International Journal of
Environmental Research and Public Health

MDPI

Review

Accuracy of Machine Learning Classification Models for the Prediction of Type 2 Diabetes Mellitus: A Systematic Survey and Meta-Analysis Approach

Micheal O. Olusanya [1,*], Ropo Ebenezer Ogunsakin [2], Meenu Ghai [3] and Matthew Adekunle Adeleke [3]

[1] Department of Computer Science and Information Technology, Sol Plaatje University, Kimberley 8300, South Africa
[2] Biostatistics Unit, Discipline of Public Health Medicine, School of Nursing & Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban 4000, South Africa
[3] Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Durban 4000, South Africa
* Correspondence: michael.olusanya@spu.ac.za

Highlights:

- We reviewed soft-computing and statistical learning methods for predicting type 2 diabetes mellitus.
- We searched for papers published between 2010 and 2021 on three academic search engines, obtaining 34 relevant documents for the final meta-analysis.
- We analyzed the data extracted, compared the results and models, discussed their performance, and highlighted the issues related to T2DM.
- Finally, the decision trees model has the best prediction performances, with excellent accuracy compared to other soft-computing models in this systematic meta-analysis.

Abstract: Soft-computing and statistical learning models have gained substantial momentum in predicting type 2 diabetes mellitus (T2DM) disease. This paper reviews recent soft-computing and statistical learning models in T2DM using a meta-analysis approach. We searched for papers using soft-computing and statistical learning models focused on T2DM published between 2010 and 2021 on three different search engines. Of 1215 studies identified, 34 with 136952 patients met our inclusion criteria. The pooled algorithm's performance was able to predict T2DM with an overall accuracy of 0.86 (95% confidence interval [CI] of [0.82, 0.89]). The classification of diabetes prediction was significantly greater in models with a screening and diagnosis (pooled proportion [95% CI] = 0.91 [0.74, 0.97]) when compared to models with nephropathy (pooled proportion = 0.48 [0.76, 0.89] to 0.88 [0.83, 0.91]). For the prediction of T2DM, the decision trees (DT) models had a pooled accuracy of 0.88 [95% CI: 0.82, 0.92], and the neural network (NN) models had a pooled accuracy of 0.85 [95% CI: 0.79, 0.89]. Meta-regression did not provide any statistically significant findings for the heterogeneous accuracy in studies with different diabetes predictions, sample sizes, and impact factors. Additionally, ML models showed high accuracy for the prediction of T2DM. The predictive accuracy of ML algorithms in T2DM is promising, mainly through DT and NN models. However, there is heterogeneity among ML models. We compared the results and models and concluded that this evidence might help clinicians interpret data and implement optimum models for their dataset for T2DM prediction.

Keywords: diagnosis; soft computing; predictive models; type 2 diabetes mellitus; meta-analysis

check for updates

Citation: Olusanya, M.O.; Ogunsakin, R.E.; Ghai, M.; Adeleke, M.A. Accuracy of Machine Learning Classification Models for the Prediction of Type 2 Diabetes Mellitus: A Systematic Survey and Meta-Analysis Approach. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14280. https://doi.org/10.3390/ijerph192114280

Academic Editor: Noël Christopher Barengo

Received: 27 September 2022
Accepted: 25 October 2022
Published: 1 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

Data mining, such as soft-computing (that is, machine learning (ML)) methods, has become essential in diagnosing T2DM and assigning management to healthcare providers [1]. ML is a subdivision of artificial intelligence that is gradually exploited within the field of

diabetic medicine. Primarily, it is how computers make sense of data and classify tasks with or without human supervision. Several ML models have been used extensively in diabetes mellitus (DM) studies to explore DM risk factors [2,3]. The ML methods, which include logistic regression (LR), artificial neural networks (ANN), and decision trees (DT), were used to predict both DM and pre-diabetes [4–9]. Other ML models, such as random forest (RF), support vector machines (SVM), k-nearest neighbors (KNN), and the naïve Bayes, have also been used in the literature [10–15].

Given the above, previous studies have distinct pragmatic shreds of evidence for each ML model [16,17]. Still, no agreement has arisen to guide the choice of precise ML models for clinical investigation in the context of diabetic medicine. The overall classification accuracy reported in each model varies from one study to another. Furthermore, ML studies conveyed the model evaluation criteria, including the area under the curve (AUC). Most significantly, an adequate boundary for AUC to be employed in clinical investigation and suitable ML models that are efficient in diabetic research have yet to be appraised. As a result of the visible success in a wide range of predictive tasks, medical researchers and clinicians have a significant interest in using ML techniques.

Nevertheless, pooled estimates for ML techniques among patients with T2DM and the trends over the years remain unknown globally. Against this background, our study pooled data from previous independent studies to determine the overall ML models' predictive ability of T2DM disease. Our findings could be helpful to clinicians, healthcare managers, and policymakers involved in the delivery of Type-2 diabetes healthcare worldwide.

2. Materials and Methods

2.1. Search Strategy and Selection Process

In this meta-analysis, we searched the scholastic databases of Web of Science, Scopus, and PubMed for relevant published articles on ML applied to health applications for T2DM. These databases were searched for English papers published between 2010 and 2021. We excluded studies published before January 2010 because most of those studies used outdated computer-aided algorithms that are currently not popular. The literature search strategy, selection of publications, data extraction, and reporting results were executed following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) (Moher et al., 2010). During a comprehensive literature search, the search terms used were: With MESH terms: ("diabetes mellitus, type 2" [All Fields] AND ("machine learning" [All Fields] OR "deep learning" [All Fields] OR "neural networks (computer)" [All Fields] OR "support vector machine" [All Fields] OR "classification" [All Fields] OR "decision trees" [All Fields] OR "cluster analysis" [All Fields] OR "principal component analysis" [All Fields] OR "data mining" [All Fields] OR "logistic models" [All Fields] OR "algorithms" [All Fields])) AND ("diagnosis" [All Fields] OR "roc curve" [All Fields] OR "area under curve" [All Fields])", ("machine learning" or "deep learning" or "artificial intelligence" or "neural network" or "support vector machine" or "classification-tree" or "regression-tree" or "decision-tree" or "random forest" or "gradient boosting" or "k-nearest neighbors" or "supervised-learning" or "unsupervised-learning". The search terms were separated or combined using Boolean operators such as "OR" or "AND". After data extraction, we summarized and reported the findings in tables and figures according to the study's objectives.

2.2. Inclusion Criteria

The inclusion criteria were original articles and clinical trials. In addition, those studies with model performance evaluation, such as accuracy, sensitivity, specificity, and area under the curve (AUC), were included.

2.3. Exclusion Criteria

Articles written in languages other than English, published before January 2010, or with study designs such as reviews, letters to editors, editorials, commentaries, expert opinions, books, book chapters, brief reports, and theses were excluded. Conference articles,

grey literature, and literature that failed to report model performance evaluation criteria were excluded.

2.4. Assessments of Methodological Quality

The quality of the individual studies was independently assessed based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. QUADAS-2 is a validated tool used to evaluate the quality of diagnostic accuracy studies by patient selection, index tests, reference standards, and the risk of bias for internal and external validity for applicability concerns of individual studies. In this meta-analysis, each article's qualities were evaluated. The two authors assessed the identified methodological quality and eligibility of articles, and disagreements among reviewers were fixed accordingly with discussion. The data extracted included the author, the publication year, the country where the study was conducted, the study design, the sample size, the prediction type, the T2DM cases, the number of participants, the sensitivity, the specificity, the impact factor of the articles (extracted from Scopus webpage), the ML models and the software deployed. In addition, we included the model that had the best overall performance in the primary analysis for the studies that proposed multiple models. We also extracted the performance of models with the best sensitivity and specificity in studies with numerous ML models to perform further sensitivity-focused and specificity-focused analyses.

2.5. Statistical Analysis

A meta-analysis was conducted for the pooled overall classification accuracy proportion; a chi-squared test was used for heterogeneity; Higgins I-squared (I^2) was used to assess the total heterogeneity/total variability among studies. For Higgins I-squared (I^2) [18,19], forest plots of over 50% were observed as an indication of heterogeneity among studies. If the estimated amount of total heterogeneity (Tau I^2) (DerSimonian and Laird, 1986) was less than 40%, the studies were considered similar. Because the extracted articles were from general populations, a random-effects meta-analysis was deemed to be taken from an inverse-variance model [20].

Additionally, a subgroup analysis was performed to investigate the heterogeneity among the studies based on the prediction type of the algorithm for T2DM and machine-learning diabetes prediction. Combining only the published studies may lead to an insignificant or biased result in the meta-analysis. Thus, this study used a funnel plot to report publication bias among the included studies [18]. The publication bias was assessed through the Begg & Eggers test and the visual inspection of the funnel plot. Meta-regression was used to explore the factors possibly contributing to the between-study heterogeneity. The extracted data were captured into an excel spreadsheet. A meta-analysis was performed via the metafor, rma, meta, and metaprop packages in R (version 4.0.3, R Core Team, Vienna, Austria); the statistical significance was expressed with a 95% Confidence Interval (CI), and p-values < 0.05 were considered statistically significant.

3. Results

3.1. Characteristics of Selected Studies

The features of the eligible studies in Table 1 showed that the application for T2DM with most of the included studies was diagnostic (38.2%, 13/34), followed by prognostic (26.5%, 9/34), nephropathy (20.6%, 7/34), screening and diagnosis (8.8%, 3/34) and risk factor analysis (5.9%, 2/34). The learning algorithm subset of artificial intelligence includes all the methods and algorithms that enable the machines to automatically learn mathematical models to extract useful information from large datasets. Thus, in terms of learning algorithm classification techniques, 23.53% (8/34) of studies applied linear regression (LR), and 23.53% (8/34) used decision trees (DT) on the diabetes patient's data, respectively. A total of 17.65% (6/34) applied an artificial neural network (ANN), 8.82% (3/34) deployed random forest (RF), and 14.71% (5/34) employed a support vector machine

(SVM). One (2.94%) example of a hybrid model, a neural network (NN), a CRISP method, and phenotyping, respectively.

3.2. Meta-Analyses Methods

The literature search of three databases (Web of Science, PubMed, and Scopus) and reference screening yielded 1215 articles. We imported all the retrieved articles into End-Note X9, identifying 945 duplicates. Out of the remaining documents, 98 were excluded because their abstracts and titles did not meet the eligibility requirements. Additionally, 172 studies were eligible for a full review, out of which 130 were excluded for not reporting the outcome variable, incomplete information, or non-relevant. A total of 42 studies were eligible for quality assessment, and, finally, 34 documents were found to qualify and were included in the final meta-analysis (Figure 1). The flow diagram in Figure 1 summarizes the reasons for excluding research articles from study inclusion following the PRISMA.

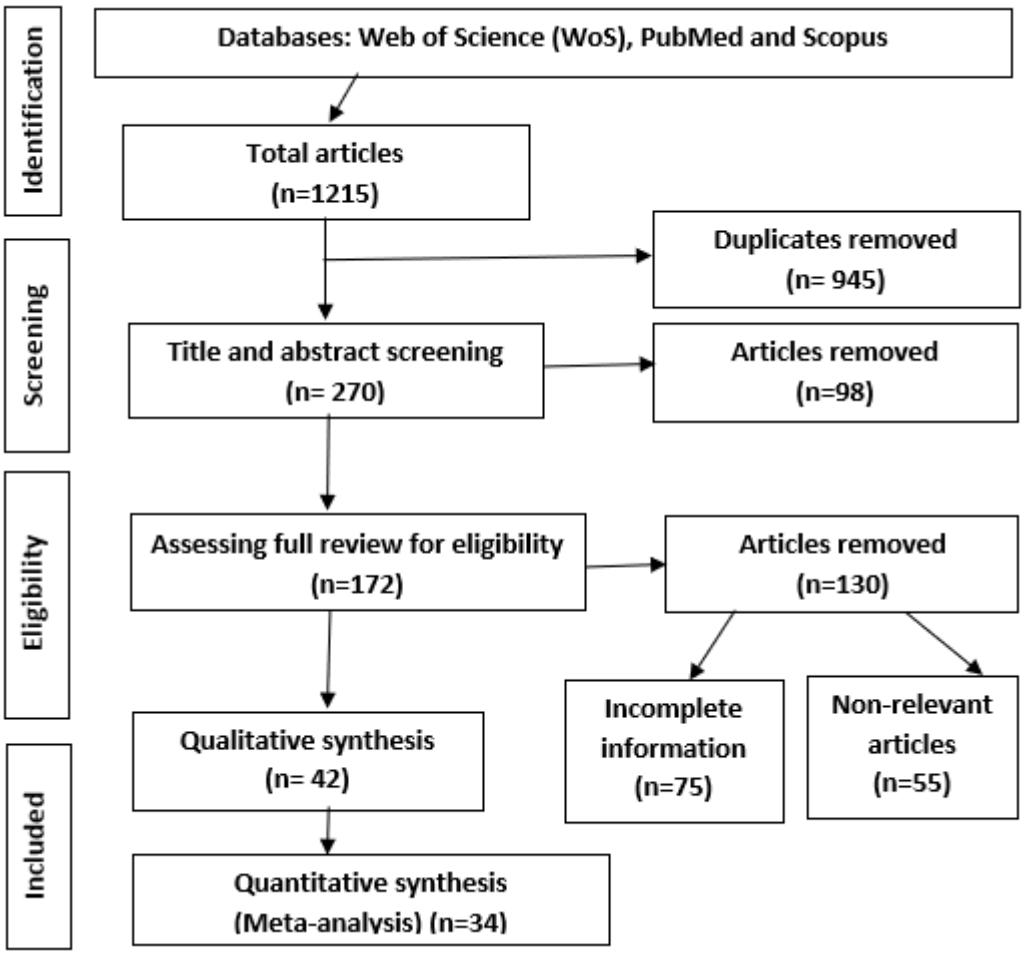


Figure 1. The process of selecting published literature according to *PRISMA* and Meta-Analyses guidelines.

Table 1. Summary of the studies used in the systematic review and meta-analysis (n = 34).

Author	Reference	Year	Diabetes Prediction	Sample Size	Sensitivity (%)	Specificity (%)	Overall Classification Accuracy (%)	Classification Technique	Country First Author	Impact Factor
Rathmann et al.	[5]	2010	Prognostic	1353			88	LR	Germany	3.11
Upadhyaya et al.	[21]	2013	Diagnostic	663	97	99	98	NN	USA	5.45
Wang et al.	[6]	2013	Diagnostic	8640	87	79	90	ANN	China	3.24
Huang et al.	[7]	2015	Nephropathy	345	85	83	85	DT	China	3.24
Kuo et al.	[8]	2020	Diagnostic	149			78	DT	China	2.38
Pei et al.	[9]	2019	Prognostic	4205			95	DT	China	2.07
Casanova et al.	[10]	2016	Prognostic	2363			82	RF	USA	2.78
Rau et al.	[2]	2016	Risk factor analysis	2060	75	75	88	ANN	Taiwan	3.63
Ramezankhani et al.	[11]	2014	Prognostic	1995	31	98	91	DT	Iran	3.24
Ramezankhani et al.	[12]	2016	Prognostic	6647	70	79	78	DT	Iran	2.38
Ramezankhani et al.	[13]	2016	Prognostic	1164	22	99	91	DT	Iran	2.79
Dugee et al.	[14]	2015	Diagnostic	1018			76	LR	Mongolia & Finland	2.57
Esmaeily et al.	[15]	2018	Diagnostic	9528	71	70	71	RF	Iran	1.51
Heydari et al.	[22]	2016	Diagnostic	2536	98	67	95	DT	Iran	0.59
Nanri et al.	[23]	2015	Prognostic	37,416	84	80	80	LR	Japan	2.78
Cichosz et al.	[24]	2014	Diagnostic	5381			85	LR	Denmark	3.30
Olivera et al.	[25]	2017	Diagnostic	3709	66	69	74	ANN	Brazil	0.13
Upadhyaya et al.	[4]	2017	Prognostic	4208	99	99	58	Phenotyping	USA	0.00
Usharani & Shanthini	[26]	2021	Nephropathy	768			79	LR	India	4.59
Rodriguez-Romero et al.	[27]	2019	Nephropathy	6777			83	RF	USA	3.99
Parashar et al.	[28]	2014	Diagnostic	768			77	SVM	China	2.5
Farahmandian et al.	[29]	2015	Diagnostic	768			81	SVM	Iran	0.00
Khashei et al.	[30]	2012	Diagnostic	768			80	SVM	Iran	0.00
Bozkurt et al.	[31]	2014	Diagnostic	768	53	89	76	ANN	India	0.68
Kumari & Chitra	[32]	2013	Diagnostic	460			78	SVM	India	1.45
Anderson et al.	[33]	2016	Screening and diagnosis	9948	80	73	75	LR	USA	2.95

Table 1. *Cont.*

Author	Reference	Year	Diabetes Prediction	Sample Size	Sensitivity (%)	Specificity (%)	Overall Classification Accuracy (%)	Classification Technique	Country First Author	Impact Factor
Alssema et al.	[34]	2011	Prognostic	18,301			74	LR	The Netherlands	7.11
Chen et al.	[35]	2015	Nephropathy	519			89	ANN	China	4.19
Marateb et al.	[36]	2014	Nephropathy	200	95	85	92	Hybrid model	Iran	3.43
Leung et al.	[37]	2013	Nephropathy	673			95	SVM	China	2.03
Chikh et al.	[38]	2012	Screening and diagnosis	768	85	92	89	CRISP	Algeria	3.06
Zheng et al.	[39]	2017	Screening and diagnosis	300			98	LR	China	3.03
Yu et al.	[40]	2016	Nephropathy	299	83	88	87	ANN	Taiwan	0.43
Meng et al.	[41]	2013	Risk factor analysis	1487	81	75	78	DT	China	1.74

3.3. Spatial Distribution of Articles and Soft-Computing Models

Machine learning techniques are popular compare to other methods due to their outstanding classification performance. The distribution of articles by year of publication is shown in Figure 2a. It was evident that publications related to the application of ML techniques in diagnosing diabetes mellitus increased significantly from 2013 to 2016. Based on the inclusion criteria, we also noted a downward trend for publications in the past four years. Many factors could be attributed to this downward trend, but we can only attribute this observation to the inclusion criteria and the disease under investigation in the current study. Thus, we cannot generalize since other researchers can apply the techniques to other diseases.

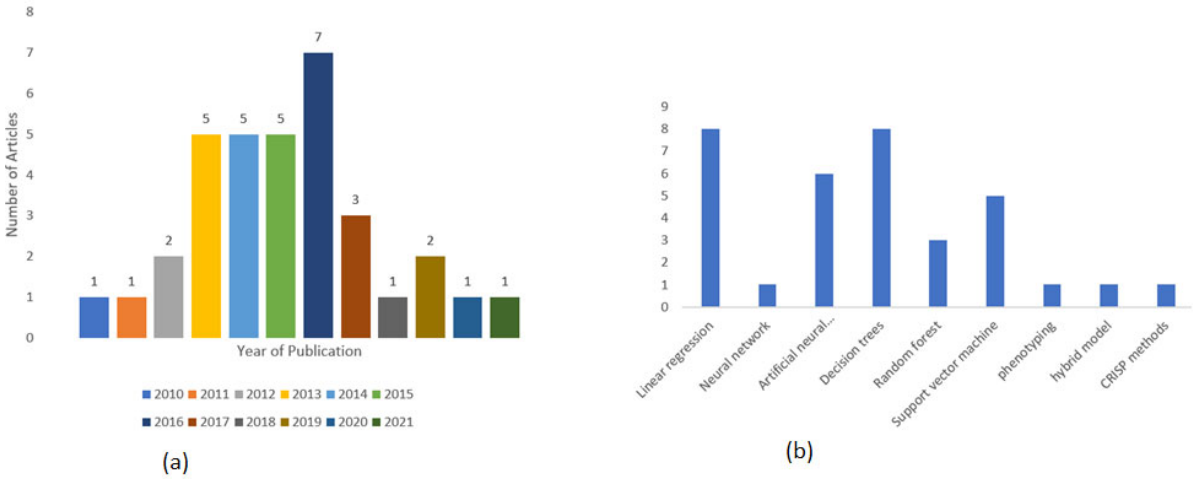


Figure 2. Distribution of articles and soft computing models in meta-analyses: (**a**) Classification of articles by year of publication; (**b**) Soft computing models in general.

Additionally, Figure 2b shows the frequency of algorithms applied specifically in ML. Based on the articles that met the inclusion criteria, decision trees are the most significantly used ML techniques in predicting T2DM. It can be said that the four most popular ML models are LR, ANN, DT, and SVM, consecutively.

Given the data sources for the included articles, Figure 3a shows the trend between the impact factor and the publication year. There were 22 studies released between 2013 and 2016 (65%). Algeria and Japan each contributed to one study (medium impact factor = 3.06 and 2.78, respectively); China and the United States each contributed to nine and five studies (medium impact factor = 2.71 and 3.03, respectively). In addition, a substantial impact study was conducted in the Netherlands [34], and another was conducted in Denmark [24]. A moderate impact factor was undertaken in Germany [5], and Brazil and Iran each contributed to a lower impact factor. Figure 3b gives an exhaustive comparison of regional differences in publications and the country's average impact factor.

Furthermore, Figure 4 shows the frequency of ML applications to health aspects of T2DM. The results showed that the most common medical application of ML for T2DM care was diagnostics, with a 38% frequency, followed by prognostics (26%).

3.4. Results of the Meta-Analysis

Proportions of Classification Accuracy

As acknowledged earlier, the meta-analysis results were based on the 34 documents that met the inclusion criteria. The summary proportion was presented as a random effect due to the heterogeneity of estimates across studies. This classification accuracy was 86% (95% CI: 82–89%). The I^2 was 99.00% (95% CI: 99.54–99.84%) of the total variance between studies. A possible reason for this high heterogeneity could be attributed to the sampling error between studies and other design aspects. Tau I^2 was 59% (95% CI: 0.39–1.13%) (SE = 0.1128). The Q statistic Q (df = 33) = 4202.3722, p-value < 0.0001, which indicated that

the included studies did share a standard effect size (Figure 5). So, we concluded that our analysis had substantial homogeneity (Figure 5).

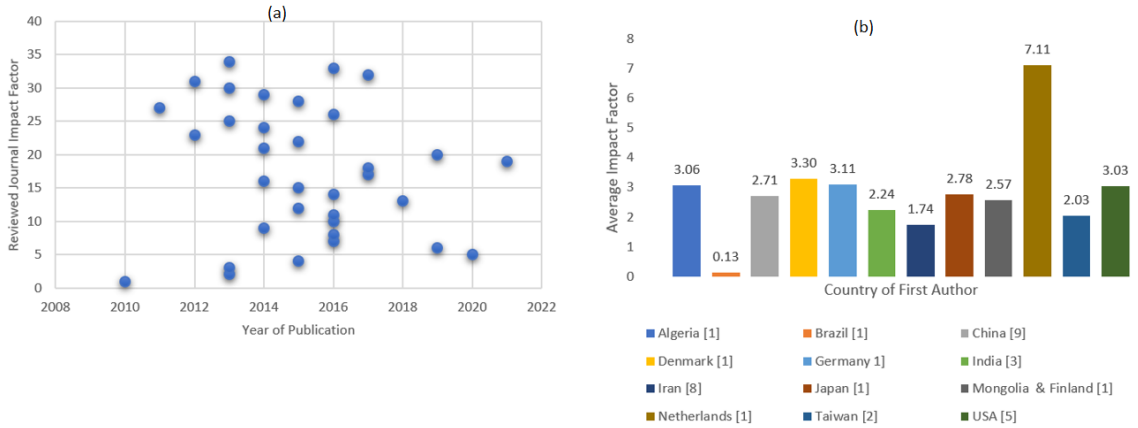


Figure 3. Graphical representation of the temporal and regional trends: (**a**) Temporal trends in publications; (**b**) Regional trends in publications.

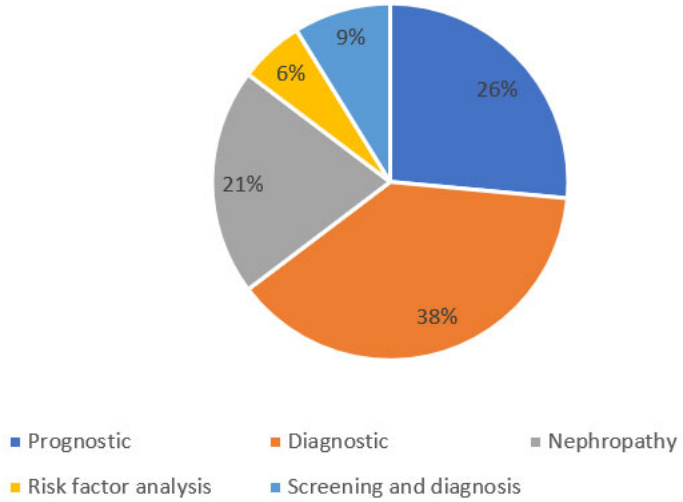


Figure 4. Frequency of machine learning applications for health aspects of type 2 diabetes mellitus.

3.5. ML Models and Diabetes Prediction

Machine learning approaches became a standard solution for processing big data analytics when the scope of theoretical knowledge of a problem is incomplete [42] and when the preliminary statistical data are unknown [43]. Because of these factors, combined with their robustness as one of the best techniques to solve non-linear geo-environmental problems, ML techniques are increasingly used in disease forecasting. In addition, different varieties exist within an ML model, and their performance varies depending on the area under investigation and the input data. Due to variations in sample size, studies, inclusion criteria, and methodology, heterogeneity examination in meta-analyses becomes inevitable. Classification diabetes prediction significantly differed between diabetes predictions. It was greatest among models with a screening and diagnosis (p = 3, proportion = 0.91, 95% CI [0.74, 0.97]) when compared to nephropathy (p = 7, proportion = 0.88, 95% CI [0.83, 0.91]), prognostic (proportion = 0.84, 95% CI [0.77, 0.90]), diagnostic (proportion = 0.84, 95% CI [0.77, 0.89]) and risk factor analysis (proportion = 0.84, 95% CI [0.76, 0.89]) (Figure 6).

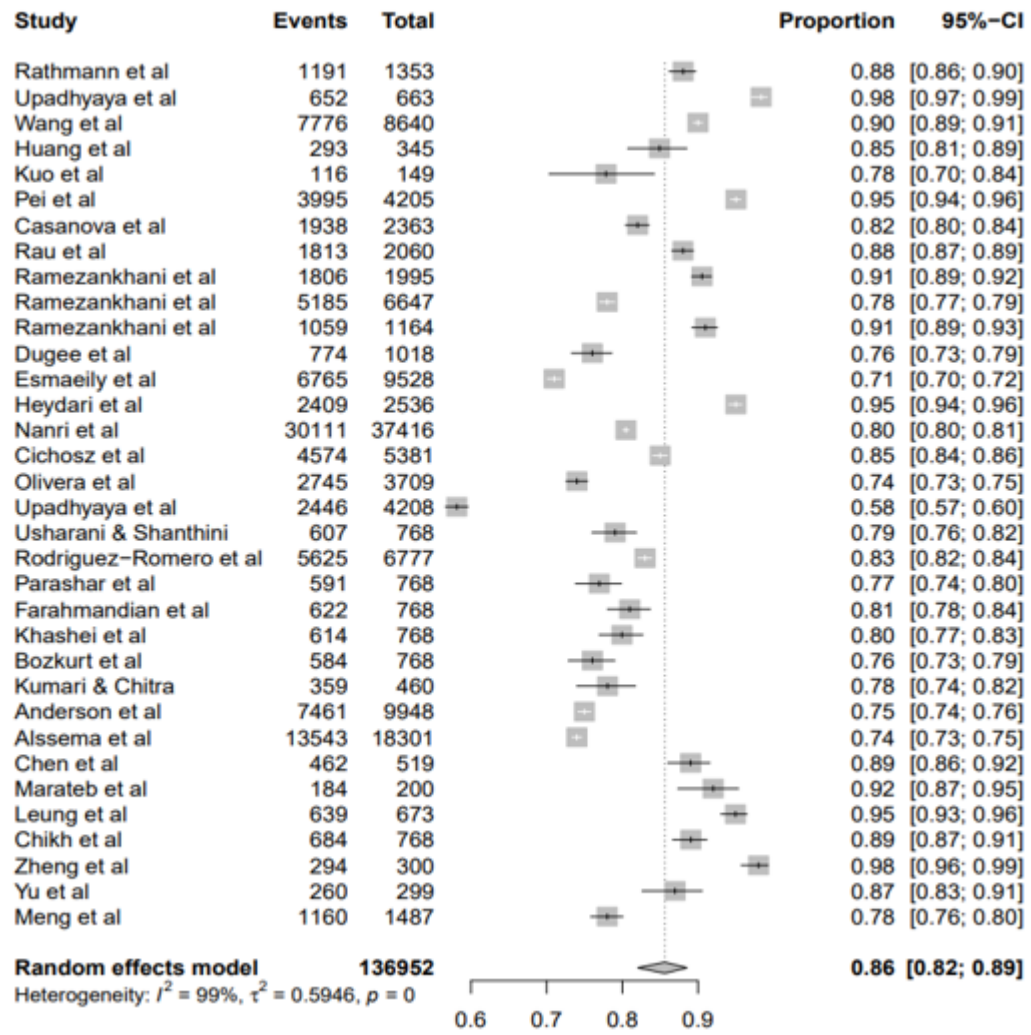


Figure 5. Forest plots showing the proportion of classification accuracy ML models for T2DM.

3.6. ML Models and Prediction of T2DM

In recent years, ML models have been more widely and increasingly applied in biomedical fields. However, given their complexity and potential clinical implications, there is an ongoing need for further research on their accuracy. The prediction performance of each soft computing approach was compared by using either the accuracy or the area under the curve (AUC) of the receiver operating characteristic curve. Based on the systematic literature, the articles that met the inclusion criteria reported the following algorithms for the prediction of T2DM: DT, hybrid model, LR, NN, phenotyping, RF and SVM, classification algorithm and combined the prediction of them into one to-increase the prediction accuracy of the algorithm. Moreover, for the prediction of T2DM, the DT and ANN models had a pooled accuracy of (p = 8, proportion = 0.88, 95% CI [0.82, 0.92]) and (p = 6, proportion = 0.85, 95% CI [0.79, 0.89]), resulting in the best approaches in these meta-analyses, respectively. We believe these findings could represent an encouraging step toward the translation to clinical prediction, diagnosis, and prognosis (Figure 7).

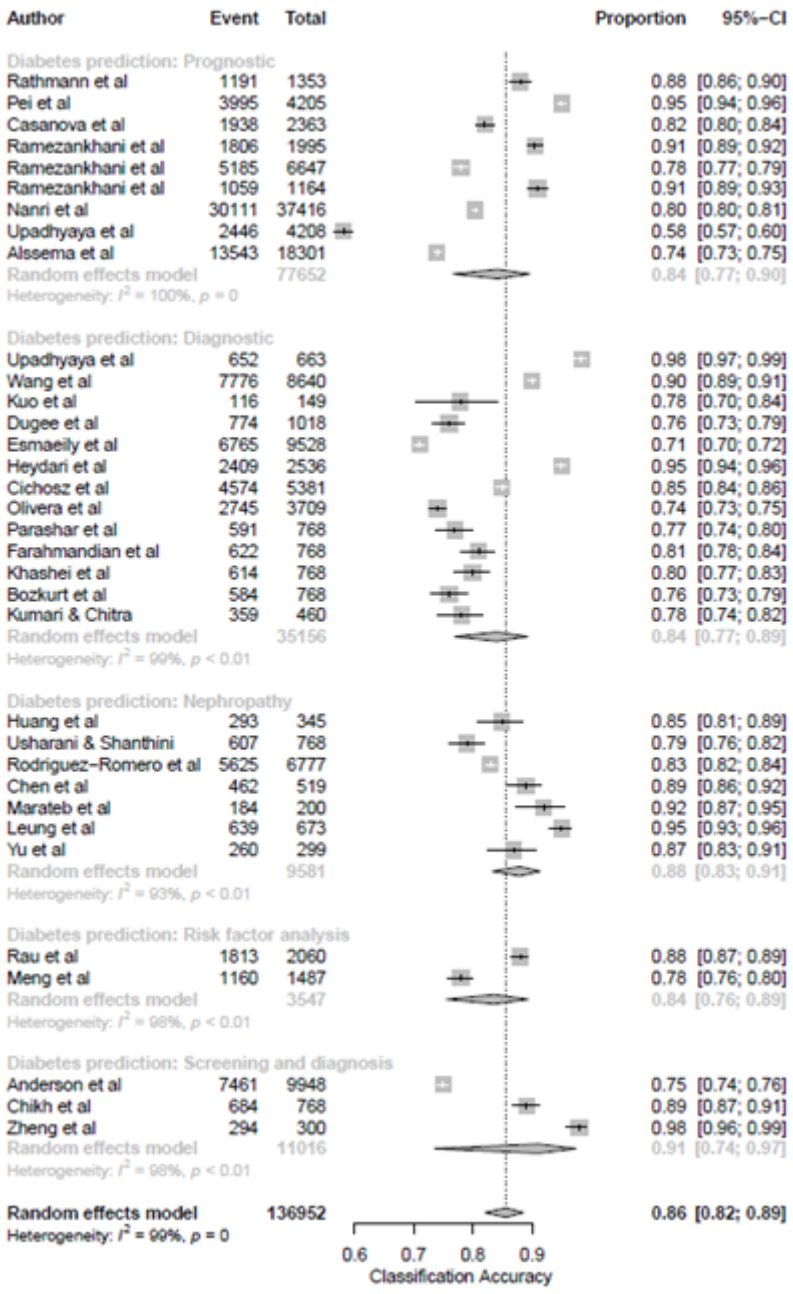


Figure 6. Subgroup analysis of classification accuracy proportions reported by studies that applied a machine learning model to predict type 2 diabetes mellitus.

Additionally, according to the "no free lunch" theorem (Wolpert et al., 1995), no single learning algorithm universally performs best across all domains. As such, several models should be tested and compared. Thus, these approaches mentioned above were further classified into a linear or non-linear model for straightforward interpretation. The purpose of this section was to compare the classification performance of linear and non-linear ML models for the prediction of T2DM. Overall, non-linear ML models outperformed linear models for the prediction of T2DM (Figure 8). This valuable relative performance information can help researchers select an appropriate non-linear ML model for their studies.

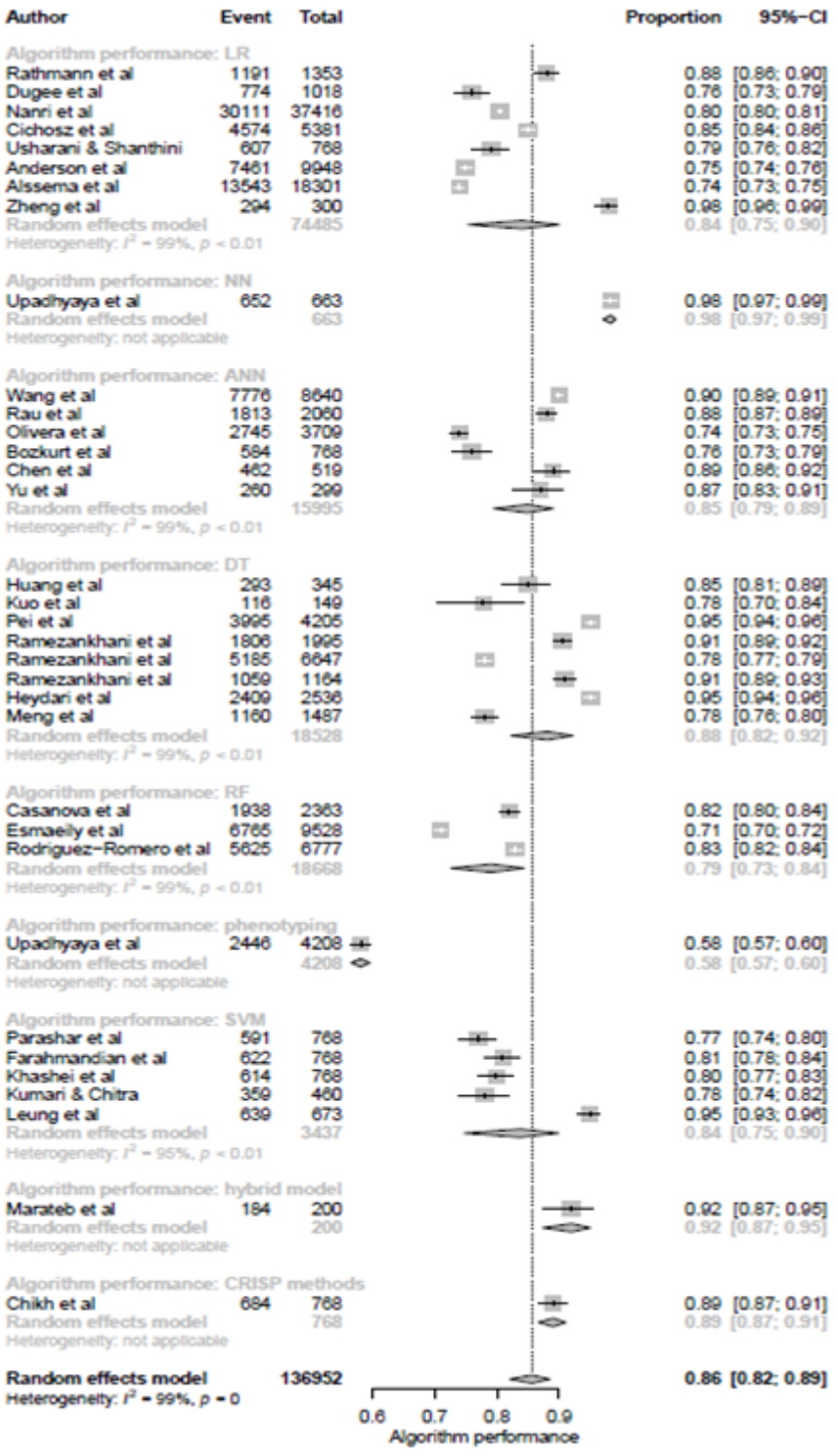


Figure 7. Subgroup analysis based on the various machine learning models for predicting type 2 diabetes mellitus.

3.7. Moderator Analysis

The meta-regression analysis in Table 2 shows that the categorical variables affirmed that the publication year and impact factor did not affect variance in the pooled estimates of classification accuracy. Application for T2DM did not significantly moderate the pooled estimates of classification accuracy, explaining 5.52% of the variance in the pooled classification accuracy proportions (Q_M = 2.24, df = 3, p = 0.6923; Q_E = 3941.8090, df = 29, p < 0.0001). Additionally, the model types significantly moderated the pooled estimates

of classification accuracy, explaining 46.76% of the variance in the pooled classification accuracy proportions (Q_M = 26.04, df = 8, p = 0.0010; Q_E = 2473.3453, df = 25, $p < 0.0001$). The moderation effects of the model types were driven mainly by the NN and phenotyping subgroup, which accounted for an average total variance in the observed proportions (NN: β = 2.36, p = 0.0004) and (phenotyping: $\beta = -1.40$, p = 0.0196), respectively. None of the other model types' subgroups were statistically significant (Table 2). However, the combined model, publication year, impact factor, and application for T2DM and model types explained more heterogeneity (I^2 = 98.49%, p = 0.007, and R^2 = 54.61%). The pooled classification accuracy proportions decreased insignificantly with an increasing publication year (p = 0.5001) and sample size (p = 0.1540) (Figures 9 and 10).

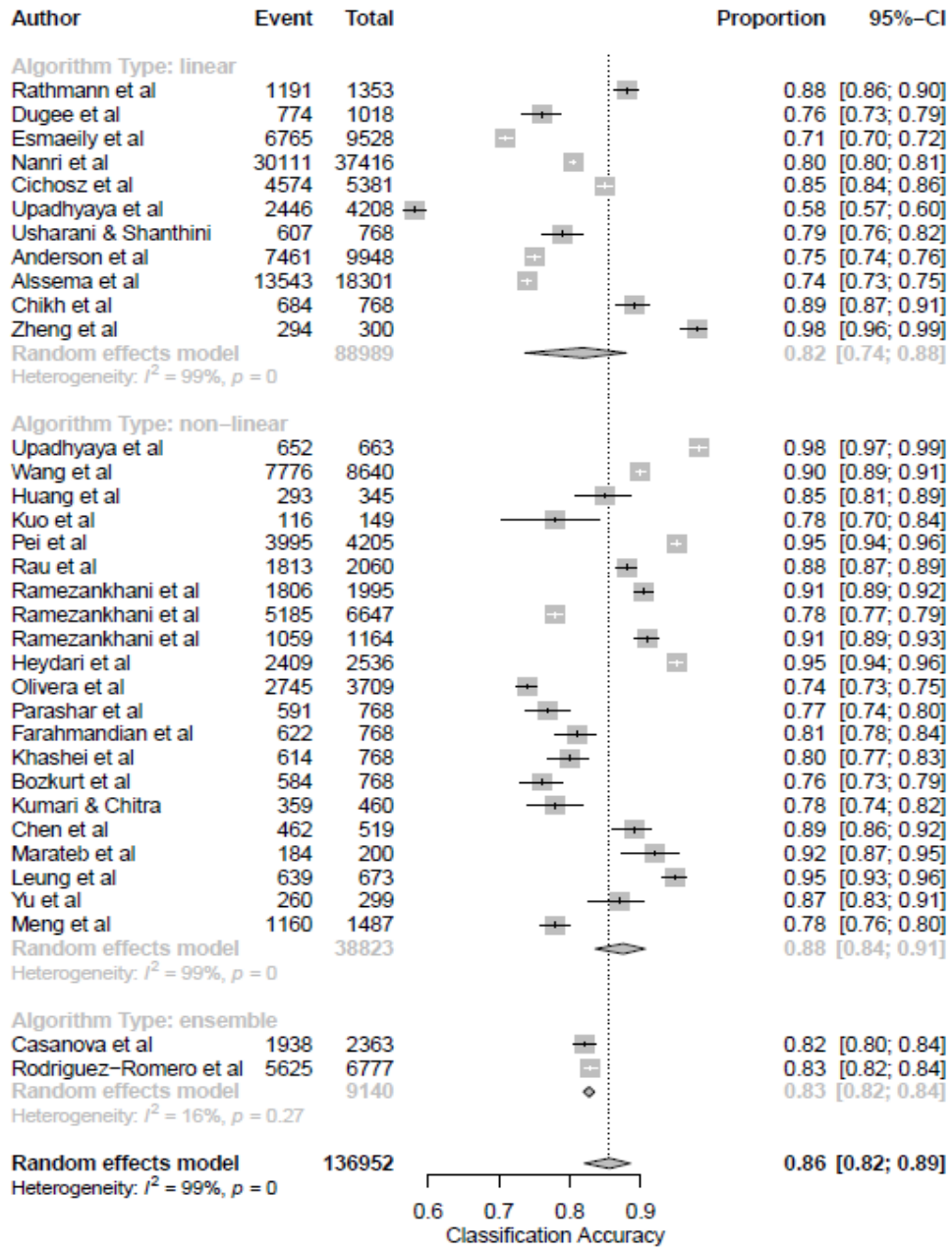


Figure 8. Subgroup analysis based on the machine learning model's performance in predicting the diagnosis of type 2 diabetes mellitus.

Table 2. Meta-analytic regression results (* implies significant value).

Model	β	SE	*p*-Values	Q_M	df	*p*-Values
Publication year	−0.0359	0.0532	0.5001	0.4546	1	0.5001
Impact factor	0.1297	0.0809	0.1086	2.5747	1	0.1086
Diabetes prediction				2.2366	4	0.6923
Diagnostic	**Ref**					
Nephropathy	0.3391	0.3516	0.3348			
Prognostic	0.0219	0.3213	0.9456			
Risk factor analysis	−0.0210	0.5617	0.9701			
Screening and diagnosis	0.5811	0.4905	0.2361			
Model types				26.0392	8	0.0010
ANN	**Ref**					
CRISP method	0.3714	0.6090	0.5420			
Decision trees	0.2786	0.3035	0.3586			
Hybrid model	0.7166	0.6523	0.2720			
Linear regression	−0.1191	0.3047	0.6959			
Neural network	2.3564	0.6708	0.0004 *			
Phenotyping	−1.3977	0.5988	0.0196 *			
Random forest	−0.3935	0.3933	0.3171			
Support vector machine	−0.0946	0.3409	0.7813			

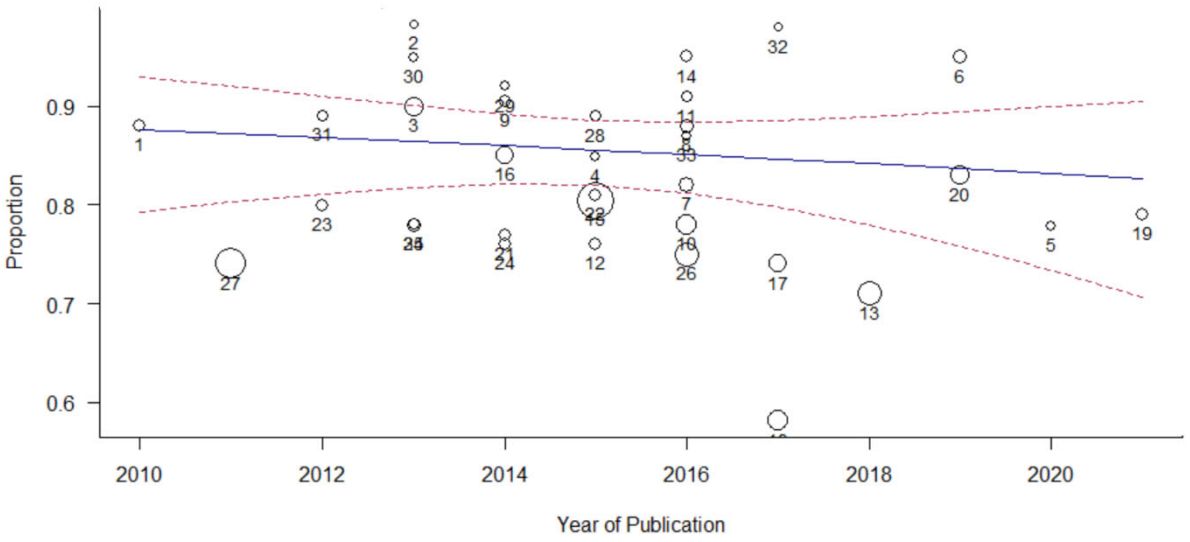


Figure 9. Meta-regression of the performance of the machine learning model in predicting type 2 diabetes mellitus. The study displays the observed effect sizes of the individual studies against the continuous variable publication year.

3.8. Evaluation of Publication Bias

A funnel plot was generated to explore the potential for publication bias. We detected no potential publication bias based on the symmetric shape of the funnel plot of the pooled model performance (Figure 11) and the Eggers' regression test's non-significant value (slope = 0.253, $p = 0.196$). Two studies (2, 6) were identified as outliers with a cut-off of ($>z^2$), and the Baujat plot showed that there was no single study that influenced the results, and each point represents the number of studies (Figure 12).

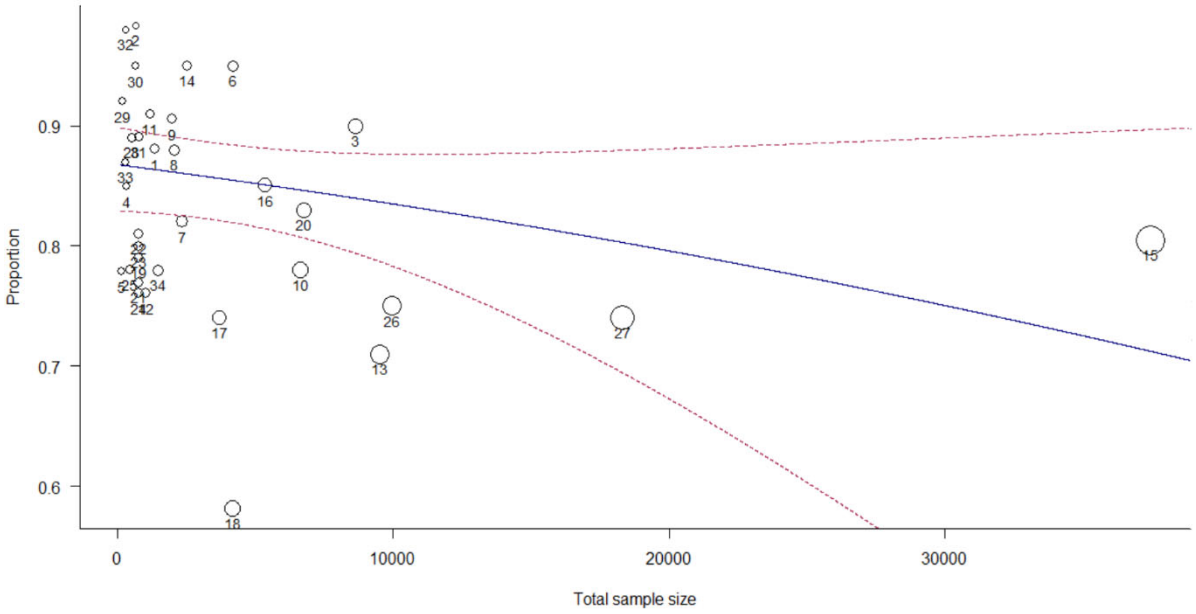


Figure 10. The study displays the observed effect sizes of the individual studies against the continuous variable sample size.

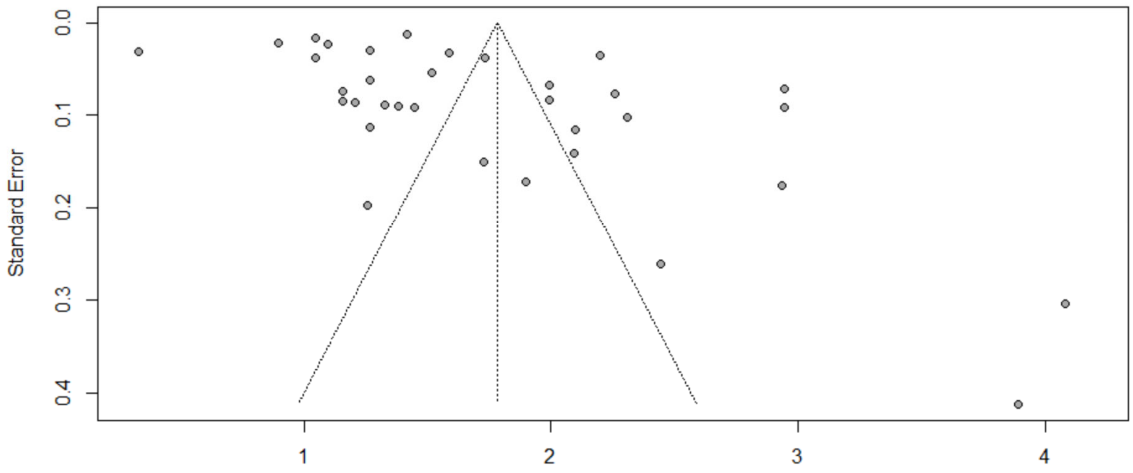


Figure 11. Funnel plot for the evaluation of potential publication bias. Each solid circle represents a study in the meta-analysis.

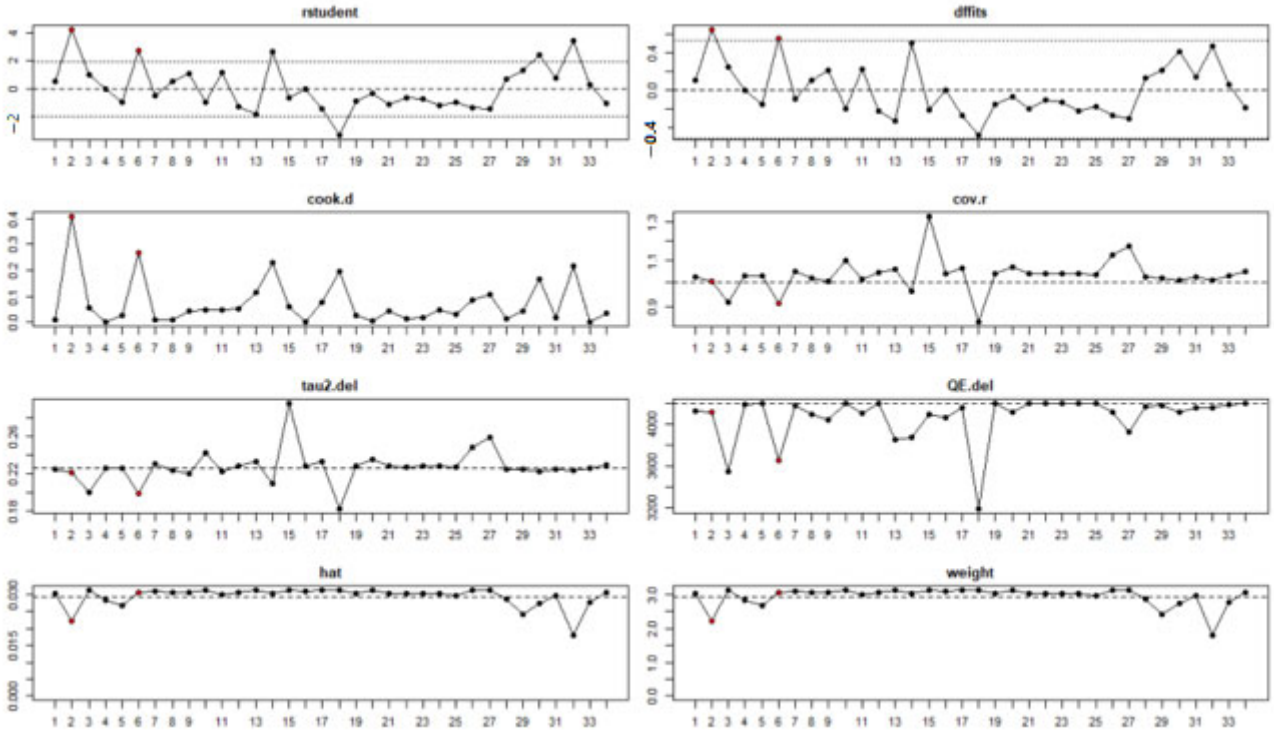


Figure 12. Baujat plot shows no single study that influenced the results.

4. Discussion

4.1. Synopsis of Evidence

In recent years, information technologies such as ML models have become essential in predicting T2DM in patients and assigning management to healthcare providers. A significant research focus has been on developing intelligent digital health interventions. To our knowledge, this is the foremost and largest innovative systematic meta-analytic approach in ML model research at a global level, which drew from a wide-ranging number of articles that included over one thousand participants reporting the ML model's prediction in T2DM disease. In this study, we evaluated the predictive performances of studies using ML prediction models for T2DM. Primary articles were chosen from the Web of Science, Scopus, and PubMed research databases. ML techniques, mixed with other perceptions presented in the learning healthcare systems method, tend to bring better care and management of T2DM to the well-being of society.

Nevertheless, when presenting novel prediction models, one should consider the predictive performance, where the strengths and weaknesses of the ML approaches need to be considered. Recently, numerous modeling methods have been used to predict T2DM and manage T2DM; thus, selecting the most appropriate ML approaches for a specific problem one is trying to solve is always challenging. In this study, we pooled various ML approaches used in previous studies related to T2DM and compared their performance in terms of accuracy. However, the publication year and impact factor did not moderate the aggregate estimates of overall classification accuracy in the meta-regression analyses. However, it is essential to note that our research was limited to the English language. The pooled models' performance predicted T2DM with an overall accuracy of 86% (95% CI: 82%, 89%), similar to the 82% pooled therapeutic outcomes in depression reported recently [44]. The current pool is slightly higher than the overall c-index of 81.2% reported from a meta-analysis study of use and performance for diabetes prediction in a local setting [45]. This disparity could be attributed to differences in the burden of the disease across study settings, the sensitivity of the diagnostic assays used during these two different periods, and the choice and characteristics of study subjects. High predictive performance was achieved by all models, with accuracy ranging from 0.58 to 0.98. Compared to other models, the DT

model performed the best, with an accuracy value of 0.88 (95% CI 0.82–0.92). However, this finding is not surprising because previous studies have revealed that the same ML model can produce diverse accuracy outcomes for the same dataset by selecting various values for the underlying hyperparameters [46,47]. Previous studies have demonstrated the significant role of the DT approach in other medical fields, such as therapeutic outcomes in depression [44] and cardiovascular diseases [48] and predicting diabetes mellitus [49]. Our results confirmed the outstanding performance of the DT method in the risk assessment of T2DM.

Additionally, we grouped the various ML models into three categories: linear, non-linear, and ensemble. The models that used non-linear algorithms to predict T2DM performed better (0.88, 95% CI 0.84–0.91) than the linear model and ensemble modeling approach. This finding is consistent with the previous comparison between linear and non-linear models for classifying thyroid modules [50]. In addition, we also observed that the models based on ML for prediction in T2DM had been mainly focused on screening and diagnostics (0.91, 95% CI [0.74, 0.97]). This observation is also supported by the previous meta-analysis that utilized the ML model for therapeutic outcomes in depression [44]. A possible reason for this finding could be the variation in the year of publication. Our results show a broad spectrum of applications of ML models dominated by predictive approaches.

4.2. Policy Implications

Since the discovery of non-infectious diseases, many scientific publications have been produced globally. The current T2DM offers a wide-ranging analysis of the research trends linked to T2DM through documents indexed in academic databases. At the same time, the findings from this systematic survey and meta-analysis have significant policy implications for evaluation and monitoring. These are adequate resources for clinicians to determine if an individual will develop type 2 diabetes mellitus in the coming time. Additionally, synthesizing individuals with T2DM is essential in assisting clinicians in designing an appropriate mechanism to protect vulnerable individuals and reduce pressure on health systems. The current ML techniques have outclassed conventional risk models in predicting T2DM. Still, individuals should be careful about changing their attitude regarding future diabetes risk after having the outcomes of a diabetes prediction test through ML techniques. In addition, ML techniques are vital to improving the predictive capacity of T2DM. Ongoing work should be carried out to build additional precise ML techniques other than the existing ones, supposing that the practicability of utilizing ML in a clinical situation would be improved compared to regular costly and time-consuming blood tests. Finally, the pooling of independent studies gives policymakers the information needed to make informed decisions in uncertain circumstances.

4.3. Limitations of the Overview Study

A wide-ranging literature search and watchful data extraction were conducted to avoid bias. However, limitations exist in our study. This systematic meta-analysis was limited to articles written in the English language. In addition, only articles written between 2010 and 2021 were included in the study. Secondly, the authors may have overlooked some valuable keywords and bibliographic sources that may contain relevant articles. Furthermore, due to the scarcity of primary studies, very few preliminary studies have been included to aggregate the accuracy of predictive models at the global level. As a result, in the future, the scope of the study may be broadened to reflect such limitations.

4.4. Concluding Remarks and Recommendations

This paper provided an in-depth study of automated T2DM prediction models. It reveals how the data mining and meta-analysis approach can be efficiently implemented in clinical medicine to obtain models that use patient-specific information to predict the end product. Critical articles were compiled from the Web of Science, Scopus, and PubMed scientific repositories. The classification models predicted outcomes for patients diagnosed

with T2DM in previously published documents (p = 34, n = 136, 952), with an overall accuracy of 86%. The pooled estimates of classification accuracy differed significantly from model to model based on applying the algorithm to T2DM ($p < 0.01$). Predictive models with screening and diagnostics had the most significant overall classification accuracy (pooled proportion = 0.91) compared to models with other algorithms for T2DM (proportion = 0.84 to 0.88).

In summary, our results on the aggregate estimates of model performance can help researchers and decision-makers undertake health technology assessments for various T2DM screening strategies. Hopefully, this analysis will benefit researchers involved in DM therapy's detection, diagnosis, self-management, and personalization. Additionally, the findings can provide an exhaustive overview of the relative performance of diverse variants of ML models for disease prediction. The implication is that it can aid researchers in selecting appropriate ML models for their studies. Finally, we recommend comparing different ML models to develop a predictive model based on our meta-analysis.

5. Conclusions

We pooled data from previous independent studies to determine the overall ML models' predictive ability of T2DM disease. This systematic review and meta-analysis show that ML models can correctly predict T2DM with good discrimination. Our findings indicated that the decision trees model has the best prediction performances, with excellent accuracy compared to other soft-computing models in this systematic meta-analysis. Moreover, this finding suggests that ML algorithms have a high capacity for advanced enhancement of predictive ability for T2DM. The results are expected to further the global research agenda, and policymakers could use the findings to strengthen medical policies in the clinical diagnosis of a patient with T2DM. This calls for the development of informing procedures for ML for intensive care medicine.

Author Contributions: M.O.O. conceived the study; R.E.O. conducted the search, selected primary studies, and extracted and analyzed the data. M.O.O., R.E.O., M.G. and M.A.A. were involved in the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw and processed data are available upon request to the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Rigla, M.; García-Sáez, G.; Pons, B.; Hernando, M.E. Artificial Intelligence Methodologies and Their Application to Diabetes. *J. Diabetes Sci. Technol.* **2017**, *12*, 303–310. [CrossRef] [PubMed]
2. Rau, H.-H.; Hsu, C.-Y.; Lin, Y.-A.; Atique, S.; Fuad, A.; Wei, L.-M.; Hsu, M.-H. Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. *Comput. Methods Programs Biomed.* **2016**, *125*, 58–65. [CrossRef] [PubMed]
3. Muhammad, L.J.; Algehyne, E.A.; Usman, S.S. Predictive supervised machine learning models for diabetes mellitus. *SN Comput. Sci.* **2020**, *1*, 1–14. [CrossRef] [PubMed]
4. Upadhyaya, S.G.; Murphree, D.H., Jr.; Ngufor, C.G.; Knight, A.M.; Cronk, D.J.; Cima, R.R.; Curry, T.B.; Pathak, J.; Carter, R.E.; Kor, D.J. Automated diabetes case identification using electronic health record data at a tertiary care facility. *Mayo Clin. Proc. Innov. Qual. Outcomes* **2017**, *1*, 100–110. [CrossRef]
5. Rathmann, W.; Kowall, B.; Heier, M.; Herder, C.; Holle, R.; Thorand, B.; Strassburger, K.; Peters, A.; Wichmann, H.E.; Giani, G.; et al. Prediction models for incident type 2 diabetes mellitus in the older population: KORA S4/F4 cohort study. *Diabet. Med.* **2010**, *27*, 1116–1123. [CrossRef]

6. Wang, C.; Li, L.; Wang, L.; Ping, Z.; Flory, M.T.; Wang, G.; Xi, Y.; Li, W. Evaluating the risk of type 2 diabetes mellitus using artificial neural network: An effective classification approach. *Diabetes Res. Clin. Pract.* **2013**, *100*, 111–118. [CrossRef]
7. Huang, G.-M.; Huang, K.-Y.; Lee, T.-Y.; Weng, J.T.-Y. An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. *BMC Bioinform.* **2015**, *16* (Suppl. S1), S5. [CrossRef]
8. Kuo, K.-M.; Talley, P.; Kao, Y.; Huang, C.H. A multi-class classification model for supporting the diagnosis of type II diabetes mellitus. *PeerJ* **2020**, *8*, e9920. [CrossRef]
9. Pei, D.; Gong, Y.; Kang, H.; Zhang, C.; Guo, Q. Accurate and rapid screening model for potential diabetes mellitus. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 1–8. [CrossRef]
10. Casanova, R.; Saldana, S.; Simpson, S.L.; Lacy, M.E.; Subauste, A.R.; Blackshear, C.; Wagenknecht, L.; Bertoni, A.G. Prediction of Incident Diabetes in the Jackson Heart Study Using High-Dimensional Machine Learning. *PLoS ONE* **2016**, *11*, e0163942. [CrossRef]
11. Ramezankhani, A.; Pournik, O.; Shahrabi, J.; Khalili, D.; Azizi, F.; Hadaegh, F. Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study. *Diabetes Res. Clin. Pract.* **2014**, *105*, 391–398. [CrossRef]
12. Ramezankhani, A.; Hadavandi, E.; Pournik, O.; Shahrabi, J.; Azizi, F.; Hadaegh, F. Decision tree-based modelling for identification of potential interactions between type 2 diabetes risk factors: A decade follow-up in a Middle East prospective cohort study. *BMJ Open* **2016**, *6*, e013336. [CrossRef]
13. Ramezankhani, A.; Pournik, O.; Shahrabi, J.; Azizi, F.; Hadaegh, F.; Khalili, D. The Impact of Oversampling with SMOTE on the Performance of 3 Classifiers in Prediction of Type 2 Diabetes. *Med. Decis. Mak.* **2014**, *36*, 137–144. [CrossRef]
14. Dugee, O.; Janchiv, O.; Jousilahti, P.; Sakhiya, A.; Palam, E.; Nuorti, J.P.; Peltonen, M. Adapting existing diabetes risk scores for an Asian population: A risk score for detecting undiagnosed diabetes in the Mongolian population. *BMC Public Health* **2015**, *15*, 938. [CrossRef]
15. Esmaily, H.; Tayefi, M.; Doosti, H.; Ghayour-Mobarhan, M.; Nezami, H.; Amirabadizadeh, A. A Comparison between Decision Tree and Random Forest in Determining the Risk Factors Associated with Type 2 Diabetes. *J. Res. Health Sci.* **2018**, *18*, e00412.
16. Baum, A.; Scarpa, J.; Bruzelius, E.; Tamler, R.; Basu, S.; Faghmous, J. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: A machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial. *Lancet Diabetes Endocrinol.* **2017**, *5*, 808–815. [CrossRef]
17. Wilkinson, J.; Arnold, K.F.; Murray, E.J.; van Smeden, M.; Carr, K.; Sippy, R.; de Kamps, M.; Beam, A.; Konigorski, S.; Lippert, C.; et al. time to reality check the promises of machine learning-powered precision medicine. *Lancet Digit. Health* **2020**, *2*, e677–e680. [CrossRef]
18. Higgins, J.P.T.; Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **2002**, *21*, 1539–1558. [CrossRef]
19. Ogunsakin, R.E.; Olugbara, O.O.; Moyo, S.; Israel, C. Meta-analysis of studies on depression prevalence among diabetes mellitus patients in Africa. *Heliyon* **2021**, *7*, e07085. [CrossRef]
20. DerSimonian, R.; Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **1986**, *7*, 177–188. [CrossRef]
21. Upadhyaya, S.; Farahmand, K.; Baker-Demaray, T. Comparison of NN and LR classifiers in the context of screening native American elders with diabetes. *Expert Syst. Appl.* **2013**, *40*, 5830–5838. [CrossRef]
22. Heydari, M.; Teimouri, M.; Heshmati, Z.; Alavinia, M. Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran. *Int. J. Diabetes Dev. Ctries.* **2015**, *36*, 167–173. [CrossRef]
23. Nanri, A.; Nakagawa, T.; Kuwahara, K.; Yamamoto, S.; Honda, T.; Okazaki, H.; Uehara, A.; Yamamoto, M.; Miyamoto, T.; Kochi, T.; et al. Correction: Development of Risk Score for Predicting 3-Year Incidence of Type 2 Diabetes: Japan Epidemiology Collaboration on Occupational Health Study. *PLoS ONE* **2018**, *13*, e0199075. [CrossRef] [PubMed]
24. Cichosz, S.L.; Johansen, M.D.; Ejskjaer, N.; Hansen, T.K.; Hejlesen, O.K. A novel model enhances HbA1c-based diabetes screening using simple anthropometric, anamnestic, and demographic information. *J. Diabetes* **2014**, *6*, 478–484. [CrossRef] [PubMed]
25. Olivera, A.R.; Roesler, V.; Iochpe, C.; Schmidt, M.I.; Vigo, Á.; Barreto, S.M.; Duncan, B.B. Comparison of ma-chine-learning algorithms to build a predictive model for detecting undiagnosed diabetes-ELSA-Brasil: Accuracy study. *Sao Paulo Med. J.* **2017**, *135*, 234–246. [CrossRef] [PubMed]
26. Usharani, R.; Shanthini, A. Neuropathic complications: Type II diabetes mellitus and other risky parameters using machine learning algorithms. *J. Ambient. Intell. Humaniz. Comput.* **2021**, 1–23. [CrossRef]
27. Rodriguez-Romero, V.; Bergstrom, R.F.; Decker, B.S.; Lahu, G.; Vakilynejad, M.; Bies, R.R. Prediction of nephropathy in type 2 diabetes: An analysis of the ACCORD trial applying machine learning techniques. *Clin. Transl. Sci.* **2019**, *12*, 519–528. [CrossRef]
28. Parashar, A.; Burse, K.; Rawat, K. A Comparative approach for Pima Indians diabetes diagnosis using lda-support vector machine and feed forward neural network. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2014**, *4*, 378–383.
29. Farahmandian, M.; Lotfi, Y.; Maleki, I. Data mining algorithms application in diabetes diseases diagnosis: A case study. *MAGNT Res. Tech. Rep.* **2015**, *3*, 989–997.
30. Khashei, M.; Eftekhari, S.; Parvizian, J. Diagnosing diabetes type II using a soft intelligent binary classification model. *Rev. Bioinform. Biom.* **2012**, *1*, 9–23.
31. Bozkurt, M.R.; Yurtay, N.; Yilmaz, Z.; Sertkaya, C. Comparison of different methods for determining diabetes. *Turk. J. Electr. Eng. Comput. Sci.* **2014**, *22*, 1044–1055. [CrossRef]
32. Kumari, V.A.; Chitra, R. Classification of diabetes disease using support vector machine. *Int. J. Eng. Res. Appl.* **2013**, *3*, 1797–1801.

33. Anderson, A.E.; Kerr, W.T.; Thames, A.; Li, T.; Xiao, J.; Cohen, M.S. Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: A cross-sectional, unselected, retrospective study. *J. Biomed. Inform.* **2016**, *60*, 162–168. [CrossRef]
34. Alssema, M.; Vistisen, D.; Heymans, M.W.; Nijpels, G.; Glümer, C.; Zimmet, P.Z.; Shaw, J.E.; Eliasson, M.; Stehouwer, C.D.; Tabák, A.G.; et al. The Evaluation of Screening and Early Detection Strategies for Type 2 Diabetes and Im-paired Glucose Tolerance (DETECT-2) update of the Finnish diabetes risk score for prediction of incident type 2 diabetes. *Diabetologia* **2011**, *54*, 1004–1012. [CrossRef]
35. Chen, J.; Tang, H.; Huang, H.; Lv, L.; Wang, Y.; Liu, X.; Lou, T. Development and validation of new glomerular filtration rate predicting models for Chinese patients with type 2 diabetes. *J. Transl. Med.* **2015**, *13*, 317. [CrossRef]
36. Marateb, H.R.; Mansourian, M.; Faghihimani, E.; Amini, M.; Farina, D. A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin. *Comput. Biol. Med.* **2014**, *45*, 34–42. [CrossRef]
37. Leung, R.K.; Wang, Y.; Ma, R.C.; Luk, A.O.; Lam, V.; Ng, M.; So, W.Y.; Tsui, S.K.; Chan, J. Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: A prospective case–control cohort analysis. *BMC Nephrol.* **2013**, *14*, 162. [CrossRef]
38. Chikh, M.A.; Saidi, M.; Settouti, N. Diagnosis of Diabetes Diseases Using an Artificial Immune Recognition System2 (AIRS2) with Fuzzy K-nearest Neighbor. *J. Med. Syst.* **2011**, *36*, 2721–2729. [CrossRef]
39. Zheng, T.; Xie, W.; Xu, L.; He, X.; Zhang, Y.; You, M.; Yang, G.; Chen, Y. A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int. J. Med. Inform.* **2016**, *97*, 120–127. [CrossRef]
40. Yu, C.S.; Liu, C.S.; Chen, R.S.; Lin, C.W. Artificial neural networks for estimating glomerular filtration rate by urinary dipstick for type 2 diabetic patients. *Biomed Eng Singap.* **2016**, *28*, 1650016.
41. Meng, X.H.; Huang, Y.X.; Rao, D.P.; Zhang, Q.; Liu, Q. Comparison of three data mining models for predicting diabetes or pre-diabetes by risk factors. *Kaohsiung J. Med. Sci.* **2013**, *29*, 93–99. [CrossRef] [PubMed]
42. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [CrossRef]
43. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Han, Z.; Pham, B.T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* **2019**, *17*, 641–658. [CrossRef]
44. Lee, Y.; Ragguett, R.M.; Mansur, R.B.; Boutilier, J.J.; Rosenblat, J.D.; Trevizol, A.; Brietzke, E.; Lin, K.; Pan, Z.; Subramaniapillai, M.; et al. Applications of machine learning algorithms to predict therapeutic outcomes in depression: A me-ta-analysis and systematic review. *J. Affect. Disord.* **2018**, *241*, 519–532. [CrossRef] [PubMed]
45. De Silva, K.; Lee, W.K.; Forbes, A.; Demmer, R.T.; Barton, C.; Enticott, J. Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis. *Int. J. Med Inform.* **2020**, *143*, 104268. [CrossRef]
46. Levy, O.; Goldberg, Y.; Dagan, I. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Trans. Assoc. Comput. Linguist.* **2015**, *3*, 211–225. [CrossRef]
47. Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; Bousquet, O. Are gans created equal? a large-scale study. *arXiv* **2017**, arXiv:1711.10337.
48. Krittanawong, C.; Virk, H.U.H.; Bangalore, S.; Wang, Z.; Johnson, K.W.; Pinotti, R.; Zhang, H.; Kaplin, S.; Narasimhan, B.; Kitai, T.; et al. Machine learning prediction in cardiovascular diseases: A meta-analysis. *Sci. Rep.* **2020**, *10*, 1–11. [CrossRef]
49. Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. Predicting Diabetes Mellitus With Machine Learning Techniques. *Front. Genet.* **2018**, *9*, 515. [CrossRef]
50. Ouyang, F.S.; Guo, B.L.; Ouyang, L.Z.; Liu, Z.W.; Lin, S.J.; Meng, W.; Huang, X.Y.; Chen, H.X.; Qiu-Gen, H.; Yang, S.M. Comparison between linear and non-linear machine-learning algorithms for the classification of thyroid nodules. *Eur. J. Radiol.* **2019**, *113*, 251–257. [CrossRef]