JOURNAL OF DIABETES RESEARCH AND MANAGEMENT
Jan - Dec 2026, VOL. 1, ISSUE 1, pp. 25-29
https://doi.org/10.61577/jdrm.2026.100006

RESEAPRO JOURNALS

NARRATIVE REVIEW

OPEN ACCESS

# Artificial intelligence in clinical endocrinology: A narrative review of predictive analytics, therapeutic optimization, and diagnostic applications

**Asjed Khan Jadoon[1], Syeda Afroz[1], Misbah[2], Asfand Yar Khaliq[3], Uswah Dawood[3] and Saadat Khan[4]**

[1]Department of Pharmacology, University of Karachi, Karachi, Pakistan

[2]Department of Operative and Conservative Dentistry, Bolan Medical Complex Hospital Unit 2, Quetta, Pakistan

[3]Sheikh Mohammad Bin Zayed Al-Nahyan Institute of Cardiology, Quetta, Pakistan

[4]Teaching Hospital, Pishin, Pakistan

## Abstract

**Background:** Endocrine disease is governed by non-linear, multi-organ feedback loops and longitudinal physiological data, which makes it a natural fit for machine learning (ML) and deep learning (DL) methods. This narrative review synthesizes current evidence on the application of artificial intelligence (AI) across five domains of clinical endocrinology: diabetes risk prediction, metformin response and the gut microbiome, continuous glucose monitoring (CGM) with closed-loop insulin delivery, thyroid nodule ultrasound, and adrenal lesion imaging.

**Methods:** PubMed, Scopus, and Web of Science were searched for studies and systematic reviews published mainly between 2020 and 2026, using combinations of terms including ML, DL, "diabetes prediction," "metformin," "gut microbiome," "continuous glucose monitoring," "closed-loop insulin," "thyroid nodule," and "adrenal incidentaloma." Priority was given to meta-analyses, randomized controlled trials, and large validated cohorts; small single-center case series were included only where no larger evidence existed for that domain, and are flagged as such. As a narrative rather than systematic review, no PRISMA flow diagram or formal record count is reported; this is stated as a limitation.

**Results:** Ensemble models (random forest, XGBoost) predicted incident type 2 diabetes with 98-99% accuracy in one large longitudinal cohort in which most participants did not develop diabetes; because accuracy can be inflated in such imbalanced outcome data, this figure should be interpreted alongside positive predictive value, which was not reported in the source study, and alongside the fact that external validation remains uncommon in this literature. DL models applied to thyroid ultrasound reached pooled sensitivities of 87-91% and specificities of 83-95% across meta-analyses covering more than 146,000 patients. Randomized trial data show that closed-loop insulin delivery significantly improves time-in-range and reduces hypoglycemia, while reinforcement-learning-based dosing remains largely at the simulation stage. ML models linking gut microbiome composition to drug pharmacokinetics achieved moderate accuracy (AUROC approximately 0.75), a distinct question from separate work linking gut microbiome signatures to type 2 diabetes risk itself. Adrenal CT radiomics models reported AUCs frequently above 0.85, though largely from small, single-center cohorts (as few as 19-40 patients).

**Conclusion:** AI shows strong, evidence-backed value in endocrine risk prediction and image-based diagnosis, though the diabetes-prediction accuracy figures above warrant cautious interpretation given class imbalance. Evidence for therapeutic personalization, particularly metformin response, is earlier-stage. Larger prospective trials, external validation, standardized outcome reporting (including PPV alongside accuracy), and improved interpretability are needed to support broader clinical adoption.

## KEYWORDS

Artificial intelligence; Machine learning; Deep learning; *Diabetes mellitus*; Metformin; Continuous glucose monitoring; Thyroid nodules; Adrenal incidentaloma

## ARTICLE HISTORY

Received 11 July 2026;
Accepted 03 August 2026; Published 10 August 2026

## Introduction

Endocrinology governs physiological systems defined by non-linear interactions, circadian rhythms, and multi-organ feedback mechanisms. This complexity, combined with the growing digitization of clinical, laboratory, and imaging data, has made endocrinology one of the more active specialties for applied AI research over the past decade [1,2]. Ensemble learning methods, convolutional neural networks, and reinforcement learning have each found distinct footholds: structured clinical data for diabetes risk prediction, imaging data for thyroid and adrenal lesion characterization, and time-series glucose data for closed-loop insulin delivery. Beyond the five domains reviewed in depth here, AI has also achieved regulatory milestones elsewhere in diabetes care, most notably the 2018 FDA authorization of an autonomous deep-learning system for diabetic retinopathy screening in primary care settings, built on earlier DL models validated against ophthalmologist grading in large retinal photograph datasets, a landmark that illustrates both the clinical feasibility and the regulatory pathway available to endocrine-AI tools that reach comparable evidentiary maturity [3,4]. Independent reproduction attempts using public data have since underscored the importance of external validation and code availability for this class of model, a theme that recurs throughout the domains reviewed below [5].

Most existing reviews of AI in endocrinology address either a single disease domain (for example, diabetic retinopathy screening or thyroid nodule classification) or provide a broad survey without distinguishing the maturity of the evidence behind each claim. The novelty of this review lies in bringing five distinct endocrine-AI application areas into a single comparative framework and explicitly grading the strength of evidence behind each: from randomized-trial-validated closed-loop insulin delivery, to externally validated thyroid imaging models, to earlier-stage, mechanistically promising but clinically unproven metformin-microbiome prediction. This review synthesizes the current, verifiable evidence across these areas, identifies where the evidence base is strongest and weakest, and proposes concrete directions for closing the gap between algorithmic performance and routine clinical use; the specific evidence gaps and proposed directions are detailed in Section (Discussion).

## Methods

This narrative review is based on a structured literature search of PubMed, Scopus, and Web of Science for English-language studies published primarily from 2020 through mid-2026. Search terms combined ML, DL

***Correspondence:** Dr. Asjed Khan Jadoon, Department of Pharmacology, University of Karachi, Karachi, Pakistan. e-mail: asjedjadoon@gmail.com

and AI with domain-specific terms: diabetes prediction, metformin, gut microbiome, CGM, closed-loop insulin, thyroid nodule, and adrenal incidentaloma. Priority was given to systematic reviews, meta-analyses, randomized controlled trials, and large prospective cohorts. Smaller single-center studies were included only for the adrenal-imaging domain, where no larger evidence base currently exists, and are explicitly flagged wherever they are discussed. Reported performance metrics, sample sizes, and confidence intervals are drawn directly from the source publications; where a source reported a percentage and another reported the equivalent decimal proportion (for example, sensitivity of 0.87 versus 87%), both are converted to a consistent percentage format in this review for readability. Studies based on simulation rather than clinical data are identified explicitly wherever relevant.

This review follows the SANRA (Scale for the Assessment of Narrative Review Articles) framework for narrative-review methodology and reporting [6]. Readers seeking a fully systematic account of any single domain covered here are directed to the domain-specific systematic reviews and meta-analyses cited [7-12].

## Results

### Diabetes prediction

A 2025 bibliometric and systematic synthesis covering 33 years of literature (1991-2024) applied PRISMA-guided screening to identify 2,351 relevant records on ML and AI in type 2 diabetes prediction from Web of Science and Scopus, documenting a marked rise in publication volume and methodological sophistication since 2015 [13]. At the level of individual cohort studies, a 10-year longitudinal analysis of 6,687 relatively healthy adults in Taiwan found that random forest and XGBoost models predicted incident diabetes with 98-99% accuracy, with free thyroxine (fT4) emerging as an informative predictor alongside conventional metabolic variables [14].

This accuracy figure requires careful interpretation. Most participants in a general, relatively healthy cohort do not go on to develop diabetes within the follow-up window, so the outcome is imbalanced; a model that is biased toward predicting the majority (non-diabetic) outcome can achieve high accuracy without necessarily identifying most true cases well. Positive predictive value, PPV, the proportion of model-flagged positive cases that are true positives, is the more clinically informative metric in this setting and was not reported in the source study [14]. In the absence of PPV, the headline accuracy figure should be read as an upper-bound indicator of overall correctness rather than as a direct measure of the model's ability to identify people who will develop diabetes. These figures are also notably higher than typical population risk scores; such single-cohort accuracy figures of this kind often decline when tested on external, more heterogeneous populations, and prospective external validation remains uncommon in this literature. Readers evaluating any future accuracy claim in this space are encouraged to specifically request PPV, sensitivity, and specificity reported together, rather than accuracy alone, particularly for imbalanced clinical outcomes [15,16].

### Metformin response prediction and gut microbiome risk signatures: two distinct questions

The literature linking the gut microbiome to metformin and to type 2 diabetes actually addresses two separate, non-interchangeable research questions, which this review treats separately to avoid conflating them.

The first question is whether gut microbiome composition can predict how an individual drug, including metformin, will be metabolized or depleted by intestinal bacteria. A ML study built on 455 curated drug-microbiota interaction records trained eleven supervised models to predict whether a given drug would be depleted or altered by gut bacteria; the best-performing model, a tuned extremely randomized trees classifier, reached an AUROC of 75.1% (+/-6.8) on a held-out set of 91 drugs [17]. This body of work supports the biological plausibility of microbiome-informed pharmacotherapy, but direct, adequately powered trials predicting individual metformin response specifically from baseline microbiome data remain limited.

The second, distinct question is whether gut microbiome composition predicts type 2 diabetes risk itself, independent of any drug. An interpretable ML framework applied to three independent Chinese cohorts (a discovery cohort and two validation cohorts, totaling 9,111 participants with 926 T2DM cases) identified a robust gut microbiome risk score consistently associated with T2DM across all three cohorts [18]. This is a risk-prediction finding, not a drug-response finding, and should not be read as evidence for microbiome-guided metformin dosing specifically. Interest in microbiome-guided pharmacotherapy more broadly stems from the observation that gut bacterial composition varies widely between individuals and can alter drug metabolism and tolerability [19]. Taken together, both strands of evidence are real and separately useful, but neither, individually or combined, yet constitutes a tool ready for metformin prescribing decisions; this remains an active area of investigation.

### Continuous glucose monitoring and closed-loop insulin delivery

The clinical evidence base is strongest here, because it includes randomized trial data. A two-center, randomized crossover trial in 36 adults with type 1 diabetes compared a model-predictive-control closed-loop system against sensor-augmented pump therapy over 12-day outpatient periods. Time in the 3.9-10.0 mmol/L range increased from 61% to 69% (P=0.006), tight-range time (3.9-7.8 mmol/L) rose from 37% to 45% (P=0.011), and time below 3.9 mmol/L fell from 3.5% to 1.6% (P=0.0021) [20]. A separate line of research explores deep reinforcement learning (DRL) for fully autonomous dosing. A 2025 proximal-policy-optimization (PPO) controller achieved a median time-in-range of 87.45% in the Simglucose simulation environment, outperforming baseline control methods in silico [21]. This DRL result is simulation-only; clinical reinforcement-learning-based systems remain earlier in development than the model-predictive-control systems already validated in randomized trials. A neural-network-based closed-loop system tested in a small clinical cohort achieved 86.1% time-in-range in a supervised hybrid mode and 72-74% in fully closed-loop, unsupervised conditions, versus 68% with usual care [22], illustrating the typical performance gap between supervised and fully autonomous operation.

### Thyroid nodule ultrasound

This is the most extensively validated imaging application in endocrine AI. A meta-analysis of 11 studies using CNN-based (VGGNet) models to classify thyroid nodules on ultrasound reported pooled sensitivity of 87% (95% CI 83-91%), specificity of 85% (95% CI 79-90%), and an AUC of 93% (95% CI 90-95%) [7]. A larger 2025 systematic review and meta-analysis spanning 41 studies separated detection tasks (pooled sensitivity 91%, specificity 85%, AUC 94%) from segmentation tasks (sensitivity 82%, specificity 95%, AUC 91%) [8,9]. The most externally robust estimate comes from a 2025 meta-analysis restricted to studies with independent external test sets: across 27 studies covering 146,332 patients and more than 600,000 ultrasound images, pooled sensitivity was 87% (95% CI 84-89%) and specificity was 83% (95% CI 79-86%), with an AUC of 91.9% (95% CI 90.0-93.2%) [10]. Performance declines modestly but remains strong when models are tested on external rather than internal validation sets, which supports the credibility of this literature. Diagnostic performance in this domain is reported consistently as sensitivity, specificity, and AUC across the cited sources, so no unit standardization was required beyond converting decimal proportions to percentages for readability. It is worth noting that AI-based classification is being evaluated against, and sometimes alongside, established structured ultrasound risk-stratification systems such as ACR TI-RADS and EU-TIRADS, which themselves show meaningful performance variation across validation cohorts and are not a fixed gold standard [23-25]; AI models in this space should therefore be benchmarked against contemporary TI-RADS performance rather than against an idealized reference standard.

### Adrenal incidentaloma and lesion characterization

Evidence in this domain is real but comes largely from smaller, single-center studies; both findings below are reported here with their sample sizes made explicit at first mention, given how small these cohorts are relative to the other domains reviewed. An early radiomics study using CT texture analysis and unsupervised K-means clustering in a cohort of only 19 patients (9 adenomas, 10 carcinomas) correctly predicted malignancy in 7 of 8 adrenocortical carcinomas [26]; a finding of this kind from 19 patients is hypothesis-generating rather than confirmatory and requires replication in larger cohorts before any clinical inference is drawn. A separate random-forest model built on enhanced CT features in a cohort of 40 patients achieved an AUC of 85% (sensitivity 84.2%, specificity 71.4%)



https://doi.org/10.61577/jdrm.2026.100006

for adrenal tumor pathology classification [26]; a broader meta-analysis of adrenal radiomics reported a comparable pooled discriminative performance across studies of similar scale [27], and a more recent single-center study similarly combined original radiomics data with a PRISMA-based systematic review to contextualize CT texture findings in adrenal lesion characterization [25]. A 2026 narrative review of AI in adrenal imaging found that segmentation models now approach Dice coefficients of 0.90 and that several lesion-characterization studies report AUCs above 0.90, with multimodal models (imaging plus clinical data) generally outperforming imaging alone [11]. The same review is explicit that most of this evidence comes from retrospective, single-center, small or selected cohorts, which limits how confidently these figures generalize [11,12].

## Summary of key quantitative findings

**Table 1.** Summary of reported AI performance metrics across the endocrine domains reviewed.

| Domain | Key finding | Source |
|---|---|---|
| Type 2 diabetes prediction | Random forest and XGBoost reached 98-99% accuracy predicting incident T2DM over 10 years in a Taiwanese cohort of 6,687 adults (of whom a minority developed diabetes); free thyroxine (fT4) emerged as an informative feature. Because diabetes onset was the minority outcome in this cohort, accuracy alone can overstate performance, and positive predictive value was not reported (see Section Diabetes prediction). | Liu et al., [14] |
| Diabetes/AI literature volume | A 33-year (1991-2024) PRISMA-guided bibliometric synthesis identified 2,351 relevant records on ML/AI in T2DM prediction across Web of Science and Scopus. | Kiran et al., [13] |
| Metformin response prediction (drug-microbiota depletion) | A supervised ML model (tuned extremely randomised trees) trained on 455 drug-microbiota interaction records predicted susceptibility of drugs, including metformin-relevant compounds, to gut-microbial depletion with AUROC 75.1% (+/-6.8). | Klunemann et al., [17] |
| Gut microbiome signatures for T2DM risk (distinct endpoint) | An interpretable ML framework applied to three independent Chinese cohorts (n=9,111) identified a gut microbiome risk score consistently associated with T2DM, a distinct question from drug-microbiota depletion above. | Gou et al., [18] |
| Thyroid nodule ultrasound (CNN/VGGNet) | Meta-analysis of 11 studies: pooled sensitivity 87% (95% CI 83-91%), specificity 85% (95% CI 79-90%), AUC 93% (95% CI 90-95%). | Meta-analysis, [7] |
| Thyroid nodule DL, external validation | 27 studies, 146,332 patients, over 600,000 images: pooled sensitivity 87% (95% CI 84-89%), specificity 83% (95% CI 79-86%), AUC 91.9% (95% CI 90.0-93.2%). | Systematic review, [10] |
| Closed-loop insulin delivery (RCT) | Randomized crossover trial (n=36, type 1 diabetes): time-in-range rose from 61% to 69% (P=0.006) versus sensor-augmented pump therapy; time below 3.9 mmol/L fell from 3.5% to 1.6%. | Haidar et al., [20] |
| Deep reinforcement learning insulin control | A PPO-based deep reinforcement learning controller achieved 87.45% median time-in-range in the Simglucose simulation environment; clinical testing has not yet been reported. | Zhao et al., [21] |
| Adrenal lesion classification | A random-forest model on enhanced CT distinguished adrenal tumor pathology with AUC 85% (sensitivity 84.2%, specificity 71.4%); note this and the K-means finding below both derive from very small cohorts (n=40 and n=19 respectively) and require external validation before generalization. | Li et al., [26] |

## Comparative performance across domains

Figure 1 summarizes the best available reported sensitivity, specificity, and discrimination metric (AUC, AUROC, or accuracy, as applicable) across the five domains reviewed, drawn directly from the source studies cited in Table 1. These are fundamentally different statistical quantities: AUC and AUROC describe a model's discrimination across all possible decision thresholds, whereas accuracy is a single-threshold measure that, as discussed in diabetes prediction section, can be misleading under class imbalance. The bars are plotted on a common 0-100% axis purely as a visual layout convenience to allow domains to be viewed side by side; this figure is not intended to imply that, for example, an AUC of 93% for thyroid nodule classification is quantitatively equivalent to or better than an accuracy of 98.5% for diabetes prediction. Each bar is labeled with its specific metric type, and the diabetes-prediction accuracy figure in particular should be read together with the PPV caveat in Section Diabetes prediction. With that caveat, thyroid nodule classification shows the strongest and most consistent evidence across independent meta-analyses; metformin-microbiome prediction shows the most modest and earliest-stage performance.

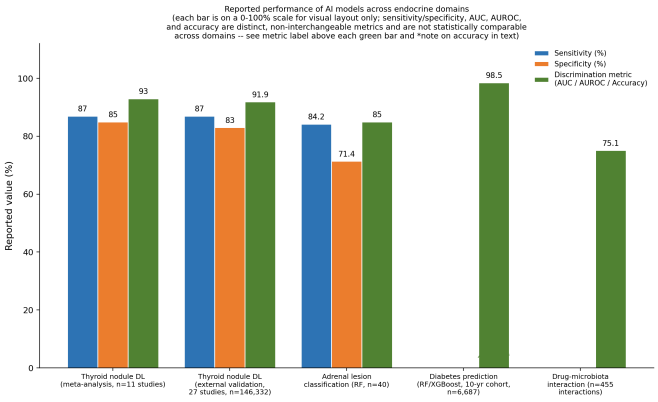


**Figure 1.** Reported diagnostic and predictive performance of AI models across endocrine domains. Bars share a common percentage axis for layout only; metric types differ by domain and are not directly comparable (see Section Comparative performance across domains).

Figure 2 presents a conceptual framework situating these applications along the endocrine care pathway, from risk prediction through diagnosis,



https://doi.org/10.61577/jdrm.2026.100006

therapeutic optimization, and monitoring, alongside the cross-cutting barriers discussed in the conclusion.

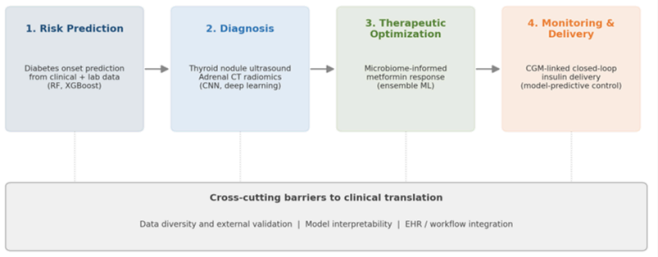


**Figure 2.** AI applications across the endocrine care continuum, with cross-cutting barriers to clinical translation.

## Discussion

The strongest, most reproducible evidence for AI in endocrinology concerns image-based diagnosis, particularly thyroid nodule classification, where large multi-study meta-analyses with external validation converge on sensitivity and specificity figures in the mid-to-high 80s and AUCs above 90% [7,10]. Diabetes risk prediction from structured clinical and laboratory data is also well studied, though the highest accuracy figures tend to come from single-cohort studies not yet externally validated at scale, and, as discussed in the diabetes prediction section, accuracy alone is an incomplete metric for the class-imbalanced outcome these studies predict [14-16]. Closed-loop insulin delivery has moved furthest into real clinical use, backed by randomized trial data showing statistically significant improvements in time-in-range and reduced hypoglycemia [20]. Reinforcement learning specifically, as distinct from the model-predictive-control algorithms already in clinical closed-loop systems, remains largely at the simulation stage [21], and this distinction should not be blurred in future reviews or clinical claims.

The weakest evidence, relative to how it is sometimes portrayed in secondary literature, concerns microbiome-guided metformin dosing specifically, as distinct from the separate and better-supported finding that gut microbiome signatures predict T2DM risk itself [17,18]. The mechanistic case for microbiome-drug interaction is credible, and ML models can predict drug-microbiota interaction with moderate accuracy, but translating that into a validated, individual-level metformin response predictor requires larger, prospective, metformin-specific cohorts that are not yet published at scale. Adrenal imaging AI sits in between: individual studies show strong discrimination, but nearly all of it comes from very small, single-center datasets (as few as 19-40 patients) that have not been tested on outside populations [26,11].

Three recurring gaps emerged across the literature reviewed here. First, most datasets underpinning diagnostic and predictive models are drawn from limited geographic and demographic populations, raising concerns about generalizability when models are deployed in more diverse clinical settings [29]. Second, model interpretability remains inconsistent: DL models frequently function as black boxes, which limits clinician trust in high-stakes decisions such as adrenal surgery referral [30]. Third, few studies report external, prospective validation; most performance figures, particularly in diabetes prediction and adrenal imaging, come from single-center retrospective cohorts, and, per Section 2, this narrative review itself does not report a systematic screening yield, which is a limitation of the present synthesis rather than of the underlying primary literature.

To address these gaps, this review proposes three concrete directions. Federated learning approaches, in which models are trained across multiple institutions without sharing raw patient data, have already been shown in other areas of medicine to approach the performance of centrally pooled data while preserving data privacy, and could improve dataset diversity in endocrine AI without requiring centralized data pooling [31]. Standardized, mandatory reporting of external validation metrics, and of PPV alongside accuracy for imbalanced outcomes, would allow more honest comparison across studies. Wider integration of explainable AI techniques, such as SHAP-based feature attribution, into clinical decision-support interfaces would give endocrinologists a transparent, quantified basis for weighing AI-generated recommendations rather than treating them as unexplained outputs [30,32].

Addressing these barriers, namely data diversity, interpretability, and workflow integration, will likely determine how quickly these tools move from published performance metrics to routine clinical decision support.

## Conclusion

AI applications in endocrinology are not uniformly at the same stage of maturity. Thyroid ultrasound classification and closed-loop insulin delivery have real, externally validated or randomized-trial evidence behind them. Diabetes risk prediction and adrenal lesion characterization show strong single-study results that require broader validation and, in the case of diabetes prediction, more complete metric reporting given class imbalance. Gut-microbiome-based prediction addresses two genuinely distinct questions, disease risk and drug response, neither of which is yet clinically actionable for metformin prescribing specifically. Reporting these distinctions clearly, alongside concrete steps such as federated learning, standardized external validation and PPV reporting, and explainable AI integration, is necessary for the field to build justified clinical trust.

## Disclosure Statement

The authors declare no conflict of interest.

## Acknowledgements

None.

## Author Contributions

All authors contributed to the conception, literature review, drafting, and critical revision of this manuscript, and approved the final version for submission.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

1. Belkhouribchia J. Artificial intelligence is going to transform the field of endocrinology: An overview. Front Endocrinol. 2025; 16:1513929. https://doi.org/10.3389/fendo.2025.1513929
2. Giorgini F, Di Dalmazi G, Diciotti S. Artificial intelligence in endocrinology: A comprehensive review. J Endocrinol Invest. 2024;47(5):1067-1082. https://doi.org/10.1007/s40618-023-02235-9
3. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1:39. https://doi.org/10.1038/s41746-018-0040-6
4. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402-2410. https://doi.org/10.1001/jama.2016.17216
5. Voets M, Mollersen K, Bongo LA. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. PLoS One. 2019;14(6): e0217541. https://doi.org/10.1371/journal.pone.0217541
6. Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4:5. https://doi.org/10.1186/s41073-019-0064-8
7. Zhu PS, Zhang YR, Ren JY, Li QL, Chen M, Sang T, et al. Ultrasound-based deep learning using the VGGNet model for the differentiation of benign and malignant thyroid nodules: A meta-analysis. Front Oncol. 2022; 12:944859. https://doi.org/10.3389/fonc.2022.944859
8. Xu J, Xu HL, Cao YN, Huang Y, Gao S, Wu QJ, et al. The performance of deep learning on thyroid nodule imaging predicts thyroid cancer: A systematic review and meta-analysis of epidemiological studies with independent external test sets. Diabetes Metab Syndr. 2023;17(11):102891. https://doi.org/10.1016/j.dsx.2023.102891
9. Ni J, You Y, Wu X, Chen X, Wang J, Li Y. Performance evaluation of deep learning for the detection and segmentation of thyroid nodules: Systematic review and meta-analysis.



https://doi.org/10.61577/jdrm.2026.100006

J Med Internet Res. 2025;27: e73516. https://doi.org/10.2196/73516

10. Gatta E, Gatta R, Morandi R, Isoli S, Corvaglia S, Vetrugno S, et al. Machine learning for diagnosis of malignant thyroid nodules based on thyroid ultrasound: Systematic review and meta-analysis of studies with external datasets. Eur J Radiol Open. 2025; 16:100716. https://doi.org/10.1016/j.ejro.2025.100716
11. Mucha M, Roszkowska Z, Ambroziak U, Bobrowicz M, Roszczyk R. The application of artificial intelligence in adrenal imaging: Current state of knowledge, challenges, and future directions. Endocr Connect. 2026;15(7). https://doi.org/10.1530/EC-26-0267
12. Ferro M, Tataru OS, Carrieri G, Busetto GM, Falagario UG, Maggi M, et al. Artificial intelligence and radiomics applications in adrenal lesions: A systematic review. Ther Adv Urol. 2025. https://doi.org/10.1177/17562872251352553
13. Kiran M, Xie Y, Anjum N, Ball G, Pierscionek B, Russell D. Machine learning and artificial intelligence in type 2 diabetes prediction: A comprehensive 33-year bibliometric and literature analysis. Front Digit Health. 2025; 7:1557467. https://doi.org/10.3389/fdgth.2025.1557467
14. Liu YQ, Chang TW, Lee LC, Chen CY, Hsu PS, Tsan YT, et al. Use of machine learning to predict the incidence of type 2 diabetes among relatively healthy adults: A 10-year longitudinal study in Taiwan. Diagnostics. 2025;15(1):72. https://doi.org/10.3390/diagnostics15010072
15. Cabot JH, Ross EG. Evaluating prediction model performance. Surgery. 2023;174(3):723-726. https://doi.org/10.1016/j.surg.2023.05.023
16. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology. 2010;21(1):128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2
17. Klunemann M, Andrejev S, Blasche S, Mateus A, Phapale P, Devendran S, et al. Machine learning predicts drug metabolism and bioaccumulation by intestinal microbiota. Nature. 2021;597(7877):533-538. https://doi.org/10.3390/pharmaceutics13122001
18. Gou W, Ling CW, He Y, Jiang Z, Fu Y, Xu F, et al. Interpretable machine learning framework reveals robust gut microbiome features associated with type 2 diabetes. Diabetes Care. 2021;44(2):358-366. https://doi.org/10.2337/dc20-1536
19. Anwardeen NR, Naja K, Elrayess MA. Advancements in precision medicine: Multi-omics approach for tailored metformin treatment in type 2 diabetes. Front Pharmacol. 2024; 15:1506767. https://doi.org/10.3389/fphar.2024.1506767
20. Haidar A, Legault L, Raffray M, Gouchie-Provencher N, Jacobs PG, El-Fathi A, et al. Comparison between closed-loop insulin delivery system (the artificial pancreas) and sensor-augmented pump therapy: A randomized-controlled crossover trial. Diabetes Technol Ther. 2021;23(3):168-174. https://doi.org/10.1089/dia.2020.0365
21. Zhao YF, Chaw JK, Ang MC, Tew Y, Shi XY, Liu L, et al. A safe-enhanced fully closed-loop artificial pancreas controller based on deep reinforcement learning. PLoS One. 2025;20(1):e0317662. https://doi.org/10.1371/journal.pone.0317662
22. Kovatchev B, Castillo A, Pryor E, et al. Neural-net artificial pancreas: A randomized crossover trial of a first-in-class automated insulin delivery algorithm. Diabetes Technol Ther. 2024;26(6):375-382. https://doi.org/10.1089/dia.2023.0469
23. Castellana M, Grani G, Radzina M, Guerra V, Giovanella L, Deandrea M, et al. Performance of EU-TIRADS in malignancy risk stratification of thyroid nodules: A meta-analysis. Eur J Endocrinol. 2020;183(3):255-264.
24. Chatti HA, Oueslati I, Azaiez A, Marrakchi J, Boukriba S, Mizouni H, et al. Diagnostic performance of the EU TI-RADS and ACR TI-RADS scoring systems in predicting thyroid malignancy. Endocrinol Diabetes Metab. 2023;6(5): e434. https://doi.org/10.1002/edm2.434
25. Kang YJ, Stybayeva G, Lee JE, Hwang SH. Diagnostic performance of ACR and Kwak TI-RADS for benign and malignant thyroid nodules: An update systematic review and meta-analysis. Cancers (Basel). 2022;14(23):5961. https://doi.org/10.3390/cancers14235961
26. Li C, Fu Y, Yi X, Guan X, Liu L, Chen BT. Application of radiomics in adrenal incidentaloma: a literature review. Discov Oncol. 2022;13(1):112. https://doi.org/10.1007/s12672-022-00577-z
27. Zhang H, Lei H, Pang J. Diagnostic performance of radiomics in adrenal masses: A systematic review and meta-analysis. Front Oncol. 2022; 12:975183. https://doi.org/10.3389/fonc.2022.975183
28. Frisina P, Ricci P, Valentini F, Messineo D. Radiomic characterization of adrenal incidentalomas on NECT: Retrospective exploratory study and systematic review. J Imaging. 2026;12(4):151. https://doi.org/10.3390/jimaging12040151
29. Castellana M, Grani G, Radzina M, Guerra V, Giovanella L, Deandrea M, et al. Performance of EU-TIRADS in malignancy risk stratification of thyroid nodules:
a meta-analysis. Eur J Endocrinol. 2020;183(3):255-264. https://doi.org/10.1530/EJE-20-0204
30. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-453. https://doi.org/10.1126/science.aax2342
31. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020; 10:12598. https://doi.org/10.1038/s41598-020-69250-1
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017; 30:4765-4774. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html



https://doi.org/10.61577/jdrm.2026.100006