



Explainable Artificial Intelligence for Early Detection of Chronic Diseases in Saudi Healthcare: A Systematic Review

Awadalla Abdelwahid^{1*}, Aalaa Almuazel², Malaz Taglwy³, Gawahir Murad Abdelrahman Alzein⁴, Asim Abdelwahid Elnoor Ali⁵, Abdelbagi Ali Elbashir Mohammed⁶, Reem A Younis⁷, Ashruf Ishag Mohammed Elhassan⁸, Dina Mahmoud Fathi⁹, Ahazeej Gurashi¹⁰

¹Department of Obstetrics and Gynecology, Faculty of Medicine, Al Neelain University, Khartoum, Sudan. E-mail: awad336@yahoo.com

²Obstetrics and Gynecology, Specialized Medical Company Hospital [SMC1], Riyadh, Saudi Arabia. E-mail: aalaaelmuazzal@gmail.com

³Obstetrics and Gynecology, Wellness Experts Medical Center, Riyadh, Saudi Arabia. E-mail: malazbabekir990@gmail.com

⁴Obstetrics and Gynecology Consultant, Maternity and Children Hospital Alkharj, Alkharj, Saudi Arabia. E-mail: totaelzein2017@gmail.com

⁵Preventive Medicine and Public Health Consultant, Khamis Mushait Health Sector, Aseer Health Cluster, Saudi Arabia.

E-mail: asimalnoor@gmail.com

⁶Medical Specialist, Department of Internal Medicine, Aseer Cluster, Alnamas General Hospital, Saudi Arabia.

E-mail: midoahmed053@gmail.com

⁷Department of Internal Medicine, College of Medicine, Qassim University, Buraidah, Saudi Arabia. E-mail: R.younis@qu.edu.sa

⁸Community Medicine Specialist, Infection Prevention Director, Namerah General Hospital, Al Qunfudhah, Saudi Arabia.

E-mail: ashruf.ishag@gmail.com

⁹Obstetrics and Gynecology Specialist, Tabuk Health Cluster, Tabuk, KSA. E-mail: dinamahmoudfm@hotmail.com

¹⁰Obstetrics and Gynecology Senior Registrar, Haya National Hospital, Saudi Arabia. E-mail: ahazeej.g@hotmail.com

Abstract

Background: Explainable artificial intelligence (XAI) may support earlier detection of chronic disease, but Saudi-specific evidence has not been systematically characterized. Objective: To synthesize Saudi patient-level studies of AI/XAI for chronic-disease prediction and assess clinical readiness.

Methods: PubMed and supplementary publisher-site, reference, and citation searches were completed on 26 August 2026. Eligible primary studies used Saudi clinical data for prediction, classification, or risk stratification. Findings were narratively synthesized and appraised with an adapted PROBAST framework.

Results: Fifty-eight records were identified, 52 screened, nine assessed in full, and four included. Studies addressed chronic kidney disease, diabetic kidney disease, hypertension, and right-sided cardiac dysfunction. Reported discrimination was promising, with explicit SHAP-based explanations in three studies; however, all lacked independent external validation and had high overall risk of bias.

Conclusion: Saudi XAI research is promising but clinically preliminary. Multicentre external validation, calibration, fairness evaluation, and prospective impact studies are required before routine deployment.

Keywords: explainable artificial intelligence; chronic disease; early detection; machine learning; Saudi Arabia; digital health

Introduction

Chronic diseases are a major and growing challenge for the Saudi healthcare system. Diabetes mellitus, hypertension, cardiovascular disease (CVD), chronic kidney disease (CKD), obesity, and multimorbidity contribute substantially to morbidity, premature mortality, disability, reduced quality of life, and healthcare expenditure [1–7]. Many of these conditions develop gradually and remain clinically silent until complications are established. Earlier identification of high-risk individuals can support prevention, targeted screening, timely referral, treatment optimization, and patient education before irreversible organ damage occurs.

Artificial intelligence (AI) and machine-learning methods can identify complex relationships within electronic health records, laboratory results, imaging, physiological measurements, medication histories, and longitudinal clinical data [8,9]. This capability is particularly relevant to Saudi Arabia, where Vision 2030 has accelerated digital-health transformation, service integration, virtual care, and data-driven innovation [10–14]. Nevertheless, high predictive accuracy alone is insufficient for safe clinical adoption. Healthcare decisions are high-stakes, patient-specific, and ethically sensitive; clinicians therefore need to understand why a model generated a prediction, whether its reasoning is clinically plausible, and how the output should influence care [15–17].

Explainable artificial intelligence (XAI) encompasses methods that make model behaviour more transparent. These include inherently interpretable models, global feature-importance analyses, patient-level attribution

methods such as SHAP, surrogate models, rule extraction, counterfactual explanations, and clinician-facing summaries [15–17]. In chronic-disease care, explanations may reveal whether estimated risk is driven by glycaemic status, renal function, blood pressure, obesity, age, comorbidity, medication use, or other clinically meaningful variables. Appropriate explanations can facilitate clinical review, expose data artefacts or bias, support patient communication, and improve accountability.

The Saudi context requires particular attention to local validity. Healthcare is delivered across primary-care centres, specialist hospitals, regional networks, and virtual-care services, with variation in patient characteristics, disease prevalence, documentation quality, referral pathways, and available resources. Models developed elsewhere may not transport reliably to Saudi populations, and locally developed single-centre models may not generalize across regions or facility types. Explainability may help identify reliance on site-specific coding, test-ordering behaviour, missingness patterns, or referral bias; however, explanations cannot compensate for poor model design or inadequate validation.

This systematic review aimed to synthesize Saudi-specific and directly relevant evidence on AI, interpretable machine learning, and XAI for early detection, screening, prediction, and risk stratification of chronic diseases. The objectives were to identify the principal disease targets and model families, describe explanation methods, assess reported clinical relevance and methodological quality, and identify evidence gaps affecting validation, calibration, fairness, governance, and implementation in Saudi healthcare settings.

Materials and Methods

Review design and reporting standard

This systematic review followed the PRISMA 2020 statement. The review was not prospectively registered because the search and protocol were reconstructed during manuscript revision; this deviation is reported transparently. The completed review was subsequently retrospectively registered on the Open Science Framework (OSF; <https://doi.org/10.17605/OSF.IO/UAKQE>). A narrative synthesis was prespecified because the studies differed in disease target, data type, outcome, algorithm, validation, and performance measure.

Eligibility criteria

Peer-reviewed primary studies were eligible when they used Saudi patient-level or Saudi health-system data; evaluated artificial intelligence or machine learning for early detection, screening, classification, prediction, or risk stratification of a chronic disease; and reported model-performance or explanation findings. Studies using an explicit XAI method were prioritized. A multimodal deep-learning study was retained as a clinically interpretable Saudi comparator. Reviews, editorials, protocols, conference abstracts without sufficient data, non-healthcare studies, acute procedural outcomes, and studies based entirely on non-Saudi public datasets were excluded.

Information sources and search strategy

The search was conducted on 26 August 2026. PubMed was searched from inception without language restriction. Supplementary searching comprised publisher websites and backward and forward citation searching of eligible reports and relevant reviews. Scopus and Web of Science were not searched directly; they were used only when publisher pages exposed citation links. The exact PubMed query and search yield are reported in Table 1.

Table 1. Search strategy and information sources

Source	Search date	Strategy/method	Yield
PubMed/MEDLINE	26 Aug 2026	(Saudi[Affiliation] OR "Saudi Arabia"[Title/Abstract]) AND ("explainable artificial intelligence"[Title/Abstract] OR XAI[Title/Abstract] OR SHAP[Title/Abstract] OR "interpretable machine learning"[Title/Abstract]) AND (diabetes[Title/Abstract] OR "chronic kidney disease"[Title/Abstract] OR cardiovascular[Title/Abstract] OR hypertension[Title/Abstract] OR obesity[Title/Abstract] OR "metabolic syndrome"[Title/Abstract])	46
Supplementary searching	26 Aug 2026	Publisher-site searching plus backward and forward citation searching of eligible reports and relevant reviews	12

Study selection and data extraction

Search results were collated, duplicates were removed, and titles and abstracts were screened against the predefined criteria. Potentially relevant reports underwent full-text assessment, with exclusion reasons documented. Data were extracted from published reports into a structured form covering design, setting, sample, disease, predictors, outcome, algorithm, validation, explanation method, performance, calibration, subgroup analysis, deployment, and limitations. No unpublished data were imputed.

Risk-of-bias assessment

Prediction studies were appraised using an adapted PROBAST framework covering participants, predictors, outcome, and analysis. Overall risk was judged high when concerns in the analysis domain—including limited sample size, model-development optimism, incomplete calibration, or absence of independent external validation—could materially affect reported performance. Judgements are presented in Table 3.

Synthesis

Studies were grouped by disease domain, data source, model family, explanation approach, and translational readiness. Meta-analysis was not performed because outcomes and performance measures were clinically and methodologically heterogeneous. Results were synthesized narratively, emphasizing discrimination, calibration, external validation, subgroup assessment, explainability, and deployment evidence.

Results

Study selection

The search identified 58 records: 46 through PubMed and 12 through supplementary publisher-site, reference, and citation searching. Six duplicates were removed, leaving 52 records for title and abstract screening. Forty-three records were excluded. Nine full-text reports were assessed; five were excluded because they used non-Saudi patient data ($n = 3$) or addressed an ineligible outcome or clinical context ($n = 2$). Four studies were included in the narrative synthesis (Figure 1).

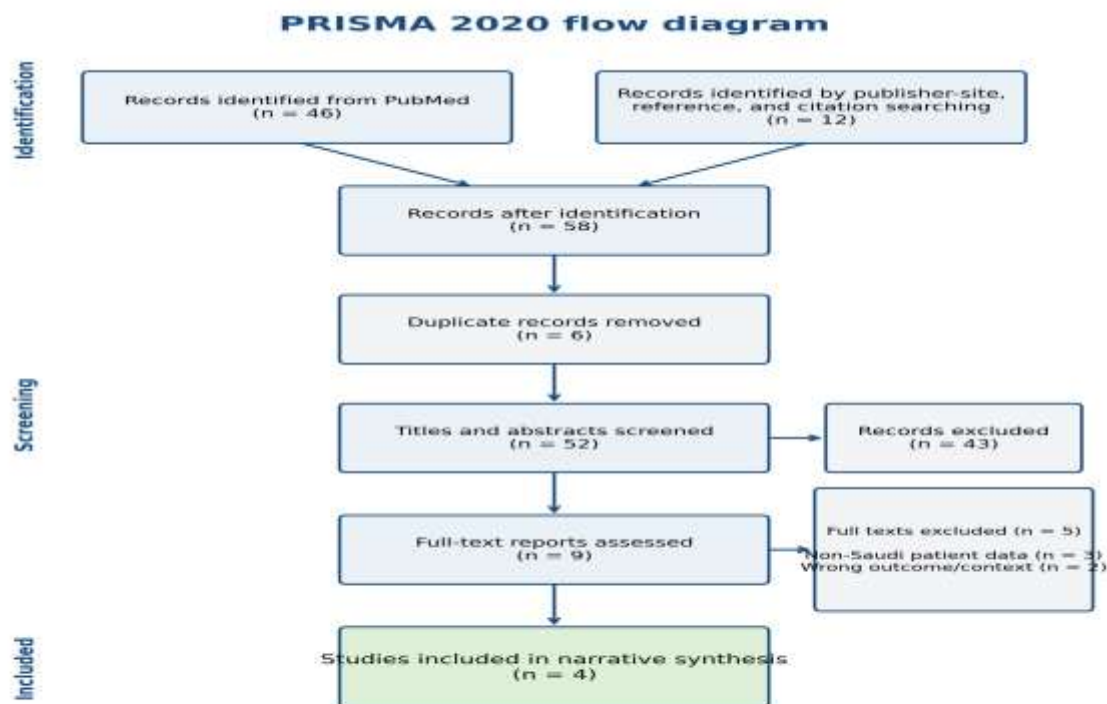


Figure 1. PRISMA 2020 flow diagram for study identification, screening, eligibility, and inclusion.

Study characteristics

The four included studies were published between 2024 and 2026 and used Saudi clinical data. Two addressed chronic kidney disease or diabetic kidney disease, one predicted hypertension from retinal images and

cardiometabolic variables, and one identified severe right-sided cardiac dysfunction in people with diabetes. Sample sizes ranged from approximately 370 participants to 18,742 adults; one retinal-imaging study included 5,016 images from 1,243 individuals. Study characteristics and principal results are summarized in Table 2.

Table 2. Characteristics and main findings of included studies

Study	Setting/sample	Target	Model/validation	Explainability	Principal findings and limitations
Alhaifi et al., 2026	Real-world Saudi laboratory data; n=746	Binary CKD and CKD stage	XGBoost/RFE; MLP/SelectKBest; internal validation	Risk score and SHAP	Binary accuracy/F1 97%; stage accuracy 85%, F1 86%. Single local dataset; no independent external validation.
Tourkmani et al., 2026	PSMMC registry, Riyadh; 18,742 adults with T2DM, 2019–2024	3-year incident/progressive DKD	Stacked LightGBM+CoxBoost; temporal internal validation	SHAP	AUROC 0.852 (95% CI 0.847–0.857); calibration slope 0.98; decision-curve net benefit. No external-centre or prospective impact validation.
Baharoon et al., 2024	Saudi National Guard system; 5,016 fundus images from 1,243 persons	Hypertension	RETFound image pathway plus age/sex fusion; internal validation	Modality-level interpretability; no formal post-hoc XAI	F1 0.771 (95% CI 0.747–0.796); diabetic vs non-diabetic F1 0.796 vs 0.466. No external validation; subgroup disparity.
Hasan et al., 2026	Retrospective Saudi diabetic cohort; approximately n=370	Severe right-sided cardiac dysfunction	Seven classifiers; random forest selected using true nested cross-validation	SHAP	AUROC 0.8602; AUPRC 0.8343; accuracy 0.8075; Brier 0.1281. Small retrospective cohort; no external validation.

Model performance and explainability

Alhaifi et al. reported 97% accuracy and F1 score for binary CKD classification and 85% accuracy with an F1 score of 86% for stage classification, using feature selection, an interpretable risk score, and SHAP. Tourkmani et al. developed the PSMMC NephraRisk ensemble in 18,742 adults with type 2 diabetes; the three-year diabetic-kidney-disease model achieved an AUROC of 0.852 (95% CI, 0.847–0.857), a calibration slope of 0.98, and an intercept of –0.012. SHAP, subgroup assessment, decision-curve analysis, and a web implementation were reported.

Baharoon et al. combined retinal images with age and sex in HyMNet. The multimodal model achieved an F1 score of 0.771 (95% CI, 0.747–0.796), compared with 0.745 for the fundus-only model; performance differed substantially between diabetic and non-diabetic subgroups (F1, 0.796 vs 0.466). The architecture enabled modality-level interpretation but did not report a formal post-hoc XAI method. Hasan et al. evaluated seven lightweight classifiers in approximately 370 Saudi adults with diabetes. A random-forest model selected through nested cross-validation achieved an AUROC of 0.8602, AUPRC of 0.8343, accuracy of 0.8075, sensitivity of 0.7733, specificity of 0.8239, and Brier score of 0.1281. SHAP identified left-atrial dimensions, diastolic-dysfunction grade, and diuretic use as influential predictors.

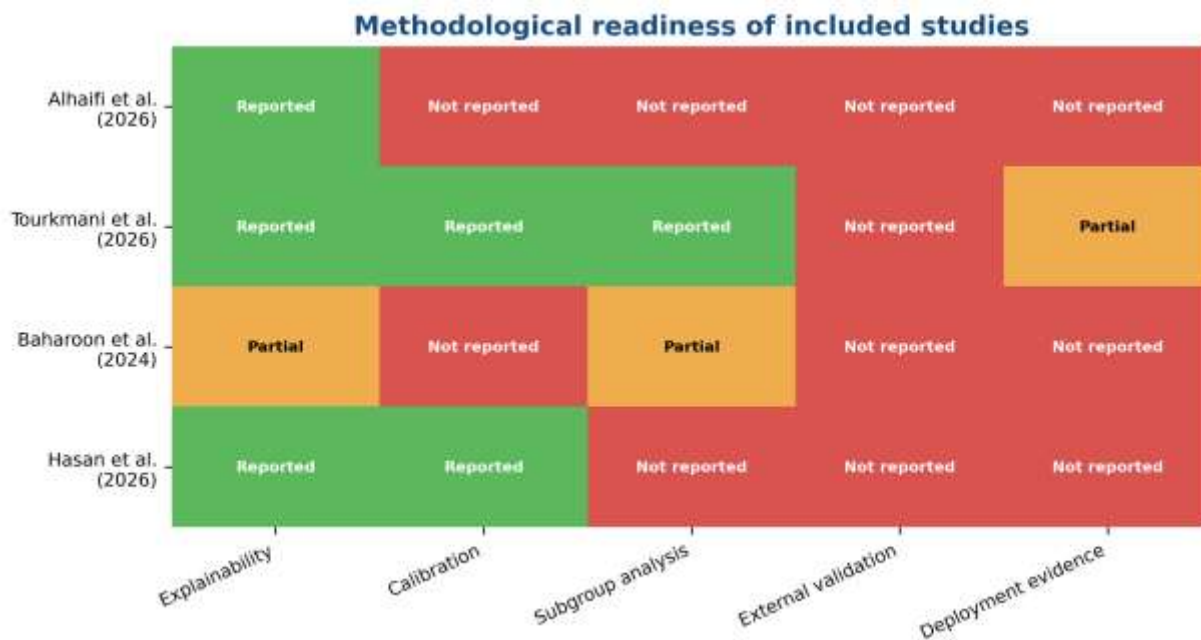
Risk of bias and clinical readiness

All studies were judged at high overall risk of bias, primarily because no model had independent external validation and all evidence came from retrospective or locally derived cohorts. Calibration was convincingly reported in two studies, subgroup assessment in two, and explicit SHAP-based explanation in three. Only the NephraRisk study described an operational web implementation; even there, prospective clinical-impact and external-centre evidence were absent. The domain judgements are shown in Table 3 and the readiness summary in Figure 2.

Table 3. Adapted PROBAST risk-of-bias assessment

Study	Participants	Predictors	Outcome	Analysis	Overall
Alhaifi et al., 2026	Some concerns	Low	Low	High	High
Tourkmani et al., 2026	Some concerns	Some concerns	Low	Some concerns	High
Baharoon et al., 2024	Some concerns	Low	Low	High	High
Hasan et al., 2026	Some concerns	Low	Low	High	High

Note. High overall risk principally reflects retrospective/local development and absence of independent external validation. These judgements are based on information available in published reports.


Figure 2. Methodological-readiness summary of included studies.

Discussion

This systematic review found only four eligible Saudi patient-level studies, indicating that the national evidence base for XAI-supported chronic-disease detection remains small. The strongest work combined clinically meaningful predictors with calibration, nested or temporal validation, SHAP-based explanations, and decision-analytic evaluation. Reported discrimination was encouraging across kidney, hypertension, and cardiac applications, but it should not be interpreted as evidence of general clinical effectiveness.

The most important common limitation was the absence of independent external validation. Locally trained models may exploit site-specific coding, referral, imaging, testing, or missingness patterns and may not transport across Saudi regions, primary-care networks, military hospitals, or tertiary centres. Explainability can expose implausible dependencies, but a persuasive explanation cannot correct biased sampling, data leakage, or miscalibration.

Clinical usefulness also requires a defined action pathway. A high-risk prediction should lead to an appropriate response such as confirmatory testing, medication review, preventive counselling, closer monitoring, or specialist referral. The NephraRisk model came closest to this translational standard by reporting calibration, decision curves, subgroup results, SHAP, and a web implementation. Nevertheless, prospective evaluation is needed to determine whether it improves decisions, timeliness, equity, safety, or patient outcomes.

The evidence also illustrates the distinction between interpretability and formal XAI. HyMNet's multimodal design permits assessment of the incremental value of retinal and demographic inputs, but it did not provide patient-level attribution comparable with SHAP. Future studies should specify whether explanations are global or local, assess their stability and fidelity, and test whether clinicians and patients understand them.

Saudi multicentre studies should prospectively validate models across sex, age, nationality, region, comorbidity burden, and facility type. Reporting should include confidence intervals, calibration plots and slopes, clinically relevant thresholds, decision-curve analysis, subgroup error rates, missing-data handling, and model-updating procedures. Governance should address privacy, cybersecurity, monitoring, drift, accountability, and withdrawal criteria.

Limitations of this review

This reconstruction used a direct PubMed search supplemented by publisher-site, reference, and citation searching; it did not include direct searches of Scopus, Web of Science, IEEE Xplore, Embase, or the Saudi Digital Library. Relevant studies may therefore have been missed. Screening and extraction were reconstructed during manuscript revision rather than performed under a prospectively registered protocol; the completed review was subsequently retrospectively registered on OSF (<https://doi.org/10.17605/OSF.IO/UAKQE>). The small number of heterogeneous studies precluded meta-analysis, and published reports did not always provide complete information for risk-of-bias appraisal. These limitations should be considered when interpreting the review.

Potential implications for practice and policy

XAI tools should be introduced only when predictions are linked to clear clinical pathways and supported by local validation, calibration, fairness assessment, privacy review, usability testing, and post-deployment monitoring. Clinicians should be trained to interpret explanations critically and to communicate AI-supported risk estimates without presenting them as deterministic conclusions. National and institutional policies should address data quality, interoperability, auditability, cybersecurity, accountability, and patient communication. Multicentre Saudi studies should test whether explanations improve clinical decisions, patient understanding, adherence, workflow efficiency, and health outcomes.

Conclusion

Saudi patient-level research on XAI for chronic-disease detection is promising but clinically preliminary. Four eligible studies reported encouraging performance in kidney disease, hypertension, and cardiac dysfunction, and three used explicit SHAP-based explanations. No study provided independent external validation, and all were judged at high overall risk of bias. Routine adoption should await multicentre validation, robust calibration, subgroup fairness assessment, prospective impact evaluation, and accountable governance.

Recommendations for future research

Future research should evaluate prospective clinical impact, implementation cost, workforce requirements, alert burden, clinic flow, and potential overuse. Studies should compare explanation formats—including ranked predictors, concise narratives, visual displays, rules, and counterfactual recommendations—with clinicians, informatics professionals, and patients. Patient-facing explanations should be culturally and linguistically appropriate and connected to feasible actions. Implementation studies should include feedback mechanisms for incorrect explanations, missing variables, confusing alerts, and workflow problems, enabling recalibration and interface improvement. Success should be defined by safer and earlier detection, appropriate referral, improved risk communication, equitable performance, and measurable improvement in preventive-care processes or outcomes.

Ethical considerations

This review used published and publicly available information and involved no direct recruitment, intervention, or access to identifiable patient data; therefore, institutional ethics approval and informed consent were not required. Future clinical use of XAI should comply with applicable ethics, privacy, cybersecurity, and data-governance requirements and should ensure transparency, accountability, protection against algorithmic bias, human oversight, and appropriate patient communication.

Conflict of interest

The authors declare no conflict of interest.

Funding

No specific funding was reported for this manuscript.

Data availability

All data supporting this review are derived from the published sources cited in the manuscript. The completed search strategies, screening decisions, data-extraction table, and risk-of-bias assessments should be provided as supplementary materials at submission.

References

1. Al-Rayes S, AlOfi R, Alanezi F, et al.: Public awareness and use of health tools provided by the portal of the Ministry of Health of Saudi Arabia. *Inform Med Unlocked*. 2023, 36:101151. <https://doi.org/10.1016/j.imu.2022.101151>
2. World Health Organization: Independent evaluation of the UN Inter-Agency Task Force on the Prevention and Control of Non-communicable Diseases: report. World Health Organization, Geneva; 2026. WHO/DGO/EVL/2025.52 <https://www.who.int/publications/i/item/WHO-DGO-EVL-2025.52>
3. Alghnam S, Bosaeed M, Aljouie A, Alshahrani SM, Alshenqeety O, Atun R, Alqahtani S: Estimating the prevalence of select non-communicable diseases in Saudi Arabia using a population-based sample: econometric analysis with natural language processing. *Ann Saudi Med*. 2024, 44:329-338. <https://doi.org/10.5144/0256-4947.2024.329>
4. Alenzi EO, Fatima W, Amara A, et al.: A systematic review of chronic diseases and their prevalence among the population of Northern Borders Province (NBP) in Saudi Arabia. *J Multidiscip Healthc*. 2023, 16:1047-1056. <https://doi.org/10.2147/JMDH.S401001>
5. Alqahtani B, Elnaggar RK, Alshehri MM, Khunti K, Alenazi A: National and regional prevalence rates of diabetes in Saudi Arabia: analysis of national survey data. *Int J Diabetes Dev Ctries*. 2023, 43:392-397. <https://doi.org/10.1007/s13410-022-01092-1>
6. Al-Rasheedi M, Alhazmi Y, AlDaiji LA, et al.: Status of diabetes mellitus in different regions of KSA and update on its management. *Front Clin Diabetes Healthc*. 2024, 5:1482090. <https://doi.org/10.3389/fcdhc.2024.1482090>
7. Karim SI, Sattar K, Ahmad T, Memon MK, Alqahtani AS, Alsubiheen AM, Alrashed FA: Cardiovascular and renal comorbidities among Saudi patients with type 2 diabetes: a cross-sectional observational study. *PLoS One*. 2025, 20. <https://doi.org/10.1371/journal.pone.0324233>
8. Du Y, Yang P, Liu Y, Deng C, Li X: Artificial intelligence in chronic disease self-management: current applications and future directions. *Front Public Health*. 2025, 13:1689911. <https://doi.org/10.3389/fpubh.2025.1689911>
9. Alasiri AA, Mohammed V: Healthcare transformation in Saudi Arabia: an overview since the launch of Vision 2030. *Health Serv Insights*. 2022, 15:11786329221121214. <https://doi.org/10.1177/11786329221121214>
10. Ministry of Health, Saudi Arabia: E-health initiative in digital transformation (Vision 2030). Ministry of Health, Kingdom of Saudi Arabia. Accessed: August 19, 2026. <https://www.moh.gov.sa/Ministry/nehs/Pages/vision2030.aspx>
11. Alghareeb E, Aljehani N: AI in health care service quality: systematic review. *JMIR AI*. 2025, 4. <https://doi.org/10.2196/69209>
12. Alqurashi MA, Alshagrawi S: Assessing the impact of artificial intelligence applications on diagnostic accuracy in Saudi Arabian healthcare: a systematic review. *Open Public Health J*. 2025, 18. <https://doi.org/10.2174/0118749445369173250203072754>
13. Bayer M, Eisawi A: Exploring the landscape of artificial intelligence in Saudi Arabia's healthcare sector: current trends and challenges. *Cureus*. 2025, 17. <https://doi.org/10.7759/cureus.84163>
14. Lundberg SM, Lee SI: A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017, 30:4765-4774. <https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
15. Rudin C: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019, 1:206-215. <https://doi.org/10.1038/s42256-019-0048-x>
16. World Health Organization: Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization, Geneva; 2021. <https://www.who.int/publications/i/item/9789240029200>
17. Bindhim NF, Senitan M, Almutairi MN, Alhadlaq LS, Alnajem SA, Alfaifi MA, Althumiri NA: Demographic, health, and behaviors profile of Saudi Arabia's aging population 2022-2023. *Front Aging*. 2025, 6:1491146. <https://doi.org/10.3389/fragi.2025.1491146>
18. Alenazi AM, Alhwoaimel NA, Alqahtani BA, Alshehri MM, Alhowimel AS, Khunti K: Prevalence of multiple long-term chronic conditions and associated disabilities among community-dwelling adults in Riyadh. *Front Public Health*. 2024, 12:1275124. <https://doi.org/10.3389/fpubh.2024.1275124>
19. Alnuwaysir N, Alturki H, Almalki B, Bin Amer L, Alotaibi R: Review of multimorbidity in Saudi Arabia: prevalence, gender differences, and common chronic diseases. *J Multimorb Comorb*. 2025, 15:26335565251367587. <https://doi.org/10.1177/26335565251367587>

20. Rahman R, Al-Borie HM: Strengthening the Saudi Arabian healthcare system: role of Vision 2030. *Int J Healthc Manag.* 2021, 14:1483-1491. <https://doi.org/10.1080/20479700.2020.1788334>
21. Albejaidi FM, Alharbi A: Analyzing the technological framework for e-health in the Kingdom of Saudi Arabia. *Proceedings of the 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom).* IEEE, New Delhi, India; 2024. 1822-1827. <https://doi.org/10.23919/INDIACom61295.2024.10498725>
22. Algerian N, Arafat M, Aldhubib A, et al.: Artificial intelligence in health care and its application in Saudi Arabia. *Int J Innov Res Med Sci.* 2022, 7:666-670. <https://doi.org/10.23958/ijirms/vol07-i11/1558>
23. Abbas M, Shoaib Khan MB, Khan AH, Bilal A, Algarni A, Sarwar R: Interpretable machine learning models for beta thalassemia prediction: an explainable AI approach for smart healthcare 5.0. *Front Med (Lausanne).* 2026, 12:1688645. <https://doi.org/10.3389/fmed.2025.1688645>
24. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D: Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019, 17:195. <https://doi.org/10.1186/s12916-019-1426-2>
25. Steyerberg EW, Vickers AJ, Cook NR, et al.: Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010, 21:128-138. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
26. Alsaleh MM, Allery F, Choi JW, Hama T, McQuillin A, Wu H, Thygesen JH: Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: a systematic review. *Int J Med Inform.* 2023, 175:105088. <https://doi.org/10.1016/j.ijmedinf.2023.105088>
27. Baharoon M, Almatar H, Alduhayan R, et al.: HyMNet: a multimodal deep learning system for hypertension prediction using fundus images and cardiometabolic risk factors. *Bioengineering (Basel).* 2024, 11:1080. <https://doi.org/10.3390/bioengineering11111080>
28. Ogunpola A, Saeed F, Basurra S, Albarrak AM, Qasem SN: Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics (Basel).* 2024, 14:144. <https://doi.org/10.3390/diagnostics14020144>
29. Rashed-Al-Mahfuz M, Haque A, Azad AKM, Alyami SA, Quinn JMW, Moni MA: Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (CKD) for use in low-cost diagnostic screening. *IEEE J Transl Eng Health Med.* 2021, 9:4900511. <https://doi.org/10.1109/JTEHM.2021.3073629>
30. Albahr A, Albahr M, Thanoon M, Binsawad M: Computational learning model for prediction of heart disease using machine learning based on a new regularizer. *Comput Intell Neurosci.* 2021, 2021:8628335. <https://doi.org/10.1155/2021/8628335>
31. Alhaifi S, Naemi FMA, Alowidi N: A risk-oriented and explainable hierarchical AI framework for chronic kidney disease classification. *Diagnostics (Basel).* 2026, 16:1157. <https://doi.org/10.3390/diagnostics16081157>
32. Khan S: Biochemical biomarker-driven deep learning framework with SHAP-based feature interpretation for diabetes classification. *Biophys Chem.* 2026, 334:107601. <https://doi.org/10.1016/j.bpc.2026.107601>
33. Tourkmani AM, Al-Harbi TJ, Alghamdi AA, Youzghadli IM, Alosaimi FS, Azzam AY: Literature-informed ensemble machine learning for three-year diabetic kidney disease risk prediction in type 2 diabetes: development, validation, and deployment of the PSMMC NephrRisk model. *Diabetes Obes Metab.* 2026, 28:1997-2026. <https://doi.org/10.1111/dom.70385>
34. Alshahrani JA, Alshahrani AM, Alshalaan AM, et al.: Rising obesity and shifting disease patterns in Saudi Arabia: a nine-year population-based analysis of chronic disease burden and multimorbidity profiles. *BMC Public Health.* 2026, 26:1163. <https://doi.org/10.1186/s12889-026-26923-2>
35. Alnomasy N, Banappagoudar SB, Alrashedi H, Mahmoud SKM, Abouhashish E, Ruksakulpiwat S: Longitudinal trends in noncommunicable disease risk factors and premature mortality in Saudi Arabia: a 33-year ecological time-series study with machine learning prediction. *J Clin Med.* 2026, 15:4387. <https://doi.org/10.3390/jcm15114387>
36. Hasan U, Algeffari M, Alhasson HF, Nayeem MA: Explainable lightweight AI for the identification of right-sided cardiac dysfunction in a Saudi Arabian diabetic cohort. *J Clin Med.* 2026, 15:4719. <https://doi.org/10.3390/jcm15124719>