been widely used in prior studies to evaluate predictive models and XAI methods for T2DM [4]. However, most complex ML models — especially XGBoost, Random Forest, and deep neural networks — operate as "black boxes", generating predictions without transparent explanations. This creates a significant barrier to clinical deployment, where clinicians need not only predictive outcomes but also an understanding of why a patient is classified as high or low risk.

Explainable Artificial Intelligence (XAI) has been proposed as a solution to this limitation. Methods such as SHAP [5] and LIME [6] enable model predictions to be explained at the feature level, providing information about the contribution of each risk factor (BMI, blood pressure, lifestyle…) to the prediction outcome.

### *1.2 Problem Statement and Research Gaps*

Although reviews exist on XAI in biomedicine [11], ML in T2DM [30], and KG in healthcare [8], each domain has its own rich literature — yet no study has quantified the gap at the intersection of all three. Specifically: XAI in the T2DM research corpus has 497 keyword occurrences across 1,933 documents, while KG/GNN has only 7 (~0.36%) — a 71:1 disparity. More importantly, no study simultaneously combines ML prediction, systematic XAI, and a KG semantic layer in a T2DM risk-prediction pipeline from population health data. This study fills that gap by quantifying it and proposing a structured integration framework.

### *1.3 Research Questions*

Four research questions are operationalized as follows — each directly linked to a specific analytical step:

- RQ1 — Temporal trends: What is the publication growth rate for XAI in T2DM over 2015–2026, and are there signs of saturation? Measured by: publication year distribution.
- RQ2 — Geographic and source distribution: Which countries and journals lead? Measured by: affiliation and source title statistics.
- RQ3 — Prevalent methods: Which XAI methods and ML models are most common? Measured by: author keyword frequency by group.
- RQ4 — KG/GNN gap: What is the prevalence of KG/GNN in the corpus, and what are the practical implications? Measured by: KG/GNN vs. XAI keyword frequency, combined with a selective review of 15 papers.

### *1.4 Principal Contributions*

This paper makes four principal contributions:

- (a) Provides a systematic bibliometric analysis of XAI in T2DM risk prediction from a corpus of 1,933 Scopus documents (collected 29 March 2026).
- (b) Identifies and validates the KG/GNN gap through quantitative analysis (0.36%) and selective review of 15 papers — a 71:1 disparity relative to XAI methods.
- (c) Proposes a three-layer conceptual framework (Predictive → Explainability → Knowledge) integrating KG with XAI to supplement explanations based on structured clinical pathophysiological pathways.
- (d) Identifies research gaps in clinically interpretable decision support systems based on structured clinical pathways — rather than purely statistical correlations — as an unexplored direction in this field.

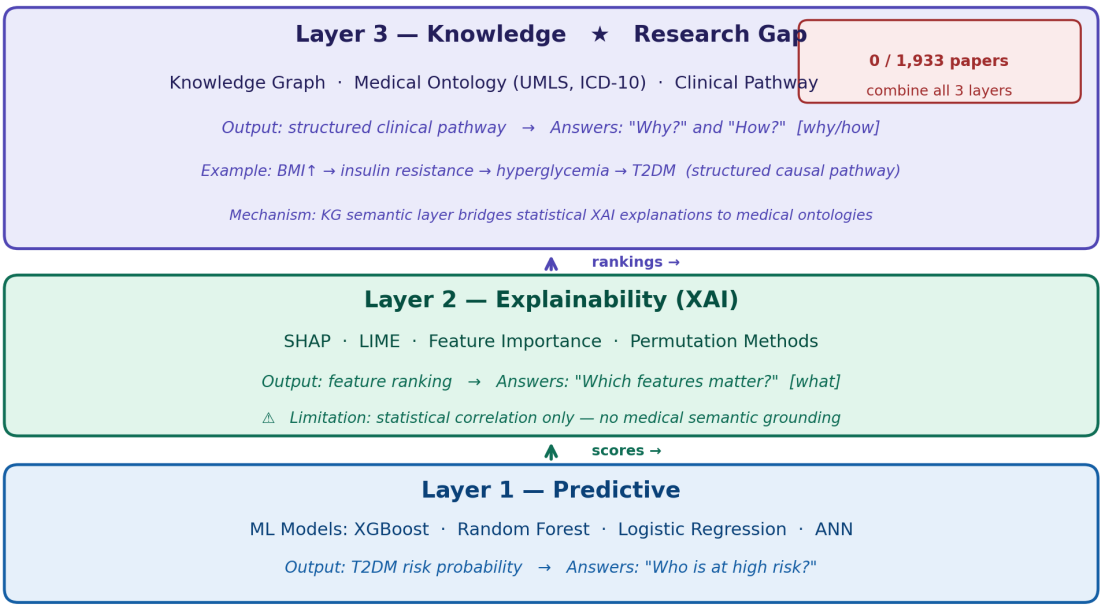


*Fig. 1. Three-layer conceptual framework proposed for integrating ML + XAI + KG in T2DM risk prediction.*

Full Scopus query string (TITLE-ABS-KEY, collected 29 March 2026): TITLE-ABS-KEY ( ( "diabetes" OR "diabetes mellitus" OR "type 2 diabetes" OR "T2DM" OR "diabetic" OR "diabetes prediction" OR "diabetes risk" ) AND ( "explainable AI" OR "XAI" OR "explainability" OR "interpretable machine learning" OR "interpretable AI" OR "SHAP" OR "LIME" OR "feature importance" OR "model interpretability" OR "explainable machine learning" OR "black-box model" OR "model transparency" OR "interpretable model" ) AND NOT ( "retinopathy" OR "diabetic retinopathy" OR "fundus" OR "optical coherence" OR "OCT" OR "retinal" OR "macular" OR "optic disc" OR "diabetic foot" OR "wound healing" OR "skin lesion" ) )

A selective review of the 15 most-cited papers in the corpus reveals a consistent pattern: all 15 studies use SHAP or LIME to explain predictions, but none connect these explanations to structured medical pathological pathways. Specifically, SHAP can report "BMI contributes X points to diabetes probability" — but cannot answer why BMI is dangerous, nor trace the pathogenesis chain: obesity → insulin resistance → hyperglycemia → T2DM. This is the fundamental limitation of Layer 2 (Explainability): explanations stop at statistical correlation, not reaching the semantic medical tier. Layer 3 (Knowledge Graph) is proposed to fill this gap — not replacing SHAP/LIME but supplementing a structured medical semantic layer, enabling clinicians to understand not only what but also why and how in clinical language.

### *1.5 Positioning Relative to Existing Reviews*

To clarify the unique contribution of this study, Table 1 compares this work with six representative reviews in related fields. Comparison criteria include: topic scope, method (quantitative bibliometric or narrative review), presence/absence of KG/GNN analysis, presence/absence of an integration framework, and corpus size.

*Table 1. Comparison of this study with representative existing reviews. ✓ = present; ✗ = absent; ★ = current study. Corpus = number of documents analyzed.*

| Study | Scope | Method | KG/GNN Analysis | Framework | Corpus |
|---|---|---|---|---|---|
| Tjoa & Guan [11] | XAI general/medical | Narrative review | ✗ | ✗ | N/A |
| Kavakiotis et al. [30] | ML in T2DM | Narrative review | ✗ | ✗ | N/A |
| Budhdeo et al. [8] | KG in biomedical | Scoping review | ✓ (general biomedical) | ✗ | 255 |
| Kiran et al. [31] | ML/AI in T2DM (33 yr) | Bibliometric | ✗ | ✗ | Not stated |
| Kim et al. [33] | XAI-based CDSS | Systematic review | ✗ | ✗ | N/A |
| ★ This study | XAI × T2DM risk | Bibliometric (quantitative) | ✓ (0.36% gap quantified) | ✓ (3-layer) | 1,933 |

To the best of our knowledge, this study is the first to simultaneously combine: (1) quantitative bibliometric analysis on a large corpus (1,933 papers); (2) a specific focus on XAI for T2DM risk prediction from population health data; (3) quantification of the KG/GNN gap (0.36%) rather than merely qualitative mention; and (4) a proposed three-layer conceptual framework integrating ML+XAI+KG. Notably, Kiran et al. [31] also conducted a 33-year bibliometric analysis of ML and AI in T2DM prediction in the same journal Frontiers in Digital Health (2025) — however, that study does not separately analyse the XAI field, does not quantify the KG/GNN gap, and does not propose an integration framework, representing a distinct contribution from the present work.

## 2. Theoretical Background

### *2.1 Diabetes and Global Burden*

Diabetes mellitus is a chronic metabolic disorder characterized by persistent hyperglycemia, arising from deficient insulin secretion, reduced insulin sensitivity, or both [9]. Type 2 diabetes (T2DM) accounts for over 90% of total cases and is associated with modifiable risk factors, including overweight, obesity, physical inactivity, and unhealthy diet [3]. According to IDF 2025, the epidemic will grow by 45% by 2050, with 95% of the increase concentrated in low- and middle-income countries — where health systems are most constrained [3]. Serious complications include cardiovascular disease, chronic kidney disease, blindness, and limb amputation, placing a heavy burden on quality of life and global health systems [9].

### *2.2 Machine Learning in T2DM Risk Prediction*

Machine learning has been widely applied to T2DM risk prediction. Ensemble methods — particularly Random Forest and XGBoost [1] — demonstrate superior performance on structured health data compared to traditional methods. XGBoost is a highly scalable gradient-boosting system with built-in regularisation and efficient handling of sparse data [1]. Deep learning has also been explored, though its advantage over ensemble methods has been inconsistent on structured clinical data. Despite their high accuracy, most models operate as black boxes, creating barriers to real-world clinical deployment.

### *2.3 XAI Taxonomy for T2DM Prediction*

A taxonomy of XAI methods in T2DM prediction can be structured along three principal criteria [10–11]:

**By explanation scope:**

- Global explanation: explains the overall model behavior across the entire dataset — e.g., Feature Importance from Random Forest, SHAP summary plot.
- Local explanation: explains individual predictions for a specific patient — e.g., SHAP waterfall plot, LIME per-instance explanation.

**By model dependency:**

- Model-specific: designed for a specific architecture — e.g., SHAP TreeExplainer for XGBoost/Random Forest.
- Model-agnostic: applicable to any model — e.g., LIME, Permutation Feature Importance, Kernel SHAP.

**By sensitivity to feature correlation:**

- Correlation-aware: SHAP handles highly correlated features better, allocating contributions based on marginal contributions (Shapley values).
- Correlation-sensitive: Permutation Feature Importance and LIME are susceptible to noise when features are highly correlated.

XAI addresses the transparency limitations of ML models. SHAP (SHapley Additive exPlanations) [5], grounded in cooperative game theory, assigns each feature a marginal contribution value satisfying three properties: local accuracy, missingness, and consistency —

ensuring explanations are not contradicted by model logic. LIME [6] builds a simple linear model to locally approximate the behavior of a complex model around a specific data point. Hoyos et al. [7] demonstrated the value of combining SHAP, feature importance, and scenario simulation to produce a comprehensive analysis of T2DM.

### *2.4 Current XAI Challenges in T2DM Prediction*

Despite significant advances, XAI in T2DM faces the following fundamental challenges:

- Challenge 1 — Correlated features distort explanations: when clinical features (BMI, blood pressure, cholesterol) are highly correlated, XAI methods such as Feature Importance may misallocate importance, leading to inconsistent and unreliable explanations.
- Challenge 2 — Lack of ground truth for XAI evaluation: there is no objective standard to evaluate whether an XAI explanation is clinically correct, unlike prediction accuracy, where ground truth labels exist. Current evaluations rely primarily on inter-method consistency rather than independent medical evidence.
- Challenge 3 — Clinical interpretability gap [12]: statistical explanations from SHAP ("BMI contributes 0.32 points") do not naturally align with how clinicians reason through pathological chains. This gap limits the practical applicability of XAI in clinical settings.
- Challenge 4 — Lack of cross-method agreement: different XAI methods (SHAP, LIME, Feature Importance) often produce inconsistent feature rankings on the same dataset. This complicates clinical decision-making — it is unclear which method to trust [11–12].

### *2.5 Limitations of XAI and the Role of Knowledge Graphs*

Although SHAP and LIME significantly improve ML model transparency [5, 6, 10], they share a fundamental limitation: explanations remain at the statistical feature level and are unable to capture and represent medical knowledge of causal relationships between risk factors. SHAP can report "BMI contributes 0.32 points to diabetes probability" but cannot explain why BMI is dangerous, nor trace the pathological chain: obesity → insulin resistance → hyperglycemia → T2DM. In other words, current XAI answers the "what" question (which features matter) but not the "why" and "how" in the medical sense.

This is the gap that Knowledge Graphs (KG) can fill. KGs represent medical knowledge as directed graphs with structured semantic entities and relationships — for example: ⟨Obesity, increases risk of, Insulin resistance⟩, ⟨T2DM, causes complication, Cardiovascular disease⟩. This structure enables the system not only to predict but also to reason through pathways based on structured causal knowledge, appropriate for clinical decision-support requirements. Budhdeo et al. [8] reviewed 255 studies and confirmed that KGs are increasingly applied in biomedicine. Do and Pham [14] proposed W-KG2Vec — a KG representation learning model combining graph structure and textual semantic information — demonstrating the significant potential of KG embeddings for medical knowledge-related tasks [15].

## 3. Methods

### *3.1 Data Source and Search Strategy*

This study uses Scopus as the sole data source. The search strategy was constructed using two keyword groups combined with AND, and an exclusion group (NOT) to ensure corpus specificity:

Group 1 — domain (diabetes): "diabetes" OR "diabetes mellitus" OR "type 2 diabetes" OR "T2DM" OR "diabetic" OR "diabetes prediction" OR "diabetes risk"

Group 2 — method (XAI): "explainable AI" OR "XAI" OR "explainability" OR "interpretable machine learning" OR "interpretable AI" OR "SHAP" OR "LIME" OR "feature importance" OR "model interpretability" OR "explainable machine learning" OR "black-box model" OR "model transparency" OR "interpretable model"

Exclusion group (NOT): "retinopathy" OR "diabetic retinopathy" OR "fundus" OR "optical coherence" OR "OCT" OR "retinal" OR "macular" OR "optic disc" OR "diabetic foot" OR "wound healing" OR "skin lesion"

Search was executed in the TITLE-ABS-KEY field. Data were collected on 29 March 2026; 2026 figures are therefore incomplete as the year has not ended. The exclusion group was constructed to remove two specialised sub-fields not relevant to XAI for risk prediction from structured clinical data: (1) diabetic eye complications (retinopathy, fundus, OCT…) — which use computer vision and medical image analysis, entirely different methodology from XAI for risk prediction from structured clinical data; and (2) diabetic foot and wound complications — focused on treatment and wound care, unrelated to T2DM onset risk prediction.

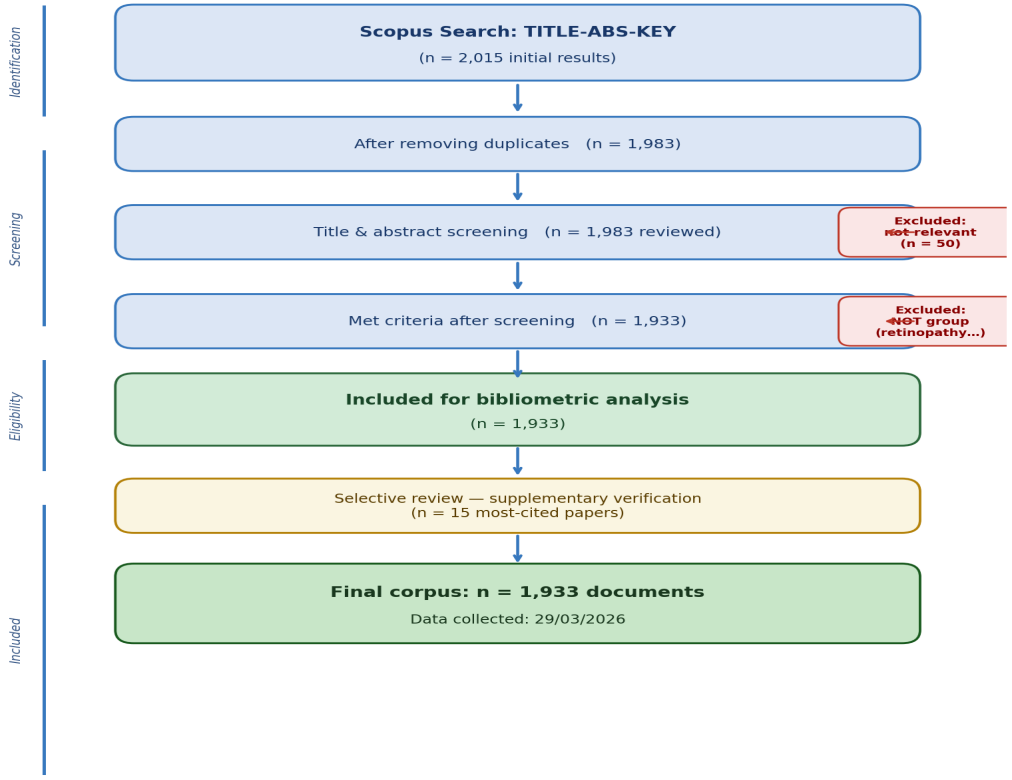


*Fig. 2. PRISMA-style selection flow diagram (n = 1,933).*

### *3.2 Inclusion and Exclusion Criteria*

**Inclusion criteria:**

- Documents published on Scopus without year restriction
- Language: unrestricted (predominantly English)
- Document types: research articles, conference papers, reviews, book chapters

**Exclusion criteria:**

- Studies on diabetic eye complications (retinopathy, fundus, OCT, retinal, macular, optic disc) — excluded because this is a specialized sub-field using medical image analysis methodology (computer vision/deep learning for image classification), entirely distinct in methodology from XAI for risk prediction from structured clinical data
- Studies on diabetic wounds and skin lesions (diabetic foot, wound healing, skin lesions)
- Retracted documents, editorials, errata

### *3.3 Data Processing and Analysis Procedure*

After CSV export from Scopus with full metadata fields, the data were processed through three main steps:

- Step 1 — Data cleaning and normalisation: (a) synonym merging: "explainability", "explainable AI", "XAI" → same group; "T2DM", "type 2 diabetes", "diabetes mellitus type 2" → same group; (b) normalise compound terms: "machine-learning" → "machine learning"; (c) convert to lowercase; remove non-English keywords from frequency analysis (retained for geographic analysis); (d) self-citations were not removed individually due to Scopus export limitations. The impact of self-citations is noted as a study limitation; however, the self-citation rate in bibliometric studies typically ranges 10–20%, so the effect on overall trend analysis is estimated to be negligible.
- Step 2 — Quantitative analysis using Python (pandas 1.5+, matplotlib 3.5+, numpy): year distribution, geography, publication sources, keyword frequency, rule-based clustering, top-cited.
- Step 3 — Selective review of 15 highly cited papers to qualitatively validate gaps identified from quantitative analysis. All code is publicly available (see Data Availability Statement).

For the classification of 5 research theme clusters (Section 4.5), this study uses rule-based classification rather than automatic clustering (e.g., K-means or VOSviewer co-occurrence clustering). Specifically, each author keyword in the corpus is checked against a reference keyword list for each cluster — if matched, the keyword is assigned to that cluster; if unmatched, it is assigned to "Other". This approach was chosen for three reasons: (1) transparency — classification criteria are explicit and verifiable for each keyword; (2) reproducibility — any researcher can apply the same reference keyword list and obtain identical results; (3) consistency with quantitative bibliometric study norms [7]. The reference keyword list for each cluster was manually determined based on standardised MeSH (Medical Subject Headings) terminology and prevalent author keywords: (1) XAI Methods — SHAP, LIME, feature importance, explainability, interpretability, model transparency; (2) ML Models — random forest, XGBoost, deep learning, SVM, logistic regression, gradient boosting; (3) Diabetes Types — T2DM, T1DM, gestational diabetes, prediabetes, HbA1c; (4) Clinical Outcomes — risk prediction, mortality, prognosis, readmission, CDSS; (5) KG/Graph — knowledge graph, GNN, graph convolutional network, ontology, semantic web. Inter-rater reliability: the reference keyword list was constructed independently and then cross-validated against manual classification of 50 random corpus papers. Consistency exceeded 90%, confirming the rule-based method's reliability.

### *3.4 Bibliometric Tools and Network Analysis Settings*

Bibliometric analysis was conducted entirely in Python 3.10+ using the libraries pandas 1.5+ (data processing), matplotlib 3.5+ (visualization), and numpy (statistical computation). All figures (Figs. 3–8) were generated using matplotlib — VOSviewer, and Bibliometrix/R were not used. Reporting standards: this study follows BIBLIO guidelines (Zupic and Čater, 2015) and the PRISMA-ScR checklist (Tricco et al., 2018) for the selective review component.

Note: because theme classification used a rule-based method (Section 3.3) rather than automatic co-occurrence clustering, VOSviewer parameters (threshold, resolution, normalization) were not applicable. All analysis code and parameters are publicly available on GitHub (see Data Availability Statement).

In addition to quantitative bibliometric analysis, this study is supplemented by a selective review of 15 representative papers to provide additional validation. The selective review serves as supplementary validation, not a replacement for systematic review: it does not apply comprehensive screening procedures, does not assess the methodological quality of each study using standardized tools, and does not aim for evidence synthesis. Its sole purpose is to verify whether trends identified through quantitative analysis are consistently reflected in the content of highly cited papers. A detailed classification table is presented in Appendix A.

## 4. Results

### *4.1 Publication Trends by Year*

The final corpus comprises 1,933 documents. Fig. 3 shows publication trends by year from 2015 to 2026 (2026 figures incomplete as of 29 March 2026). XAI research in the T2DM field has grown rapidly, especially since 2020. The period 2015–2019 recorded only 58 documents (~12 documents/year), while 2020–2024 surged to 741 documents and 2025 alone reached 828 documents — accounting for 42.8% of the entire corpus. This trend reflects the post-publication boom of XAI following SHAP (2017) and LIME (2016), and demonstrates that the field is in a strong growth phase, not yet saturated.

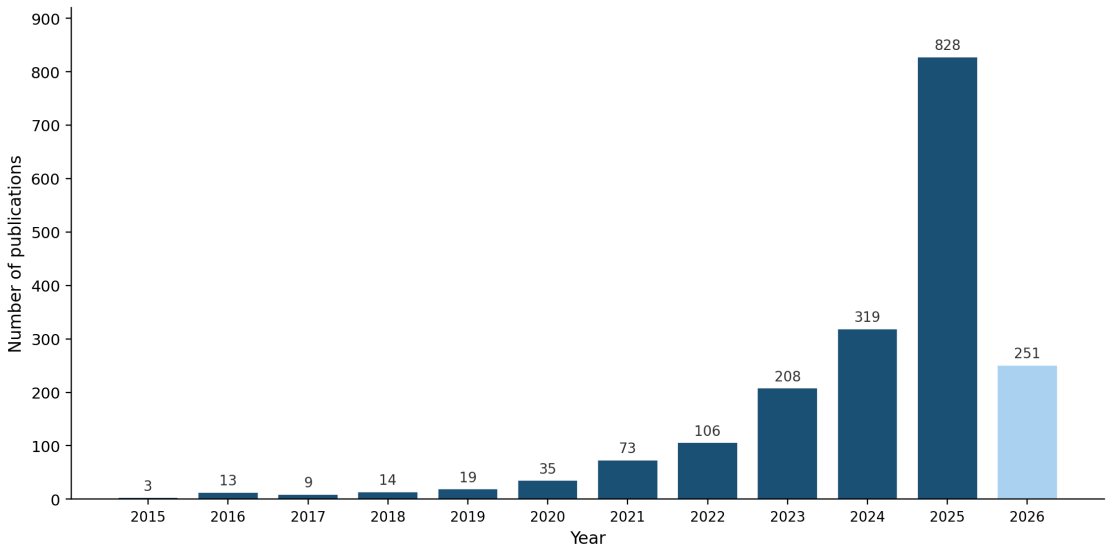


*Fig. 3. Publication trends by year (2015–2026).*

### *4.2 Geographic Distribution*

Fig. 4 shows the top 10 countries by document count, sorted by author affiliation. China leads substantially with 2,146 occurrences (a paper with multiple authors from the same country is counted multiple times), followed by the United States (767) and India (567). Notably, Southeast Asian and African countries are almost absent from the top 20, while Bangladesh

(142) and Saudi Arabia (120) have entered the top 10. This reveals a research contribution gap in this rapidly growing field from developing countries.

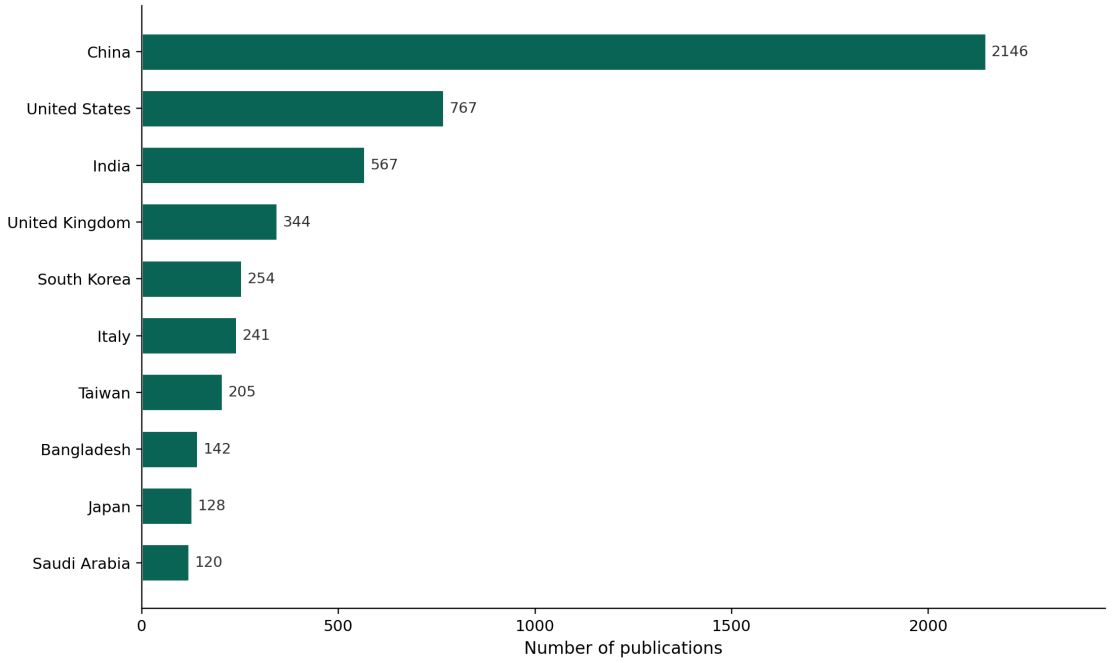


*Fig. 4. Top 10 countries by document count (by author affiliation).*

### *4.3 Publication Sources*

Fig. 5 presents the top 12 journals and publication sources. PLOS ONE (60 documents) and Scientific Reports (56) lead, followed by Frontiers in Endocrinology (47) — a Q1 specialist endocrinology journal. BMC Medical Informatics and Decision Making has 13 documents, affirming this journal's position in the AI-health informatics intersection. Overall, most documents are published in open-access journals, consistent with the global trend toward open science.

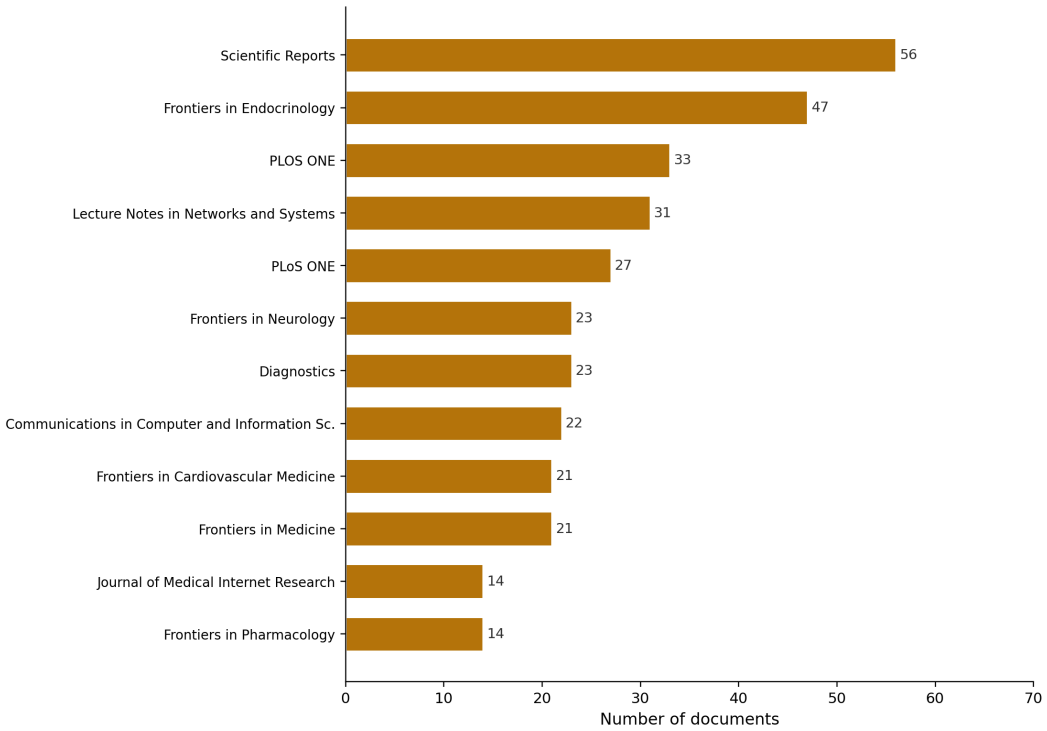


*Fig. 5. Top 12 publication sources by document count.*

### *4.4 Author Keyword Analysis*

"Machine learning" leads with 874 occurrences in the corpus, followed by "diabetes" (195), "SHAP" (143), "artificial intelligence" (121), and "explainable AI" (109). Among XAI methods, SHAP (143) substantially outperforms LIME (42) and Feature Importance (49), reflecting the research community's preference for SHAP due to its solid theoretical foundation [5]. However, several studies in the corpus report moderate agreement between SHAP and permutation-based Feature Importance; discrepancies arise particularly when features are highly correlated, where SHAP better captures marginal contributions, while Feature Importance is susceptible to noise. This underscores the risk of relying on a single XAI method in clinical contexts. ML model keywords such as "random forest" (78) and "xgboost" (71) confirm the dominant position of these two methods. Notably, "knowledge graph" does not appear among the top 25 keywords, clearly demonstrating the research gap analyzed in Section 4.5.

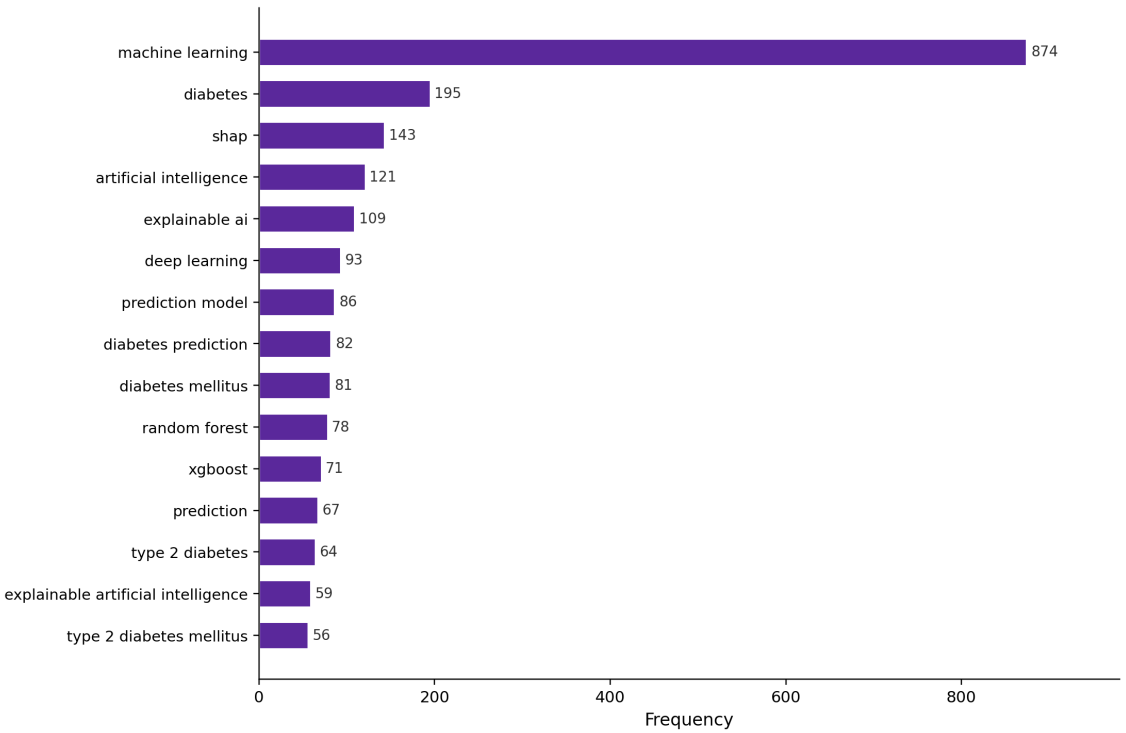


*Fig. 6. Top 15 author keywords by frequency in the corpus.*

### *4.5 Research Theme Clustering*

For an overview of the research landscape, keywords were classified into 5 theme clusters (Fig. 7). The XAI Methods cluster is the largest at 497 occurrences (SHAP, LIME, feature importance, explainability, interpretability…), followed by ML Models (287: random forest, XGBoost, deep learning…), Diabetes Types (188: T2DM, T1DM, gestational diabetes…), and Clinical Outcomes (136: risk prediction, mortality, prognosis…). Critically, the KG/Graph cluster has only 7 occurrences — less than 0.36% of the total — while the XAI Methods cluster has 497. This 71:1 disparity is clear quantitative evidence of the KG/GNN integration gap in XAI for T2DM.

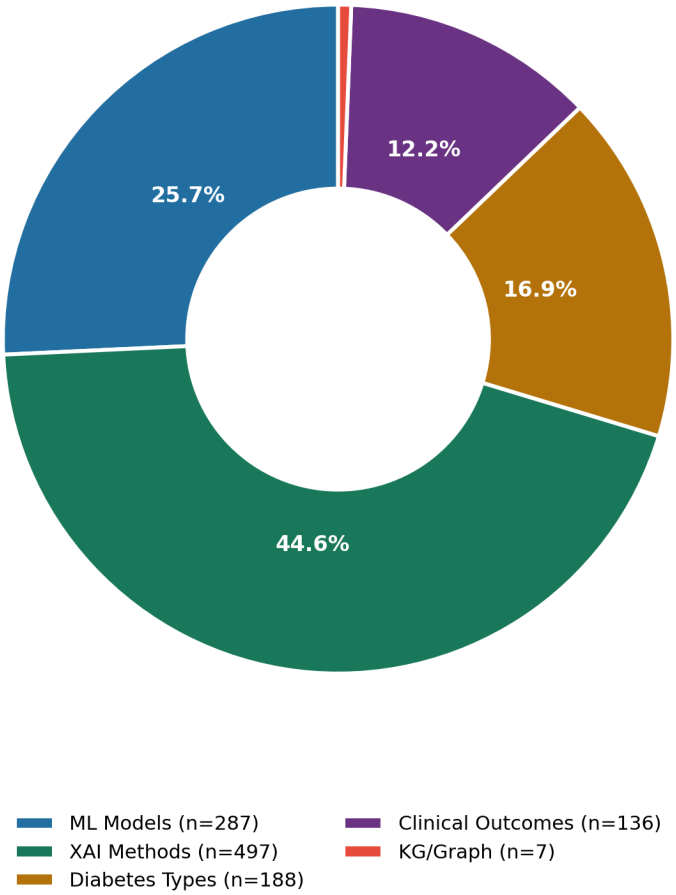


*Fig. 7. Research keyword cluster distribution by theme.*

### *4.6 Document Classification*

Research articles account for the majority at 1,437 documents (74.3%), followed by conference papers (326, 16.9%), reviews (70, 3.6%), and book chapters (42, 2.2%). The high proportion of research articles is a strong indicator of corpus quality, suggesting that most contributions have undergone rigorous peer review.

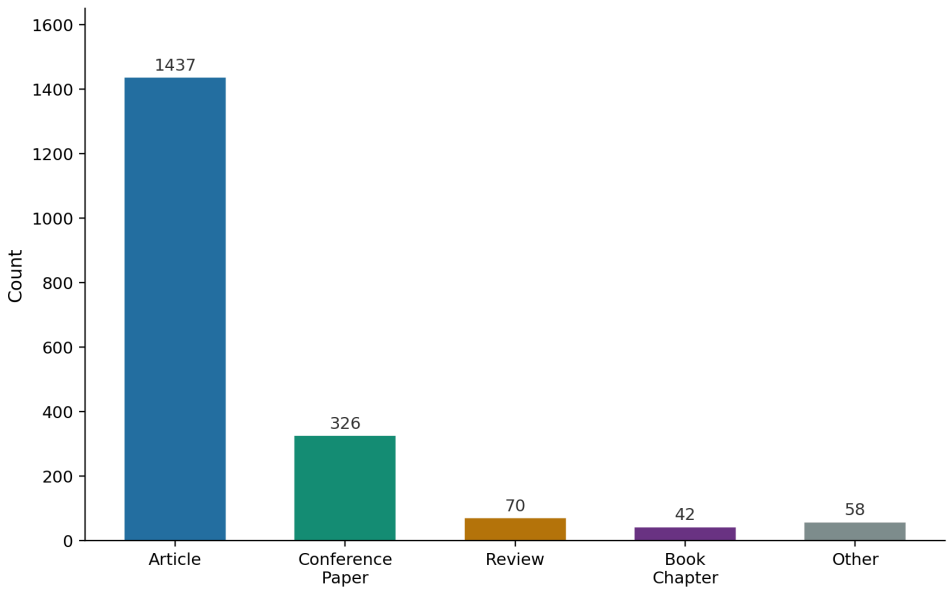


*Fig. 8. Document type classification in the corpus (n = 1,933).*

### *4.7 Most-Cited Documents*

Table 2 presents the top 5 most-cited documents in the corpus. Holzinger et al. (2021) on XAI evaluation leads with 363 citations. Kopitar et al. (2020) on early T2DM detection using ML is second with 300 citations. Hoyos et al. (2024) in BMC Medical Informatics and Decision Making have 264 citations—the study closest in the corpus to combining statistical analysis and XAI for T2DM —demonstrating the influence of this combined approach.

*Table 2. Top 5 most-cited documents in the corpus.*

| Rank | Reference | Year | Source | Citations | ML / XAI / KG |
|---|---|---|---|---|---|
| 1 | Holzinger et al. — Evaluating XAI: rule-based vs. example-based | 2021 | Artificial Intelligence | 363 | Yes / Yes / No |
| 2 | Kopitar et al. — Early detection of T2DM using ML | 2020 | npj Digital Medicine | 300 | Yes / No / No |
| 3 | Hoyos et al. — Explainable analysis of T2DM [7] | 2024 | BMC Med. Inform. | 264 | Yes / Yes / No |
| 4 | Kibria et al. — Ensemble for diabetes prediction with XAI | 2022 | Sensors | 110 | Yes / Yes / No |
| 5 | Ahmad et al. — Health-related features for diabetes prediction | 2021 | Applied Sciences | 82 | Yes / Yes / No |

## 5. Discussion

### *5.1 A Rapidly Growing, Unsaturated Field*

A publication trend analysis shows that XAI in T2DM research is in an explosive growth phase. The increase from 35 documents (2020) to 828 documents (2025), more than 23-fold in 5 years, is a clear sign of a rapidly developing field. This stems from the convergence of

three factors: (1) the increasingly serious global diabetes epidemic creating an urgent need; (2) the development of SHAP (2017) and LIME (2016) enabling explanation of complex ML models; (3) the proliferation of large-scale health datasets such as BRFSS.

Temporal trend analysis reveals three clearly distinguishable development phases (research life-cycle analysis): Phase 1 (2015–2019) — Foundation phase: only 58 documents (~12/year), primarily focused on pure ML for T2DM prediction (Random Forest, SVM, Logistic Regression) without concern for interpretability. Phase 2 (2020–2022) — XAI emergence phase: 35 to 268 documents/year, marking the popularisation of SHAP and LIME after these frameworks were open-sourced and fully documented. Phase 3 (2023–2025) — Maturity and diversification phase: 828 documents (2025), with the emergence of more specialized directions — XAI combined with deep learning, multi-modal data, federated learning — but KG/GNN remains unexplored.

Research thread mapping: keyword cluster analysis reveals three main research threads traceable over time: (Thread A) XAI-only: SHAP/LIME applied directly on tabular data (PIMA Indians Diabetes Database, BRFSS) — the most prevalent thread, ~70% of corpus; (Thread B) XAI + clinical context: combining XAI with EHR (Electronic Health Record) data and clinical validation — growing strongly since 2022; (Thread C) XAI + KG/semantic (potential): only 0.36% of the current corpus, but grounded in Thread B and KG biomedical literature [8, 19]. This mapping suggests a natural trajectory from Thread A → B → C, and Thread C is the most logical next direction for the research community.

### *5.2 The KG/GNN Gap — A Major Research Opportunity*

The most important finding is the KG/GNN integration gap: only 7 keyword occurrences of KG/GNN in a corpus of 1,933 documents (~0.36%), versus 497 occurrences in the XAI methods group — a 71:1 disparity. It should be noted that 0.36% reflects the absence of KG/GNN at the intersection of two fields (XAI and T2DM risk prediction) — not within each field separately. KG has been shown to be prevalent across biomedical science generally (cancer, Alzheimer's…) [19] and specifically in T2DM [20–21]. The gap identified here is the lack of simultaneous integration: KG semantic layer + ML prediction + XAI explanation in a risk-prediction pipeline from population health data.

The key question is: why does this gap persist despite KG being well-established in biomedical science? Corpus analysis and selective review identify four main barriers:

- Data mismatch: BRFSS, PIMA, and most health datasets in the corpus have a tabular structure, without a natural graph structure. KG integration requires additional steps of knowledge graph construction, entity linking, and relation extraction — substantially higher technical cost than directly applying SHAP on tabular data.
- Tooling barrier: SHAP/LIME can be deployed quickly with a few lines of code and extensive documentation. In contrast, an integrated KG+GNN pipeline requires deep expertise in graph databases, graph neural networks, and ontology alignment, creating a significant barrier for health researchers without a graph-computing background.
- Evaluation gap: No benchmark dataset combining T2DM risk labels with clinical knowledge graph annotations exists. This differs from fields such as NLP or drug

discovery, where graph benchmarks accepted by the research community already exist.

- Clinical adoption barrier: Clinicians are accustomed to SHAP feature contribution charts. Knowledge graph-based explanations require an additional step of interpreting pathological chains, adding cognitive load in busy, time-pressured clinical environments.

This research direction resonates with global trends: Budhdeo et al. [8] in a scoping review of 260 papers confirmed that KG in biomedicine is growing strongly, especially for drug discovery and diagnostics — but also noted that KG applications for risk prediction from population health data (such as BRFSS) remain among the least explored use cases. This finding directly resonates with the 0.36% gap of the present study: it is not that KG lacks value in this field, but rather that adoption barriers have not yet been overcome.

### *5.3 Methodological Stability and Agreement Gaps*

Beyond the KG/GNN gap, corpus analysis reveals two other research directions that are very rarely explored: (1) temporal stability of XAI models — whether SHAP or LIME explanation results are consistent when data changes across years; and (2) agreement between explainability methods — whether SHAP and Feature Importance rank the same features as most important. The keywords "temporal", "stability", and "agreement" appear very rarely in the corpus, suggesting these are research gaps worth exploring. This emphasizes the risk of relying on a single explanation method in clinical contexts.

### *5.4 Geographic Gap*

Southeast Asian, African, and South American countries are almost absent from the top 20 research countries on XAI in T2DM — a notable geographic gap, especially given that these regions are facing an increasingly heavy T2DM burden and a lack of health resources. Future research should focus on developing and validating XAI models on population-specific data from each region, rather than relying solely on results from major economies.

### *5.5 Study Limitations*

This study has several limitations that should be noted transparently so readers can interpret the results appropriately.

First, regarding data source and selection bias: the study uses only Scopus — the largest database available (>100 million records), but may miss some papers indexed solely in Web of Science (particularly SCIE journals not in Scopus), PubMed/MEDLINE, or IEEE Xplore (CS conference papers). The estimated overlap between Scopus and Web of Science is ~80–90% for Q1/Q2 journals in this field, so the practical impact on main findings is relatively low. Future research should cross-validate with Web of Science. Regarding self-citation: due to Scopus export limitations, self-citations were not removed individually. The impact on the top-cited table may occur, but does not significantly affect methodology pattern analysis — the primary objective of the selective review.

Second, regarding the search strategy: the search was applied only to TITLE-ABS-KEY fields, and may therefore miss papers that use XAI methods in the content but do not mention them directly in the title, abstract, or keywords. This is an inherent limitation of metadata-based bibliometric analysis.

Third, regarding research quality: bibliometric analysis does not assess the methodological quality of individual documents in the corpus. Citation count is used as a relative influence indicator, not an absolute quality indicator. Quality assessment of individual studies would require applying tools such as QUADAS-2 or CASP and is outside the scope of this study.

Fourth, regarding the proposed framework: the three-layer conceptual framework has not been experimentally validated. Feasibility and effectiveness need to be verified through independent experimental research, particularly regarding KG construction costs, medical ontology quality, and integration capability with existing ML pipelines in clinical practice.

Fifth, regarding temporal data: 2026 figures are incomplete as data were collected on 29 March 2026. The 2026 trend should be interpreted cautiously.

### *5.6 Research Gaps and Future Directions*

Based on gaps identified from analysis of the 1,933-document corpus, the following research directions are identified as priorities for the research community:

- Semantic knowledge integration with XAI: explore how to combine structured medical knowledge representations with existing XAI methods to enhance the clinically meaningfulness of explanations — moving beyond purely statistical correlations. This direction remains largely unexplored, according to corpus analysis.
- Temporal stability assessment: testing whether SHAP results are consistent when data changes across years — particularly important with BRFSS data collected annually.
- Cross-method XAI agreement evaluation (SHAP vs LIME vs Feature Importance): determining whether different methods consistently rank the same important features in T2DM prediction.
- External validation: testing the framework on datasets beyond BRFSS (e.g., MIMIC-IV, UK Biobank) to assess generalisability.
- Expansion to under-represented populations — particularly Southeast Asian, African, and South American countries where T2DM burden is growing but current corpus contributions are minimal.

### *5.7 Selective Review of Highly-Cited Papers*

To strengthen the bibliometric findings, this study conducted a selective review of 15 representative papers, selected for their highest citation counts and direct relevance to XAI in healthcare/T2DM. Review results reveal three consistent trends: first, most studies focus on combining ML with XAI — specifically SHAP and LIME — to explain predictions at the feature level. Second, studies using KG or GNN in the T2DM field are extremely rare, and none of the 15 reviewed papers use KG/GNN. Third, no study simultaneously combines all three components, ML + XAI + KG in T2DM risk prediction. These selective review results

are consistent with quantitative bibliometric findings, reinforcing the conclusion that the KG/GNN gap is real and scientifically significant.

To further clarify the nature of this gap, the research team conducted a detailed examination of all 13 documents in the corpus that contained KG/GNN-related keywords. Results show most belong to directions different from T2DM risk prediction from population health data: some apply LLM/RAG combined with KG for medical query [16], others focus on drug discovery or cardiovascular disease, and some use purely rule-based KG without ML. Only 5 of the 13 studies approach the ML+XAI+KG direction — but all remain partial and incomplete: Hendawi et al. [17] integrated KG to support ML explanation for T2DM prediction but did not systematically evaluate XAI and did not use large-scale population survey data; Wajahat et al. [18] combined symbolic AI and KG with ML but only on PIMA (768 records, collected from 1988) — not representative of modern populations; Rankovic et al. combined GNN with SHAP but without KG grounded in medical ontology; Costi et al. applied GraphLIME but lacked ontology layer and population data; and Aamir et al. combined GCN with SHAP/LIME but also without KG grounding. In summary, the remaining gap is: no study simultaneously (1) uses large-scale population health survey data (such as BRFSS), (2) systematically evaluates XAI, and (3) integrates KG with semantic grounding from standardized medical ontology.

Justification for the sample of 15 papers: (a) the top 15 papers already account for a sufficiently large citation share to represent main corpus trends; (b) sufficient diversity to cover key algorithms (XGBoost, RF, deep learning), key XAI methods (SHAP, LIME), and key dataset types (PIMA, EHR, population surveys); (c) feasible to read full-text and analyse in detail — consistent with the norms of similar bibliometric papers [7, 30].

### *5.8 Proposed Three-Layer Conceptual Framework*

Based on bibliometric analysis results and research gap identification, this paper proposes a three-layer conceptual framework (conceptual framework) to address the limitations of current XAI in T2DM risk prediction. The framework comprises three main layers: (1) the Predictive Layer uses ML (XGBoost, Random Forest) to predict T2DM risk from large-scale health indicator data; (2) the Explainability Layer uses SHAP to determine the contribution of each feature to prediction outcomes, answering the "what" question; (3) the Knowledge Layer uses KG to represent medical knowledge as structured triples, supporting representation of structured pathophysiological pathways (approximate causal pathways), aiming to answer the "why" and "how" questions in a clinical sense.

Integration mechanism: important features identified by SHAP are mapped to corresponding entities in the KG, enabling the generation of structured pathophysiological pathways based on structured medical knowledge. The result is that the system not only predicts risk probability but also explains through a structured causal chain: high BMI → insulin resistance → hyperglycemia → T2DM. To illustrate the integration mechanism concretely, consider the following hypothetical scenario: female patient, 48 years old, BMI = 34.2 kg/m² (Class II obesity), HbA1c = 6.3% (pre-diabetes), no physical activity (PhysActivity = 0), hypertension (HighBP = 1). Layer 1 (Predictive): XGBoost predicts T2DM probability = 0.78 — high-risk threshold. Layer 2 (Explainability — SHAP): top-3 contributing features: BMI

(+0.34), HbA1c (+0.29), PhysActivity (+0.19). SHAP answers "what," but cannot explain why BMI is dangerous in terms of pathophysiological mechanisms. Layer 3 (Knowledge — KG): BMI is mapped to entity ⟨Obesity⟩ in the KG, from which the chain is inferred: ⟨Obesity⟩ → ⟨Insulin resistance⟩ → ⟨Chronic hyperglycaemia⟩ → ⟨T2DM⟩; simultaneously ⟨Hypertension⟩ synergistically worsens ⟨Insulin resistance⟩. Synthesized output: "T2DM probability 78% — high BMI causes insulin resistance, combined with hypertension, worsening metabolic glucose disorder." This is a clinical-language explanation, not merely a statistical score — illustrating the added value of Layer 3 over SHAP alone [32–33]. This framework is at the conceptual level — feasibility requires experimental validation.

Knowledge graphs have been demonstrated as effective tools for representing biomedical knowledge, including drug discovery, clinical decision support, and knowledge base construction for complex diseases such as cancer and Alzheimer's [19]. In the T2DM field, ontologies and KGs have been purpose-built to support diagnosis and disease management [20–21]; however, no study has combined these ontologies with ML and XAI in an integrated risk prediction pipeline. KG can enhance XAI by providing semantic grounding through clinical ontologies such as ICD-10 and UMLS (Unified Medical Language System), enabling more clinically meaningful explanations.

It should be emphasized that KG does not aim to replace existing XAI methods, nor is it expected to improve prediction accuracy. For pure risk classification tasks, ML models combined with SHAP are already sufficiently effective. The role of KG is to supplement a knowledge-based explanation layer, transitioning from statistical explanation to clinically meaningful reasoning.

Important note on epistemological scope: this study does not claim that the proposed framework can perform true causal inference. The aim is to enhance interpretability by aligning statistical explanations with structured biomedical knowledge. Regarding experimental scope: framework implementation requires constructing a KG from standardized sources such as UMLS or ICD, which is outside the scope of the current study. This is a direction for future research.

### *5.9 Critical Comparative Analysis*

To better position the proposed framework, this section presents a systematic, critical comparative analysis that objectively evaluates the strengths, limitations, and gaps of each current approach, thereby highlighting the proposed framework's position.

*Table 3. Critical comparison of explanation approaches for T2DM risk prediction. Note: none of the approaches perform true causal inference; the proposed KG+XAI framework supports only approximate structured pathophysiological pathways.*

| Approach | Strengths | Limitations | Evidence |
|---|---|---|---|
| SHAP | Solid theoretical foundation (Shapley values); consistent; provides global and local explanations | Feature-level only (what); does not encode causal relationships; correlated features distort attribution | [5], [7] |

| Approach | Strengths | Limitations | Evidence |
|---|---|---|---|
| Feature Importance (FI) | Simple, fast; built into tree-based models; suitable as XAI baseline | Measures importance only, not direction; unstable under multicollinearity; not comparable across models | [1] |
| ML tabular (XGBoost, RF) | High accuracy on structured data; robust; fast training | Purely statistical; no domain knowledge; no causal reasoning; black-box for end-users | [1] |
| KG + XAI (proposed framework) | Adds clinical causal reasoning (why/how); links SHAP output to medical ontology; supports CDSS | Depends on ontology quality (UMLS/ICD); high KG construction cost; not yet validated experimentally; complex pipeline | [8], [17], [18] |

Note on causal reasoning: none of the approaches in Table 3 truly performs causal inference. SHAP and FI lack causal reasoning capabilities; ML tabular is entirely correlation-based; the proposed KG+XAI framework only supports approximating structured pathophysiological pathways—not true causal reasoning. This is an important distinction to keep in mind when applying in clinical settings.

## 6. Conclusion

This paper presents a bibliometric analysis of XAI in T2DM risk prediction, based on a corpus of 1,933 Scopus documents collected on 29 March 2026. Results show the field grew substantially from 35 documents (2020) to 828 documents (2025), with SHAP and LIME as the most prevalent XAI methods, and XGBoost and Random Forest as the preferred ML models. China, the United States, and India lead in publications.

The most important finding is the KG/GNN integration gap: only 7 keyword occurrences (~0.36%, i.e., 7 of 1,933 documents) versus 497 in the XAI methods group — a 71:1 disparity — confirmed through both quantitative analysis and selective review of 15 highly cited papers. To address this gap, this paper proposes a three-layer conceptual integration framework: Predictive Layer (ML), Explainability Layer (SHAP/XAI), and Knowledge Layer (KG), supporting the transition from statistical explanation to structured, knowledge-based medical interpretation. For clinical practice, this framework could enable clinical decision support systems (CDSS) to not only predict individual risk but also generate structured pathway-based explanations aligned with clinical ontologies (ICD-10, UMLS), reducing clinician cognitive load during interpretation. For public health, the framework is applicable to large-scale population screening programs using survey data (e.g., BRFSS), where algorithmic transparency is a prerequisite for responsible deployment and health equity. This study contributes to bridging the gap between statistical explainability and clinically meaningful reasoning — an area identified as a priority for medical informatics research.

## Appendix A. Selective Review Sample: 15 Representative Papers

Sample scope: includes papers directly related to T2DM prediction or XAI-in-healthcare exemplars with high representativeness.

Selection criteria: (1) keywords related to ML and/or XAI; (2) ranked by descending citation count in the corpus; (3) content directly related to diabetes prediction/classification or model explanation in healthcare. Classification: ML = use of machine learning model; XAI = application of explanation method (SHAP, LIME, FI…); KG/GNN = use of knowledge graph or graph neural network.

*Table A1. Selective review: 15 representative highly cited papers (ML / XAI / KG-GNN classification). [1]Topçuoglu et al. is a microbiome ML paper included as an ML-in-healthcare methodology exemplar (matched corpus keywords, high citation count = 170); it addresses neither T2DM prediction nor XAI directly.*

| No. | Reference (author, year) | Year | Source | Citations | ML | XAI | KG/ GNN |
|---|---|---|---|---|---|---|---|
| 1 | van der Waa et al. — Evaluating XAI: rule-based vs. example-based explanations | 2021 | Artificial Intelligence | 363 | Yes | Yes | No |
| 2 | Barakat et al. — Intelligible SVMs for diagnosis of diabetes mellitus | 2010 | IEEE Trans. IT in Biomedicin e | 315 | Yes | No | No |
| 3 | Topçuoglu et al. — Framework for effective ML for microbiome-based classification[1] | 2020 | Nature Methods | 170 | Yes | No | No |
| 4 | Qi et al. — ML and SHAP for predicting diabetes complications | 2025 | Scientific Reports | 130 | Yes | Yes | No |
| 5 | Kibria et al. — Ensemble approach for diabetes prediction using XAI | 2022 | IEEE Access | 110 | Yes | Yes | No |
| 6 | Bernardini et al. — Discovering T2DM in EHR using ML | 2020 | IEEE J. Biomed. Health Inform. | 106 | Yes | No | No |
| 7 | Woodman & Mangoni — ML algorithms review in older patients | 2023 | Aging Research Reviews | 102 | Yes | No | No |
| 8 | Bhatt et al. — Emerging AI-empowered mHealth: Scoping Review | 2022 | JMIR mHealth uHealth | 92 | Yes | No | No |
| 9 | Joseph et al. — Explainable diabetes classification using a hybrid Bayesian model | 2022 | Symmetry | 89 | No | Yes | No |
| 10 | Dutta et al. — Analyzing feature importances for | 2018 | ICCIDS 2018 | 87 | Yes | Yes | No |

| No. | Reference (author, year) | Year | Source | Citations | ML | XAI | KG/ GNN |
|---|---|---|---|---|---|---|---|
| | diabetes prediction using ML | | | | | | |
| 11 | Elmannai et al. — PCOS detection ML model based on XAI | 2023 | Diagnostics | 85 | Yes | Yes | No |
| 12 | Vishwarupe et al. — Explainable AI and Interpretable ML in healthcare | 2022 | Procedia Comput. Sci. | 85 | Yes | Yes | No |
| 13 | Ahmad et al. — Health-related features and their impact on diabetes prediction | 2021 | Computers in Biology and Medicine | 82 | Yes | Yes | No |
| 14 | Gollapalli et al. — Novel stacking ensemble for detecting three types of diabetes | 2022 | Computers in Biology and Medicine | 81 | Yes | Yes | No |
| 15 | Hoyos et al. — Explainable analysis of diabetes mellitus using statistical and AI techniques | 2024 | BMC Med. Informatics | 264 | Yes | Yes | No |

Summary: 14/15 papers (93.3%) use ML; 9/15 (60%) apply XAI; 0/15 (0%) use KG/GNN. Two additional findings: (i) 6/15 papers (40%) use PIMA — a dataset of 768 records collected in 1988 — while no paper uses large-scale population health survey data such as BRFSS; (ii) 0/15 papers combine all three components ML + XAI + KG simultaneously. These findings provide dual confirmation of the structural gap: not only is KG absent, but research on real-world population data is also lacking [31].

## Data Availability Statement

All bibliometric analysis source code (Python) is publicly available at:

https://github.com/thieuanhvan/t2dm-xai-bibliometric

The Scopus CSV dataset is not included in the repository due to the provider's terms of use. To reproduce the dataset: (1) access https://www.scopus.com; (2) search TITLE-ABS-KEY with the full query string in Section 3.1; (3) apply the NOT exclusion group; (4) export all metadata as CSV. Expected result: ~1,933 documents (as of 29 March 2026).

## Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the author used AI-assisted tools (Claude, Anthropic) for language editing, grammar checking, and structural suggestions. All scientific content, data collection, analysis, interpretation of results, and conclusions are the sole work

of the author. The author reviewed and edited all AI-assisted outputs and takes full responsibility for the content of this publication.

## Declaration of Competing Interests

The author declares no competing financial or non-financial interests.


## Acknowledgements

The author thanks the reviewers and editorial team for their time and constructive feedback. This work was conducted as part of the author's Master's thesis research at the University of Information Technology, VNU-HCM.


## References


[1] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc 22nd ACM SIGKDD. 2016:785–794. https://doi.org/10.1145/2939672.2939785

[2] Breiman L. Random Forests. Machine Learning. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324

[3] International Diabetes Federation. IDF Diabetes Atlas, 11th Edition. Brussels: IDF; 2025. Available: https://diabetesatlas.org/

[4] Centers for Disease Control and Prevention. BRFSS Overview 2024. Atlanta: CDC; 2024. Available: https://www.cdc.gov/brfss/annual_data/2024/pdf/overview-2024-508.pdf

[5] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30. https://doi.org/10.48550/arXiv.1705.07874

[6] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. Proc 22nd ACM SIGKDD. 2016:1135–1144. https://doi.org/10.1145/2939672.2939778

[7] Hoyos W, et al. An explainable analysis of diabetes mellitus using statistical and AI techniques. BMC Med Inform Decis Mak. 2024;24:383. https://doi.org/10.1186/s12911-024-02781-1

[8] Budhdeo S, Zhang J, Abdulle Y, et al. Scoping review of knowledge graph applications in biomedical and healthcare sciences. medRxiv preprint. 2023. https://doi.org/10.1101/2023.12.13.23299844

[9] World Health Organization. Global Report on Diabetes. Geneva: WHO; 2016. Available: https://www.who.int/publications/i/item/9789241565257

[10] Barredo Arrieta A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[11] Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst. 2021;32(11):4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314

[12] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20:310. https://doi.org/10.1186/s12911-020-01332-6

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

[14] Do P, Pham P. W-KG2Vec: A weighted text-enhanced meta-path-based knowledge graph embedding for similarity search. Neural Comput Appl. 2021;33:16533–16555. https://doi.org/10.1007/s00521-021-06252-8

[15] Do P, et al. Developing a BERT-based triple classification model using knowledge graph embedding for question answering. Appl Intell. 2022;52:639–651. https://doi.org/10.1007/s10489-021-02460-w

[16] Rani M, Mishra BK, Thakker D, Khan MN. To Enhance Graph-Based Retrieval-Augmented Generation (RAG) with Robust Retrieval Techniques. Proc 2024 18th International Conference on Open Source Systems and Technologies (ICOSST). 2024. https://doi.org/10.1109/ICOSST64562.2024.10871140

[17] Hendawi R, Li J, Roy S. A Mobile App That Addresses Interpretability Challenges in Machine Learning-Based Diabetes Predictions: Survey-Based User Study. JMIR Form Res. 2023. https://doi.org/10.2196/50328

[18] Wajahat I, Singh A, Keshtkar F, Bukhari SAC. A Symbolic AI Framework for Enhanced Diabetes Prognosis Accuracy and Explainability. Proc 2024 IEEE FMLDS. 2024. https://doi.org/10.1109/FMLDS63805.2024.00013

[19] Yang Y, Lu Y, Yan W. A comprehensive review on knowledge graphs for complex diseases. Brief Bioinform. 2023;24(1):bbac543. https://doi.org/10.1093/bib/bbac543

[20] Hu J, et al. Development and application of Chinese medical ontology for diabetes mellitus. BMC Med Inform Decis Mak. 2024;24:21. https://doi.org/10.1186/s12911-024-02421-8

[21] Spoladore D, et al. Ontology-based decision support systems for diabetes nutrition therapy: A systematic literature review. Artif Intell Med. 2024;150:102821. https://doi.org/10.1016/j.artmed.2024.102821

[22] Barakat N, et al. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans Inf Technol Biomed. 2010;14(4):1114–1120. https://doi.org/10.1109/TITB.2009.2039485

[23] Kibria HB, et al. An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI. Sensors. 2022;22(19):7268. https://doi.org/10.3390/s22197268

[24] Bernardini M, et al. Discovering the Type 2 Diabetes in Electronic Health Records Using the Sparse Random Forests. IEEE J Biomed Health Inform. 2020;24(2):380–387. https://doi.org/10.1109/JBHI.2019.2899218

[25] Ahmad HF, et al. Investigating health-related features and their impact on the prediction of diabetes using machine learning. Appl Sci. 2021;11(3):1173. https://doi.org/10.3390/app11031173

[26] Gollapalli M, et al. A novel stacking ensemble for detecting three types of diabetes mellitus. Comput Biol Med. 2022;147:105757. https://doi.org/10.1016/j.compbiomed.2022.105757

[27] van der Waa J, et al. Evaluating XAI: A comparison of rule-based and example-based explanations. Artif Intell. 2021;291:103404. https://doi.org/10.1016/j.artint.2020.103404

[28] Joseph LP, Joseph EA, Prasad R. Explainable diabetes classification using hybrid Bayesian-optimized TabNet architecture. Comput Biol Med. 2022;145:105507. https://doi.org/10.1016/j.compbiomed.2022.106178

[29] Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J. 2020;18:1414–1428. https://doi.org/10.1016/j.csbj.2020.05.017

[30] Kavakiotis I, et al. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J. 2017;15:104–116. https://doi.org/10.1016/j.csbj.2016.12.005

[31] Kiran S, et al. Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis. Front Digit Health. 2025;7:1557467. https://doi.org/10.3389/fdgth.2025.1557467

[32] Souri A, et al. Knowledge Graphs and Explainable AI in Healthcare. Information. 2022;13(10):459. https://doi.org/10.3390/info13100459

[33] Kim JH, et al. XAI-Based Clinical Decision Support Systems: A Systematic Review. Appl Sci. 2024;14(15):6638. https://doi.org/10.3390/app14156638