



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	ARTIFICIAL INTELLIGENCE FOR PREDICTING CHRONIC KIDNEY DISEASE PROGRESSION: A REVIEW OF CURRENT EVIDENCE AND GLOBAL CLINICAL IMPLICATIONS
----------------------	--

DOI	https://doi.org/10.31435/ijitss.3(51).2026.6502
------------	---

RECEIVED	26 July 2026
-----------------	--------------

ACCEPTED	21 September 2026
-----------------	-------------------

PUBLISHED	29 September 2026
------------------	-------------------

LICENSE



The article is licensed under a **Creative Commons Attribution 4.0 International License**.

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

ARTIFICIAL INTELLIGENCE FOR PREDICTING CHRONIC KIDNEY DISEASE PROGRESSION: A REVIEW OF CURRENT EVIDENCE AND GLOBAL CLINICAL IMPLICATIONS

Krzysztof Maksym Kasparek (Corresponding Author, Email: krzysztof.kasparek@protonmail.com)
Poznan University of Medical Sciences, Poland
ORCID ID: 0009-0007-1554-0535

Julia Obersztyn
University Clinical Hospital in Poznań 49 Przybyszewskiego Street, 60-355 Poznań, Poland
ORCID ID: 0009-0001-7547-4556

Jakub Maksymilian Bajer
Poznan University of Medical Sciences, Student Scientific Society, Poland
ORCID ID: 0009-0009-8392-4463

Hubert Kulawinek
University Clinical Hospital in Poznań 49 Przybyszewskiego Street, 60-355 Poznań, Poland
ORCID ID: 0009-0005-3178-0833

Grzegorz Kasparek
Jagiellonian University 12 Świętej Anny Street, 31-008 Kraków, Poland
ORCID ID: 0009-0009-0626-6751

Adrianna Szczecińska
University Clinical Hospital in Poznań 49 Przybyszewskiego Street, 60-355 Poznań, Poland
ORCID ID: 0009-0004-1714-0832

Wiktoria Glanert
Provincial Hospital in Poznań, Juraszów 7/19, 60-479 Poznań Poland
ORCID ID: 0009-0005-2868-4170

Aleksandra Mliczek
University Clinical Hospital in Poznań 49 Przybyszewskiego Street, 60-355 Poznań, Poland
ORCID ID: 0009-0000-9883-3307

Piotr Mrozek
University Clinical Hospital in Poznań 49 Przybyszewskiego Street, 60-355 Poznań, Poland
ORCID ID: 0009-0004-1317-3610

Marta Czech
Poznań University of Medical Sciences, Poland
ORCID ID: 0009-0004-9561-9005

ABSTRACT

Chronic Kidney Disease (CKD) affects approximately 800 million people globally and represents a growing public health and economic burden. Traditional prognostic tools, such as the Kidney Failure Risk Equation, demonstrate limited discriminative accuracy in early disease stages (G1–G3), where kidney failure events are inherently rare.

Aim: The primary objective of this review is to evaluate the performance of artificial intelligence (AI) and machine learning (ML) models in predicting CKD progression.

Methods: A narrative review approach was used to synthesise evidence published between January 2022 and June 2026, identified through searches of PubMed/MEDLINE, Embase, and Web of Science. Ten studies comprising diverse international cohorts were included.

Key findings: AI/ML models demonstrated superior predictive accuracy compared with traditional benchmarks, achieving a pooled Area Under the Curve (AUC) of 0.87–0.89. Long Short-Term Memory recurrent neural networks, leveraging longitudinal electronic health record data, attained AUROC scores of 0.957–0.967. Tree-based ensemble methods, including Random Forest and XGBoost, showed robust performance across heterogeneous populations. However, a consistent imbalance between high model specificity (0.87–0.92) and relatively low sensitivity (0.43–0.68) was identified, primarily attributable to class imbalance in training datasets. Explainable AI frameworks, particularly SHAP and LIME, successfully addressed the interpretability barrier, identifying eGFR slope, proteinuria, and haemoglobin as dominant predictors of renal decline.

Conclusions: These findings support a transition toward proactive, precision-based nephrology. Global clinical adoption, however, requires standardisation of predictor criteria, optimisation of model sensitivity, broader external validation across diverse ethnic populations, and targeted solutions for implementation in low- and middle-income healthcare settings.

KEYWORDS

Chronic Kidney Disease, Machine Learning, Artificial Intelligence, Disease Progression, Predictive Modelling, End-stage Kidney Disease

CITATION

Kasperek, K. M., Obersztyn, J., Bajer, J. M., Kulawinek, H., Kasperek, G., Szczecińska, A., Glanert, W., Mliczek, A., Mrozek, P., & Czech, M. (2026). Artificial intelligence for predicting chronic kidney disease progression: A review of current evidence and global clinical implications. *International Journal of Innovative Technologies in Social Science*, 3(51). [https://doi.org/10.31435/ijtss.3\(51\).2026.6502](https://doi.org/10.31435/ijtss.3(51).2026.6502)

COPYRIGHT

© The author(s) 2026. This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

1.1 Global CKD

Chronic Kidney Disease (CKD) represents an escalating global public health crisis, currently affecting between 8% and 16% of the world's population (Lei et al., 2022). Recent estimates indicate that almost 800 million adults worldwide live with the condition, and its prevalence has increased by 29.3% over the past three decades (Bikbov et al., 2020; Pethő et al., 2024). In 2023 alone, CKD was responsible for approximately 1.48 million deaths, ranking as the 9th leading cause of mortality globally (Mark et al., 2025). This burden is projected to worsen significantly, with the World Health Organisation predicting that CKD will become the fifth most common chronic disease by 2040 (Borg et al., 2023). The rise in cases is primarily driven by ageing populations and the increasing incidence of diabetes and hypertension, particularly in low- and middle-income countries that are often least equipped to manage the disease (Kovesdy, 2022; Murray et al., 2012). Beyond its high mortality rate, CKD profoundly impacts quality of life, contributing to an estimated 35.8 million disability-adjusted life years (DALYs) globally in 2017 (Bikbov et al., 2020). The economic implications are equally staggering; for example, in the United States, Medicare spending on CKD beneficiaries alone exceeded \$120 billion in 2017, representing approximately 34% of total Medicare fee-for-service expenditures (Saran et al., 2020). Because CKD is often asymptomatic in its early stages—a phenomenon frequently described as a "silent disease"—many patients are diagnosed only after significant kidney function is lost (Stevens et al., 2024). Consequently, there is an urgent global focus on leveraging artificial intelligence and machine learning for earlier identification and risk stratification to delay disease progression and reduce the massive socioeconomic burden (Sabanayagam et al., 2025).

1.2 Techniques and methods of artificial intelligence used in CKD

The landscape of Artificial Intelligence (AI) and Machine Learning (ML) for predicting chronic kidney disease (CKD) progression encompasses a broad spectrum of techniques, ranging from classical statistical classifiers to advanced deep learning architectures (Saif et al., 2023). Traditional methods, such as Logistic Regression (LR) and Support Vector Machines (SVM), are frequently utilised as high-performing benchmarks due to their ability to handle qualitative risk stratification (Isaza-Ruget et al., 2024). However, recent evidence highlights the superiority of tree-based ensemble models, particularly Random Forest (RF) and its survival variant, Random Survival Forest (RSF), which excel in managing non-linear interactions and missing clinical data (Pan & Tong, 2024). High-performance gradient boosting frameworks like XGBoost, LightGBM (LGBM), and Gradient Boosting Decision Trees (GBDT) have also demonstrated exceptional accuracy, often achieving Area Under the Curve (AUC) scores exceeding 0.90 in diverse global cohorts (Lei et al., 2022).

To address the complex temporal dynamics of kidney function deterioration, researchers have pioneered the use of Deep Learning, specifically Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) and Bidirectional LSTM (BLSTM) units. These models utilise longitudinal electronic health records (EHR) to capture disease trajectories over time rather than relying on cross-sectional snapshots (Zhu et al., 2023). Furthermore, Deep Ensemble models utilise majority voting techniques to fuse diverse classifiers—such as Convolutional Neural Networks (CNN) and LSTMs—thereby minimising the influence of individual weak learners and maximising overall predictive precision (Saif et al., 2023). To enhance clinical utility, practitioners employ the Synthetic Minority Over-sampling Technique (SMOTE) to balance datasets and improve model sensitivity, while Explainable AI (XAI) frameworks like SHAP and LIME are integrated to resolve the "black-box" problem. These XAI tools provide clinicians with transparent, patient-specific insights by illustrating exactly how biomarkers like eGFR or haemoglobin influence an individual's risk score (Chouit et al., 2026). These techniques can be broadly categorised according to their learning paradigm and clinical application, as presented in Table 1.

Table 1. Classification of AI/ML models used in CKD prognosis research.

Learning Category	Model Type	Description & Clinical Characteristics	Example Algorithms from Sources
Supervised Learning	Classical Classification	Used to assign patients into predefined risk groups (e.g., "progression" vs. "stable"). These models often generate qualitative risk scores.	Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes, Decision Trees (DT), k-Nearest Neighbors (k-NN).
	Regression Models	Focus on predicting continuous variables, such as estimating future eGFR values or the specific probability of an outcome event.	Elastic Net, Lasso Regression, Ridge Regression, Logistic Regression.
	Time-to-Event (Survival Analysis)	Specialised models that predict not just if an event will occur (e.g., dialysis), but when it is likely to happen by estimating survival functions.	Random Survival Forest (RSF), Cox Penalty (P-Cox), Gradient Boosting Model (GBM).
	Ensemble Methods	Combine multiple "weak" classifiers into a single robust model to improve stability and minimise the influence of individual learner errors.	XGBoost, LightGBM (LGBM), Gradient Boosting Decision Tree (GBDT), Deep Ensembles.
Deep Learning	Sequential / Time-Series	Recurrent architectures designed to process longitudinal data from electronic health records (EHRs), capturing temporal disease trajectories over years.	Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BLSTM).
	Spatial / Feature Extraction	Often used to capture spatial patterns in clinical data or for processing medical imaging (retinal fundus images or histology).	Convolutional Neural Networks (CNN).
Unsupervised Learning	Clustering & Embedding	Used for finding hidden patterns or identifying disease phenotypes without predefined labels. Useful for risk stratification in diverse populations.	Clustering-based spatial analysis (CluSA), Disease Embedding, Convolutional Autoencoders.
Explainable AI (XAI)	Interpretability Frameworks	Applied to "black-box" models to provide transparent, patient-specific explanations for clinical decision-making.	SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations).

Several analytical themes emerge from the existing literature that form the focus of the present review. Tree-based ensemble methods, particularly Random Forest and XGBoost, have attracted considerable attention for their applicability to structured clinical CKD datasets, while recurrent architectures such as LSTM RNNs have been proposed specifically to address the longitudinal nature of kidney function decline – capturing disease trajectories over time rather than relying on single-point assessments. Meta-analytical evidence has further highlighted a recurring tension between model specificity and sensitivity, attributable in part to class imbalance in training datasets where progression events are inherently rare. Finally, the growing adoption of Explainable AI (XAI) frameworks, particularly SHAP and LIME, reflects a broader clinical shift toward interpretable, patient-specific risk assessment that can be meaningfully integrated into nephrology practice. These themes are systematically examined in the Results and Discussion sections that follow.

2. Methodology

2.1 Study Design

This study employs a narrative review design to synthesise contemporary evidence on the application of artificial intelligence (AI) and machine learning (ML) in predicting the progression of chronic kidney disease (CKD). A narrative rather than systematic approach was selected for two principal reasons. First, the primary aim of this work is to provide a structured conceptual synthesis of a rapidly evolving and methodologically heterogeneous body of literature, rather than to derive a pooled quantitative estimate of effect size from comparable trials. Second, the substantial heterogeneity characterising the included studies – encompassing differences in AI model architectures, outcome definitions, patient populations, data sources, and follow-up durations – renders formal statistical pooling of primary performance data methodologically inappropriate. Where pooled estimates are cited, these are drawn from published systematic reviews and meta-analyses that applied formal meta-analytical methods to comparable subsets of the literature (Lei et al., 2022; Pan & Tong, 2024).

2.2 Search Strategy

A structured electronic search was conducted across three major bibliographic databases: PubMed/MEDLINE, Embase, and Web of Science. The search was restricted to studies published between January 2022 and June 2026. This time window was deliberately selected to reflect the contemporary landscape of AI-driven CKD research – a period defined by the increasing availability of large-scale electronic health record (EHR) datasets, the widespread adoption of gradient boosting frameworks, and the proliferation of deep learning architectures capable of processing longitudinal clinical data at the patient level.

Search queries were constructed using a combination of controlled vocabulary terms and free-text keywords applied in various Boolean combinations. Terms included: "chronic kidney disease", "CKD progression", "kidney failure", "end-stage renal disease", "end-stage kidney disease", "machine learning", "artificial intelligence", "deep learning", "recurrent neural network", "LSTM", "random forest", "gradient boosting", "XGBoost", "prediction model", "risk stratification", and "explainable artificial intelligence". Reference lists of identified review articles and meta-analyses were additionally hand-searched to identify eligible studies not captured by the primary database queries.

2.3 Eligibility Criteria

Studies were considered eligible for inclusion if they met all of the following criteria: (1) they employed at least one AI or ML algorithm, whether supervised, ensemble-based, or deep learning in architecture; (2) the algorithm was applied to a primary outcome pertaining to CKD progression, defined as advancement in disease staging, a clinically meaningful decline in estimated glomerular filtration rate (eGFR) of $\geq 25\%$ or $\geq 40\%$ from baseline, onset of end-stage kidney disease (ESKD), or the requirement for renal replacement therapy (RRT) including dialysis or transplantation; (3) the study population consisted of adult patients (aged ≥ 18 years) with a confirmed pre-existing diagnosis of CKD; and (4) at least one quantitative model performance metric was reported, such as the Area Under the Receiver Operating Characteristic

Curve (AUROC/AUC), concordance index (C-index), Matthews Correlation Coefficient (MCC), sensitivity, or specificity.

Studies were excluded if they focused exclusively on: CKD screening or diagnosis in populations without a prior CKD diagnosis; acute kidney injury (AKI) as a distinct clinical entity from progressive CKD; post-renal transplant outcomes; non-quantitative or purely qualitative AI applications; or paediatric populations. Studies reporting only internal validation without any form of external, geographic, or temporal validation were retained given the early developmental stage of the field, but this limitation is acknowledged in the synthesis.

2.4 Data Extraction and Synthesis

For each eligible study, the following data were extracted: first author and year of publication; country and healthcare setting; sample size; AI model types evaluated; comparator benchmark applied (where applicable); primary outcome definition; reported performance metrics; duration of follow-up; and principal findings regarding model accuracy and predictor importance. These data are presented in summary form in Table 2 to facilitate transparent cross-study comparison. A total of ten studies meeting all eligibility criteria were included in the final synthesis.

Given the narrative design, results were synthesised thematically rather than statistically. Four principal analytical themes guided the synthesis, identified inductively from the included literature: (1) the comparative accuracy of AI/ML models relative to established clinical benchmarks; (2) the longitudinal advantage conferred by time-series deep learning architectures over cross-sectional models; (3) the sensitivity-specificity imbalance and the class imbalance mechanisms underlying it; and (4) the role of Explainable AI frameworks in bridging the gap between algorithmic output and clinically actionable interpretation. These themes are addressed in the Results and Discussion sections that follow.

3. Results

The characteristics and predictive performance of the included studies are summarised in Table 2.

Table 2. Summary of included studies: AI/ML models, populations, and predictive performance metrics.

Author/Year	Country/n	AI Model(s)	Comparator	Primary Outcome	Performance (AUC/Sens/Spec)	Follow-up	Key Findings
Klinrisk (Tangri/Ferguson et al. 2022, 2024)	Canada / n=14,464 (trials); n=177,000+ (dev)	Random Forest (Survival)	KDIGO heatmap, refitted KFRE	eGFR decline or ESKD	AUC: 0.88 (3y) / Brier: 0.056	2-4 years	Superior to KDIGO and KFRE; maintains accuracy across SGLT2i treatment groups.
Zhu et al. (2023)	USA / n=82,667	LSTM RNN	Cox (static/dynamic), RF, LightGBM	Progression (Stages II/III to IV/V)	AUC: 0.967 / Sens: 0.76* / MCC: 0.72	1 year	LSTM RNN significantly outperforms baseline models by utilising longitudinal EHR data.
Isaza-Ruget et al. (2024)	Colombia / n=2,143	RSF, GBM, P-Cox	KFRE (as reference)	RRT, stage progression	C-index: 0.89 (RRT)	5 years	Time-to-event model effectively predicts the onset and timing of progression in established CKD.
Kanda et al. (2023)	Japan / n=26,906	Random Forest, GBDT, XGBoost	KFRE	ESKD or death	AUC: 0.93 / HR: 104.9	3 years	Developed a public web-based calculator; RF models were statistically superior to KFRE.
Lee et al. (2022)	Taiwan / n=11,661	GBDT, LGBM, RF, XGBoost	KFRE	ESRD (in Sepsis survivors)	AUC: 0.879 / Sens: 0.96 / Spec: 0.61	3.5 years	First model specific to sepsis survivors; outperformed conventional scoring systems for long-term

							ESRD.
Su et al. (2022)	Taiwan / n=858	RF, XGBoost, LR, SVM, GNB	N/A	Renal failure/ESRD	AUC: 0.99 (1y) / Sens: 0.96 / Spec: 0.91	1-5 years	SMOTE data balancing significantly improved performance; RF suitable for all CKD stages.
Bai et al. (2022)	China / n=748	LR, NB, RF, DT, KNN	KFRE	ESKD	AUC: 0.81 (RF) / Sens: 0.79 (LR) / Spec: 0.83	6.3 years	ML models showed greater sensitivity for screening than KFRE in the Chinese population.
Li et al. (2025)	China / n=451	XGBoost, RF, Ridge	N/A	ESRD (within 6 months)	AUC: 0.85 (Ext) / Acc: 0.79	25 weeks	XGBoost provided the most reliable risk estimates for short-term progression in stage 4 CKD.
Duong et al. (2025)	Taiwan / n=11,488	LGBM, RF, XGBoost, ANN	N/A	$\geq 25\%$ eGFR decline	AUC: 0.82 (LGBM) / Sens: 0.70	1 year	LightGBM emerged as top performer; identified baseline eGFR and BUN as top predictors.
Chouit et al. (2026)	UAE / n=491	XGBoost, RF	N/A	CKD prediction	AUC: 0.904 / Sens: 0.60 / Spec: 0.94	N/A	Framework uses SHAP and LIME to solve the "black-box" barrier in clinical environments.

***Note. Zhu et al. reported an F1-score of 0.76 and MCC of 0.72 for their primary metrics.**

Meta-analytical data reveal that AI and ML models achieve a pooled AUC ranging from 0.87 to 0.89, with strong pooled specificity (0.87–0.92) and lower pooled sensitivity (0.43–0.68) (Lei et al., 2022; Pan & Tong, 2024). Among specific architectures, RNN models utilising longitudinal EHR data attained AUROC scores of 0.967 for predicting CKD progression from stages II/III to IV/V within one year, while deep ensemble models combining CNN and LSTM-BLSTM units reached accuracy levels as high as 0.993 for short-term risk detection (Saif et al., 2023; Zhu et al., 2023). Tree-based methods demonstrated robust performance across diverse cohorts: the Klinrisk random forest model achieved an AUC of 0.88 at 3 years in external validation within the CANVAS/CREDENCE trial populations, and Random Forest classifiers integrated with SMOTE reached AUC values of 0.96 for early-stage and 0.97 for advanced-stage CKD progression prediction (Su et al., 2022; Tangri et al., 2024). Feature importance analyses across included studies consistently identified baseline eGFR, proteinuria (UPCR/UACR), and haemoglobin as the most frequently reported predictors of renal decline (Isaza-Ruget et al., 2024; Khalid et al., 2024; Su et al., 2022). Lipid-related markers, including apolipoprotein A1 and HDL-C, were identified as significant predictors in studies focused on advanced-stage CKD, with SHAP analyses indicating a protective directional effect consistent with the reverse epidemiology observed in this population (Chouit et al., 2026; Su et al., 2022). The application of XAI frameworks, including SHAP and LIME, was reported in 6 studies, providing variable importance rankings and patient-specific risk explanations (Chouit et al., 2026; Duong et al., 2025; Lee et al., 2022; Su et al., 2022; Ye et al., 2026).

4. Discussion

The collective evidence synthesised from studies published between 2022 and 2026 confirms that AI and ML models significantly outperform established clinical benchmarks – most notably the Kidney Failure Risk Equation (KFRE) and KDIGO risk heatmaps – in identifying patients at risk of CKD progression across a broad range of international cohorts and clinical settings (Lei et al., 2022; Pan & Tong, 2024; Tangri et al., 2024). This superiority is particularly pronounced in early-stage CKD (Stages 1–3), where conventional prognostic tools frequently underperform due to the inherently low frequency of kidney failure events within these populations, which limits both the statistical power of equations derived from predominantly late-stage

cohorts and their discriminative resolution when applied upstream in the disease course (Ferguson et al., 2022; Tangri et al., 2024).

The explanation for this advantage is both architectural and statistical. Classical risk instruments such as KFRE were calibrated on datasets enriched with advanced-stage patients and rely on a small, fixed set of predictors – typically age, sex, eGFR, and urinary albumin-to-creatinine ratio – combined according to parametric, log-linear assumptions. This structural simplicity, while appropriate for rapid bedside use, is inherently incapable of capturing the high-dimensional, non-linear interactions between the numerous biochemical, clinical, and demographic variables that collectively determine the trajectory of an individual patient's kidney function. AI models impose no such linearity constraints. Ensemble methods such as Random Forest and XGBoost are specifically designed to model complex variable interactions, accommodate non-normal distributions, and handle missing clinical data – a pervasive challenge in real-world EHR environments where laboratory tests are ordered inconsistently across patients and clinical episodes (Pan & Tong, 2024; Su et al., 2022). This combination of architectural flexibility and practical robustness to data imperfections explains the consistently superior AUC values observed across diverse global cohorts – ranging from 0.81 in Chinese patient populations (Bai et al., 2022) to 0.93 in large Japanese registry datasets (Kanda et al., 2023) – largely irrespective of the specific healthcare system or EHR data structure in which they were trained.

A critically important dimension of this advantage is the capacity to incorporate longitudinal data. The transition from cross-sectional "snapshot" risk assessment to the analysis of temporal disease trajectories, enabled by Recurrent Neural Network (RNN) architectures with Long Short-Term Memory (LSTM) units, represents a conceptually significant advance in CKD prognostication (Saif et al., 2023; Zhu et al., 2023). The biological rationale for this approach is compelling: CKD progression is not a static phenomenon adequately captured by a single set of laboratory measurements, but rather a dynamic process characterised by a rate and pattern of kidney function decline that varies over months and years in response to disease activity, intercurrent illness, treatment adherence, and pharmacological interventions. A patient's eGFR recorded at a single clinic visit conveys substantially less prognostic information than the slope, variability, and temporal shape of their eGFR trajectory recorded across multiple encounters. LSTM networks are architecturally designed to exploit precisely this type of sequential dependency: gated memory cells allow the model to retain and weight information from distant prior time-steps, enabling recognition that a sustained, monotone downward eGFR trend carries fundamentally different risk implications from a transient dip followed by stabilisation – a distinction that is clinically important but invisible to any cross-sectional model. The empirical evidence supports this theoretical argument convincingly: LSTM RNN models trained on longitudinal EHR data achieved AUROC scores of 0.957 to 0.967 for predicting one-year CKD stage progression, substantially exceeding comparator static models including dynamic Cox regression, cross-sectional Random Forest, and LightGBM applied to single time-points (Zhu et al., 2023). This performance gap carries direct practical implications: it suggests that the routine collection and long-term preservation of sequential laboratory data within EHR systems is not merely administratively convenient but is a genuine prerequisite for deploying the most powerful category of predictive tools currently available to nephrologists.

The emergence of Explainable AI (XAI) frameworks – most prominently SHAP and LIME – addresses one of the most consequential barriers to real-world clinical adoption: the opacity inherent in complex algorithmic decision-making (Chouit et al., 2026; Ye et al., 2026). The "black-box" problem in clinical AI extends well beyond academic discomfort with opacity; it carries concrete regulatory, professional, and ethical consequences. Clinicians and healthcare institutions bear legal and professional accountability for patient care decisions, and a risk score generated by a neural network operating without an accompanying explanation of how that score was derived provides insufficient grounds for altering treatment intensity, scheduling earlier nephrology review, or initiating pre-emptive dialysis planning. XAI methods resolve this barrier by attributing a quantifiable, directional contribution to each input variable for each individual prediction. In clinical practice, a SHAP output accompanying a high-risk classification would indicate to a nephrologist, for a specific patient, which variables are principally driving the model's risk estimate – for example, a sharply declining eGFR slope, elevated proteinuria, and worsening anaemia – information that is simultaneously algorithmically derived and clinically interpretable within the framework of recognised renal pathophysiology. The practical utility of this transparency is considerable: it enables the clinician to assess whether the model's reasoning is coherent, to integrate the AI output with findings from the clinical consultation, and to communicate the basis of risk-based management decisions to the patient in an intelligible manner.

Feature importance analyses conducted using XAI methods across the reviewed studies consistently identified three dominant predictors of renal decline: baseline eGFR trajectory, proteinuria measured as either UPCR or UACR, and haemoglobin concentration (Isaza-Ruget et al., 2024; Khalid et al., 2024; Su et al., 2022). The biological coherence of these rankings reinforces confidence in the models' validity. Progressive eGFR decline reflects the cumulative, irreversible loss of functional nephron mass; sustained proteinuria drives tubulointerstitial fibrosis through direct tubular toxicity from filtered protein – a mechanism central to the pathogenesis of progressive CKD regardless of aetiology; and declining haemoglobin, a consequence of reduced erythropoietin synthesis by the diseased kidney combined with uraemic inhibition of erythropoiesis, serves as a sensitive marker of advancing tubulointerstitial damage (Ali et al., 2020; Chaudhuri et al., 2021; Pethő et al., 2024). The correspondence between algorithmically derived feature rankings and these well-characterised pathomechanisms constitutes an important form of face validity, indicating that the AI systems are detecting genuine biological signals rather than exploiting spurious statistical correlations in the training data.

A more nuanced finding concerns the role of lipid-related markers, particularly apolipoprotein A1 and HDL cholesterol, as significant predictors in models trained on advanced-stage CKD populations. Rather than increasing risk, elevated levels of these markers were associated with a protective directional effect on CKD progression in SHAP analyses (Chouit et al., 2026; Su et al., 2022). This finding reflects the well-documented phenomenon of reverse epidemiology in end-stage renal disease: whereas elevated HDL cholesterol confers cardiovascular protection in the general population, in advanced CKD low lipid levels frequently signal protein-energy wasting, chronic malnutrition, and systemic inflammation – a metabolic phenotype associated with higher mortality that is pathophysiologically distinct from genuinely low lipid levels in a metabolically healthy individual (Li et al., 2025; Ye et al., 2026). The capacity of AI models guided by XAI tools to detect and correctly interpret this paradoxical biological relationship – consistent with what experienced nephrologists encounter in advanced CKD clinics – is a further demonstration of the clinical coherence these methods can achieve when appropriately developed and validated.

Despite these advances, several persistent limitations must be addressed before AI-based CKD risk stratification can achieve the widespread clinical integration its performance metrics suggest it merits. The most consistently reported limitation across independent meta-analyses is the pronounced asymmetry between model specificity and sensitivity. Pooled estimates indicate that while AI models achieve strong specificity of 0.87–0.92 – correctly identifying patients unlikely to progress – sensitivity is substantially lower, typically ranging from 0.43 to 0.68, meaning that between one third and more than half of patients who will ultimately progress to kidney failure are incorrectly classified as low risk (Lei et al., 2022; Pan & Tong, 2024). The clinical implications of this false-negative burden are serious. Patients classified as low risk who are progressing will not receive timely nephrology review, optimisation of renoprotective therapy, or preparation for renal replacement therapy, and the window of opportunity during which interventions most effectively slow disease progression may be missed irreversibly.

The primary structural cause of this sensitivity deficit is class imbalance in training datasets: CKD progression to ESKD is, by definition, a relatively infrequent event within any clinical cohort, particularly among early-stage patients, and models trained on datasets where the majority of outcome labels are negative develop a systematic tendency to assign low-risk classifications by default. Several methodological strategies have been applied to mitigate this effect. SMOTE generates synthetic examples of the minority class by interpolating between existing positive instances in multidimensional feature space, rebalancing the training dataset and improving model sensitivity without discarding majority-class observations (Chouit et al., 2026; Su et al., 2022). However, SMOTE introduces its own risks: interpolated synthetic examples may not faithfully represent the genuine biological heterogeneity of the progression phenotype, and aggressive oversampling in small datasets can lead to overfitting that inflates validation performance without improving clinical utility. Cost-sensitive learning approaches – which impose higher misclassification penalties on false negatives during model training – offer an alternative that does not require data augmentation. Future studies should systematically evaluate and report both approaches, and the field would benefit from clearer guidance on reporting standards: sensitivity, specificity, and positive predictive value should be reported alongside AUC as co-primary performance metrics, as AUC alone can conceal clinically significant false-negative rates in imbalanced datasets.

A further obstacle to the integration and comparability of evidence is the substantial heterogeneity in outcome definitions, predictor sets, follow-up periods, and performance metrics across studies. Primary outcomes range from a $\geq 25\%$ eGFR decline over one year (Duong et al., 2025) to a $\geq 40\%$ decline over three

years (Tangri et al., 2024), categorical CKD stage advancement, or composite endpoints incorporating both ESKD and all-cause mortality. Predictor sets vary from models relying exclusively on routinely available biochemical and demographic variables to those incorporating specialist biomarkers unavailable in many healthcare settings. Performance metrics span AUROC, C-index, MCC, and Brier scores, each capturing different aspects of model calibration and discrimination. This methodological fragmentation means that headline performance figures across studies are rarely directly comparable, and the field lacks the standardised reporting infrastructure that would allow clinicians, health technology assessors, and regulators to make informed comparisons between competing models (Khalid et al., 2024; Pan & Tong, 2024). Developing consensus-driven minimum reporting standards for AI-based CKD prognostic studies – analogous to the TRIPOD guidelines for conventional prediction models – represents a priority for methodological work in the coming years.

The pharmacological context in which CKD risk models are applied also demands greater attention. The widespread adoption of SGLT2 inhibitors – agents with demonstrated renoprotective effects established independently of their glucose-lowering properties in the CANVAS Program and CREDENCE trial – has substantially altered the natural trajectory of CKD progression in treated populations relative to the historical cohorts on which most currently available models were trained. A model calibrated on pre-SGLT2i data may systematically overestimate progression risk in patients receiving these agents, generating false-positive alerts that erode clinical trust and contribute to the alert fatigue that already limits the effectiveness of EHR-embedded clinical decision support tools. The external validation of the Klinrisk Random Forest model within the CANVAS and CREDENCE trial populations demonstrated that its discriminative performance was preserved across both SGLT2i-treated and untreated groups, establishing an important proof of principle that AI models can be designed to maintain accuracy within contemporary pharmacological management contexts (Tangri et al., 2024). However, this remains an exception in the published literature, and a systematic programme of re-validation of existing models in cohorts receiving current standard-of-care nephrology management – including SGLT2 inhibitors, finerenone, and renin-angiotensin system blockade – is necessary before these tools can be relied upon in routine clinical practice.

Lastly, the pathway from a research-validated model to an embedded clinical decision support tool involves implementation challenges that extend well beyond statistical performance. Successful clinical integration requires bidirectional EHR data pipelines capable of supplying model inputs in real time, computational infrastructure supporting near-real-time risk computation at the population scale, and CDSS interface design that presents risk information in a format that is actionable, concise, and does not overwhelm clinicians with alerts of variable clinical significance. The publicly accessible web-based risk calculator represents a pragmatic and immediately deployable intermediate step that democratises access to ML-based risk estimation without requiring institutional EHR integration – a meaningful contribution to bridging the gap between research and practice, particularly in settings where embedded CDSS implementation is not currently feasible (Kanda et al., 2023).

4.1 Global and Regional Considerations

The clinical promise of AI-driven CKD risk stratification carries particular significance for low- and middle-income countries (LMICs), where the burden of undiagnosed and untreated CKD is both disproportionately large and disproportionately underappreciated within global health policy. Awareness gaps remain striking even in settings where screening infrastructure exists: in many countries, fewer than 15% of patients living with CKD are aware of their diagnosis, and treatment gaps for identified patients range from 68.7% to 97.8% across different national contexts (Betts et al., 2021). This diagnostic invisibility is compounded by the asymptomatic nature of early-stage CKD, which ensures that the majority of affected individuals remain undetected until substantial, often irreversible nephron loss has already occurred. In principle, AI-based risk stratification applied to routinely collected clinical data – blood pressure, basic biochemistry, urinalysis – could identify high-risk individuals years before they would otherwise enter clinical awareness, providing the temporal window necessary for interventions capable of slowing or arresting progression.

In practice, however, the same infrastructural constraints that motivate the search for AI-based solutions in resource-limited settings are precisely those that obstruct their deployment. The highest-performing AI architectures reviewed in this paper – particularly LSTM RNN models – require comprehensive longitudinal EHR data recorded across multiple clinical encounters and extended follow-up

periods. Many LMIC health systems lack both the digital infrastructure to generate such data at scale and the interoperability standards that would allow records generated at different facilities and by different providers to be aggregated into the continuous longitudinal profiles these models require (Saran et al., 2020). This creates a structural paradox: the populations that would benefit most from AI-based risk stratification are often the least equipped to generate the data needed to train and validate such models locally. Consequently, models validated in Taiwanese, Canadian, or other high-income cohorts cannot be assumed to generalise to LMIC populations, where differences in disease aetiology, comorbidity patterns, and laboratory testing availability may substantially alter model performance. Algorithmic bias mitigation strategies and federated learning approaches – which allow models to be trained across institutions without centralising sensitive patient data – represent promising but still largely unproven solutions to this generalisability gap.

The economic case for AI-based CKD risk stratification, where implementation is technically feasible, appears potentially compelling. Modelling undertaken in the United States context estimated that AI-driven risk stratification for early-stage diabetic kidney disease could generate undiscounted savings of approximately \$1.052 billion over five years per 100,000 patients evaluated, driven primarily by slowing disease progression and thereby deferring the onset of the most resource-intensive treatment phases – haemodialysis and transplantation (Betts et al., 2021). Whether comparable returns on investment are achievable in resource-constrained healthcare systems is an open empirical question. Health economic modelling calibrated to LMIC cost structures, healthcare utilisation patterns, drug pricing, and willingness-to-pay thresholds is a significant gap in the current literature, and such analyses should be considered a priority for future research to ensure that the economic case for AI adoption is evaluated on terms relevant to the settings where the potential public health impact is greatest. Importantly, any such modelling must account for the full cost of implementation, including upfront investment in digital infrastructure, data governance frameworks, staff training, and the ongoing costs of model maintenance and recalibration – expenditures that are not incurred by the research groups that produce the published models but fall entirely on the health systems that deploy them.

The integration of AI risk tools into EHR platforms and clinical decision support infrastructure represents the horizon toward which the field is moving, with the potential to deliver automated, real-time risk stratification at the population scale, prompting earlier nephrology referral and enabling targeted pharmacological intervention before irreversible nephron loss occurs. Realising this potential equitably – across income levels, healthcare systems, and ethnic populations – will require sustained collaboration between AI scientists, nephrologists, health informaticians, health economists, ethicists, patients, and policymakers, as well as a deliberate commitment from the research community to prioritise external validation in underrepresented populations rather than continued refinement of models in the high-resource settings where they are easiest to develop (Khalid et al., 2024; Pan & Tong, 2024).

5. Conclusions

In conclusion, the current body of research demonstrates that AI and ML models represent a transformative auxiliary tool for predicting CKD progression with a high degree of diagnostic accuracy. These advanced analytical approaches, particularly tree-based ensembles (Random Forest, XGBoost) and deep learning architectures (LSTM RNN), systematically outperform traditional clinical standards by leveraging routinely collected longitudinal laboratory data. By facilitating earlier risk detection and more precise stratification, especially in the historically difficult-to-predict early stages, AI empowers clinicians to move CKD management from a reactive to a proactive model of care, potentially delaying or preventing the lifetime occurrence of kidney failure. The development of web-based risk-prediction systems further bridges the gap between research and real-world clinical practice, providing accessible tools for busy medical environments (Kanda et al., 2023). However, achieving global clinical adoption requires addressing persistent challenges, including the standardisation of predictor criteria, the optimisation of model sensitivity through improved data-balancing methods, and the establishment of rigorous prospective validation frameworks. As Explainable AI (XAI) continues to bridge the transparency gap and build clinical trust by aligning model logic with renal biology, the future focus must remain on the seamless integration of these models into hospital workflows. This shift toward precision nephrology holds the promise of significantly reducing the global socioeconomic burden of end-stage renal disease and substantially improving long-term patient outcomes.

Acknowledgements: This research has not received any administrative or technical support.

Conflicts of Interest: The authors declare no conflict of interest.

AI Usage Disclosure: The authors utilised AI to evaluate clinical narratives for logical fallacies and to enhance the manuscript's academic language for better clarity. Following the use of this tool, the authors rigorously reviewed and revised the text, assuming ultimate responsibility for the final published content.

REFERENCES

1. Ali, I., Chinnadurai, R., Ibrahim, S. T., Green, D., & Kalra, P. A. (2020). Predictive factors of rapid linear renal progression and mortality in patients with chronic kidney disease. *BMC Nephrology*, 21(1), 345. <https://doi.org/10.1186/s12882-020-01982-8>
2. Bai, Q., Su, C., Tang, W., & Li, Y. (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. *Scientific Reports*, 12(1), 8377. <https://doi.org/10.1038/s41598-022-12316-z>
3. Betts, K., Song, J., Faust, E., Yang, K., Du, Y., Kong, S., & Singh, R. (2021). Medical costs for managing chronic kidney disease and related complications in patients with chronic kidney disease and type 2 diabetes. *The American Journal of Managed Care*, 27(Suppl 20), S369–S374. <https://doi.org/10.37765/ajmc.2021.88807>
4. Bikbov, B., Purcell, C. A., Levey, A. S., Smith, M., Abdoli, A., Abebe, M., Adebayo, O. M., Afarideh, M., Agarwal, S. K., Agudelo-Botero, M., Ahmadian, E., Al-Aly, Z., Alipour, V., Almasi-Hashiani, A., Al-Raddadi, R. M., Alvis-Guzman, N., Amini, S., Andrei, T., Andrei, C. L., ... Vos, T. (2020). Global, regional, and national burden of chronic kidney disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 395(10225), 709–733. [https://doi.org/10.1016/S0140-6736\(20\)30045-3](https://doi.org/10.1016/S0140-6736(20)30045-3)
5. Borg, R., Carlson, N., Søndergaard, J., & Persson, F. (2023). The growing challenge of chronic kidney disease: An overview of current knowledge. *International Journal of Nephrology*, 2023, 1–8. <https://doi.org/10.1155/2023/9609266>
6. Chaudhuri, S., Long, A., Zhang, H., Monaghan, C., Larkin, J. W., Kotanko, P., Kalaskar, S., Kooman, J. P., Van Der Sande, F. M., Maddux, F. W., & Usvyat, L. A. (2021). Artificial intelligence enabled applications in kidney disease. *Seminars in Dialysis*, 34(1), 5–16. <https://doi.org/10.1111/sdi.12915>
7. Chouit, E. M., Rachdi, M., Bellafkih, M., & Raouyane, B. (2026). Interpretable machine learning for chronic kidney disease prediction: Insights from SHAP and LIME analyses. *PLOS One*, 21(2), e0343205. <https://doi.org/10.1371/journal.pone.0343205>
8. Duong, T. T., Nguyen, M. T., Liao, C. T., Le, N. H., Phan, T. P., Huang, C. W., Hsu, J. C., & Nguyen, A. P. A. (2025). Personalized prediction of chronic kidney disease progression in patients with chronic kidney disease stages 3–5: A multicenter study using the machine learning approach. In m. S. Househ, z. U. A. Tariq, m. Al-Zubaidi, u. Shah, & e. Huesing (eds), *Studies in Health Technology and Informatics*. IOS Press. <https://doi.org/10.3233/SHTI251022>
9. Ferguson, T., Ravani, P., Sood, M. M., Clarke, A., Komenda, P., Rigatto, C., & Tangri, N. (2022). Development and external validation of a machine learning model for progression of CKD. *Kidney International Reports*, 7(8), 1772–1781. <https://doi.org/10.1016/j.ekir.2022.05.004>
10. Isaza-Ruget, M. A., Yomayusa, N., González, C. A., H., C. A., De Oro V., F. A., Cely, A., Murcia, J., Gonzalez-Velez, A., Robayo, A., Colmenares-Mejía, C. C., Castillo, A., & Conde, M. I. (2024). Predicting chronic kidney disease progression with artificial intelligence. *BMC Nephrology*, 25(1), 148. <https://doi.org/10.1186/s12882-024-03545-7>
11. Kanda, E., Epureanu, B. I., Adachi, T., & Kashiara, N. (2023). Machine-learning-based web system for the prediction of chronic kidney disease progression and mortality. *PLOS Digital Health*, 2(1), e0000188. <https://doi.org/10.1371/journal.pdig.0000188>
12. Khalid, F., Alsadoun, L., Khilji, F., Mushtaq, M., Eze-odurukwe, A., Mushtaq, M. M., Ali, H., Farman, R. O., Ali, S. M., Fatima, R., & Bokhari, S. F. H. (2024). Predicting the progression of chronic kidney disease: A systematic review of artificial intelligence and machine learning approaches. *Cureus*. <https://doi.org/10.7759/cureus.60145>
13. Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: An update 2022. *Kidney International Supplements*, 12(1), 7–11. <https://doi.org/10.1016/j.kisu.2021.11.003>

14. Lee, K.-H., Chu, Y.-C., Tsai, M.-T., Tseng, W.-C., Lin, Y.-P., Ou, S.-M., & Tarng, D.-C. (2022). Artificial intelligence for risk prediction of End-Stage renal disease in sepsis survivors with chronic kidney disease. *Biomedicines*, 10(3), 546. <https://doi.org/10.3390/biomedicines10030546>
15. Lei, N., Zhang, X., Wei, M., Lao, B., Xu, X., Zhang, M., Chen, H., Xu, Y., Xia, B., Zhang, D., Dong, C., Fu, L., Tang, F., & Wu, Y. (2022). Machine learning algorithms' accuracy in predicting kidney disease progression: A systematic review and meta-analysis. *BMC Medical Informatics and Decision Making*, 22(1), 205. <https://doi.org/10.1186/s12911-022-01951-1>
16. Li, J., Du, X., Zhang, R., Li, X., Xu, J., Song, X., Zhao, Y., Liu, L., Xu, G., & Fan, Y. (2025). Machine learning models for predicting short-term progression in patients with stage 4 chronic kidney disease: A multi-center validation study. *Scientific Reports*, 15(1), 39285. <https://doi.org/10.1038/s41598-025-23037-4>
17. Mark, P. B., Stafford, L. K., Grams, M., Lees, J. S., Ong, K. L., & Coresh, J. (2025). Global, regional, and national burden of CKD in adults: Updated systematic analyses of the Global Burden of Disease Study: FR-OR093. *Journal of the American Society of Nephrology*, 36(10S). <https://doi.org/10.1681/ASN.2025dzpbxjpn>
18. Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V., Abraham, J., Ackerman, I., Aggarwal, R., Ahn, S. Y., Ali, M. K., AlMazroa, M. A., Alvarado, M., Anderson, H. R., ... Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380(9859), 2197–2223. [https://doi.org/10.1016/S0140-6736\(12\)61689-4](https://doi.org/10.1016/S0140-6736(12)61689-4)
19. Pan, Q., & Tong, M. (2024). Artificial intelligence in predicting chronic kidney disease prognosis. A systematic review and meta-analysis. *Renal Failure*, 46(2), 2435483. <https://doi.org/10.1080/0886022X.2024.2435483>
20. Pethő, Á. G., Tapolyai, M., Csongrádi, É., & Orosz, P. (2024). Management of chronic kidney disease: The current novel and forgotten therapies. *Journal of Clinical & Translational Endocrinology*, 36, 100354. <https://doi.org/10.1016/j.jcte.2024.100354>
21. Sabanayagam, C., Banu, R., Lim, C., Tham, Y. C., Cheng, C.-Y., Tan, G., Ekinici, E., Sheng, B., McKay, G., Shaw, J. E., Matsushita, K., Tangri, N., Choo, J., & Wong, T. Y. (2025). Artificial intelligence in chronic kidney disease management: A scoping review. *Theranostics*, 15(10), 4566–4578. <https://doi.org/10.7150/thno.108552>
22. Saif, D., Sarhan, A. M., & Elshennawy, N. M. (2023). Deep-kidney: An effective deep learning framework for chronic kidney disease prediction. *Health Information Science and Systems*, 12(1), 3. <https://doi.org/10.1007/s13755-023-00261-8>
23. Saran, R., Robinson, B., Abbott, K. C., Bragg-Gresham, J., Chen, X., Gipson, D., Gu, H., Hirth, R. A., Hutton, D., Jin, Y., Kapke, A., Kurtz, V., Li, Y., McCullough, K., Modi, Z., Morgenstern, H., Mukhopadhyay, P., Pearson, J., Pisoni, R., ... Shahinian, V. (2020). US Renal Data System 2019 annual data report: Epidemiology of kidney disease in the United States. *American Journal of Kidney Diseases*, 75(1), A6–A7. <https://doi.org/10.1053/j.ajkd.2019.09.003>
24. Stevens, P. E., Ahmed, S. B., Carrero, J. J., Foster, B., Francis, A., Hall, R. K., Herrington, W. G., Hill, G., Inker, L. A., Kazancioğlu, R., Lamb, E., Lin, P., Madero, M., McIntyre, N., Morrow, K., Roberts, G., Sabanayagam, D., Schaeffner, E., Shlipak, M., ... Levin, A. (2024). KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International*, 105(4), S117–S314. <https://doi.org/10.1016/j.kint.2023.10.018>
25. Su, C.-T., Chang, Y.-P., Ku, Y.-T., & Lin, C.-M. (2022). Machine learning models for the prediction of renal failure in chronic kidney disease: A retrospective cohort study. *Diagnostics*, 12(10), 2454. <https://doi.org/10.3390/diagnostics12102454>
26. Tangri, N., Ferguson, T. W., Bamforth, R. J., Leon, S. J., Arnott, C., Mahaffey, K. W., Kotwal, S., Heerspink, H. J. L., Perkovic, V., Fletcher, R. A., & Neuen, B. L. (2024). Machine learning for prediction of chronic kidney disease progression: Validation of the klinrisk model in the CANVAS program and CREDENCE trial. *Diabetes, Obesity and Metabolism*, 26(8), 3371–3380. <https://doi.org/10.1111/dom.15678>
27. Ye, B., Zhang, X., Zhu, W., Xiao, Z., & Huang, S. (2026). Interpretable machine learning model based on routine metabolic laboratory indices to identify advanced chronic kidney disease. *Frontiers in Endocrinology*, 17, 1776419. <https://doi.org/10.3389/fendo.2026.1776419>
28. Zhu, Y., Bi, D., Saunders, M., & Ji, Y. (2023). Prediction of chronic kidney disease progression using recurrent neural network and electronic health records. *Scientific Reports*, 13(1), 22091. <https://doi.org/10.1038/s41598-023-49271-2>