healthcare

MDPI

Systematic Review

Type 2 Diabetes Prediction Model in China: A Five-Year Systematic Review

Juncheng Duan * and Norshita Mat Nayan *

Institute of IR4.0, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
* Correspondence: p145106@siswa.ukm.edu.my (J.D.); norshitaivi@ukm.edu.my (N.M.N.)

Abstract

Background: China has the largest number of patients with type 2 diabetes (T2D) worldwide, and the chronic complications and economic burden associated with T2D are becoming increasingly severe. Developing accurate and widely applicable risk prediction models is of great significance for the early identification of and intervention in high-risk populations. However, current Chinese models still have many shortcomings in terms of methodological design and clinical application. **Objective:** This study conducts a systematic review and narrative synthesis of existing risk prediction models for type 2 diabetes in China, aiming to identify issues with existing models and provide references with which Chinese scholars can develop higher-quality risk prediction models. **Methods:** This study followed the PRISMA guidelines to conduct a systematic search of the literature related to T2D risk prediction models in China published in English journals from October 2019 to October 2024. The databases included PubMed, CNKI and Web of Science. Included studies had to meet criteria such as clear modeling objectives, detailed model development and validation processes, and a focus on non-diabetic populations in China. A total of 20 studies were ultimately selected and comprehensively analyzed based on model type, variable selection, validation methods, and performance metrics. **Results:** The 20 included studies employed various modeling methods, including statistical and machine learning approaches. The AUC values of the models ranged from 0.728 to 0.977, indicating overall good predictive capability. However, only one study conducted external validation, and 45% (9/20) of the studies binned continuous variables, which may have reduced the models' generalization ability and predictive performance. Additionally, most models did not include key variables such as lifestyle, socioeconomic factors, and cultural background, resulting in limited data representativeness and adaptability. **Conclusions:** Chinese T2DM risk prediction models remain in the developmental stage, with issues such as insufficient validation, inconsistent variable handling, and incomplete coverage of key influencing factors. Future research should focus on strengthening multicenter external validation, standardizing modeling processes, and incorporating multidimensional social and behavioral variables to enhance the clinical utility and cross-population applicability of these models. Registration ID: CRD420251072143.

Keywords: China; type 2 diabetes; risk prediction model; machine learning; generalization ability; external validation

check for updates

Academic Editor: Chrysi Koliaki

Received: 31 May 2025
Revised: 3 August 2025
Accepted: 12 August 2025
Published: 15 August 2025

Citation: Duan, J.; Nayan, N.M. Type 2 Diabetes Prediction Model in China: A Five-Year Systematic Review. *Healthcare* **2025**, *13*, 2007. https://doi.org/10.3390/healthcare13162007

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Diabetes is one of the most serious and prevalent chronic diseases today, with type 2 diabetes (T2D) accounting for more than 90% of all cases [1]. As T2D progresses, patients

face a high risk of complications, including blindness, renal failure, myocardial infarction, stroke and premature death, with an 84% higher risk of heart failure compared to non-diabetic patients [2]. According to the International Diabetes Federation's Diabetes Atlas, 11th edition (2025), an estimated 589 million adults aged 20–79 years had diabetes globally in 2024 (≈11.1% of the adult population, or 1 in 9), and this number is projected to rise to 853 million by 2050. In the same year, diabetes caused 3.4 million deaths globally and caused at least USD 1 trillion in annual healthcare costs. In China, the burden of diabetes is growing even more dramatically, with the number of adults (aged 20–79 years) with diabetes surging from 22.6 million in 2000 to 90 million in 2011, reaching 148 million in 2024, and projected to climb to 168.3 million by 2050. As one of 37 countries in the Western Pacific region, China currently has the largest number of adults with diabetes in the world and the second highest diabetes-related healthcare expenditure in the world [3], a trend that underscores the urgent need to develop precise diabetes prevention strategies tailored to the Chinese population. The clinical symptoms of type 2 diabetes are usually less pronounced. As a result, the disease may be diagnosed several years after its onset, by which time complications have already developed. Thus, delayed diagnosis is a key factor affecting overall disease manageability and risk of complications [4]. The timely screening and management of people at risk for diabetes is important in reducing the incidence of diabetes [5].

The development of effective early prediction models for T2D can help to identify at-risk individuals and provide valuable insights for clinical decision-making, enabling the early implementation of targeted prevention strategies [6]. Globally, the development of diabetes risk prediction models has undergone a gradual evolution from traditional statistical methods to artificial intelligence algorithms. Early models mainly used methods such as logistic regression to construct concise risk scoring tools (e.g., FINDRISC, QDiabetes, and Framingham models), emphasizing interpretability among variables and simplicity of clinical application [7]. With the improvement in computing power and the application of big data technology, a series of machine learning models have emerged in recent years, such as Random Forest, XGBoost, Support Vector Machines (SVMs), and Deep Neural Networks (DNNs), which have demonstrated significant advantages in the modeling of variable interactions, nonlinear relationship capture, and prediction performance [8]. Li [9] highlighted alterations in the TGF-β/Smad, NF-κB, PI3K/AKT and AMPK pathways in diabetic cardiomyopathy, providing a rationale for integrating cardiovascular, molecular and microbial biomarkers into multimodal hybrid models for T2D risk prediction. These personalized prediction tools have shown promising results [10,11], all of which are widely used in epidemiology, clinical screening, and dynamic risk assessment.

However, despite the continuous development of modeling methods, current T2D risk prediction models in China still have major shortcomings in terms of generalization ability and clinical applicability. First, most studies lacked external validation, and models were only tested in the original dataset or similar populations, limiting their ability to generalize to different regions and populations [12]. Second, most studies treat continuous variables (e.g., BMI, blood pressure, etc.) in artificial groups, which facilitates clinical interpretation, but this results in a significant loss of variable information, reduces model prediction accuracy, and is prone to introducing nonlinear relationship bias [13]. In addition, there is a general lack of standardization in the treatment of missing data, often using simple filling or sample deletion, and sensitivity analyses are rarely performed, leading to potential result bias [14]. More notably, social determinants of health (SDOH, e.g., education level, income level, healthcare accessibility, urban–rural differences, etc.) have a significant impact on the risk of developing T2D and management outcomes in Chinese populations. This impact has been grossly underestimated in most of the studies, despite the fact that

there is sufficient evidence to show that these factors have important correlations with metabolic chronic diseases [15]. Sung and Lee [16] systematically reviewed the relationship between multiple SDOH and T2D in Asia, emphasizing the need for culturally appropriate interventions and longitudinal studies; Hu [17] constructed and compared five machine learning models based on SDOH data from 26,298 adults in Fujian Province, revealing the relative importance of each SDOH variable in risk prediction; Lan [18] analyzed the impact of SDOH on self-management behaviors in 495 patients with T2D in Zhejiang Province, and found that education, residential environment, and social support were significantly associated with self-care; Zhao [19] conducted an urban–rural difference study in 3225 older adults in Yunnan Province to quantify the role of SDOH factors such as lifestyle and residential environment on T2D and glucose tolerance abnormalities. Meanwhile, more and more studies have begun to explore the predictive value of molecular and microbial markers. Chang [20] systematically evaluated the hypoglycemic effect of Myrica rubra pomace polyphenols (MRPP) in db/db mice, and found that MRPP exerted multiple physiological effects through the PI3K/AMPK signaling pathway and the remodeling of gut flora, providing a new idea of introducing molecular and microbial signatures into a T2D risk model. The problems mentioned above are particularly prominent in China, and the number of relevant studies in China is currently very limited. There is an urgent need to create models suitable for Chinese populations [21].

The main contributions of this study are as follows. First, it systematically reviews and synthesizes research on type 2 diabetes(T2D) risk prediction models developed in China over the past five years, with a focus on model construction methods, key predictive factors, and performance outcomes. Second, it provides a critical evaluation of existing models, highlighting common issues such as data heterogeneity, insufficient external validation, and the inappropriate handling of continuous variables. Third, the study compares traditional statistical approaches with machine learning-based methods, exploring their respective strengths, limitations, and application contexts. Finally, it proposes directions for future research, emphasizing the need to enhance external validation, incorporate social determinants of health variables, and improve data standardization to strengthen both the generalization ability and clinical applicability of the models.

2. Methods

2.1. Search Strategy

This study was conducted in accordance with the latest PRISMA guidelines 2020 [22]. A PRISMA checklist is provided in Supplementary File S2 to ensure comprehensive reporting. This systematic review was pre-designed and registered in the PROSPERO database prior to the commencement of this study (Registration ID: CRD420251072143). To ensure the comprehensiveness and scientific validity of this systematic literature review, a search was conducted in November 2024 in the three major electronic databases, PubMed, Web of Science, and China National Knowledge Infrastructure (CNKI), covering biomedical sciences, computer sciences, and multidisciplinary cross-cutting fields. Through the combined use of these databases, we were able to retrieve a wide range of studies related to type 2 diabetes prediction and early intervention models in China. In terms of search strategy, we designed a comprehensive keyword combination to ensure the accuracy and breadth of the search. The core search terms include 'type 2 diabetes' ('Type 2 Diabetes' OR 'T2D' OR 'Diabetes Mellitus Type 2') and terms related to prediction models (e.g., 'Prediction Model' OR 'Predictive Model' OR 'Risk Model'). To ensure the geographical and population relevance of the literature, keywords related to China (e.g., 'China' OR 'Chinese') were also added to the search. Boolean logic operators were used extensively during the search process, e.g., the AND operator was used to connect different topics such as T2D,

predictive modeling, and China to ensure that the retrieved literature contained both of these key elements, while the OR operator was used to extend the search to cover different terminological expressions that might appear. See Supplementary File S1 for detailed search terms. With this strategy, we were able to find literature related to prediction models and the early intervention of type 2 diabetes in China from multiple dimensions, avoiding missing potentially important studies.

2.2. Inclusion/Exclusion Criteria

To ensure that the timeframe of the studies reflects the latest advances in recent years, we set the timeframe of the search from 1 October 2019 to 1 October 2024. This period not only covers research on predictive models and early intervention strategies for type 2 diabetes over the past five years, but also ensures that we have access to the most recent research findings. The choice of this period is based on the observation of the trend of the application of emerging technologies such as machine learning and artificial intelligence in the field of type 2 diabetes, as these technologies have been widely used in prediction modeling and have gradually matured during this period, especially in countries such as China [23], which has a large population base and abundant health data resources. Additionally, the quantity and quality of related studies have increased significantly [24]. In order to screen for literature of high quality and relevance, we set strict inclusion and exclusion criteria:

Inclusion Criteria: (1) The study population was a Chinese population without diabetes mellitus at baseline. (2) The study was directed at constructing predictive models (excluding diagnostic models) for the risk of developing T2D, and the process of model development, validation, and evaluation was described in detail. (3) The type of study design was a cross-sectional, case–control, or cohort study. (4) The study's purpose was explicitly focused on the development or validation.

Exclusion Criteria: (1) The study was not conducted on the Chinese population. (2) Studies looked at specific high-risk populations (e.g., obese, hypertensive, etc.). (3) Studies had a predictive endpoint of a combined model of multiple diseases that included T2D, but this was not its only outcome (e.g., combined prediction of cardiovascular disease). (4) Studies focused on predictive models of T2D complications (e.g., retinopathy, nephropathy, etc.). (5) Documents were abstracts of international conferences only, without the full text. (6) Studies were conducted at the molecular, cellular, and genetic levels. (7) Publications were duplicates.

2.3. Risk of Bias and Applicability Assessment

To ensure the reliability of the review results, we conducted a systematic assessment of the 20 included studies using the PROBAST tool, covering the four major bias domains—participants, predictors, outcome, and analysis. Additionally, three applicability domains—participants, predictors, and outcomes—were assessed for fit. Each domain was labeled with "+" (low risk of bias/high applicability), "−" (high risk of bias/low applicability), or "?" (unclear), and the overall risk of bias and overall suitability were synthesized based on the ratings of each domain [25]. This process, which is described in more detail in Section 3.4. "Literature quality assessment," ensures that the quality of the study is fully understood and that the quality of the study can be assessed by considering it at the time of synthesis. This process ensures a comprehensive understanding of the quality of the research and helps to improve the scientific validity and reliability of the conclusions by taking into account potential systematic errors in the synthesis.

2.4. Data Synthesis

Two researchers independently screened the literature, extracted information, and cross-checked results; if there was disagreement, a third party was consulted. For the initial screening of the literature, the title and abstract were read first, and after excluding obviously irrelevant literature, the full text was further read through to finalize the inclusion based on the inclusion and exclusion criteria [26]. For literature management, we use literature management software (e.g., EndNote 21.5 or Zotero 7.0.16) to organize and filter the retrieved literature. After completing the literature search and screening, we conducted a systematic data extraction of the included literature. The purpose of data extraction was to collect relevant information from each piece of literature in order to compare and synthesize different studies [27]. Specifically, we designed a standardized data extraction form, including the following core elements: The first element was the basic information in the studies, such as authors, year of publication, and location of the study. This was used in order to clarify the background and spatial–temporal characteristics of each piece of the literature. The second element included the objectives and methodology of the studies, with a focus on extracting specific designs regarding the prediction models of T2D, such as the type of model, selection of predictor variables, data sources, sample size, and validation methods of the model [28]. In addition, the extraction also pays special attention to the demographic characteristics of the study population, such as age, gender, health status, etc., which may affect the applicability and prediction effect of the model [29]. Given the substantial heterogeneity among the included studies in terms of study design, predictor selection, validation methods, and performance metrics, this review only employed narrative synthesis and did not conduct a quantitative meta-analysis.

3. Results

3.1. Literature Screening Process and Results

The search yielded 1080 relevant studies, which were screened layer by layer, resulting in the inclusion of 20 studies. The literature screening process is shown in Figure 1.

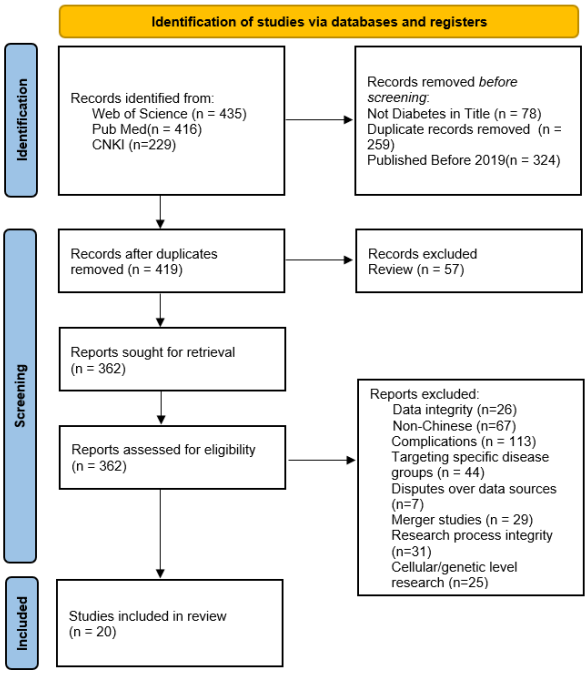


Figure 1. PRISMA flow diagram of article search and selection.

After screening the titles and abstracts, 362 documents proceeded to the full-text assessment phase. According to the pre-defined exclusion criteria in Methods Section 2.2, we reviewed each of these full texts and finally excluded 342 studies. The exclusion categories and corresponding reasons are shown in Supplementary File S1.

Figure 2 shows the five-year trend in published papers. The figure illustrates the popularity of predictive modeling in medical research. Of these 20 articles, the highest number of studies were published in 2024, with 6 (30.0%). This was followed 2020 and 2023, with 4 (20.0%) published. Finally, there was a decreasing number of studies published in 2022, 2021, and 2019, with 3 (15.0%), 2 (10.0%), and 1 (5.0%) published in successive years, respectively. The figure shows that the number of studies on early prediction models for type 2 diabetes first increased over time, gradually decreased after 2020, and then increased year by year, with a general trend of gradual incremental increase, indicating that the topic has become increasingly important to researchers in recent years.

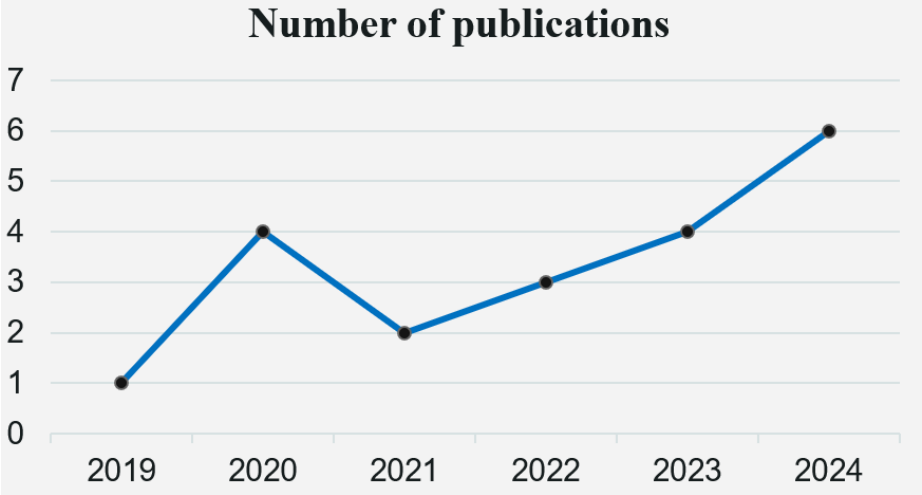


Figure 2. Year-wise distribution of publications relevant to studies.

3.2. Basic Characteristics of the Included Literature

All 20 included studies were conducted in China; 3 was prospective, and the remaining 17 were retrospective. Sample sizes (excluding missing data) ranged from 936 to 4,075,431 and the number of patients who experienced outcome events ranged from 99 to 301,347.

As shown in Table 1, fasting blood glucose (FBG) $\geq$ 7.0 mmol/L was used as an observational endpoint in 17 studies, glycosylated hemoglobin (HbA1c) $\geq$ 6.5% was used in 9 studies, 2 h postprandial glucose (2h-PG) $\geq$ 11.1 mmol/L was used in 8 studies, and random blood glucose $\geq$ 11.0 mmol/L was used in 2 studies.

These studies, published between 2019 and 2024, reflect recent advances in T2D risk prediction research in China. The high frequency of FBG $\geq$ 7.0 mmol/L as an endpoint suggests its leading role in diabetes screening and diagnosis. Several studies have included multiple diagnostic criteria simultaneously, highlighting the multifactorial nature of T2D risk assessment. The wide range of sample sizes and data sources, from national health databases to local hospitals, further enhances the generalizability of the findings.

Table 1. Basic characteristics of included studies on risk prediction models for T2D in China.

First Author	Year of Publication (Year)	Country	Research Type	Age of the Research Subjects (Years)	Sample Source	Sample Size	Number of Patients with the Occurrence of Endpoint Events	Observation Endpoint	References
Xu et al.	2022	China	Retrospective study	12~94	Nanjing Drum Tower Hospital Health Management Center	15,166	623	①②⑤	[30]
Lin et al.	2024	China	Retrospective study	≥60	Nanjing Shengrun Hospital	D:1564 V:671	99	②	[31]
Wang et al.	2023	China	Retrospective study	≥18	Monitoring Data on Chronic Disease Risk Factors Among Residents of Dongguan City	4106	149	①③④⑤	[32]
Liu et al.	2024	China	Retrospective study	—	National Health Examination Center Database Suzhou	D: 32,372 V: 13,875	D: 411 V: 205	①②③⑥	[33]
Yang et al.	2024	China	Retrospective study	≥20	University First Affiliated Hospital Health Checkup Center	D: 3221 V: 1381	760	①②③	[34]
Tong et al.	2023	China	Retrospective study	≥18	Sichuan Provincial People's Hospital	980	513	①②③④⑥	[35]
Shao et al.	2020	China	Retrospective study	20~80	China Health and Nutrition Survey Guangzhou	D: 4498 V: 1525	D: 257 V: 92	①②③④⑤	[36]
Jiang et al.	2024	China	Retrospective study	50~75	Haizhu District Grassroots Community Service Management Information System	252,176	—	①	[37]
Li et al.	2023	China	Retrospective study	≥18	national physical examination (NPE) project	4,075,431	301,347	①③⑤	[38]
Wang et al.	2021	China	Retrospective study	≥18	2018 Health Checkup Data for All Residents of Ili Kazakh Autonomous Prefecture, Xinjiang	D: 366,523 V: 91,630	D: 30,758 V: 7577	①③	[39]
Dong et al.	2022	China	Retrospective study	18~84	Department of Health, Government of the Hong Kong Special Administrative Region	1857	280	①②④⑤	[40]
Liu et al.	2022	China	Retrospective study	≥65	Wuhan Elderly Health Screening Data	127,031	—	①	[41]
Hu et al.	2019	China	Prospective study	—	Retired employees of Dongfeng Motor Corporation (DMC) in Shiyan City, Hubei Province, China	4833	171	①②⑤	[42]
Yang et al.	2024	China	Retrospective study	43~102	Health check-up data for middle-aged and elderly people in Hongguang Street, Pidu District, Chengdu	7602	434	①②④	[43]
Long et al.	2024	China	Prospective study	≥18	Jinchuan Group Staff Hospital	D: 22,025 V: 9438	—	①⑤	[44]

Table 1. *Cont.*

First Author	Year of Publication (Year)	Country	Research Type	Age of the Research Subjects (Years)	Sample Source	Sample Size	Number of Patients with the Occurrence of Endpoint Events	Observation Endpoint	References
Ma et al.	2020	China	Retrospective study	—	Beijing Huazhao Yisheng Health Checkup Data	D: 4754 V: 2375	—	①	[45]
Miao et al.	2020	China	Retrospective study	—	Physical examination data from a hospital in China	936	—	—	[46]
Ma et al.	2020	China	Prospective study	38~88	A cohort survey of chronic cardiovascular diseases in Fangshan District, Beijing, China	3127	187	①③④⑤	[47]
Ouyang et al.	2021	China	Retrospective study	≥18	Southern Hospital Health Management Center	36,292	2244	①⑤	[48]
Wu et al.	2023	China	Retrospective study	18~80	Fujian Shishi Community Health Center	165,263	—	—	[49]

Note:—indicates no special requirements; D = modeling cohort, V = validation cohort, T2D = type 2 diabetes; ① indicates fasting blood glucose (FBG) ≥ 7.0 mmol/L, ② indicates glycated hemoglobin (HbA1c) ≥ 6.5%, ③ indicates postprandial 2 h blood glucose (2 h-PG) ≥ 11.1 mmol/L, ④ indicates receiving hypoglycemic treatment, ⑤ indicates self-reported diabetes, ⑥ indicates random blood glucose ≥ 11.1 mmol/L; meeting any of the following criteria is sufficient to diagnose T2D.

3.3. Basic Features Included in the Prediction Model

3.3.1. Establishment and Validation of the Model

In terms of variable selection methods, among the 20 relevant studies [42] included, the methods of variable selection showed diversified characteristics, which included both traditional statistical methods and some modern machine learning techniques. Overall, 13 studies first conducted univariate analysis to initially screen variables related to diabetes using *t*-tests, chi-square tests, and other means. This link is particularly common in retrospective studies, facilitating the rapid targeting of factors that may have predictive power [50], and multifactor analysis was also applied in 14 studies, mainly based on multivariate logistic regression or Cox proportional risk modeling, to identify independent predictors under the control of other variables and to quantify the relative contribution of each variable to the risk of diabetes incidence. In the further modeling stage, 5 studies used stepwise regression for variable selection, and 6 studies used LASSO regression. For some complex models, Jiang [37] applied SelectKBest, RFE, the Boruta algorithm, and SHAP value interpretation to assess the contribution of variables to the predictive model, taking into account both precision and interpretability. Yang [43], on the other hand, combined meta-analysis and AUC sorting to optimize variables from the evidence base, and these methods were combined with VIF to exclude covariates to construct a representative prediction model.

Overall, 11 studies maintained the continuity of continuous variables, and 9 studies converted all continuous variables into categorical variables. The AUC values of the included models ranged from 0.728 to 0.977, proving that the models in the 20 studies had good predictive performance. In terms of model validation, 0 studies validated the models externally only, 19 studies validated the models internally only, and only 1 study validated the models using a combination of internal and external validation. In addition, 4 studies used the HL goodness-of-fit test to assess calibration. Overall, 14 studies considered model overfitting and calibrated the model accordingly, as shown in Table 2.

Table 2. Basic characteristics of development and validation included risk prediction models for T2D in China.

First Author	Modeling Methods	Variable Selection Methods	Methods for Handling Continuous Variables	AUC (95%CL)	Verification Method	Calibration Method
Xu et al.	LR	LASSO regression	Maintain continuity	0.865 (0.847, 0.865)	Internal verification	Calibration curve
Lin et al.	LR	LASSO regression	Maintain continuity	D: 0.824 (0.765, 0.883) V: 0.809 (0.732, 0.886)	Internal verification	Calibration curve
Wang et al.	LR, DT (CART, C4.5), BPNN, SVM, DNN	Univariate analysis, stepwise selection	Maintain continuity	Model 1: 0.962 Model 2: 0.906 Model 3: 0.888 Model 4: 0.977 Model 5: 0.911 Model 6: 0.845	Internal verification	—
Liu et al.	XGBoost, SVM, LR, RF	Univariate and multivariate analysis	Maintain continuity	Model 1: D: 0.986 V: 0.812 Model 2: D: 0.896 V: 0.668 Model 3: D: 0.914 V: 0.913 Model 4: D: 0.998 V: 0.838	Internal verification	Hosmer–Lemeshow test calibration curve
Yang et al.	LR	Univariate and multivariate analysis, stepwise selection	Convert to categorical variable	0.800 (0.770, 0.829)	Internal verification	Calibration curve
Tong et al.	RF, MLP, XGBoost, LGBM, CB	Univariate and multivariate analysis, LASSO regression	Maintain continuity	Model 1: 0.840 Model 2: 0.816 Model 3: 0.848 Model 4: 0.852 Model 5: 0.850	Internal verification	Calibration curve
Shao et al.	LR	Univariate and multivariate analysis, LASSO regression	Maintain continuity	Model 1: 0.788 (0.761, 0.816) Model 2: 0.807 (0.780, 0.834) Model 3: 0.905 (0.879, 0.932) Model 4: 0.882 (0.853, 0.912)	Internal validation and external validation	Calibration curve and bootstrap resampling
Jiang et al.	RF, KNN XGBoost VC	—	Maintain continuity	—	Internal verification	Calibration curve and bootstrap resampling
Li et al.	CART, LGBM, RF, XGBoost TabNet, MLP, LR	Univariate and multivariate analysis	Convert to categorical variable	Model 1: 0.884 Model 2: 0.881 Model 3: 0.873 Model 4: 0.912 Model 5: 0.876 Model 6: 0.875 Model 7: 0.816	Internal verification	Calibration curve
Wang et al.	LR	Univariate, multivariate analysis, LASSO regression	Convert to categorical variable	D: Male: 0.894 Woman: 0.816 V: Male: 0.865 Woman: 0.815	Internal verification	Hosmer–Lemeshow test calibration curve
Dong et al.	LR, XGBoost	Univariate and multivariate analysis, stepwise selection	Convert to categorical variable	Model 1: 0.812 (0.769, 0.853) Model 2: 0.822 (0.779, 0.863)	Internal verification	Hosmer–Lemeshow test Calibration curve
Liu et al.	LR, DT, RF, XGBoost	Univariate analysis, LASSO regression	Convert to categorical variable	Model 1: 0.760 Model 2: 0.728 Model 3: 0.777 Model 4: 0.780	Internal verification	Calibration curve
Hu et al.	Cox	Univariate and multivariate analysis	Convert to categorical variable	D: 0.850 V: 0.830	Internal verification	—

Table 2. *Cont.*

First Author	Modeling Methods	Variable Selection Methods	Methods for Handling Continuous Variables	AUC (95%CL)	Verification Method	Calibration Method
Yang et al.	LR	Univariate and multivariate analysis	Convert to categorical variable	0.794 (0.771, 0.816)	Internal verification	Calibration curve
Long et al.	Cox	Univariate and multivariate analysis	Convert to categorical variable	D: 3 year: 0.783 5 year: 0.825 7 year: 0.842 V: 3 year: 0.782 5 year: 0.805 7 year: 0.807	Internal verification	Calibration curve
Ma et al.	RF, LR, SVM, DT, Naive Bayes (NB)	Multivariate analysis	Maintain continuity	Model 1: 0.931 Model 2: 0.903 Model 3: 0.813 Model 4: 0.776 Model 5: 0.858	Internal verification	—
Miao et al.	SVM (PSO-FWSVM)	Multivariate analysis	Maintain continuity	—	Internal verification	—
Ma et al.	LR	Multivariate analysis and stepwise selection	Convert to categorical variable	Original model: 0,878 (0.853, 0.903) Model 1: 0.880 (0.856, 0.903) Model 2: 0.880 (0.855, 0.903) Model 3: 0.879 (0.854, 0.903)	Internal verification	Hosmer–Lemeshow test Calibration curve
Ouyang et al.	LR, LGBM	Univariate and multivariate analysis, stepwise selection	Maintain continuity	Model1: 0.906 Model2: 0.910	Internal verification	—
Wu et al.	LR LGBM	—	Maintain continuity	—	Internal verification	—

Overall, 14 techniques were extracted from the SLR. As can be seen from the results in Figure 3, machine learning (ML) was the preferred choice, with only a few researchers exploring mathematical and deep learning (DL) approaches. Overall, 18 of the studies (90.0%, 18/20) chose ML to construct prediction models for diabetes progression, while 2 articles (10.0%, 2/20) considered mathematics, and 3 papers used DL (15.0%, 3/20). Figure 3 shows the 12 most commonly used models in the field of diabetes progression prediction. The most commonly used method was logistic regression (LR) (75.0%, 15/20). This is followed by Extreme Gradient Boosting (XGB) (30.0%, 6/20) and Random Forest (RF) (30.0%, 6/20), then Decision Trees (DT), Support Vector Machines (SVM), and Light Gradient Boosting Machines (LGBM), which have almost equal distributions (20.0%, 4/20), and then Multilayer Perceptron (MLP) and Cox Proportional Risk (COX), which are also almost equally distributed (10.0%, 2/20). These are followed by Artificial Neural Networks (ANN), K-Nearest Neighbor Algorithm (KNN), Deep Neural Networks (DNN) and Naive Bayes (NB), which are also almost equally distributed (5.0%, 1/20). Table 3 is a fully enriched comparison table used for all twelve modeling methods, with detailed strengths and weaknesses and classic supporting references for each:

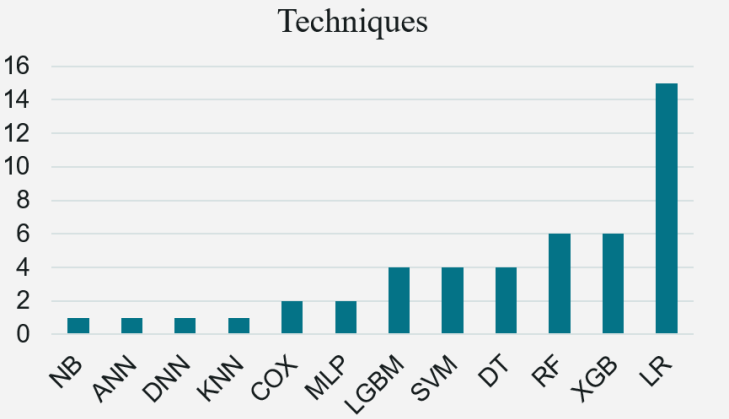


Figure 3. The frequency of prediction models applied by selected articles.

Table 3. Comparative analysis of screened models.

Method	Usage Count	Strengths	Weaknesses	Reference
LR	15	• Simple, fast and interpretable • Low computational cost • Naturally handles binary outcomes	• Assumes linear log-odds relationship • Cannot capture complex nonlinearity without manual feature engineering • Sensitive to outliers and high-leverage points	[30,31,34,51,52]
XGBoost	6	• Highly efficient and scalable gradient boosting • Ability to be flexible and adjust to mission needs	• Demands careful hyperparameter tuning • Less transparent than traditional models	[33,35,53]
RF	6	• Robust to overfitting • Easily handle high-dimensional data and effectively cover complex relationships between variables	• Less interpretable than single decision trees • Prediction of large datasets is slow and costly	[35,37,54]
DT	4	• Easy to visualize and interpret • Handles both numeric and categorical inputs	• Prone to overfitting without pruning • Unstable, small data changes can alter splits	[32,55]
SVM	4	• Robust to overfitting; excels in high-dimensional spaces • Handles complex, nonlinear patterns via kernel trick	• Computationally expensive on large datasets • Difficult to interpret model parameters • Highly sensitive to kernel choice and hyperparameters	[32,33,56]
LGBM	4	• Fast training and low memory footprint • Better able to handle large datasets • Excellent predictive performance	• Can overfit on smaller datasets • Less community support and tools than XGBoost	[48,49,57]
MLP	2	• Flexible at modeling complex, nonlinear relationships • Once trained, enables fast, real-time predictions	• Prone to overfitting without sufficient data • Demands careful tuning of network architecture and hyperparameters	[38,58]

Table 3. *Cont.*

Method	Usage Count	Strengths	Weaknesses	Reference
COX	2	• Produces interpretable hazard ratios • Widely used in survival analysis for medical research • Ability to effectively process incident event data	• Assumes proportional hazards, which may not hold • Limitations in dealing with complex relationships between predictors	[44,55]
ANN	1	• Automatic learning and fitting of complex nonlinear relationships • Flexible network structure, can adjust the number of layers and nodes for different problems	• Black box model, poor interpretability, not easy to understand the internal decision-making mechanism. • Highly sensitive to the amount of data and hyper-parameter settings, easy to overfitting when the sample is insufficient.	[59]
DNN	1	• Learns hierarchical features, capturing deep nonlinear pattern • Automatically extracts complex interactions without manual feature engineering	• Prone to overfitting in complex architectures • Computationally expensive and requires large amounts of data • 'Black-box' nature makes interpretation very difficult	[60]
KNN	1	• Simple and non-parametric behavior • Can handle multiclass classification problems • Robust to outliers and nonlinear relationships	• Computationally expensive for large datasets • Requires careful selection of the number of neighbors (k) and distance metric • Sensitive to irrelevant features and high-dimensional data	[61]
NB	1	• Simple and computationally efficient • Performs well with high-dimensional data	• Assumes proportional hazards, which may not hold • Not well-suited for capturing complex relationships	[45,62]

3.3.2. Performance of Predictive Factors in the Model and Research Limitations

A compilation and analysis of the predictors from these 20 studies revealed that the predictors included in the models ranged from 8 to 47. The predictors were categorized into three main groups: demographic factors, physical examination indicators, and laboratory tests. Among them, demographic factors such as age, gender, and family history of diabetes were more common, while physical examination indicators such as body mass index (BMI) and waist circumference were more common. Laboratory indicators are common, such as FBG, HbA1c, triglycerides (TG), etc. Most of the studies were based on physical examination or laboratory test data, and widely included traditional biological variables such as age, gender, BMI, fasting blood glucose (FBG), blood lipids (e.g., TG, HDL-C, LDL-C), and indicators of liver and renal function (e.g., ALT, CREA, BUN). These were used as predictors. However, lifestyle variables such as diet, exercise, smoking, alcohol consumption, and sleep, which are closely related to the development of T2D, were only systematically included in a few studies (e.g., Shao et al. and Jiang et al.); key risk factors recommended by the guidelines, such as family history, a history of gestational diabetes mellitus, HbA1c, and OGTT, were omitted from a number of studies or failed to be introduced due to missing data. This limitation in variable coverage makes the predictive power of the model likely to be limited in reality [63]. In addition, about two-thirds of the studies were based on single-city, single-unit, or single-center data, with poorly representative samples and bias problems such as unbalanced sex ratios or incomplete physical examination data, as shown in Table 4.

Table 4. Predictors, presentation and limitations of included risk prediction models for T2D in China.

First Author	Predictor Factor	Limitations
Xu et al.	Gender, Age, BMI, ALT, CREA, CHOL, HDL, GLU, MCHC, WBC,	Predicting the risk of type 2 diabetes solely based on laboratory data does not include factors such as diet, exercise, or genetics, which have been proven to be closely related to type 2 diabetes. Single-center data sources, lack of external validation.
Lin et al.	Age, Gender, BMI, FBG, ALT, ALT/AST, BUN, TG, Hb	Single-center data sources, lack of external validation, exclusion of key variables (such as family history and history of gestational diabetes), and the model's applicability being limited to the elderly population.
Wang et al.	Age, drinking, cereals, potatoes, beans, fruits, eggs, milk, poultry, fish, DBP, FPG, TC, TG, HDL-C, LDL-C	Single-center data sources, lack of external validation. Failure to incorporate common disease risk factors such as genetics and self-care conditions (physical activity, sleep duration, etc.) into the model.
Liu et al.	FPG, Age, TG, ALT, BMI, CR, DBP, gender, family	Lack of external validation, single source of data, non-inclusion of key indicators such as HbA1c.
Yang et al.	Gender, Age, BMI, Blood Glucose, HDL-C, LDL-C, Fatty liver, ALT/AST	Few women were included, resulting in an imbalanced male-to-female ratio. No external validation was conducted. The follow-up period was short.
Tong et al.	FBG, previous HbA1c values, having a rational and reasonable diet, health status scores, type of manufacturers of metformin, interval of measurement, EQ-5D scores, occupational status, Age	The sample size is small, and there are recall biases for some variables.
Shao et al.	Model 1: Age, gender, race, BMI, waist circumference, hypertension Model 2: Model 1 + diet (calories, carbohydrates, protein), exercise, sleep duration Model 3: Model 2 + FPG, HbA1c, TG, LDL, HDL Model 4: FPG, HbA1c, TG, LDL, HDL	The data only comes from the China Health and Nutrition Survey (CHNS), which has issues with re-gional and sample representativeness, and there is a lack of further external validation to assess the model's general applicability.
Jiang et al.	BMI, age, systolic BP, diastolic BP, staple food, exercise frequency, exercise time	The feature variables are not comprehensive enough.
Li et al.	Gender, age, ethnicity, EH, SS, HTN, CAD, PDM, WC, BMI, WBC, PLT, FBG, ECG, TC, TG, LDL-C, HDL-C	Using cross-sectional data cannot establish causal relationships, and the high heterogeneity and missing rates of health check-up data affect the model's test effectiveness.
Wang et al.	Age, FHOT, WC, TC, TG, BMI, HDLc, and history of hypertension.	It is not possible to analyze causal relationships from cross-sectional data, the regional limitations of data sources affect generalizability, and the model may not cover all risk factors for type 2 diabetes, which could lead to prediction bias.
Dong et al.	Model 1: Age, BMI, WHR, smoking status, sleep duration, vigorous recreational activity time per week, and fruit consumption per week Model 2: age, BMI, WHR, SBP, waist circumference, sleep duration, smoking status, and vigorous recreational activity time per week	This study did not include key risk factors such as family history of diabetes and gestational diabetes history, and the validation was limited to the same population sample, which restricted the comprehensiveness of the results.

Table 4. *Cont.*

First Author	Predictor Factor	Limitations
Liu et al.	Age, gender, education, marital status, hypertension, fatty liver, exercise, current smoking, BMI, WC, SBP, DBP, FPG, TC, TG, HDL-C, LDL-C, ALT, AST, TBIL, SCR, BUN, and SUA	Selection bias, omission of certain key risk factors (such as HbA1c and insulin), failure to use OGTT may lead to diagnostic bias, only internal validation was conducted and external validation is lacking.
Hu et al.	Age, gender, BMI, waist circumference, blood pressure, fasting blood glucose, lipid profile (TC, TG, HDL-C, LDL-C), serum uric acid, smoking and drinking status, physical activity, history of hypertension, and family history of diabetes.	Insufficient sample representativeness, lack of important predictive factors, internal validation only, short follow-up period, and potential bias in some self-reported data.
Yang et al.	Age, gender, BMI, waist circumference, triglycerides, HDL-C, smoking status, drinking status, history of hypertension, and family history of diabetes	Insufficient sample representativeness, exclusion of certain key risk factors, limitations of diagnostic methods, internal validation only, and potential biases in lifestyle data.
Long et al.	Sex, age, body mass index, alcohol consumption, alcohol abstinence, hypertension, triglycerides, HDL-C, glutamyl transferase, family history of diabetes mellitus, cholecystitis, gallbladder agenesis.	No external validation; no inclusion of lifestyle variables such as diet and exercise; single source of data.
Ma et al.	Forty-seven characteristics such as blood lipids, urinalysis, liver function, blood pressure, age, gender, and height	Single source of data, lack of external validation, many missing datasets.
Miao et al.	BMI, family history of diabetes, diastolic blood pressure, fasting blood glucose, total cholesterol, triglycerides, LDL, heart rate	Single data source, lack of external validation, risk of bias due to sample imbalance, weak interpretability of features.
Ma et al.	Smoking, history of lipid-lowering drug use, 2h-PG, FPG, BMI, family history of diabetes mellitus, abnormal blood pressure markers, history of hypertension drug use	Lack of external validation and lack of extrapolation; low number of incidence and inaccurate prediction of high risk; continuous variables all categorized for treatment, which may reduce prediction accuracy.
Ouyang et al.	Sex, age, BMI, waist circumference, heart rate, systolic blood pressure, diastolic blood pressure, FBG, uric acid, 4 biochemical indicators, 2 liver function indicators, 2 renal function indicators, and 17 routine blood tests, totaling 34 study indicators, were used as independent variables.	Single source of data, lack of external validation, failure to assess the calibration ability of the model, failure to include variables such as lifestyle behaviors, possible retrospective bias.
Wu et al.	Only 42 characteristics were noted, but no specific	No external validation, no reported AUC, ROC curves, lack of model calibration assessment, lack of model interpretability.

3.4. Literature Quality Assessment

In order to systematically evaluate the methodological quality of the 20 included early T2D risk prediction model studies, we applied the PROBAST tool to assess risk of bias across four domains (participants, predictors, outcome, and analysis) and applicability across three dimensions (participants, predictors, and outcome). The results are summarized in Table 5: only Shao et al. (2020) [36] was rated as low risk of bias ("+") in all four domains; the remaining 19 studies exhibited high risk of bias ("−") in at least the analysis domain, yielding an overall high risk of bias classification ("−"). With respect to applicability, all studies received a "+" rating for each dimension, indicating that the selected models broadly align with the characteristics and clinical context of the general Chinese adult population, and thus possess high applicability and strong potential for implementation.

Table 5. Evaluation of risk of bias and applicability of the included literature.

Study	ROB					Applicability			
	Participants	Predictors	Outcome	Analysis	Overall ROB	Participants	Predictors	Outcome	Overall Applicability
Xu et al.	+	+	+	−	−	+	+	+	+
Lin et al.	+	+	+	−	−	+	+	+	+
Wang et al.	+	+	+	−	−	+	+	+	+
Liu H et al.	+	+	+	−	−	+	+	+	+
Yang et al.	+	+	+	−	−	+	+	+	+
Tong et al.	+	+	+	−	−	+	+	+	+
Shao et al.	+	+	+	+	+	+	+	+	+
Jiang et al.	+	?	+	−	−	+	+	+	+
Li et al.	+	+	+	−	−	+	+	+	+
Wang et al.	+	?	+	−	−	+	+	+	+
Dong et al.	+	+	+	−	−	+	+	+	+
Liu et al.	+	+	+	−	−	+	+	+	+
Hu et al.	+	?	+	−	−	+	+	+	+
Yang et al.	+	+	+	−	−	+	+	+	+
Long et al.	+	+	+	−	−	+	+	+	+
Ma et al.	+	+	−	−	−	+	+	+	+
Miao et al.	+	+	?	−	−	+	+	+	+
Ma et al.	+	+	+	−	−	+	+	+	+
Ouyang et al.	+	+	+	−	−	+	+	+	+
Wu et al.	+	+	+	−	−	+	+	+	+

Note ROB indicates risk of bias; − indicates high risk of bias/low applicability; + indicates low risk of bias/high applicability; ? indicates unclear.

4. Discussion

In this paper, we conducted a systematic evaluation of Chinese studies on T2D prediction models. After a staged screening process, 20 studies were finally adopted. Its key findings are shown in Table 6 below. The AUC values of the included models ranged from 0.728 to 0.977, respectively, indicating that the models had good predictive effects. Although the number of studies on T2D risk prediction models in China has continued to grow in recent years, and advanced modeling methods such as machine learning have been gradually introduced; overall, the relevant studies are still in the developmental stage, and have not yet formed a mature system that is widely applicable to clinical practice. All included studies were at high risk of bias due to a variety of factors including optimism bias, the irrational treatment of missing data, the irrational treatment of continuous variables, unstandardized model assessment, and a lack of external validation [64]. This shows that research on T2D prediction modeling in China is still in its developmental stage.

Table 6. Summary of findings.

Aspect	Main Findings	Recommendation
Predictor diversity	Predictor variables focused on biological indicators, lack of SDOH (lifestyle factors, socioeconomic factors, etc.)	Add multidimensional predictors
Continuous variable processing	Dichotomizing or categorizing continuous variables leads to information loss and exacerbates bias.	Adopt methods that preserve continuity
Risk of bias assessment	Overall risk of bias is high, affecting model reliability	Strengthen risk of bias control and implement pre-registration and standardized development processes

Table 6. *Cont.*

Aspect	Main Findings	Recommendation
Model Validation and Application	Validation is mostly focused internally, lacking external multicenter validation and clinically friendly deployment tools	Conducted multicenter external validation and developed easy-to-use interfaces such as line charts, WeChat applets, EHR plug-ins, etc.
Statistics vs. machine learning	Adoption of Machine Learning Methods Growing Rapidly, but Interpretability and Clinical Embeddedness Remain to be Improved	Exploring Interpretable Hybrid Models and Optimizing for Clinical Needs

4.1. Homogenization of Predictors

As can be seen from Table 3, T2D risk prediction models usually include common predictors such as age, gender, body mass index (BMI), waist circumference, and fasting blood glucose (FBG). On the one hand, this reflects the important early warning value of the above variables in the pathogenesis of T2D, and suggests that clinical staff should pay great attention to the dynamic changes and comprehensive assessment of these indicators in daily screening and management. On the other hand, it also reflects that the current T2D risk prediction models have obvious homogenization in the selection of variables, resulting in the limited differentiation and applicability of the models. Therefore, future studies urgently need to further explore and integrate new personalized risk factors, such as glycated hemoglobin a (HbA1c), a family history of diabetes mellitus, education level, dietary structure, exercise level, sleep quality, mental health status, and socioeconomic factors, on the basis of the traditional predictors in order to improve the predictive performance and relevance of the models, and to explore new personalized predictors that can help to break through existing developmental 'bottlenecks'. These efforts improve the predictive performance of the model, enhance individualized treatment, and promote the development of T2D risk prediction in the direction of precision and individualization.

4.2. Treatment of Continuous Variables and High Risk of Bias

In the development of T2D risk prediction models, the choice between discretizing inherently continuous predictors (e.g., age, BMI, blood pressure) into categories ("binning") or retaining their continuous form is fundamental, as it directly affects information retention, statistical power, model calibration, and interpretability. Altman and Royston showed that dichotomizing a continuous variable at an arbitrary cut-point can reduce statistical power by up to one-third and introduce artificial threshold effects that impair calibration [65]. By contrast, Harrell recommends preserving continuous predictors and, when needed, applying methods such as restricted cubic splines or fractional polynomials to capture nonlinearity, thereby maximizing information use and improving both discrimination and calibration—provided that sample size and events-per-variable requirements are satisfied (Table 7) [66].

Although binning can be useful when well-established clinical thresholds exist or rapid risk stratification is desired, we advocate that, whenever sample size and events-per-variable criteria allow, researchers should retain continuous predictors and employ semi-parametric approaches to fully leverage data variability, enhance discrimination, and achieve more reliable calibration. Appropriate regularization or resampling (e.g., bootstrap) should accompany such models to guard against overfitting.

Table 7. Comparison of binning vs. continuous use of predictors in T2D risk prediction models.

Aspect	Binning	Continuous
Information Retention	Loses within-bin variability	Retains full numeric detail
Statistical Power	Substantially reduces statistical power when categorizing	Preserves full variability, maximizing power
Model Calibration	Risk estimates "jump" at bin boundaries, hindering smooth calibration	Spline- or polynomial-based fits yield smoother, more accurate calibration
Interpretability	Easy to explain cut-points and risk groups	Requires interpretation of coefficients or spline functions
Overfitting Risk	Simpler structure may reduce overfitting	Complex fits need regularization or cross-validation to avoid overfit
Sample Size Needs	Lower requirements but must ensure balanced counts per bin	Requires larger sample and EPV ≥ 10 to support reliable estimation

On the other hand, there is a high risk of bias in the current T2D prediction studies, most of which only used a single randomization for division into training and test sets without multiple resampling or multicenter external validation, and often did not provide sufficient description of the method of dealing with the missing values, and did not report the calibration curves, Brier scores, or decision curve analyses, which made it difficult to assess the generalization ability and stability of the models. Although most of the studies were highly relevant to the general Chinese adult population in terms of participants, predictors, and outcomes, the practical generalizability of model performance remains limited due to insufficient sample sizes or event-variable ratios, inconsistent variable screening criteria, and short follow-up periods.

4.3. Model Validation and Application

The validation process of a model is the main part of assessing performance. A T2D risk prediction model is used to verify whether the model is reliable and generalizable. Of the 20 studies included in this review, the majority of them only performed internal validation, and only one study used a combination of internal and external validation, with most of the internal validation using, e.g., the division of training and test sets, bootstrap resampling, or cross-validation to assess the discriminative and calibrative ability of the model. Among them, AUC (area under the curve) is the most commonly used assessment metric to measure the discriminative ability of predictive models. Regarding external validation, a critical limitation of current Chinese T2D risk prediction models is the general lack of external validation. External validation involves evaluating a model's performance in an independent dataset distinct from that used for development, which is indispensable for assessing its reproducibility and transportability to new patient populations. This process ensures that key metrics and calibration remain robust [67]. The TRIPOD statement for the transparent reporting of prediction models further identifies external validation as a key requirement to guard against overoptimistic performance estimates and to support clinical applicability [68]. But most models are still limited to data partitioning within the same population or region and lack validation support across populations and regions now. Failure to conduct external validation exposes models to overfitting, whereby spurious associations specific to the development cohort degrade predictive accuracy in other settings [69]. Empirical investigations have demonstrated that model discrimination often declines substantially upon external testing; for instance, Nieboer found that changes in c-statistic values between development and validation cohorts can exceed 0.1, undermining confidence in risk stratification and decision-making [70]. Without rigorous external validation, risk models may generate misleading predictions, erode clinician trust, hinder uptake in clinical guidelines, and ultimately compromise patient care and resource allocation.

In terms of practical application, only a few models have been constructed as column-line graph tools or have visual interfaces to facilitate healthcare professionals to directly assess individual diabetes risk; no studies have reported that the models have been embedded in real clinical pathways or followed up research to observe their long-term intervention effects. Future research should strengthen the following directions after the model has been established: firstly, we must conduct external validation studies in multicenter settings and different populations as most of the current research only stays at the level of internal validation, and external validation studies should be strengthened in the future, especially the assessment of the model generalization ability based on multicenter, large-sample, and heterogeneous population data. In order to ensure the reliability and generalizability of the T2D risk prediction model in different regions and populations in China, it is recommended to perform the following in the multicenter validation design:

1. Select representative centers, urban and rural medical institutions in each of the five major regions, east, west, south, north, and central, to ensure a variety of geographic regions and medical levels.

2. The operation process in each region should be unified. It is necessary to formulate a concise operation manual or document for this. Unified training, unified fasting blood glucose, BMI, blood pressure and other core indicators of the measurement time, method, and instrument can greatly reduce the systematic and human-caused bias.

3. Measure the number of events required by each center according to the principle of "10 new diabetes events correspond to 1 predictor", and set aside 10% of the lost visit rate.

4. Use a unified electronic data collection platform to upload data, and conduct regular calibration and random checks.

5. Each center independently calculates and summarizes the data from the AUC and calibration curves, and then combines the data and re-evaluates the overall performance of the model to identify regional differences and guide model optimization.

Additionally, it is necessary to develop risk assessment tools that are easy to deploy and use, such as apps, WeChat applets, or scoring modules, integrated into electronic medical record systems; additionally, scholars must to evaluate the intervention value of the model's predictions in real health management in conjunction with prospective follow-up visits. This will help to meet the needs of medical staff and community diabetes patients for the use of risk prediction tools for the onset of T2D.

4.4. Comparison of Traditional Statistical Methods and Machine Learning Prediction Methods

Based on the comparative analysis of the strengths and weaknesses of the models used in the studies included in Table 2, it can be seen that traditional statistical methods and machine learning-based prediction methods have their own features and strengths, which are suitable for different data characteristics and research purposes. Traditional statistical methods, such as logistic regression and the Cox Proportional Hazards Model, have good interpretability and can clearly quantify the independent associations between predictors and disease risk [10]. These methodological models are simple in structure, computationally efficient, and widely accepted in the clinical field, helping researchers and physicians to understand the model inference process [71]. However, traditional statistical methods usually assume linear relationships between variables and have limited modeling power when confronted with high-dimensional data, complex interactions between variables, or nonlinear patterns [72].

In contrast, machine learning-based prediction methods (e.g., Random Forest, XGBoost, Support Vector Machines, Neural Networks, etc.) have powerful pattern recognition and automatic feature extraction capabilities. Machine learning methods are able to capture higher-order interactions and nonlinear features in the data without the need for preset

variable relationships, and show better prediction performance in large-sample, multi-feature environments [73]. Some algorithms (e.g., LASSO, Elastic Net, Boruta) can also automatically complete variable selection and dimensionality reduction to improve model generalization. However, at the same time, the interpretability of machine learning models is weak, especially in deep learning models, which easily become 'black boxes' and face certain obstacles in the promotion and application in clinical practice [74]. In addition, machine learning methods have higher requirements on data quantity, data quality and parameter tuning, and without a strict validation strategy, there is also risk of overfitting and unstable results [75].

Overall, traditional statistical methods are suitable for scenarios with medium sample sizes, a limited number of variables, and studies that focus more on causal inference or clinical interpretability, while machine learning methods are more suitable for tasks with large scale datasets, complex variables, and studies that prioritize prediction accuracy. In the future, the development of T2D risk prediction models can be based on the rational selection of methods based on specific study designs, or combining the advantages of statistical and machine learning methods to develop hybrid modeling strategies with both efficient prediction and interpretability, in order to better serve the needs of clinical screening and individualized management [76].

4.5. Contrast with TRIPOD Reporting Standards

TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) is a checklist of 22 items that ensures studies developing or validating risk prediction models fully report critical elements, such as justification of sample size, strategies for handling missing data, discrimination metrics, calibration measures, decision curve analysis, and validation strategies. This allows readers to assess the validity and applicability of the study. In our review, Ouyang [48] did not describe how missing data were handled, in violation of TRIPOD Item 9; Wu [49] only reported discrimination metrics without presenting calibration-in-the-large or calibration slope, as required by Item 14a,b; Ma [45] provided neither justification of the sample size nor number of outcome events, in line with Item 8; and Miao and Zhu [46] only conducted internal validation, with no assessment of independent cohorts, falling short of Item 10. We therefore recommend that future articles on Chinese T2D prediction models rigorously adhere to the full TRIPOD checklist to guarantee comprehensive documentation of study design, analysis methods, and validation procedures. This will enhance reproducibility and facilitate clinical implementation.

4.6. Early Intervention Strategies

Early intervention strategies for high-risk individuals with or at risk for type 2 diabetes encompass lifestyle modification, pharmacotherapy, and community-based support designed to interrupt or delay dysglycemia during the critical window before overt disease develops. In the Finnish Diabetes Prevention Study, personalized counseling to reduce caloric and saturated-fat intake while engaging in at least four hours of moderate-intensity aerobic exercise per week yielded a 58% reduction in diabetes incidence over three years among participants with impaired glucose tolerance [77]. The U.S. Diabetes Prevention Program demonstrated that intensive lifestyle intervention achieved a 58% risk reduction over a mean 2.8-year follow-up compared with a 31% reduction with metformin, highlighting the paramount importance of behavioral change in early prevention [78]. Similarly, the Indian Diabetes Prevention Programme (IDPP-1) showed that lifestyle modification alone delayed progression to type 2 diabetes by nearly three years, with adjunctive low-dose metformin further enhancing outcomes [79]. Collectively, these landmark trials underscore that tailored dietary adjustments, regular physical activity, and judicious use of pharmacological

support—reinforced through community health education and behavioral coaching—can significantly reduce progression to type 2 diabetes and provide a robust evidence base for implementing tiered early intervention frameworks.

4.7. Perspective for Clinical Practice

International evidence underscores the pivotal role of dedicated case managers in T2D care. In the Netherlands, a narrative review of nurses specializing in lifestyle medicine demonstrated marked improvements in glycemic control and cardiometabolic outcomes under nurse-led lifestyle interventions [80]. In Riyadh, a retrospective follow-up study of 3060 patients with poorly controlled T2D who were managed by a multidisciplinary team led by case managers achieved a 15% reduction in mean HbA1c over six months, alongside significant decreases in LDL-C, total cholesterol, and blood pressure [81]. A Saudi Arabian randomized parallel-group trial involving a senior family physician, a clinical pharmacy specialist, a dietitian, a diabetic educator, a health educator, and a social worker reported a 27.1% relative decrease in HbA1c and significant improvements in fasting blood glucose, lipid profiles, and blood pressure over a median 10-month follow-up [82]. Additionally, a Chinese RCT of a nurse-led, integrative medicine–based, structured education program for individuals with newly diagnosed T2D showed a 0.32% reduction in HbA1c and significant improvements in self-management behaviors and self-efficacy at 12 weeks [83].

Translating these international models into the Chinese context will require several coordinated steps. It is essential to establish a formal certification pathway for lifestyle medicine case manager nurses so that they receive standardized training in behavior-change counseling and chronic disease management. Case managers must then be fully integrated into multidisciplinary primary care teams, working alongside physicians, dietitians, and pharmacists with clearly defined roles and workflows. Finally, robust performance dashboards should be implemented to continuously monitor key metrics such as HbA1c levels, complication rates, and healthcare utilization, thereby enabling data-driven quality improvement.

5. Implications for Future Research

Building on the issues and recommendations from the preceding discussion, future Chinese type 2 diabetes risk prediction studies should first expand predictor selection to include lifestyle, socioeconomic and behavioral determinants; second, they should employ modeling approaches that balance interpretability with predictive performance; third, they should preserve continuous variables using spline or other smooth methods to avoid information loss; fourth, they should shift validation toward robust multicenter external testing combined with comprehensive calibration and decision curve analyses; and fifth, model outputs should be integrated into user-friendly tools (e.g., column-line graphs or web-based calculators) and combined with early intervention strategies to maximize prevention. Finally, based on the international practice insights from Section 4.7 regarding lifestyle medicine case management nurses and multidisciplinary teams led by case managers, future research should also assess the feasibility, implementation pathways, and actual impact on patient outcomes of these models in China's primary healthcare settings. By strictly adhering to these principles, future work will improve reproducibility, clinical applicability, and ultimately the accuracy of diabetes risk stratification.

6. Limitation

In this systematic evaluation of risk prediction models for type 2 diabetes in China, certain limitations remain. First, this study only included the literature published before October 2024, which may not reflect the latest scientific advances, and as medical research continues to evolve, subsequent findings may influence the current conclusions. Second,

due to differences in data processing methods between the original studies, it was difficult for us to perform in-depth analyses at the predictor level, potentially affecting the accurate assessment of the predictive ability of the model; therefore, future studies should adopt standardized methods of data collection and analysis to improve the consistency and comparability of the studies. In addition, this study was primarily based on data from a Chinese population, potentially limiting applicability to other populations and making our findings not necessarily fully representative of individuals from different regions and ethnicities. Despite these limitations, this study provides important insights into the current state of risk prediction models for type 2 diabetes in China. Future studies should expand the literature screening, update the data, optimize the statistical methods, and improve the representation of different populations in order to obtain more comprehensive and reliable results.

7. Conclusions

In conclusion, the increasing prevalence of type 2 diabetes in China requires the development of robust and effective predictive models that can be adapted to the unique demographic and cultural characteristics of this population. This systematic assessment revealed significant gaps in current predictive modeling efforts, particularly in the areas of missing external validation, the incomplete coverage of key variables, unstandardized treatment of continuous variables, and weak bias control strategies. Future research must focus on improving modeling methods, enhancing the treatment of continuous variables, and incorporating multidimensional correlated data into predictive frameworks. By addressing these challenges, we can improve the accuracy and applicability of predictive models for T2D, thereby facilitating early detection and intervention strategies. By drawing on international clinical practice experience, we can promote the application of models and case management approaches in local clinical settings. These measures are essential for managing this growing public health crisis. Ultimately, these efforts will help to improve health and reduce the economic burden associated with diabetes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/healthcare13162007/s1.

Author Contributions: Conceptualization, J.D.; methodology, J.D. and N.M.N.; writing—original draft preparation, J.D.; writing—review and editing, J.D. and N.M.N.; supervision, N.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ALT	Alanine Aminotransferase
ANN	Artificial Neural Network
AST	Aspartate Aminotransferase
AUC	Area Under the Curve
BMI	Body Mass Index
BPNN	Back Propagation Neural Network

BUN	Blood Urea Nitrogen
C4.5	C4.5 Decision Tree Algorithm
CAD	Coronary Artery Disease
CART	Classification and Regression Tree
CB	Calibration Belt
CHOL	Cholesterol
COX	Cox Proportional Hazards Model
CR	Creatinine
CREA	Creatinine
DBP	Diastolic Blood Pressure
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
ECG	Electrocardiogram
EH	Essential Hypertension
FBG	Fasting Blood Glucose
FPG	Fasting Plasma Glucose
GLU	Glucose
HB	Hemoglobin
HBA1C	Hemoglobin A1c
HDL	High-Density Lipoprotein
HDL-C	High-Density Lipoprotein Cholesterol
HTN	Hypertension
IDF	International Diabetes Federation
KNN	K-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LDL-C	Low-Density Lipoprotein Cholesterol
LGBM	Light Gradient Boosting Machine
LR	Logistic Regression
MCHC	Mean Corpuscular Hemoglobin Concentration
ML	Machine Learning
MLP	Multilayer Perceptron
PDM	Prediabetes Mellitus
PLT	Platelets
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF	Random Forest
RFE	Recursive Feature Elimination
SBP	Systolic Blood Pressure
SCR	Serum Creatinine
SHAP	SHapley Additive exPlanations
SS	Salt Sensitivity
SUA	Serum Uric Acid
SVM	Support Vector Machine
T2D	Type 2 Diabetes
TABNET	Tabular Neural Network
TBIL	Total Bilirubin
TC	Total Cholesterol
TG	Triglycerides
VC	Variable Combination
VIF	Variance Inflation Factor
WBC	White Blood Cell
WC	Waist Circumference
WHR	Waist-to-Hip Ratio
XGB	Extreme Gradient Boosting

SDOH	Social Determinants of Health
EPV	Events Per Variable

References

1. Heald, A.H.; Stedman, M.; Davies, M.; Livingston, M.; Alshames, R.; Lunt, M.; Rayman, G.; Gadsby, R. Estimating life years lost to diabetes: Outcomes from analysis of National Diabetes Audit and Office of National Statistics data. *Cardiovasc. Endocrinol. Metab.* **2020**, *9*, 183–185. [CrossRef]
2. American Diabetes Association. Standards of medical care in diabetes—2020 abridged for primary care providers. *Clin. Diabetes: A Publ. Am. Diabetes Assoc.* **2020**, *38*, 10–38. [CrossRef]
3. International Diabetes Federation. *IDF Diabetes Atlas*, 11th ed.; International Diabetes Federation: Brussels, Belgium, 2025; Available online: https://diabetesatlas.org/ (accessed on 2 August 2025).
4. Cavan, D. Why screen for type 2 diabetes? *Diabetes Res. Clin. Pract.* **2016**, *121*, 215–217. [CrossRef] [PubMed]
5. Li, G.; Zhang, P.; Wang, J.; Gregg, E.W.; Yang, W.; Gong, Q.; Li, H.; Li, H.; Jiang, Y.; An, Y.; et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: A 20-year follow-up study. *Lancet* **2008**, *371*, 1783–1789. [CrossRef] [PubMed]
6. Lowe, W.L.; Bain, J.R. "Prediction is very hard, especially about the future": New biomarkers for type 2 diabetes? *Diabetes* **2013**, *62*, 1384–1385. [CrossRef] [PubMed]
7. Janghorbani, M.; Adineh, H.; Amini, M. Evaluation of the Finnish Diabetes Risk Score (FINDRISC) as a screening tool for the metabolic syndrome. *Rev. Diabet. Stud.* **2013**, *10*, 283–292. [CrossRef]
8. Petridis, P.D.; Kristo, A.S.; Sikalidis, A.K.; Kitsas, I.K. A review on trending machine learning techniques for type 2 diabetes mellitus management. *Informatics* **2024**, *11*, 70. [CrossRef]
9. Li, W.; Liu, X.; Liu, Z.; Xing, Q.; Liu, R.; Wu, Q.; Hu, Y.; Zhang, J. The signaling pathways of selected traditional Chinese medicine prescriptions and their metabolites in the treatment of diabetic cardiomyopathy: A review. *Front. Pharmacol.* **2024**, *15*, 1416403. [CrossRef]
10. Nazirun, N.N.N.; Wahab, A.A.; Selamat, A.; Fujita, H.; Krejcar, O.; Kuca, K. Prediction models for type 2 diabetes progression: A systematic review. *IEEE Access* **2024**, *12*, 161595–161619. [CrossRef]
11. Bini, S.A. Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care? *J. Arthroplast.* **2018**, *33*, 2358–2361. [CrossRef]
12. Negi, A.; Jaiswal, V. A first attempt to develop a diabetes prediction method based on different global datasets. In *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*; IEEE: New York, NY, USA, 2016; pp. 237–241. [CrossRef]
13. Shaik, T.; Tao, X.; Higgins, N.; Li, L.; Gururajan, R.; Zhou, X.; Acharya, U.R. Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *WIREs Data Min. Knowl. Discov.* **2023**, *13*, e1485. [CrossRef]
14. Bell, M.L.; Fiero, M.; Horton, N.J.; Hsu, C.-H. Handling missing data in RCTs; a review of the top medical journals. *BMC Med. Res. Methodol.* **2014**, *14*, 118. [CrossRef]
15. Asgari, S.; Khalili, D.; Hosseinpanah, F.; Hadaegh, F. Prediction models for type 2 diabetes risk in the general population: A systematic review of observational studies. *Int. J. Endocrinol. Metab.* **2021**, *19*, e109206. [CrossRef] [PubMed]
16. Sung, K.; Lee, S. Social determinants of health and type 2 diabetes in Asia. *J. Diabetes Investig.* **2025**, *16*, 971–983. [CrossRef] [PubMed]
17. Hu, G.; Lin, L.; Hu, X.; Zheng, Y.; Liu, X.; Xu, Z.; He, Y.; Zhang, Y. Machine learning-based diagnosis of type 2 diabetes mellitus using social determinants of health. *Mol. Cell. Biomech.* **2025**, *22*, 1461. [CrossRef]
18. Lan, X.; Ji, X.; Zheng, X.; Ding, X.; Mou, H.; Lu, S.; Ye, B. Socio-demographic and clinical determinants of self-care in adults with type 2 diabetes: A multicenter cross-sectional study in Zhejiang province, China. *BMC Public Health* **2025**, *25*, 397. [CrossRef]
19. Zhao, Y.; Li, H.-F.; Wu, X.; Li, G.-H.; Golden, A.R.; Cai, L. Rural-urban differentials of prevalence and lifestyle determinants of pre-diabetes and diabetes among the elderly in southwest China. *BMC Public Health* **2023**, *23*, 603. [CrossRef]
20. Chang, G.; Tian, S.; Luo, X.; Xiang, Y.; Cai, C.; Zhu, R.; Cai, H.; Yang, H.; Gao, H. Hypoglycemic effects and mechanisms of polyphenols from *Myrica rubra* pomace in type 2 diabetes (db/db) mice. *Mol. Nutr. Food Res.* **2025**, *69*, e202400523. [CrossRef]
21. Dong, W.; Wan, E.; Bedford, L.; Wu, T.; Wong, C.; Tang, E.; Lam, C. Prediction models for the risk of cardiovascular diseases in Chinese patients with type 2 diabetes mellitus: A systematic review. *Public Health* **2020**, *186*, 144–156. [CrossRef]
22. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]
23. Alam, A.; Sohel, A.; Hasan, K.M.; Islam, M.A. Machine learning and artificial intelligence in diabetes prediction and management: A comprehensive review of models. *J. Next-Gen Eng. Syst.* **2024**, *1*, 107–124. [CrossRef]

24. Sun, Y.; Gregersen, H.; Yuan, W. Chinese health care system and clinical epidemiology. *Clin. Epidemiol.* **2017**, *9*, 167–178. [CrossRef] [PubMed]
25. Chen, R.; Wang, S.F.; Zhou, J.C.; Sun, F.; Wei, W.W.; Zhan, S.Y. Introduction of the Prediction model Risk Of Bias Assessment Tool: A tool to assess risk of bias and applicability of prediction model studies. *Chin. J. Epidemiol.* **2020**, *41*, 776–781. [CrossRef]
26. Gusenbauer, M.; Gauster, S.P. How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide. *Technol. Forecast. Soc. Change* **2025**, *212*, 123833. [CrossRef]
27. Lorenzetti, D.L.; Ghali, W.A. Reference management software for systematic reviews and meta-analyses: An exploration of usage and usability. *BMC Med. Res. Methodol.* **2013**, *13*, 141. [CrossRef]
28. Xu, W.; Zhou, Y.; Jiang, Q.; Fang, Y.; Yang, Q. Risk prediction models for diabetic nephropathy among type 2 diabetes patients in China: A systematic review and meta-analysis. *Front. Endocrinol.* **2024**, *15*, 1407348. [CrossRef]
29. Bozkurt, S.; Cahan, E.M.; Seneviratne, M.G.; Sun, R.; Lossio-Ventura, J.A.; Ioannidis, J.P.A.; Hernandez-Boussard, T. Reporting of demographic data and representativeness in machine learning models using electronic health records. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1878–1884. [CrossRef]
30. Xu, T.; Yu, D.; Zhou, W.; Yu, L. A nomogram model for the risk prediction of type 2 diabetes in healthy eastern China residents: A 14-year retrospective cohort study from 15,166 participants. *EPMA J.* **2022**, *13*, 397–405. [CrossRef]
31. Lin, Y.; Shen, Y.; He, R.; Wang, Q.; Deng, H.; Cheng, S.; Liu, Y.; Li, Y.; Lu, X.; Shen, Z. A novel predictive model for optimizing diabetes screening in older adults. *J. Diabetes Investig.* **2024**, *15*, 1403–1409. [CrossRef]
32. Wang, S.; Chen, R.; Wang, S.; Kong, D.; Cao, R.; Lin, C.; Luo, L.; Huang, J.; Zhang, Q.; Yu, H.; et al. Comparative study on risk prediction model of type 2 diabetes based on machine learning theory: A cross-sectional study. *BMJ Open* **2023**, *13*, e069018. [CrossRef]
33. Liu, H.; Dong, S.; Yang, H.; Wang, L.; Liu, J.; Du, Y.; Liu, J.; Lyu, Z.; Wang, Y.; Jiang, L.; et al. Comparing the accuracy of four machine learning models in predicting type 2 diabetes onset within the Chinese population: A retrospective study. *J. Int. Med. Res.* **2024**, *52*, 3000605241253786. [CrossRef]
34. Yang, J.; Liu, D.; Du, Q.; Zhu, J.; Lu, L.; Wu, Z.; Zhang, D.; Ji, X.; Zheng, X. Construction of a 3-year risk prediction model for developing diabetes in patients with pre-diabetes. *Front. Endocrinol.* **2024**, *15*, 1410502. [CrossRef] [PubMed]
35. Tong, Y.-T.; Gao, G.-J.; Chang, H.; Wu, X.-W.; Li, M.-T. Development and economic assessment of machine learning models to predict glycosylated hemoglobin in type 2 diabetes. *Front. Pharmacol.* **2023**, *14*, 1216182. [CrossRef] [PubMed]
36. Shao, X.; Wang, Y.; Huang, S.; Liu, H.; Zhou, S.; Zhang, R.; Yu, P.; Hu, C. Development and validation of a prediction model estimating the 10-year risk for type 2 diabetes in China. *PLoS ONE* **2020**, *15*, e0237936. [CrossRef] [PubMed]
37. Jiang, L.; Xia, Z.; Zhu, R.; Gong, H.; Wang, J.; Li, J.; Wang, L. Diabetes risk prediction model based on community follow-up data using machine learning. *Prev. Med. Rep.* **2023**, *35*, 102358. [CrossRef]
38. Li, L.; Cheng, Y.; Ji, W.; Liu, M.; Hu, Z.; Yang, Y.; Wang, Y.; Zhou, Y. Machine learning for predicting diabetes risk in western China adults. *Diabetol. Metab. Syndr.* **2023**, *15*, 165. [CrossRef]
39. Wang, Y.; Zhang, Y.; Wang, K.; Su, Y.; Zhuge, J.; Li, W.; Wang, S.; Yao, H. Nomogram model for screening the risk of type II diabetes in western Xinjiang, China. *Diabetes, Metab. Syndr. Obesity: Targets Ther.* **2021**, *14*, 3541–3553. [CrossRef]
40. Dong, W.; Tse, T.Y.E.; Mak, L.I.; Wong, C.K.H.; Wan, Y.F.E.; Tang, H.M.E.; Chin, W.Y.; Bedford, L.E.; Yu, Y.T.E.; Ko, W.K.W.; et al. Non-laboratory-based risk assessment model for case detection of diabetes mellitus and pre-diabetes in primary care. *J. Diabetes Investig.* **2022**, *13*, 1374–1386. [CrossRef]
41. Liu, Q.; Zhang, M.; He, Y.; Zhang, L.; Zou, J.; Yan, Y.; Guo, Y. Predicting the risk of incident type 2 diabetes mellitus in Chinese elderly using machine learning techniques. *J. Pers. Med.* **2022**, *12*, 905. [CrossRef]
42. Hu, H.; Wang, J.; Han, X.; Li, Y.; Miao, X.; Yuan, J.; Yang, H.; He, M. Prediction of 5-year risk of diabetes mellitus in relatively low risk middle-aged and elderly adults. *Acta Diabetol.* **2020**, *57*, 63–70. [CrossRef]
43. Yang, H.; Yuan, L.; Wu, J.; Li, X.; Long, L.; Teng, Y.; Feng, W.; Lyu, L.; Xu, B.; Ma, T.; et al. Construction of a predictive model for diabetes mellitus type 2 in middle-aged and elderly populations based on the medical checkup data of National Basic Public Health Service. *Sichuan Da Xue Xue Bao. Yi Xue Ban = J. Sichuan University. Med. Sci. Ed.* **2024**, *55*, 662–670. [CrossRef]
44. Long, X.; Hua, H.; Wu, Y.; Zhang, W.; Yin, C.; Li, N.; Cheng, N. Construction and validation of a risk prediction model for diabetes incidence. *J. Lanzhou Univ. (Med. Ed.)* **2024**, *50*, 70–78. [CrossRef]
45. Ma, W.; Wang, K.; Yu, B.; Feng, C.; Ji, J. Comparative study of diabetes risk prediction models based on physical examination data. *Mod. Inf. Technol.* **2020**, *4*, 72–75. [CrossRef]
46. Miao, Q.; Zhu, Y. Diabetes prediction model based on PSO-FWSVM. *Comput. Digit. Eng.* **2020**, *48*, 993–998. [CrossRef]
47. Ma, Y.; Che, Q.; Zheng, Q.; Chen, S.; Zhou, Z.; Yang, J.; Wu, Y.; Wu, T.; Hu, Y.; Zhang, L.; et al. Common evaluation methods of prediction model for risk of type 2 diabetes mellitus. *Chin. J. Prev. Control. Chronic Dis.* **2020**, *28*, 94–100. [CrossRef]
48. Ouyang, P.; Li, X.; Leng, F.; Lai, X.; Zhang, H.; Yan, C.; Wang, C.; Bai, Y.; Xing, Z.; Liu, X.; et al. Application of machine learning algorithms in predicting diabetes risk in a physical examination population. *Chin. J. Dis. Control. Prev.* **2021**, *25*, 849–853. [CrossRef]

49. Wu, H.; Chen, S.; Chen, Z.; Yang, Y.; Zeng, C.; Wu, S.; Su, X. Study on diabetes prediction model based on LightGBM model. *China Health Stand. Manag.* **2023**, *14*, 64–67. [CrossRef]
50. Yang, S. Study on key biological indicators of diabetes based on statistical tests. *J. Clin. Nurs. Res.* **2024**, *8*, 267–273. [CrossRef]
51. Deberneh, H.M.; Kim, I. Prediction of type 2 diabetes based on machine learning algorithm. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3317. [CrossRef] [PubMed]
52. Tarumi, S.; Takeuchi, W.; Qi, R.; Ning, X.; Ruppert, L.; Ban, H.; Robertson, D.H.; Schleyer, T.; Kawamoto, K. Predicting pharmacotherapeutic outcomes for type 2 diabetes: An evaluation of three approaches to leveraging electronic health record data from multiple sources. *J. Biomed. Inform.* **2022**, *129*, 104001. [CrossRef]
53. Hatmal, M.M.; Alshaer, W.; Mahmoud, I.S.; Al-Hatamleh, M.A.I.; Al-Ameer, H.J.; Abuyaman, O.; Zihlif, M.; Mohamud, R.; Darras, M.; Al Shhab, M.; et al. Investigating the association of CD36 gene polymorphisms (rs1761667 and rs1527483) with T2DM and dyslipidemia: Statistical analysis, machine learning based prediction, and meta-analysis. *PLoS ONE* **2021**, *16*, e0257857. [CrossRef] [PubMed]
54. Ngufor, C.; Van Houten, H.; Caffo, B.S.; Shah, N.D.; McCoy, R.G. Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c. *J. Biomed. Inform.* **2019**, *89*, 56–67. [CrossRef] [PubMed]
55. Shin, J.; Lee, J.; Ko, T.; Lee, K.; Choi, Y.; Kim, H.-S. Improving Machine Learning Diabetes Prediction Models for the Utmost Clinical Effectiveness. *J. Pers. Med.* **2022**, *12*, 1899. [CrossRef] [PubMed]
56. Kopitar, L.; Kocbek, P.; Cilar, L.; Sheikh, A.; Stiglic, G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci. Rep.* **2020**, *10*, 11981. [CrossRef]
57. Yuk, H.; Gim, J.; Min, J.K.; Yun, J.; Heo, T.-Y. Artificial intelligence–based prediction of diabetes and prediabetes using health checkup data in Korea. *Appl. Artif. Intell.* **2022**, *36*, 2145644. [CrossRef]
58. Hegde, H.; Shimpi, N.; Panny, A.; Glurich, I.; Christie, P.; Acharya, A. Development of non-invasive diabetes risk prediction models as decision support tools designed for application in the dental clinical environment. *Inform. Med. Unlocked* **2019**, *17*, 100254. [CrossRef]
59. Syed, A.H.; Khan, T. Machine learning-based application for predicting risk of type 2 diabetes mellitus (T2DM) in Saudi Arabia: A retrospective cross-sectional study. *IEEE Access* **2020**, *8*, 199539–199561. [CrossRef]
60. Oh, R.; Lee, H.K.; Pak, Y.K.; Oh, M.-S. An Interactive Online App for Predicting Diabetes via Machine Learning from Environment-Polluting Chemical Exposure Data. *Int. J. Environ. Res. Public Health* **2022**, *19*, 5800. [CrossRef]
61. Gollapalli, M.; Alansari, A.; Alkhorasani, H.; Alsubaii, M.; Sakloua, R.; Alzahrani, R.; Al-Hariri, M.; Alfares, M.; AlKhafaji, D.; Al Argan, R.; et al. A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM. *Comput. Biol. Med.* **2022**, *147*, 105757. [CrossRef]
62. Islam, S.; Qaraqe, M.K.; Belhaouari, S.B.; Abdul-Ghani, M.A. Advanced techniques for predicting the future progression of type 2 diabetes. *IEEE Access* **2020**, *8*, 120537–120547. [CrossRef]
63. Deberneh, H.M.; Kim, I.; Park, J.H.; Cha, E.; Joung, K.H.; Lee, J.S.; Lim, D.S. 1233-P: Prediction of type 2 diabetes occurrence using machine learning model. *Diabetes* **2020**, *69* (Suppl. 1), 1233. [CrossRef]
64. Navarro, C.L.A.; Damen, J.A.A.; Takada, T.; Nijman, S.W.J.; Dhiman, P.; Ma, J.; Collins, G.S.; Bajpai, R.; Riley, R.D.; Moons, K.G.M.; et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review. *BMJ* **2021**, *375*, n2281. [CrossRef] [PubMed]
65. Altman, D.G.; Royston, P. The cost of dichotomising continuous variables. *BMJ* **2006**, *332*, 1080. [CrossRef]
66. Harrell, F.E., Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2015. [CrossRef]
67. Ramspek, C.L.; Jager, K.J.; Dekker, F.W.; Zoccali, C.; van Diepen, M. External validation of prognostic models: What, why, how, when and where? *Clin. Kidney J.* **2021**, *14*, 49–58. [CrossRef] [PubMed]
68. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med.* **2015**, *13*, 148–158. [CrossRef]
69. Mansmann, U.; Ön, B.I. The validation of prediction models deserves more recognition. *BMC Med.* **2025**, *23*, 166. [CrossRef] [PubMed]
70. Nieboer, D.; van der Ploeg, T.; Steyerberg, E.W.; Collins, G. Assessing discriminative performance at external validation of clinical prediction models. *PLoS ONE* **2016**, *11*, e0148820. [CrossRef]
71. Iwagami, M.; Matsui, H. Introduction to clinical prediction models. *Ann. Clin. Epidemiol.* **2022**, *4*, 72–80. [CrossRef]
72. Hanf, M.; Guégan, J.-F.; Ahmed, I.; Nacher, M. Disentangling the complexity of infectious diseases: Time is ripe to improve the first-line statistical toolbox for epidemiologists. *Infect. Genet. Evol.* **2014**, *21*, 497–505. [CrossRef]
73. Zhang, L.; Wang, Y.; Niu, M.; Wang, C.; Wang, Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: The Henan Rural Cohort Study. *Sci. Rep.* **2020**, *10*, 4406. [CrossRef]
74. Teng, Q.; Liu, Z.; Song, Y.; Han, K.; Lu, Y. A survey on the interpretability of deep learning in medical diagnosis. *Multimedia Syst.* **2022**, *28*, 2335–2355. [CrossRef] [PubMed]

75. Chen, H.; Chen, J.; Ding, J. Data evaluation and enhancement for quality improvement of machine learning. In *2020 IEEE 20th International Conference on Software Quality, Reliability and Security (QRS)*; IEEE: New York, NY, USA, 2020; p. 13. [CrossRef]
76. Zaid, M.M.A.; Mohammed, A.A. Hybrid models in diabetes prediction: A review of techniques, performance, and potential. *J. Al-Qadisiyah Comput. Sci. Math.* **2024**, *16*, 298–308. [CrossRef]
77. Tuomilehto, J.; Lindström, J.; Eriksson, J.G.; Valle, T.T.; Hämäläinen, H.; Ilanne-Parikka, P.; Keinänen-Kiukaanniemi, S.; Laakso, M.; Louheranta, A.; Rastas, M.; et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* **2001**, *344*, 1343–1350. [CrossRef]
78. Knowler, W.C.; Barrett-Connor, E.; Fowler, S.E.; Hamman, R.F.; Lachin, J.M.; Walker, E.A.; Nathan, D.M.; Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **2002**, *346*, 393–403. [CrossRef]
79. Ramachandran, A.; Snehalatha, C.; Mary, S.; Mukesh, B.; Bhaskar, A.D.; Vijay, V.; Indian Diabetes Prevention Programme (IDPP). The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia* **2006**, *49*, 289–297. [CrossRef]
80. Cangelosi, G.; Mancin, S.; Pantanetti, P.; Nguyen, C.T.T.; Palomares, S.M.; Biondini, F.; Sguanci, M.; Petrelli, F. Lifestyle medicine case manager nurses for type two diabetes patients: An overview of a job description framework—A narrative review. *Diabetology* **2024**, *5*, 375–388. [CrossRef]
81. Alshowair, A.; Altamimi, S.; Alshahrani, S.; Almubrick, R.; Ahmed, S.; Tolba, A.; Alkawai, F.; Alruhaimi, F.; Alsafwani, E.; AlSuwailem, F.; et al. Effectiveness of case manager–led multi-disciplinary team approach on glycemic control amongst T2DM patients in primary care in Riyadh: A retrospective follow-up study. *J. Prim. Care Community Health* **2023**, *14*, 21501319231204592. [CrossRef]
82. Tourkmani, A.M.; Abdelhay, O.; Alkhashan, H.I.; Alaboud, A.F.; Bakhit, A.; Elsaid, T.; Alawad, A.; Alobaikan, A.; Alqahtani, H.; Alqahtani, A.; et al. Impact of an integrated care program on glycemic control and cardiovascular risk factors in patients with type 2 diabetes in Saudi Arabia: An interventional parallel-group controlled study. *BMC Fam. Pract.* **2018**, *19*, 1. [CrossRef]
83. Yu, X.; Chau, J.P.C.; Huo, L.; Li, X.; Wang, D.; Wu, H.; Zhang, Y. The effects of a nurse-led integrative medicine-based structured education program on self-management behaviors among individuals with newly diagnosed type 2 diabetes: A randomized controlled trial. *BMC Nurs.* **2022**, *21*, 217. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.