J. Mantas et al. (Eds.)

doi:10.3233/SHTI230582

Predicting Long-Term Type 2 Diabetes with Artificial Intelligence (AI): A Scoping Review

Salleh SONKO[a,1], Fathima LAMYA[a], Mahmood ALZUBAIDI[a], Hurmat SHAH[a], Tanvir ALAM[a], Zubair SHAH[a] and Mowafa HOUSEH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

Abstract. Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder that affects a significant portion of the global population. Artificial intelligence (AI) has emerged as a promising tool for predicting T2DM risk. To provide an overview of the AI techniques used for long-term prediction of T2DM and evaluate their performance, we conducted a scoping review using PRISMA-ScR. Of the 40 papers included in this review, 23 studies used Machine Learning (ML) as the most common AI technique, with Deep Learning (DL) models used exclusively in four studies. Of the 13 studies that used both ML and DL, 8 studies employed ensemble learning models, and SVM and RF were the most used individual classifiers. Our findings highlight the importance of accuracy and recall as validation metrics, with accuracy being used in 31 studies, followed by recall in 29 studies. These discoveries emphasize the critical role of high predictive accuracy and sensitivity in detecting positive T2DM cases.

Keywords. Artificial Intelligence (AI), Machine Learning, Deep Learning, Type-2 Diabetes Mellitus (T2DM)

1. Introduction

Type 2 diabetes mellitus (T2DM) is the most common type of diabetes, accounting for around 90% of all diabetes cases [1]. It develops over many years and is usually diagnosed in adults (but more and more in children, teens, and young adults). Risk factors for T2DM include obesity, physical inactivity, family history, and an unhealthy diet. Symptoms of T2DM may include increased thirst, frequent urination, blurred vision, fatigue, and slow healing of cuts and wounds. T2DM can be prevented or delayed with healthy lifestyle changes, such as: losing weight, eating healthy food, and being active.

The State of Qatar has one of the highest diabetes prevalence rates globally and is ranked among the top ten countries for type 2 diabetes mellitus (T2DM) prevalence. This high prevalence has been largely attributed to several well-established risk factors, including physical inactivity, unhealthy diet, and obesity. According to recent projections, the prevalence of T2DM in Qatar is expected to increase from 16.7% in 2016 to at least 24.0% by 2050, indicating that one in four Qataris will have diabetes by 2050

[1] Corresponding Author: Dr. Mowafa Househ, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; E-mail: mhouseh@hbku.edu.qa.

if no preventive measures are taken. This growth in prevalence is expected to further burden affected individuals and their families, healthcare systems, and society as a whole [2].

Moreover, in Qatar, T2DM is responsible for about 50% of end-stage renal disease requiring dialysis, 50% of all acute coronary syndrome cases, and 70% of stroke and Transient Ischemic Attack (TIA) cases. The annual cost of diabetes in Qatar has been projected to increase from QR 1.8 bn in 2015, to QR 5 bn in 2035, and QR 8.4 bn by 2055 [3].

The high prevalence of diabetes in Qatar, its significant burden on individuals and healthcare systems, and its considerable economic impact underscore the need for a scoping review. Such a review would offer a comprehensive overview of the AI techniques used to predict the onset of T2DM and assess their feasibility. Additionally, it could identify research gaps and directions for future studies.

2. Methods

This scoping review aimed to evaluate the use of Artificial Intelligence (AI) in Long-Term prediction of Type 2 Diabetes Mellitus (T2DM) without any restrictions on demographic features such as age, ethnicity, etc. The study followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA-ScR) [4] guidelines (refer to online Appendix A[2]) and included studies published from 2019 onwards in English language only. PubMed, Google Scholar, and IEEE were the search databases used, with search terms related to the population, intervention, and outcome (refer to online Appendix B[2]). The study selection process was divided into three main phases, namely identification, screening, and eligibility phases, and extracted data included the author's family name, title of the paper, specific type of algorithm used, type of dataset used in the model, validation metrics, etc. The two authors independently performed the study selection and data extraction, with disagreements resolved by discussion. The extracted data was then synthesized using a narrative approach.

3. Results

The scoping review retrieved 1,765 articles from three databases and imported them into RefWorks Citation Manager. After removing 549 duplicates, the remaining 1,216 papers were imported into Rayyan. Of these, 83 duplicate records were identified and deleted following careful evaluation of titles and abstracts. The screening process involved an initial exclusion of 696 papers based on reading the title alone, followed by the reading titles and abstracts, resulting in a further 354 exclusions. After reviewing the full text of the remaining 83 records, 43 were excluded because their publication year was before 2019. This was done because a high number of publications in this field were observed in 2019, and the trend was anticipated to continue. The inclusion criteria remained; studies published from 2019 onwards. Ultimately, 40 studies were included in the scoping review. The study selection process flowchart can be accessed via the GitHub link.

[2]https://github.com/Salboy06/A-Scoping-review-LongTerm-T2DM-prediction-using-AI

Of the final 40 papers retained for our review, 34 (85%) were journal articles, and 6 (15%) were conference papers. All the papers were published, and the main objective of the AI models was prediction. In addition, we investigated the year of publication and the countries that had the highest number of publications during that year. Our analysis revealed that in 2019, eight countries had an equal number of publications. In 2020, nine publications were recorded, and China had the highest number of publications (3; 33%), followed by India (2; 22%). Furthermore, in 2021, out of seven publications, India had the highest number of publications (2; 29%). Similarly, in 2022, out of six published articles, India had the highest number of publications (2; 33%). As of the present year, 2023, only two publications have been recorded so far, with Iran and Oman being the only countries to publish (refer to online Appendix C[3]).

Furthermore, we explored the Artificial Intelligent model used for predicting the onset of Type-2 Diabetes Miletus and discovered that most studies used Machine Learning (ML) 23(58%), while 13(33%) studies used both Machine Learning and Deep Learning (DL) and only 4(10%) studies explored the potential of DL models in T2DM prediction (refer to online Appendix D[3]).

In addition, we examined 13 studies that incorporated both Machine Learning (ML) and Deep Learning (DL) techniques. The majority of these studies, 8(44%) utilized Ensemble Learning Methods. Support Vector Machines (SVM) and Random Forest (RF) were the second most used models, with each appearing in 3(17%) of the studies Neural Networks (NN) were the third most utilized model, appearing in 2(11%) of the studies. Upon examining the four studies that exclusively employed Deep Learning (DL) techniques, we observed that 50% of these studies utilized Deep Neural Networks (DNN) as the primary classifier. The remaining 50% used Recurrent Neural Networks (RNN) as the classifier.

The reviewed studies employed a total of 17 validation metrics, with accuracy being the most commonly used 31(77.5%), followed by recall 29(72%), area under the curve (AUC) 23(57.5%), and specificity 22(55.0%) (refer to online appendix E[2]). Noteworthy, recall (sensitivity) is crucial in T2DM prediction, as misclassification of positive cases can have severe consequences. Thus, we evaluated the recall metric separately. High recall scores are essential in identifying as many positive cases as possible for early intervention and management of T2DM and its associated complications. In evaluating the 40 retained studies based on recall, we found that ensemble learning models were the most frequently used (16 instances), with an average recall score of 85.08%. Random Forest (RF) and Support Vector Machine (SVM) were the second most evaluated models based on recall, appearing in 12 instances with an average score of 80.71% and 72.77%, respectively (refer to online Appendix F[3]). These findings underscore the significance of sensitivity in the development of accurate and effective T2DM predictive models.

4. Discussion

Our scoping review revealed that the majority of studies (58%) utilized Machine Learning (ML) techniques, while 33% employed both ML and Deep Learning (DL) techniques, and only 10% exclusively explored the potential of DL models. However, the use of DL techniques for predicting the onset of T2DM is still relatively underexplored, as mentioned earlier, with only 10% of the reviewed studies employing

[3]https://github.com/Salboy06/A-Scoping-review-LongTerm-T2DM-prediction-using-AI

DL models exclusively. Nonetheless, we discovered that out of the four studies that exclusively used DL models, an average accuracy of 81.57% is commendable, indicating that DL models need to be further investigated in this area of research.

In terms of the validation metrics used, our review found that accuracy was the most commonly used metric (77.5%), followed by recall (72%), area under the curve (57.5%), and specificity (55.0%). It is understandable that recall was the second most frequently used validation metric for T2DM predictive tools since these tools need to be evaluated in terms of their sensitivity to avoid misclassification of positive incidents. Additionally, it is crucial for predictive tools to have a high recall score, which can assist in early intervention and management of T2DM and its complications. We found that ensemble learning models had the highest average recall score (85.08%) among the 40 papers analyzed. Thus, we strongly believe that ML models that use ensemble learning as their classifiers, in addition to DL models, could lead to the development of a generalized predictive tool. Our review has some limitations, including the possibility of bias in the Pima dataset used in most studies, the complexity of DL models for physicians to interpret, the potential for missing vital information in studies published before 2019, and the potential for important research in non-English languages to be overlooked.

5. Conclusion

Artificial intelligence (AI) shows great potential in predicting type 2 diabetes mellitus (T2DM) and identifying high-risk individuals. Our scoping review highlights that Machine Learning (ML) is the most commonly used AI technique, with limited studies using Deep Learning (DL) models. Our findings emphasize the importance of accuracy and recall as performance metrics in T2DM prediction, stressing the need for high sensitivity in identifying positive cases. Further research, especially exploring the feasibility of DL models, could develop more effective T2DM prediction tools, benefiting both patients and healthcare professionals.

References

[1] 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 [Internet]. 2022 -5-31 [cited Apr 23, 2023];12(6):905. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9224915/

[2] Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract [Internet]. 2011 -12 [cited Feb 6, 2023];94(3):311-21

[3] Catalano PM, McIntyre HD, Cruickshank JK, McCance DR, Dyer AR, Metzger BE, Lowe LP, Trimble ER, Coustan DR, Hadden DR, Persson B, Hod M, Oats JJN. The hyperglycemia and adverse pregnancy outcome study: Associations of GDM and obesity with pregnancy outcomes. Diabetes Care [Internet]. 2012 -04 [cited Feb 6, 2023];35(4):780-6.

[4] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine. 2018 Sep 4;169(7):467–73.