Artificial Intelligence for Diabetes Complication Prediction: A Systematic Review of Current Applications and Future Directions

Francesca Pescol, MSc[1] iD**, Pietro Bosoni, PhD[1]** iD**,
Stefania Ghilotti, MD[2], Pasquale De Cata, MD[2],
Lucia Sacchi, PhD[1]** iD**, and Riccardo Bellazzi, PhD[1]** iD

Abstract

Importance and aims: Diabetes can lead to microvascular and macrovascular complications. Modeling the complex relationships between risk factors has motivated the use of Artificial Intelligence (AI) to develop predictive models. Recent advancements, including foundation models and generative AI, have significantly changed how this technology is applied across various contexts. In this review, we summarize the current state of research on AI for predictive diabetes complications, investigating the present and future implications of these innovations.

Methods: We conducted the literature search on PubMed, Scopus, Ovid MEDLINE, CINAHL, and IEEE databases. Our analysis focused on predicted complications, population characteristics, use of AI-based approaches, models' performance, predictor variables, and feature importance evaluation results.

Results: The 49 studies selected in our analysis considered different conditions as prediction outcomes. Eye-related complications were included in 29 studies (59%), emerging as the most frequent predicted diseases. Among the 48 studies employing AI algorithms specifically for the prediction task, 26 (54%) developed only Machine Learning models, 4 (8%) only Deep Learning models, and 18 (38%) applied both approaches. Foundation models and recent AI innovations included in the query were not used by any study. Moreover, only five studies (10%) dealt with unstructured data (signals and images). In the feature importance evaluation, age and glycated hemoglobin consistently emerged as important predictors.

Conclusions: Despite the extensive existing literature on AI for predicting diabetes complications, several emerging challenges persist. These include the effective utilization of unstructured data and the integration of recent advancements introduced by foundation models and generative AI.

Keywords

artificial intelligence, diabetes complications, risk prediction, foundation models

Introduction

Diabetes-related complications affect both microvascular and macrovascular systems, increasing the clinical and financial impact of this condition. The long-term damage to the macrovascular system remains the leading cause of death in people with diabetes. On the other hand, microvascular complications involving kidneys, eyes, and nerves significantly impact morbidity and mortality, leading to severe outcomes, such as amputations, end-stage renal disease, and vision loss.[1] In this context, predictive models of diabetes progression could enable earlier interventions, personalized treatments, and suitable resource allocation.[2] Beyond well-recognized clinical risk factors of diabetes complications,

[1]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy
[2]Department of General Medicine, Istituti Clinici Scientifici Maugeri, Pavia, Italy

Corresponding Author:
Francesca Pescol, MSc, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, Pavia 27100, Italy.
Email: francesca.pescol01@universitadipavia.it

such as the disease duration, the presence of specific comorbidities and the genetic predisposition, other biological, lifestyle, and socioeconomic aspects contribute to the clinical progression of the disease. However, predicting the onset of complications considering this multitude of factors is complex, limited by the heterogeneous nature of these factors and by the variability of their collection across clinical centers. Artificial Intelligence (AI) has been increasingly used for this purpose, due to its ability to handle large and complex data sets.[3] The AI has brought several advantages compared to the existing scores for predicting diabetes complications derived using traditional statistical models.[4] By providing more advanced and sophisticated solutions, such as non-linear models trained on a larger number of variables, AI may be better suited to handle the complexity of the relationships among predictors.[3] Furthermore, generative AI models have provided powerful solutions to challenges like data scarcity, patient variability, and personalization by modeling, simulating, and generating high-quality synthetic data.[5] In this context, the advent of foundation models marked a significant milestone. Trained on massive (typically hundreds of millions to billions of samples) and heterogeneous data sets and using general-purpose objectives through a large-scale self-supervised learning approach, this last generation of AI models can be adapted for a wide range of tasks and used directly or after further fine-tuning on specific data sets, reducing the need for an extensive target data source to obtain task-specific model.[4,6] Among the variety of tasks, foundation models are often employed as generative AI models to create new content, such as text or images. Large Language Models (LLMs), as high-capacity foundation models for text generation, mostly based on the groundbreaking Transformer architecture, played a pivotal role in establishing this new paradigm of AI.[6] In the diabetes domain, some recent studies proved the potential impact of these large pre-trained models. Li et al exploited an LLM within an image-language system for primary diabetes care. The LLM demonstrated comparable performance to primary care physicians in providing individualized management recommendations for people with diabetes.[7] Lutsker et al developed GluFormer, a foundation model for CGM data, designed to learn glycemic risk patterns. GluFormer outperformed traditional metrics in risk-based patient stratification.[8] Finally, Rancati et al exploited TimeGPT model for blood glucose prediction in pediatric patients with type 1 diabetes. The performance achieved by TimeGPT against the state-of-the-art models demonstrated its promising generalization capability.[9]

Considering the increasing application of AI in developing predictive models of diabetes complications and the recent technological innovations of AI models, we conducted a systematic review to summarize the current state of research on the topic. We aimed at highlighting the advantages and limitations of AI applications in this field, investigating the impact of technological innovations, and outlining future perspectives. Our research encompassed both microvascular and macrovascular complications and, from a methodological perspective, considered traditional AI models (Machine Learning—ML—and Deep Learning—DL—techniques), as well as recent developments (foundation models, generative AI, LLMs).

Methods

Study Design

We conducted the literature search on PubMed, Scopus, Ovid MEDLINE, CINAHL, and IEEE databases. Our query was structured in three main sections (diabetes complications, methods, prediction task), linked by an AND Boolean operator:

- "Diabetes Complications" (MeSH term) AND
- "Artificial Intelligence" (MeSH term) OR "Machine Learning" (MeSH term) OR "Neural Network, Computer" (MeSH term) OR "foundation models" OR "generative AI" OR "transformer models" OR "pre-trained models" OR "large language models" AND
- "Predict* Models" OR "Risk Assessment" (MeSH term).

We used the truncation symbol "*" for "Predict" to include all terms beginning with this root, such as "predictive" and "prediction." The queries were applied to the title, abstract, and keywords fields. To select the studies, we conducted a two-phase screening procedure, based on the abstract (first screening) and the full text (second screening). When possible (in PubMed and Ovid MEDLINE databases and for terms existing in MeSH), MeSH terminologies were included in the query. Consequently, due to the possible inclusion of MeSH terms and the different modalities required to exploit this feature, queries slightly vary among databases. The specific query used for each database is detailed in the Supplementary Material.

Eligibility Criteria

We were interested in studies implementing and/or applying AI models to predict diabetes complications in patients diagnosed with diabetes. To ensure a robust assessment of the use of these techniques, it was necessary to select studies that provided a comprehensive description of both the methods and the population characteristics. Based on this general requirement, we defined exclusion criteria. Specifically, in the first screening phase, we excluded studies where:

- The population included individuals not diagnosed with diabetes (reason 1).
- The population was limited to patients with diabetes and additional conditions, such as studies focusing on

diabetes complication progression once the complication had already been diagnosed (reason 2).

- The objective did not include the prediction of the onset of the complication but only the identification of its risk factors (reason 3).
- The objective focused on predicting diabetes or classifying its status (reason 4).
- The predicted complication was not specific to diabetes, such as studies considering death, depression, or COVID-19 as prediction outcomes (reason 5).

In addition, we excluded reviews (reason 6) and retracted studies (reason 7). For the second screening, we added more detailed exclusion criteria, based on specific aspects observable from the full presentation of the studies. These conditions involved:

- The absence of information on the population (reason 8).
- The use of secondary data from a previous systematic review, instead of specific population data sets (reason 9).
- The use of proprietary tools, ie, algorithms whose implementation is owned and whose usage is controlled by a company (reason 10).

Moreover, we excluded studies where the methodology was not clearly presented, eg, where authors did not provide sufficient detailed information about the implementation strategies of the AI models, such as training procedures or validation strategies (reason 11). Finally, we could not include studies for which we were unable to retrieve the full text, even through our institutional access (reason 12).

Data Extraction and Analysis

The analysis of the selected studies was conducted based on eight main categories of characteristics describing the approaches used. Specifically, we collected information on:

1. Publication year, location, including both the country of the first author's affiliation and the population's country (*general information*).
2. The predicted complications (*prediction outcome*).
3. Population characteristics, including type of diabetes, age (distinguishing pediatric and adult patients and considering adults when age was higher than 18), and percentage of females (*population*).
4. The AI models employed, the reproducibility of the implemented approaches and the evaluation strategy, distinguishing between internal and external validation (*models*).
5. The predictor variables used in the final model (*predictor variables*).
6. The Receiver Operating Characteristic Area Under the Curve (ROC AUC) value achieved by the best

model, if ROC AUC was selected among the performance metrics (*evaluation metrics*).
7. Whether a feature importance evaluation was performed and the first three predictors selected from the feature ranking (*feature importance*).

We analyzed absolute and relative frequencies for categorical variables and computed descriptive statistics for continuous variables, including mean, standard deviation, median, minimum, maximum, and interquartile range (IQR). The analysis was carried out using Python 3.12.

Results

The total number of records extracted by the queries across the databases was 188. Among them, we identified 34 duplicates. The specific number of records obtained from each database, together with the screening procedure results, is reported in the PRISMA diagram in Figure 1. The diagram indicates the number of studies excluded for each criterion (reasons 1-12) in both the first screening and the second screening. The final number of reviewed full papers was 49. The complete list of these studies is provided in Supplementary Table 1. We will detail the results of our analysis considering each of the categories defined in section "Data extraction and analysis" (from *general information* to *feature importance*).

General Information

The included studies were published between 2010 and 2025, mainly in China and the United States. Figure 2 shows the study frequency by year, while Figure 3 illustrates the frequency by country. A general increasing trend can be observed over the years, with a noticeable peak in 2024. The frequency indicated for 2025 is limited to March 2025, when the records' extraction was performed.

Prediction Outcome

Two main groups were identified according to the number of prediction outcomes: 34 studies (69%) considered one complication as the target condition,[10-43] while 15 (31%) addressed the prediction of multiple complications.[44-58] In this second group, the mean number of prediction outcomes was 3.3 ($\pm$ 0.85 standard deviation). To evaluate the distribution of complications as prediction objective in studies, we grouped the specific conditions predicted in studies into five main categories: Cardiovascular Complication, Metabolic Complication, Diabetic Neuropathy, Eye Complication, and Renal Complication, as shown in Supplementary Table 2. Table 1 reports how frequently each complication was targeted in predictive modeling across the included studies, considering both the total number of studies (n = 49) and the number of studies that selected a single complication as the outcome (single-outcome studies, n = 34). Most studies
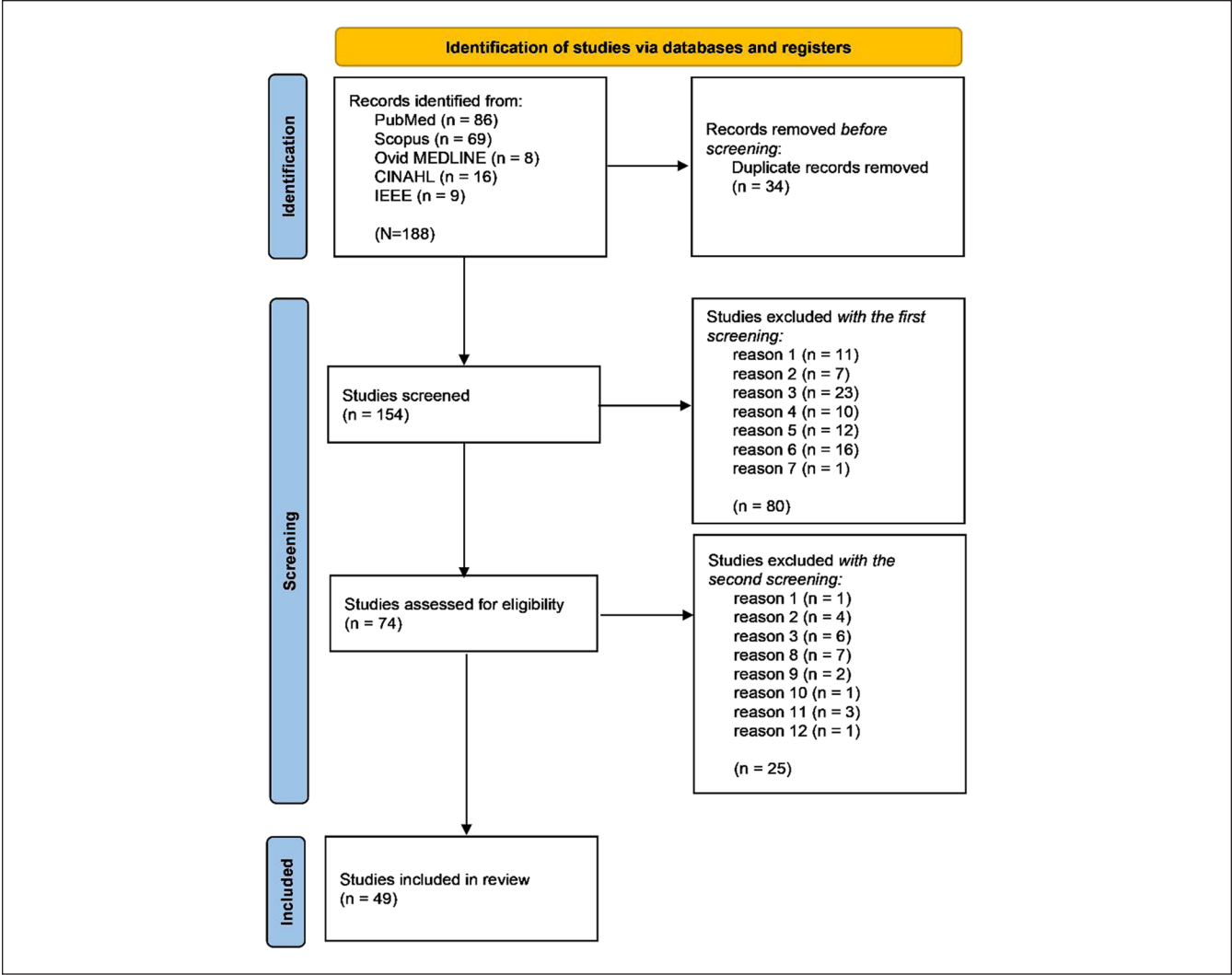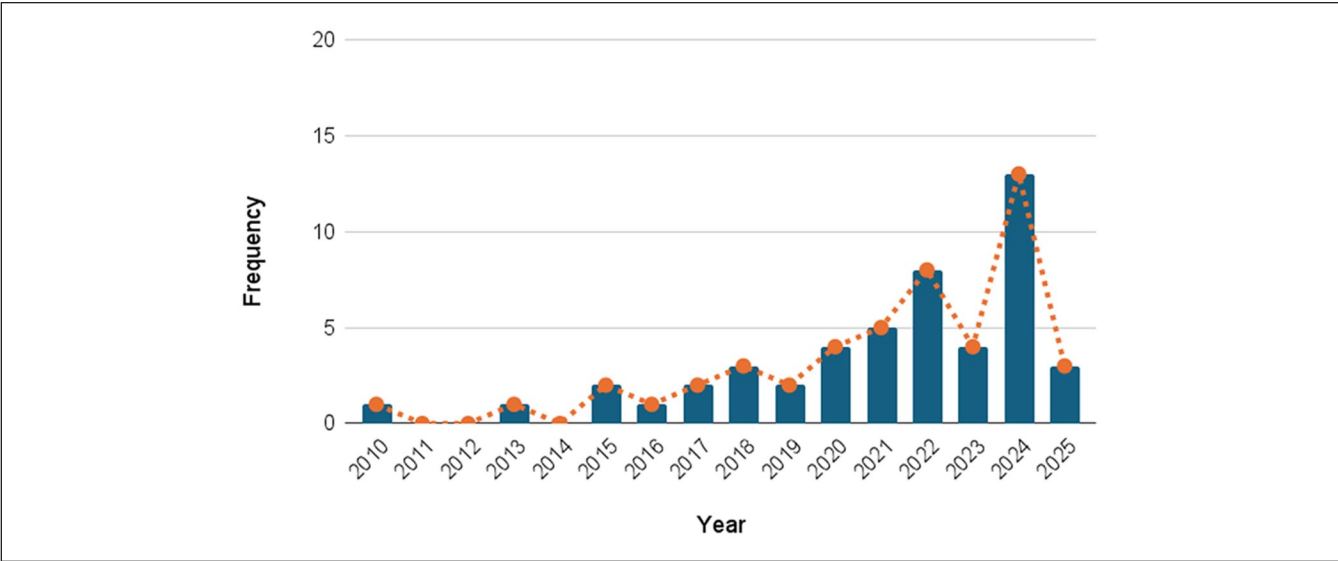
Figure 1. PRISMA diagram.

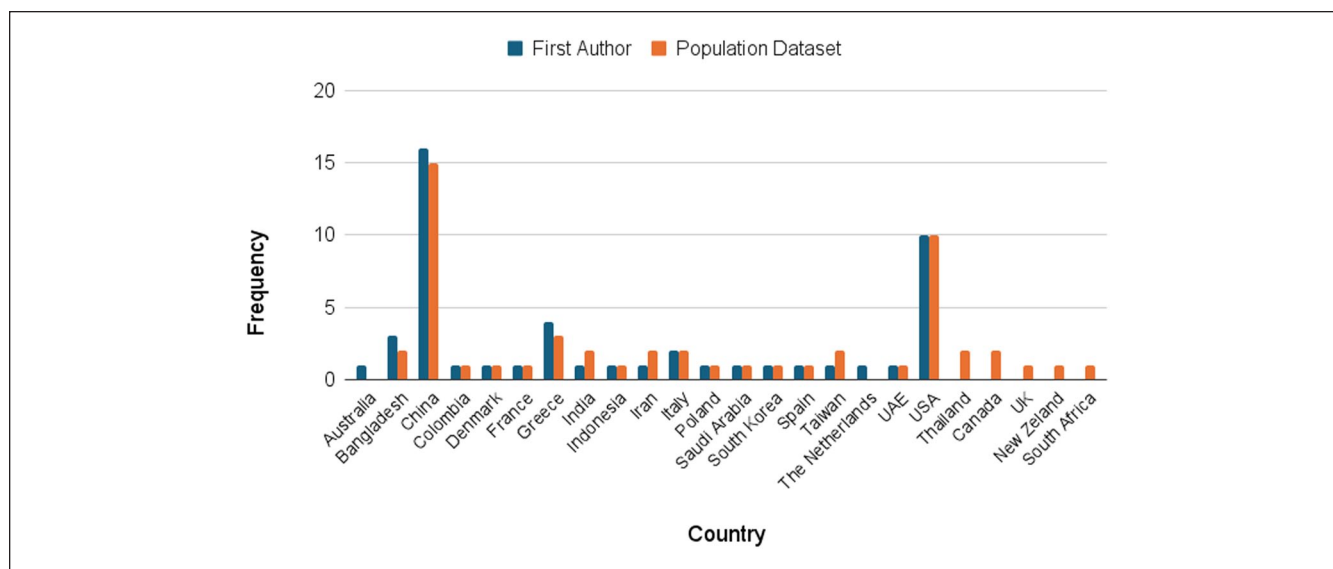**Figure 2.** Study frequency by year.

Figure 3. Study frequency by country, considering the author's affiliation and the population origin.

Table 1. Absolute and Percentage (in Parentheses) Frequency of Complication Categories Over the Total Number of Studies (n = 49) and the Number of Single-Outcome Studies (n = 34).

Complication category	Frequency over the total number of studies	Frequency over the number of single-outcome studies
Cardiovascular complication	17 (34.69%)	6 (17.65%)
Metabolic complication	5 (10.20%)	1 (2.94%)
Diabetic Neuropathy	13 (26.53%)	2 (5.88%)
Eye complication	29 (59.18%)	16 (47.06%)
Renal complication	20 (40.82%)	9 (26.47%)

focused on the diseases within the Eye Complication category, while the Metabolic Complication category was the least investigated.

Population

Results concerning the analysis of population characteristics are presented in Table 2. The number of patients included varied considerably across studies, as indicated by the standard deviation in Table 2. However, we observed that most studies (73%) employed fewer than 10 000 patients, and 50% of them fewer than 1000. This suggests that the average number of patients reported in Table 2 was inflated by a few studies with particularly large samples. Among them, Ljubic et al[53] used the largest population data set, including more than 1 million patients. Although not all studies reported information on the type of diabetes, the percentage of females, or the patients' age, we observed that most studies focused on adult patients (>18 years) with type 2 diabetes.

Only three studies dealt with type 1 diabetes,[10,44,51] and only two studies involved pediatric patients.[44,57]

Models

The most recent AI models included in queries (foundation models, LLMs, or generative AI models) were not reported by any of the studies. Traditional ML and DL algorithms constituted the most common general approaches used for predicting diabetes complications. In addition, seven studies (14%) also explored traditional statistical methods for this purpose, employing Cox proportional hazards regression models and adding Kaplan-Meier analysis (in four studies). In six out of these seven studies, traditional statistical methods were compared with or used in addition to AI models.[17,20,29,31,52,57] Basu et al,[45] instead, employed only statistical approaches. This study was anyway included in the query results due to its use of an AI-based feature selection method. Among the studies exploiting AI methods (n = 48), the majority developed task- and population-specific models (ie, the training procedure was performed on their own data set, and it was targeted to a specific task). Only two studies adopted pre-trained models.[36,40] Bora et al[40] proposed the validation on an external data set of a pre-trained Convolutional Neural Network (CNN) to predict diabetic retinopathy from fundus photographs. The network training (on a set of about 600 000 fundus photographs) and its first validation were presented in another study by the same author.[31] Zanelli et al[36] applied a pre-trained CNN for processing signals from photoplethysmography and pulse wave velocity technologies, again for the prediction of diabetic retinopathy. In this way, they performed an external validation of the CNN network that was pre-trained on the same

Table 2. Statistics Describing the Main Characteristics of Study Populations.

Variable	Statistic	Value
No. patients	Mean	62 631.76
	Standard deviation	276 571.61
	Median	2000
	Maximum	1 910 674
	Minimum	55
	IQR (25th-75th)	728-11 789
No. studies indicating the patients' type of diabetes	Frequency[a]	36 (73.47%)
No. studies specifically on type 2 diabetes	Frequency (over no. studies indicating the patients' type of diabetes)	26 (72.22%)
No. studies indicating the patients' age	Frequency	36 (73.47%)
No. studies specifically on adult patients[b]	Frequency (over no. studies indicating the patients' age)	34 (94.4%)
No. studies indicating the patients' gender distribution	Frequency	31 (63.27%)
Female proportion (%)	Mean (across no. studies indicating the patients' gender distribution)	49.25%
	Standard deviation (across no. studies indicating the patients' gender distribution)	8.29%

Frequencies are indicated with absolute and percentage (in parentheses) values.
[a]Unless otherwise specified, the frequencies are calculated based on the total number of studies (n = 49).
[b]Patients are considered adults if older than 18.

Table 3. Statistics Describing the Number of AI Models Employed in Studies.

Statistic	Value
Mean	4.02
Standard Deviation	2.41
Median	4
Maximum Value	10
Minimum Value	1
IQR (25th-75th)	2-5

types of data (900 signals) and presented in a previous study by the same author.[59] Despite leveraging pre-training, the focus on domain-specific tasks, the same input modality, and the use of relatively small data sets for the training procedure preclude the models proposed in both studies from being considered foundation models.[36,40]

In general, we found that 26 studies (54% of those employing AI) developed only ML models, 4 (8%) only DL models, and 18 (38%) exploited both approaches. General statistics describing the number of AI models employed in studies are shown in Table 3. To evaluate the frequency of model usage, we grouped the ML and DL algorithms used in the studies into high-level categories, as presented in Table 4. Figure 4 shows the percentage frequency of each category, relative to the number of studies exploiting AI models (n = 48). The same information is reported for the specific AI models in Figure 5, considering only models that were used

in at least two studies. The Ensemble Method category emerged as the most used, primarily due to the high frequency of the Random Forest (RF) model. Among studies where models' implementation was performed (n = 47), 24 (51%) provided sufficient information to ensure reproducibility (hyperparameters' values). Regarding the validation strategy chosen by the studies, we observed that 36 studies (73%) performed only an internal validation of the implemented approaches, 12 (24%) used an additional external validation data set to compare the performances achieved in the internal validation (exploiting both types of validation). As mentioned, the study conducted by Bora et al[40] presented the results of a pre-trained CNN model on another external data set, adopting only the external validation strategy.

Predictor Variables

To identify potential patterns in predictors' selection, we grouped the variables considered in the studies into ten categories (demographic data, basic clinical data, comorbidity, glucose measure, biochemical parameter, pharmacological data, international coding standard, omics data, imaging data, and signal). Moreover, we further grouped the studies as using structured or unstructured data based on the variables' type. Imaging data and signal categories were classified as unstructured data, while the predictors included in all the other categories were classified as structured data. Some examples of predictors for each category are reported in Supplementary Table 3. Textual data, such as medical reports

Table 4. ML and DL Categories Used to Group the AI Models Employed in Studies.

Model	Category	ML/DL
Logistic Regression (LR)	-[a]	ML
Linear Discriminant Analysis (LDA)	Bayesian Method	
Naive Bayes (NB)		
Decision Tree (DT)	-	
Random Forest (RF)	Ensemble Method	
Gradient Boosting Machine (GBM) and related eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), CatBoost		
Adaptive Boosting (AdaBoost)		
Support Vector Machine (SVM)	-	
K-Nearest-Neighbor (KNN)	-	
K-Means	Clustering Method	
Hierarchical Clustering (HC)		
Principal Component Analysis (PCA)	-	
Feedforward Neural Network (FNN)	Feedforward Network	DL
Multilayer Perceptron (MLP)		
Artificial Neural Network (ANN)		
Deep Neural Network (DNN)		
Recurrent Neural Network (RNN)	Recurrent Network	
Long-Short Term Memory (LSTM)		
Gated Recurrent Unit (GRU)		
Convolutional Neural Network (CNN)	-	
Wavelet Neural Network (WNN)	Wavelet Network	
Hybrid Wavelet Neural Network (HWNN)		
Self-Organizing Map (SOM)	-	

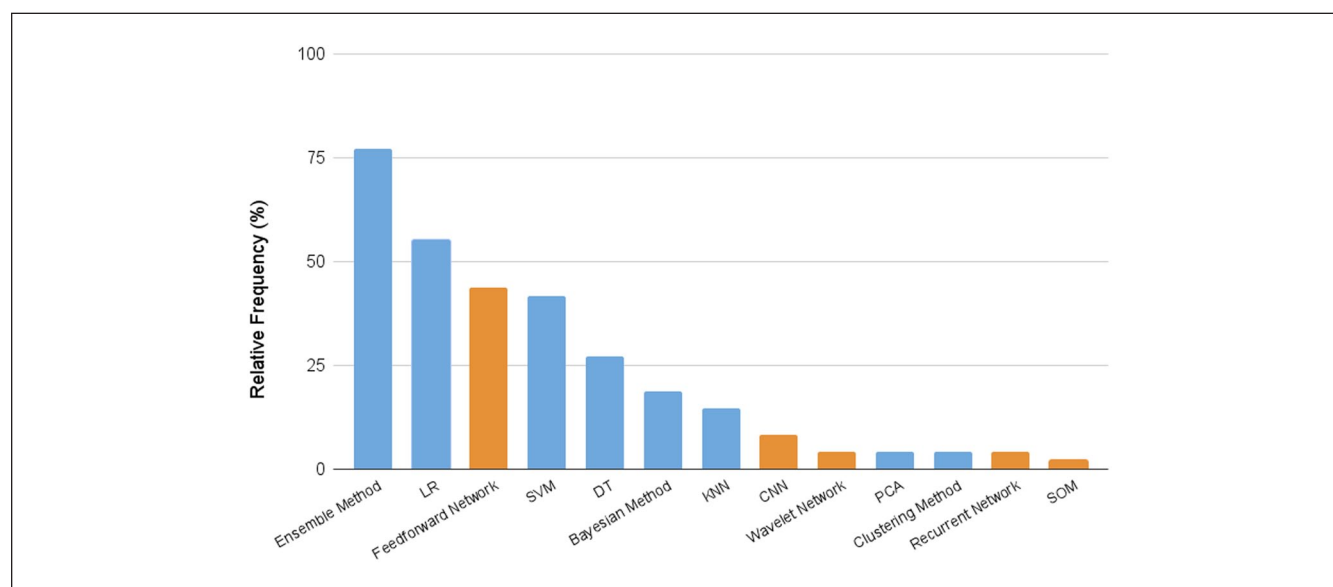[a]In these cases, the models were not suited to be grouped with others in a common category.

Figure 4. Percentage frequency of the defined ML and DL model categories over the number of studies exploiting AI methods (n = 48). Light blue bars refer to ML models; orange bars to DL models.

or clinical notes, were not explored in any of the studies. This finding aligns with the absence of Natural Language Processing (NLP) techniques or the use of LLMs. The heat map shown in Figure 6 provides an overview of the use of the different data categories for each of the complications predicted in the considered studies. The most used data
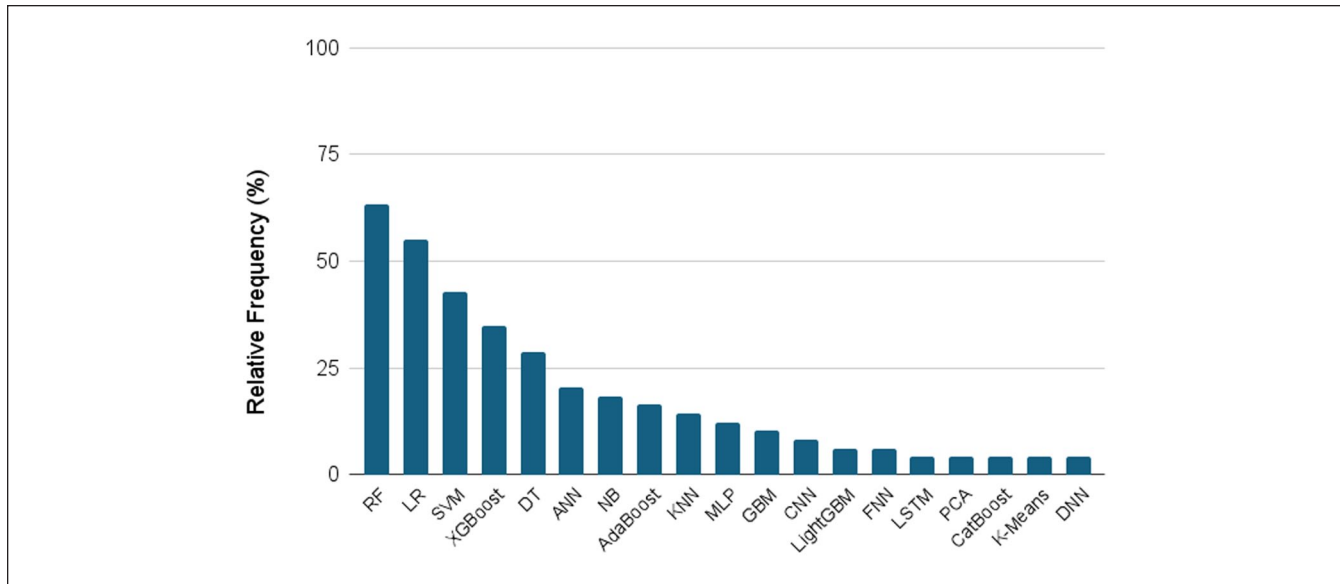
Figure 5. Percentage frequency of the specific AI models over the number of studies exploiting AI methods (n = 48). Only models employed in at least two studies are considered.
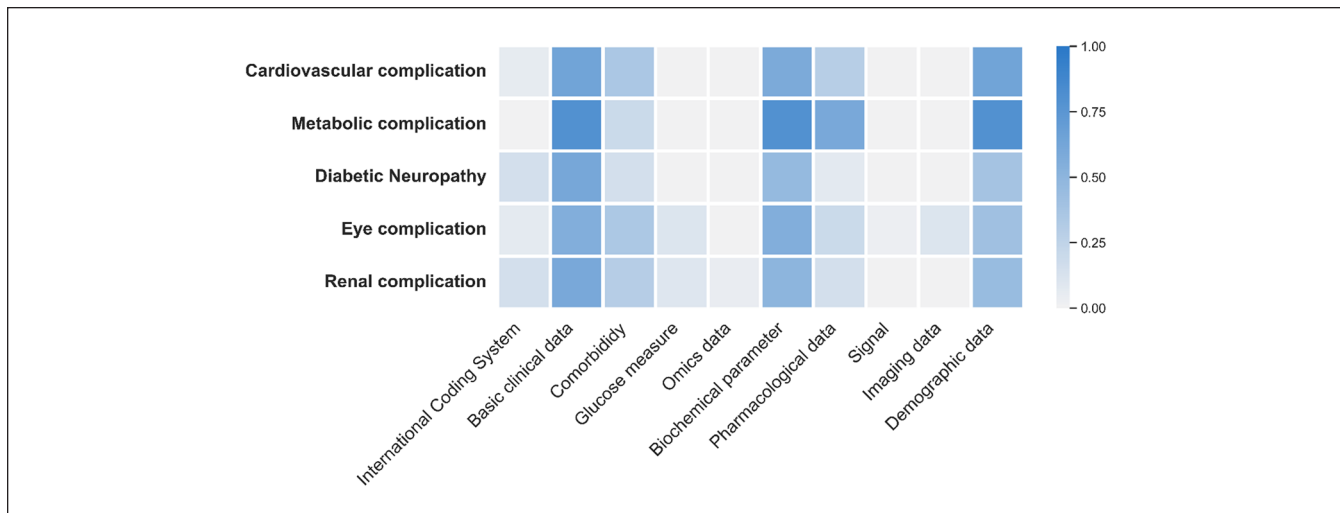

Figure 6. Heat map of the categories used to group the predictor variables (x-axis) employed in studies by complication category (y-axis). The color bar represents the relative frequency of predictors, ranging from 0 (indicating the predictor was never used) to 1 (indicating the predictor was used in all cases).

categories included basic clinical information, demographic characteristics, and biochemical parameters, while omics data and unstructured data were rarely used. Specifically, we observed that only four studies (8%) exploited the predictors from the unstructured data type group.[12,32,36,40]

Evaluation Metrics

Descriptive statistics (mean, standard deviation, median, 25th and 75th percentiles) of the ROC AUC values reported in the studies as the best metric performance achieved by the

AI models are shown in Table 5. We compared these results across the different categories of complications and the two validation strategies (internal and external). Concerning the external validation strategy, the mean and standard deviation values are shown only for Eye complication and Renal complication categories, as external validation results were available from only a single study for each of the remaining categories. We statistically investigated the difference in ROC AUC values between the two validation frameworks (internal and external) for Eye and Renal complication categories using the Mann-Whitney U test. The results (*P*-values)
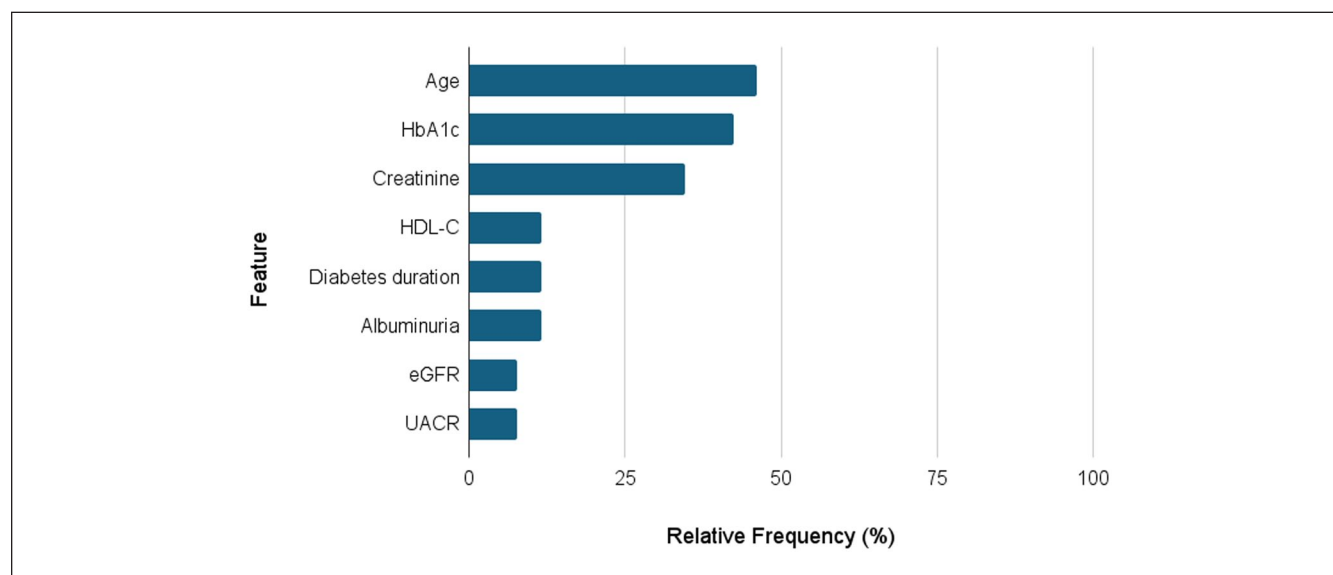
Table 5. Mean, Standard Deviation, Median, 25th and 75th Percentiles of the Best ROC AUC Values Achieved by AI Models for Each Complication Category and Each Type of Validation Strategy (Internal or External).

Complication category	Validation	No. studies	ROC AUC (mean ± standard deviation)	ROC AUC (median, 25th-75th percentiles)	P-value (internal vs external validation)
Cardiovascular complication	Internal	10	0.80 ± 0.11	0.77 (0.73-0.90)	-
	External	1	-[a]	-	
Metabolic complication	Internal	3	0.82 ± 0.03	0.83 (0.81-0.84)	-
	External	-	-	-	
Diabetic Neuropathy	Internal	8	0.87 ± 0.07	0.87 (0.83-0.91)	-
	External	1	-	-	
Eye complication	Internal	20	0.86 ± 0.07	0.87 (0.81-0.91)	**0.0305[b]**
	External	6	0.79 ± 0.05	0.78 (0.76-0.83)	
Renal complication	Internal	13	0.86 ± 0.08	0.84 (0.82-0.91)	0.0603
	External	5	0.75 ± 0.11	0.72 (0.71-0.82)	

The statistics are computed over the number of studies predicting each specific complication and exploiting the ROC AUC between performance metrics (indicated in "no. studies" column). Results (P-values) of the Mann–Whitney U test comparing ROC AUC values between internal and external validation are also reported.
[a]When the number of studies are one or null, we did not report the statistics, and we did not perform the statistical comparison.
[b]We highlighted in bold significant P-values (considering a significant level of .05).

Figure 7. The percentage frequency of the eight variables identified in the top-three predictors ranking from the feature importance investigation in at least two studies. The percentage is computed based on the number of studies conducting this type of investigation (n = 26). HbA$_{1c}$ refers to the glycated hemoglobin, HDL-C to High-Density Lipoprotein Cholesterol, eGFR to estimated glomerular filtration rate, and UACR to urinary albumin-creatinine ratio.

are shown in Table 5. This analysis generally demonstrated a good predictive ability of AI models, with a median ROC-AUC of 0.84 across all complication categories under internal validation. However, performance decreased in external validation, with 0.76 as the median ROC-AUC value, as confirmed by the overall test comparing ROC AUC values in internal and external evaluation without distinguishing the specific complication categories (P = .0039).

The comparison between AI models and traditional statistical approaches was possible only for two among the six studies employing both methodologies. Only Belur Nagaraj et al[29] and Aminian et al[52] provided AUC values achieved by Cox proportional hazards regression models. In both cases, AI models performed better.

Feature Importance

Twenty-six studies (53%) conducted a feature importance investigation, identifying the variables that most influenced the AI models in predicting complications. Figure 7 shows

the eight variables most frequently included in the top-three predictors ranking, based on the feature importance investigations. These eight variables resulted in the first three positions of feature importance ranking in at least two studies. For each one, the relative percentage frequency can be observed. Age and glycated hemoglobin (HbA_{1c}) emerged as the most important variables in almost 50% of studies conducting this type of investigation (n = 26).

Discussion

In this systematic review, we aimed to identify key methodological trends concerning the use of AI to predict diabetes complications and explore the use of recent technological innovations in this field. The 49 studies included in our analysis outlined a wide variety of conditions employed as prediction outcomes, demonstrating an established interest in applying these approaches to different disease domains. The five complication categories defined to group all the predicted conditions were well-represented in the studies, except for the Metabolic complication, which was much less investigated. The general good ROC-AUC performances reported by the studies confirm that these approaches can play an effective role in the early detection of dangerous diabetes progression. However, from a methodological perspective, several limitations can be identified in the current body of research on this topic. First, we need to consider that the reported results mostly refer to small samples of patients. Internal validation strategies, which were the most adopted, require splitting the data set into at least two subsets, ensuring that model validation is performed on a portion of data that is independent from the training set. This procedure impacts the sample size used for validation and reduces the robustness of the results and their generalizability. The lower performances observed in the external validations support this consideration. Moreover, despite the large number of ML and DL models experimented by the studies, the implemented pipelines appear highly consistent. This potentially stems from the almost exclusive use of structured data. Although the implemented approaches demonstrated the high capacity of AI models in handling different clinical and semantic domain data sets, their high potential for processing complex and unstructured data has yet to be fully exploited. Among the included studies, only a small subset developed or applied models for image and signal processing. Moreover, no study investigated the use of textual data, which still represents a relevant source of clinical information. In this context, NLP strategies and LLMs, which did not emerge in the proposed methodologies, could introduce a novel methodological perspective. These considerations underscore that, despite the rich existing literature in this field, several future directions remain to be explored to fully unlock the potential of AI models. A notable need is to leverage data sources that remain underutilized due to their unstructured nature. In this regard, foundation models present a promising solution, as they can be applied without the extensive, task-specific training that typically relies on large, labeled data sets.

Another important consideration regarding the analyzed studies concerns the limited attention given to the practical application of the models in clinical settings. Studies tended to focus primarily on technical development, often neglecting how the proposed models could be used in real-world practice. The lack of focus on clinical impact is further reflected in the absence of evaluations regarding model reliability and trustworthiness. Demonstrating the actual utility of AI models to support their integration into real-world clinical workflows requires these targeted evaluations, together with transparent reporting of models' hyperparameters (provided in only 51% of studies), accessible data, and systematic use of high-dimensional and unstructured sources.

Abbreviations

AdaBoost, adaptive boosting; AI, artificial intelligence; ANN, artificial neural network; CNN, convolutional neural network; DL, deep learning; DNN, deep neural network; DT, decision tree; eGFR, estimated glomerular filtration rate; FNN, feedforward neural network; GBM, gradient boosting machine; GRU, gated recurrent unit; HbA_{1c}, glycated hemoglobin; HC, hierarchical clustering; HDL-C, high-density lipoprotein cholesterol; HWNN, hybrid wavelet neural network; IQR, interquartile range; KNN, K-nearest-neighbor; LDA, linear discriminant analysis; LightGBM, light gradient boosting machine; LLMs, large language models; LR, logistic regression; LSTM, long-short term memory; ML, machine learning; MLP, multilayer perceptron; NB, Naïve Bayes; NLP, natural language processing; PCA, principal component analysis; RF, random forest; RNN, recurrent neural network; ROC AUC, receiver operating characteristic area under the curve; SOM, self-organizing map; SVM, support vector machine; UACR, urinary albumin-creatinine ratio; WNN, wavelet neural network; XGBoost, eXtreme gradient boosting.

Declaration of Conflicting Interests

Funding

ORCID iDs

Francesca Pescol   https://orcid.org/0009-0003-2384-1945
Pietro Bosoni   https://orcid.org/0000-0002-1431-6044
Lucia Sacchi   https://orcid.org/0000-0002-1390-9825
Riccardo Bellazzi   https://orcid.org/0000-0002-6974-9808

Supplemental Material

Supplemental material for this article is available online.

References

1. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16(7):377-390. doi:10.1038/s41581-020-0278-5.

2. Kanbour S, Harris C, Lalani B, et al. Machine learning models for prediction of diabetic microvascular complications. *J Diabetes Sci Technol*. 2024;18(2):273-286. doi:10.1177/19322968231223726.

3. Tan KR, Seng JJB, Kwan YH, et al. Evaluation of machine learning methods developed for prediction of diabetes complications: a systematic review. *J Diabetes Sci Technol*. 2023;17(2):474-489. doi:10.1177/19322968211056917.

4. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-265. doi:10.1038/s41586-023-05881-4.

5. Vehi J, Mujahid O, Beneyto A, Contreras I. Generative artificial intelligence in diabetes healthcare. *iScience*. 2025;28(8):113051. doi:10.1016/j.isci.2025.113051.

6. Rossi S, Rossi M, Mukkamala RR, Thatcher JB, Dwivedi YK. Augmenting research methods with foundation models and generative AI. *Int J Inf Manage*. 2024;77:102749. doi:10.1016/j.ijinfomgt.2023.102749.

7. Li J, Guan Z, Wang J, et al. Integrated image-based deep learning and language models for primary diabetes care. *Nat Med*. 2024;30(10):2886-2896. doi:10.1038/s41591-024-03139-8.

8. Lutsker G, Sapir G, Shilo S, et al. From glucose patterns to health outcomes: a generalizable foundation model for continuous glucose monitor data analysis. *arXiv:2408.11876*. 2025. doi:10.48550/arXiv.2408.11876.

9. Rancati S, Bosoni P, Schiaffini R, et al. Exploration of foundational models for blood glucose forecasting in type-1 diabetes pediatric patients. *Diabetology*. 2024;5(6):584-599. doi:10.3390/diabetology5060042.

10. Skevofilakas M, Zarkogianni K, Karamanos BG, Nikita KS. A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus. Paper presented at 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology; 2010. doi:10.1109/IEMBS.2010.5626245.

11. Zong GW, Wang WY, Zheng J, et al. A metabolism-based interpretable machine learning prediction model for diabetic retinopathy risk: a cross-sectional study in Chinese patients with type 2 diabetes. *J Diabetes Res*. 2023;2023:3990035. doi:10.1155/2023/3990035.

12. Chilukoti SV, Shan L, Tida VS, Maida AS, Hei X. A reliable diabetic retinopathy grading via transfer learning and ensemble learning with quadratic weighted kappa metric. *BMC Med Inform Decis Mak*. 2024;24(1):37. doi:10.1186/s12911-024-02446-x.

13. Huang GM, Huang KY, Lee TY, Weng J. An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. *BMC Bioinformatics*. 2015;16(suppl 1):S5. doi:10.1186/1471-2105-16-S1-S5.

14. Dalakleidi K, Zarkogianni K, Thanopoulou A, Nikita K. Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications. *Expert Syst*. 2017;34(6):e12214. doi:10.1111/exsy.12214.

15. David SK, Rafiullah M, Siddiqui K. Comparison of different machine learning techniques to predict diabetic kidney disease. *J Healthc Eng*. 2022;2022:7378307. doi:10.1155/2022/7378307.

16. Zarkogianni K, Athanasiou M, Thanopoulou AC. Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication. *IEEE J Biomed Health Inform*. 2018;22(5):1637-1647. doi:10.1109/JBHI.2017.2765639.

17. Afarideh M, Aryan Z, Ghajar A, et al. Complex association of serum alanine aminotransferase with the risk of future cardiovascular disease in type 2 diabetes. *Atherosclerosis*. 2016;254:42-51. doi:10.1016/j.atherosclerosis.2016.09.009.

18. Yun C, Tang F, Gao Z, et al. Construction of risk prediction model of type 2 diabetic kidney disease based on deep learning. *Diabetes Metab J*. 2024;48(4):771-779. doi:10.4093/dmj.2023.0033.

19. Wang R, Miao Z, Liu T, et al. Derivation and validation of essential predictors and risk index for early detection of diabetic retinopathy using electronic health records. *J Clin Med*. 2021;10(7):1473. doi:10.3390/jcm10071473.

20. Segar MW, Usman MS, Patel KV, et al. Development and validation of a machine learning-based approach to identify high-risk diabetic cardiomyopathy phenotype. *Eur J Heart Fail*. 2024;26(10):2183-2192. doi:10.1002/ejhf.3443.

21. Wu G, Hu Y, Zhu Q, et al. Development and validation of a simple and practical model for early detection of diabetic macular edema in patients with type 2 diabetes mellitus using easily accessible systemic variables. *J Transl Med*. 2024;22(1):523. doi:10.1186/s12967-024-05328-y.

22. Yang P, Yang B. Development and validation of predictive models for diabetic retinopathy using machine learning. *PLoS ONE*. 2025;20(2):e0318226. doi:10.1371/journal.pone.0318226.

23. Tian Z, Zhang J, Fan Y, et al. Diabetic peripheral neuropathy detection of type 2 diabetes using machine learning from TCM features: a cross-sectional study. *BMC Med Inf Decis Mak*. 2025;25(1):90. doi:10.1186/s12911-025-02932-w.

24. Oh E, Yoo TK, Park EC. Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. *BMC Med Inform Decis Mak*. 2013;13:106. doi:10.1186/1472-6947-13-106.

25. Wu R, Shu Z, Zou F, et al. Identifying myoglobin as a mediator of diabetic kidney disease: a machine learning-based cross-sectional study. *Sci Rep*. 2022;12(1):2141. https://www.nature.com/articles/s41598-022-25299-8. Accessed April 17, 2025.

26. Saleh E, Błaszczyński J, Moreno A, et al. Learning ensemble classifiers for diabetic retinopathy assessment. *Artif Intell Med*. 2018;85:50-63. doi:10.1016/j.artmed.2017.09.006.

27. Nanda R, Nath A, Patel S, Mohapatra E. Machine learning algorithm to evaluate risk factors of diabetic foot ulcers and its severity. *Med Biol Eng Comput*. 2022;60(8):2349-2357. doi:10.1007/s11517-022-02617-w.

28. Nabrdalik K, Kwiendacz H, Irlik K, et al. Machine learning identifies metabolic dysfunction-associated steatotic liver disease in patients with diabetes mellitus. *J Clin Endocrinol Metab*. 2024;109(8):2029-2038. doi:10.1210/clinem/dgae060.

29. Belur Nagaraj S, Pena MJ, Ju W, Heerspink HL; BEAt-DKD Consortium. Machine-learning-based early prediction of end-stage renal disease in patients with diabetic kidney disease using

clinical trials data. *Diabetes Obes Metab*. 2020;22(12):2479-2486. doi:10.1111/dom.14178.

30. Wan X, Zhang R, Wang Y, et al. Predicting diabetic retinopathy based on routine laboratory tests by machine learning algorithms. *Eur J Med Res*. 2025;30(1):183. doi:10.1186/s40001-025-02442-5.

31. Bora A, Balasubramanian S, Babenko B, et al. Predicting risk of developing diabetic retinopathy using deep learning. *Lancet Digital Health*. 2021;3(1):e10-e19. doi:10.1016/S2589-7500(20)30250-8.

32. Nghiem N, Wilson N, Krebs J, Tran T. Predicting the risk of diabetes complications using machine learning and social administrative data in a country with ethnic inequities in health: Aotearoa New Zealand. *BMC Med Inform Decis Mak*. 2024;24:274. doi:10.1186/s12911-024-02678-x.

33. Wang X, Wang W, Ren H, Li X, Wen Y. Prediction and analysis of risk factors for diabetic retinopathy based on machine learning and interpretable models. *Heliyon*. 2024;10(9). doi:10.1016/j.heliyon.2024.e29497.

34. Dong Z, Wang Q, Ke Y, et al. Prediction of 3-year risk of diabetic kidney disease using machine learning based on electronic medical records. *J Transl Med*. 2022;20(1):143. doi:10.1186/s12967-022-03339-1.

35. Li W, Song Y, Chen K, et al. Predictive model and risk analysis for diabetic retinopathy using machine learning: a retrospective cohort study in China. *BMJ Open*. 2021;11(11):e050989. doi:10.1136/bmjopen-2021-050989.

36. Zanelli S, Eveilleau K, Ammi M, Hallab M, El Yacoubi MA. Risk assessment of diabetic retinopathy with machine and deep learning models with PPG signals and PWV. Paper presented at 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2023. doi:10.1109/EMBC40787.2023.10340176.

37. Kuang Y, Cheng Z, Zhang J, Yang C, Zhang Y. Risk factors and predictive model construction for lower extremity arterial disease in diabetic patients. *PLoS ONE*. 2024;19(12):e0314862. doi:10.1371/journal.pone.0314862.

38. Maniruzzaman M, Islam MM, Rahman MJ, Hasan MAM, Shin J. Risk prediction of diabetic nephropathy using machine learning techniques: a pilot study with secondary data. *Diabetes Metab Syndr*. 2021;15(5):102263. doi:10.1016/j.dsx.2021.102263.

39. Xiao L, Tang L, Kuang W, et al. Risk prediction of integrated traditional Chinese and western medicine for diabetes retinopathy based on optimized gradient boosting classifier model. *Medicine (Baltimore)*. 2024;103(51):e40896. doi:10.1097/MD.0000000000040896.

40. Bora A, Tiwari R, Bavishi P, et al. Risk stratification for diabetic retinopathy screening order using deep learning: a multicenter prospective study. *Transl Vis Sci Technol*. 2023;12(12):11. doi:10.1167/tvst.12.12.11.

41. Song X, Waitman LR, Hu Y, Yu ASL, Robins D, Liu M. Robust clinical marker identification for diabetic kidney disease with ensemble feature selection. *J Am Med Inform Assoc*. 2019;26(3):242-253. doi:10.1093/jamia/ocy165.

42. Muharram AP, Tahapary DL, Lestari YD, Sarayar R, Dirjayanto VJ. Supervised learning models for early detection of albuminuria risk in type-2 diabetes mellitus patients. Paper presented at 2023 10th International Conference on Advanced

Informatics: Concept, Theory and Application (ICAICTA); 2023. doi:10.1109/ICAICTA59291.2023.10390334.

43. Zhao Y, Li X, Li S, et al. Using machine learning techniques to develop risk prediction models for the risk of incident diabetic retinopathy among patients with type 2 diabetes mellitus: a cohort study. *Front Endocrinol (Lausanne)*. 2022;13:876559. doi:10.3389/fendo.2022.876559.

44. Liu W, Hu X, Fang Y, et al. Development and external validation of an algorithm for self-identification of risk for microvascular complications in patients with type 1 diabetes. *Diabetes Obes Metab*. 2025;27(2):740-749. doi:10.1111/dom.16068.

45. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol*. 2017;5(10):788-798. doi:10.1016/S2213-8587(17)30221-8.

46. Mosa ASM, Thongmotai C, Islam H, Paul T, Hossain KSMT, Mandhadi V. Evaluation of machine learning applications using real-world EHR data for predicting diabetes-related long-term complications. *J Bus Anal*. 2022;5(2):141-151. doi:10.1080/2573234X.2021.1979901.

47. Wu Y, Dong D, Zhu L, Luo Z, Liu Y, Xie X. Interpretable machine learning models for detecting peripheral neuropathy and lower extremity arterial disease in diabetics: an analysis of critical shared and unique risk factors. *BMC Med Inf Decis Mak*. 2024;24(1):200. doi:10.1186/s12911-024-02595-z.

48. Shiren Y, Jiangnan Y, Xinhua Y, Xinye N. Interpretable prediction model for assessing diabetes complication risks in Chinese sufferers. *Diabetes Res Clin Pract*. 2024;209:111560. doi:10.1016/j.diabres.2024.111560.

49. Dagliati A, Marini S, Sacchi L, et al. Machine learning methods to predict diabetes complications. *J Diabetes Sci Technol*. 2017;12(2):295-302. doi:10.1177/1932296817706375.

50. Colmenares-Mejía CC, Rincón-Acuña JC, Cely A, et al. Multivariable prediction model of complications derived from diabetes mellitus using machine learning on scarce highly unbalanced data. *Int J Diabetes Dev Ctries*. 2024;44(3):528-538. doi:10.1007/s13410-023-01264-7.

51. Al-Sari N, Kutuzova S, Suvitaival T, et al. Precision diagnostic approach to predict 5-year risk for microvascular complications in type 1 diabetes. *EBioMedicine*. 2022;80:104032. doi:10.1016/j.ebiom.2022.104032.

52. Aminian A, Zajichek A, Arterburn DE, et al. Predicting 10-year risk of end-organ complications of type 2 diabetes with and without metabolic surgery: a machine learning approach. *Diabetes Care*. 2020;43(4):852-859. doi:10.2337/dc19-2057.

53. Ljubic B, Hai AA, Stanojevic M, et al. Predicting complications of diabetes mellitus using advanced machine learning algorithms. *J Am Med Inf Assoc*. 2020;27(9):1343-1351. doi:10.1093/jamia/ocaa120.

54. Munna MTA, Alam MM, Allayear SM, Sarker K, Ara SJF. Prediction model for prevalence of type-2 diabetes complications with ANN approach combining with K-fold cross validation and K-means clustering. In: Arai K, Bhatia R, eds. *Advances in Information and Communication. Vol. 69. Lecture Notes in Networks and Systems*. Springer International Publishing; 2020:1031-1045. doi:10.1007/978-3-030-12388-8_71.

55. Munna MdTA Alam MM, Allayear SM, Sarker K, Ara SJF. Prediction model for prevalence of type-2 diabetes complications with ANN approach combining with K-fold cross validation and K-means clustering. In: Arai K, Kapoor S, Bhatia R, eds. *Advances in Intelligent Systems and Computing*. Springer International Publishing; 2019:451-467. doi:10.1007/978-3-030-03402-3_31.

56. Nicolucci A, Romeo L, Bernardini M, et al. Prediction of complications of type 2 diabetes: a machine learning approach. *Diabetes Res Clin Pract*. 2022;190:110013. doi:10.1016/j.diabres.2022.110013.

57. Lagani V, Chiarugi F, Manousos D, et al. Realization of a service for the long-term risk assessment of diabetes-related complications. *J Diabetes Complications*. 2015;29(5):691-698. doi:10.1016/j.jdiacomp.2015.03.011.

58. Jian Y, Pasquier M, Sagahyroon A, Aloul F. Using machine learning to predict diabetes complications. Paper presented at 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART); 2021. doi:10.1109/BioSMART54244.2021.9677649.

59. Zanelli S, El Yacoubi MA, Hallab M, Ammi M. Transfer learning of CNN-based signal quality assessment from clinical to non-clinical PPG signals. Paper presented at 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021. doi:10.1109/EMBC46164.2021.9629640.