

[Home](#) [Journal of Diabetes & Metabolic Disorders](#) [Article](#)

Explainable machine learning model incorporating social determinants of health to predict chronic kidney disease in type 2 diabetes patients

Research article Published: 09 May 2025

Volume 24, article number 115, (2025) [Cite this article](#)**Journal of Diabetes & Metabolic Disorders**[Aims and scope](#)[Submit manuscript](#)

Md. Mohaimenul Islam, Tahmina Nasrin Poly, Arinzechukwu Nkemdirim Okere & Yao-Chin Wang 

 **144** Accesses [Explore all metrics](#) →

Abstract

Background and objectives

Social determinants of health (SDOH) play a critical role in the onset and progression of chronic kidney disease (CKD). Despite the well-established role of SDOH, previous studies have not fully incorporated these factors in predicting CKD in Type 2 diabetes patients. To bridge this gap, this study aimed to develop and evaluate the machine learning (ML) models that incorporate SDOH to enhance CKD risk prediction in Type 2 diabetes patients.

Methods



Data were obtained from the 2023 Behavioral Risk Factor Surveillance System (BRFSS), a national survey that collects comprehensive health-related data from adults across the United States. Missing data were addressed using the K-nearest neighbor imputation method, and the Synthetic Minority Oversampling Technique (SMOTE) was applied to balance class distributions. Potential predictive features were selected using correlation coefficient analysis. The dataset was partitioned into training (80%) and testing (20%) subsets, with a 3-fold cross-validation strategy applied to the training data. Seven ML models were developed for CKD risk prediction, including logistic regression (LR), decision tree (DT), K-nearest neighbor (KNN), random forest (RF), eXtreme Gradient Boosting (XGBoost), and an artificial neural network (ANN). Model performance was evaluated using multiple metrics, including the area under the receiver operating characteristic curve (AUROC), precision, recall, F1 score, accuracy, and false positive rate.

Results

The study included 19,912 Type 2 diabetes patients (weighted sample size: 818,878), among whom 2,924 (weighted 13.92%) had CKD, and 16,988 (weighted 86.08%) did not. Over half of the CKD group (50.4%) were aged 65 or older. The proportion of female patients was higher in both groups, comprising 53.8% of the CKD group and 50.5% of the non-CKD group. Among the ML models evaluated, the RF model demonstrated the highest predictive performance for CKD, with an AUROC of 0.89 (95% CI: 0.88 – 0.90), followed by the DT model (0.84, 95% CI: 0.83 – 0.85) and XGBoost (0.83, 95% CI: 0.82 – 0.84). The RF model achieved an accuracy of 0.81 (95%CI: 0.81 – 0.81), a precision of 0.79 (95%CI: 0.79 – 0.79), a recall of 0.85 (95%CI: 0.85 – 0.85), and an F1 score of 0.82 (95%CI: 0.82 – 0.82). Additionally, the RF model exhibited strong calibration, reinforcing its reliability as a predictive tool for CKD risk in individuals with Type 2 diabetes.

Conclusion

The study findings underscore the potential of ML models, particularly the RF model, in accurately predicting CKD among individuals with Type 2 diabetes. This approach not only enhances the precision of CKD prediction but also highlights the importance of addressing social and environmental disparities in disease prevention and management. Leveraging ML models with SDOH can lead to earlier interventions, more personalized treatment plans, and improved health outcomes for vulnerable populations.

 This is a preview of subscription content, [log in via an institution](#)  to check access.

Access this article

[Log in via an institution](#)

Subscribe and save

✓ Springer+ Basic

\$34.99 /Month

Get 10 units per month

Download Article/Chapter or eBook

1 Unit = 1 Article or 1 Chapter

Cancel anytime

[Subscribe now](#) →

Buy Now

[Buy article PDF USD 39.95](#)

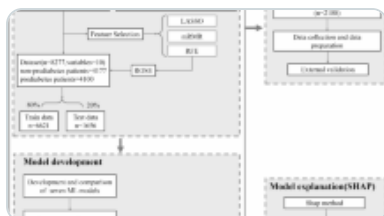
Price excludes VAT (USA)

Tax calculation will be finalised during checkout.

Instant access to the full article PDF.

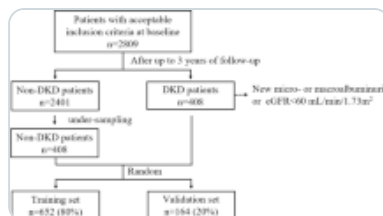
[Institutional subscriptions](#) →

Similar content being viewed by others



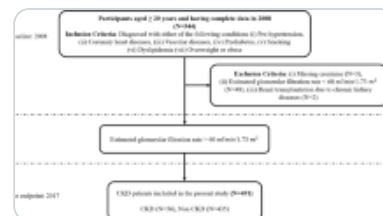
Interpretable machine learning method to predict the risk of pre-diabetes using a...

Article | Open access
26 March 2025



Prediction of 3-year risk of diabetic kidney disease using machine learning based on...

Article | Open access
26 March 2022



Investigation on explainable machine learning models to predict chronic kidney...

Article | Open access
14 February 2024

Explore related subjects

Discover the latest articles and news from researchers in related subjects, suggested using machine learning.

[Diabetic nephropathy](#)

[Disease Models](#)

[Linear Models and Regression](#)

[Machine Learning](#)

[Predictive medicine](#)

[Type 2 diabetes](#)

Data availability

The data are available upon reasonable request.

References

1. (NIDDK), N.I.o.D.a.D.a.K.D. Kidney Disease Statistics for the United States. 2023 [cited 2024; Available from: <https://www.niddk.nih.gov/health-information/health-statistics/kidney-disease>.

2. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Supplements*. 2022;12(1):7–11.

[Article](#) [Google Scholar](#)

3. Bello AK, et al. Complications of chronic kidney disease: current state, knowledge gaps, and strategy for action. *Kidney Int Supplements*. 2017;7(2):122–9.

[Article](#) [Google Scholar](#)

4. Loghman-Adham M. Medication noncompliance in patients with chronic disease: issues in dialysis and renal transplantation. *Am J Managed Care*. 2003;9(2):155–73.

[Google Scholar](#)

5. Kishi S, Kadoya H, Kashihara N. Treatment of chronic kidney disease in older populations. *Nat Rev Nephrol*. 2024;20(9):586–602.

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

6. Zafer A, Baloch S. Aging, chronic kidney disease, and psychological resilience. *Indus J Biosci Res*. 2023;1(2):63–88.

[Google Scholar](#)

7. Li X, Luo JK. Perceptions and attitudes of kidney supportive care among elderly patients with advanced chronic kidney disease and dialysis healthcare professionals in China: a qualitative study. *BMC Nephrol*. 2023;24(1):316.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

8. Anders H-J, et al. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat Rev Nephrol*. 2018;14(6):361–77.

[Article](#) [CAS](#) [PubMed](#) [Google Scholar](#)

9. Locatelli F, Vecchio LD, Pozzoni P. The importance of early detection of chronic kidney disease. *Nephrology Dialysis Transplantation*. 2002;17(suppl 11):8–7.

[Article](#) [Google Scholar](#)

10. Pereira BJ. Overcoming barriers to the early detection and treatment of chronic kidney disease and improving outcomes for end-stage renal disease. *Am J Managed Care*. 2002;8(4 SUPP):122–35.

[Google Scholar](#)

11. Wouters OJ, et al. Early chronic kidney disease: diagnosis, management and models of care. *Nat Rev Nephrol*. 2015;11(8):491–502.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

12. Major RW, et al. Cardiovascular disease risk factors in chronic kidney disease: A systematic review and meta-analysis. *PLoS ONE*. 2018;13(3):e0192895.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

13. Suckling R, Gallagher H. Chronic kidney disease, diabetes mellitus and cardiovascular disease: risks and commonalities. *J Ren Care*. 2012;38:4–11.

[Article](#) [PubMed](#) [Google Scholar](#)

14. Couser WG, et al. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney Int*. 2011;80(12):1258–70.

[Article](#) [PubMed](#) [Google Scholar](#)

15. de Oliveira Salerno PRV, et al. Social and Environmental Determinants of Health and Cardio-Kidney-Metabolic Syndrome-Related Mortality. JAMA Network Open. 2024;7(9):e2435783–e2435783.

[Article](#) [Google Scholar](#)

16. Ozieh MN, et al. The cumulative impact of social determinants of health factors on mortality in adults with diabetes and chronic kidney disease. BMC Nephrology. 2021;22:1–10.

[Article](#) [Google Scholar](#)

17. Vo A, et al. The association between social determinants of health and population health outcomes: Ecological analysis. JMIR Public Health Surveill. 2023;9(1):e44070.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

18. Foundation NK. Social Determinants of Health and Chronic Kidney Disease. 2023 [cited 2024 October 2]; Available from: <https://www.kidney.org/kidney-topics/social-determinants-health-and-chronic-kidney-disease>.

19. Nicholas SB, Kalantar-Zadeh K, Norris KC. Socioeconomic disparities in chronic kidney disease. Adv Chronic Kidney Dis. 2015;22(1):6–15.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

20. Arenas MD, et al. Social determinants of health influencing the choice of dialysis modality in advanced chronic kidney disease: Need of an interdisciplinary approach. Nefrología (English Edition). 2024;44(4):560–7.

[Article](#) [Google Scholar](#)

21. Segar MW, et al. Machine learning-based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with

heart failure. JAMA Cardiol. 2022;7(8):844–54.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

22. Abbott EE, et al. Leveraging Artificial Intelligence and Data Science for Integration of Social Determinants of Health in Emergency Medicine: Scoping Review. JMIR Med Inform. 2024;12:e57124.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

23. Nelson DE, et al. Reliability and validity of measures from the Behavioral Risk Factor Surveillance System (BRFSS). Sozial-und Praventivmedizin. 2001;46:S3–42.

[PubMed](#) [Google Scholar](#)

24. Brown DW, et al. Diabetes mellitus and health-related quality of life among older adults: Findings from the behavioral risk factor surveillance system (BRFSS). Diabetes Res Clin Pract. 2004;65(2):105–15.

[Article](#) [PubMed](#) [Google Scholar](#)

25. Xie Z, et al. building risk prediction models for type 2 diabetes using machine learning techniques. Preventing Chronic Dis. 2019;16:E130.

[Google Scholar](#)

26. Okeke F, Ugwuoke UT. Assessing the relationship between sleep duration and the prevalence of chronic kidney disease among veterans in the united states: a 2022 behavior risk factor surveillance system (brfss) cross-sectional study. Cureus. 2024;16(9):e68538.

[PubMed](#) [PubMed Central](#) [Google Scholar](#)

27. Peng S, et al. Detecting lifestyle risk factors for chronic kidney disease with comorbidities: association rule mining analysis of web-based survey data. J Med

Internet Res. 2019;21(12):e14204.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

28. Pujiyanto U, Wibawa AP, Akbar MI. K-nearest neighbor (k-NN) based missing data imputation. In 2019 5th International Conference on Science in Information Technology (ICSITech) 2019. pp. 83–88. IEEE.
<https://doi.org/10.1109/ICSITech46713.2019.8987530>.

29. Zhang S. Nearest neighbor selection for iteratively kNN imputation. J Syst Softw. 2012;85(11):2541–52.

[Article](#) [Google Scholar](#)

30. Kim M, et al. Learning-based adaptive imputation method with kNN algorithm for missing power data. Energies. 2017;10(10):1668.

[Article](#) [Google Scholar](#)

31. Agresti A. Categorical data analysis. John Wiley & Sons; 2013.

32. Kumar A, Abirami S. Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method. Inf Sci. 2018;460:23–41.

[Google Scholar](#)

33. Elreedy D, Atiya AF. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Inf Sci. 2019;505:32–64.

[Article](#) [Google Scholar](#)

34. Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Mach Learn.

2024;113(7):4903–23.

[Article](#) [Google Scholar](#)

35. LaValley MP. Logistic regression. *Circulation*. 2008;117(18):2395–9.

[Article](#) [PubMed](#) [Google Scholar](#)

36. Ryo M, Rillig MC. Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere*. 2017;8(11):e01976.

[Article](#) [Google Scholar](#)

37. Biau G. Analysis of a random forests model. *J Machine Learning Res*. 2012;13:1063–95.

[Google Scholar](#)

38. Rigatti SJ. Random forest. *J Insur Med*. 2017;47(1):31–9.

[Article](#) [PubMed](#) [Google Scholar](#)

39. Suthaharan S. Support vector machine. In: *Machine learning models and algorithms for big data classification*. Integrated Series in Information Systems. 2016, vol 36. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7641-3_9.

40. Hearst MA, et al. Support vector machines. *IEEE Intell Syst Appl*. 1998;13(4):18–28.

[Article](#) [Google Scholar](#)

41. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016. pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.

42. Bella A, et al. Calibration of machine learning models. In Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. 2010. pp. 128–146. IGI Global. <https://doi.org/10.4018/978-1-60566-766-9.ch006>.

43. Nattino G, Pennell ML, Lemeshow S. Assessing the goodness of fit of logistic regression models in large samples: a modification of the Hosmer–Lemeshow test. Biometrics. 2020;76(2):549–60.

[Article](#) [PubMed](#) [Google Scholar](#)

44. Ekanayake I, Meddage D, Rathnayake U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). Case Stud Construction Mater. 2022;16:e01059.

[Article](#) [Google Scholar](#)

45. Aldrees A, et al. Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches. J Water Process Eng. 2024;58:104789.

[Article](#) [Google Scholar](#)

46. Li Z. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. Comput Environ Urban Syst. 2022;96:101845.

[Article](#) [Google Scholar](#)

47. Parisineni SRA, Pal M. Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic shap explanations. Int J Data Sci Analytics. 2024;18(4):457–66.

[Article](#) [Google Scholar](#)

48. Hamrahian SM, Falkner B. Hypertension in chronic kidney disease. Hypertension: from basic research to clinical practice. 2017:307–25.

https://doi.org/10.1007/5584_2016_84.

49. Nawaz S, et al. Obesity and chronic kidney disease: a current review. Obes Sci Pract. 2023;9(2):61–74.

[Article](#) [PubMed](#) [Google Scholar](#)

50. Cuddapah GV, et al. Complications in diabetes mellitus: social determinants and trends. Cureus. 2022;14(4). <https://doi.org/10.7759/cureus.24415>.

51. Ulasi II, Awobusuyi O, Nayak S, Ramachandran R, Musso CG, Depine SA, Aroca-Martinez G, Solarin AU, Onuigbo M, Luyckx VA, Ijoma CK. Chronic kidney disease burden in low-resource settings: regional perspectives. In Seminars in Nephrology. 2022. Vol. 42, No. 5, p. 151336. Saunders.

<https://doi.org/10.1016/j.semnephrol.2023.151336>.

52. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health. 2013;38:976–93.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

53. Wolfe MK, McDonald NC, Holmes GM. Transportation barriers to health care in the United States: findings from the national health interview survey, 1997–2017. Am J Public Health. 2020;110(6):815–22.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

54. Dean EB, French MT, Mortensen K. Food insecurity, health care utilization, and health care expenditures. Health Serv Res. 2020;55:883–93.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

55. Levi R, Bleich SN, Seligman HK. Food insecurity and diabetes: overview of intersections and potential dual solutions. *Diabetes Care*. 2023;46(9):1599–608.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

56. Gucciardi E, et al. The intersection between food insecurity and diabetes: a review. *Curr Nutr Rep*. 2014;3:324–32.

[Article](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

57. Prasad GR. Metabolic syndrome and chronic kidney disease: Current status and future directions. *World J Nephrol*. 2014;3(4):210.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

58. Wang N, Zhang C. Oxidative Stress: A Culprit in the Progression of Diabetic Kidney Disease. *Antioxidants*. 2024;13(4):455.

[Article](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

59. Allen A, et al. Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2022;10(1):e002560.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

60. Zou Y, et al. Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease. *Ren Fail*. 2022;44(1):562–70.

[Article](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

61. Hosseini Sarkhosh SM, Hemmatabadi M, Esteghamati A. Development and validation of a risk score for diabetic kidney disease prediction in type 2 diabetes

patients: a machine learning approach. J Endocrinol Investigation. 2023;46(2):415–23.

[Article](#) [CAS](#) [Google Scholar](#)

62. Mihai S, et al. Inflammation-related mechanisms in chronic kidney disease prediction, progression, and outcome. J Immunol Res. 2018;2018(1):2180373.

[PubMed](#) [PubMed Central](#) [Google Scholar](#)

63. Shen Y, et al. The Association between Symptoms of Depression and Anxiety, Quality of Life, and Diabetic Kidney Disease among Chinese Adults: A Cross-Sectional Study. Int J Environ Res Public Health. 2022;20(1):475.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

64. Cogley C, et al. Chronic kidney disease and severe mental illness: addressing disparities in access to health care and health outcomes. Clin J Am Soc Nephrol. 2022;17(9):1413–7.

[Article](#) [CAS](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

65. Pourriyahi H, et al. Loneliness: an immunometabolic syndrome. Int J Environ Res Public Health. 2021;18(22):12162.

[Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

66. Ahmed M, Cerda I, Maloof M. Breaking the vicious cycle: The interplay between loneliness, metabolic illness, and mental health. Front Psych. 2023;14:1134865.

[Article](#) [Google Scholar](#)

67. Tang R, et al. Loneliness, social isolation and incident chronic kidney disease among patients with diabetes. General Psychiatry. 2024;37(3):e101298.

Funding

None.

Author information

Authors and Affiliations

College of Pharmacy, The University of Iowa, Iowa, IA, USA

Md. Mohaimenul Islam & Arinzechukwu Nkemdirim Okere

Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

Tahmina Nasrin Poly

Department of Emergency, Min-Sheng General Hospital, Taoyuan, Taiwan

Yao-Chin Wang

Graduate Institute of Injury Prevention and Control, College of Public Health, Taipei Medical University, Taipei, Taiwan

Yao-Chin Wang

Corresponding author

Correspondence to [Yao-Chin Wang](#).

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Conflict of interest

The authors declared that they have no conflict of interest.

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Below is the link to the electronic supplementary material.

[Supplementary file1 \(DOCX 247 KB\)](#)

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

[Reprints and permissions](#)

About this article

Cite this article

Islam, M.M., Poly, T.N., Okere, A.N. *et al.* Explainable machine learning model incorporating social determinants of health to predict chronic kidney disease in type 2 diabetes patients. *J Diabetes Metab Disord* **24**, 115 (2025). <https://doi.org/10.1007/s40200-025-01621-9>

Received

23 January 2025

Accepted

13 April 2025

Published

09 May 2025

DOI

<https://doi.org/10.1007/s40200-025-01621-9>

Keywords

[Machine learning](#)

[XGBoost](#)

[Random forest](#)

[Social determinants of health](#)

[Chronic kidney disease](#)

[Type 2 diabetes](#)