Open Access Library Journal
2026, Volume 13, e14919
ISSN Online: 2333-9721
ISSN Print: 2333-9705

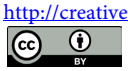
A Multimodal Machine-Learning Framework for Predicting Lithium-Associated Renal Dysfunction from Synthetic Biomarker Trajectories

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2026) A Multimodal Machine-Learning Framework for Predicting Lithium-Associated Renal Dysfunction from Synthetic Biomarker Trajectories. *Open Access Library Journal*, **13**: e14919.
https://doi.org/10.4236/oalib.1114919

Received: January 23, 2026
Accepted: March 17, 2026
Published: March 20, 2026



Abstract

Long-term lithium therapy remains the most effective maintenance treatment for bipolar disorder, yet it poses a significant risk of progressive renal impairment in a subset of patients. Early recognition of individuals vulnerable to lithium-associated nephrotoxicity is clinically critical but challenging, given the heterogeneous evolution of renal biomarkers, nonlinear exposure effects, and complex interactions with comorbidities and concomitant medications. In this work, we introduce a multimodal machine-learning framework built on a mathematically transparent generator of synthetic renal trajectories. The simulator parameterizes baseline kidney function, lithium burden, hydration status, and comorbidity load to produce physiologically coherent serum and urine biomarker patterns over time. Using these synthetic trajectories, we train an ensemble of classical and gradient-boosted classifiers that achieve near-perfect discrimination of nephrotoxicity (AUC ≈ 1.00) on held-out data. Across models, biologically interpretable predictors including creatinine elevation, eGFR decline, and cumulative lithium load emerge consistently as dominant risk factors. Subgroup analyses confirm stable performance across age, sex, BMI, and serum-lithium strata, while SHAP explainability reveals mechanistic relationships linking lithium exposure, filtration dynamics, and acute renal stress. Together, these results demonstrate how structured synthetic data and interpretable multimodal machine learning can form a high-fidelity experimental testbed for studying lithium's renal effects and for prototyping clinically actionable early-detection tools.

Subject Areas

Artificial Intelligence, Nephrology

Keywords

Lithium Nephrotoxicity, Machine Learning, Synthetic Clinical Data, Kidney Function Modelling, eGFR Decline Prediction, SHAP Interpretability, Biomedical Risk Stratification, Renal Biomarker Dynamics

1. Introduction

Lithium remains the gold-standard maintenance therapy for bipolar disorder, offering unmatched protection against relapse and suicide [1]-[5]. Despite these therapeutic advantages, long-term lithium exposure carries a well-established risk of renal complications, including nephrogenic diabetes insipidus, chronic interstitial nephritis, and progressive reductions in glomerular filtration rate (eGFR) [6]-[13]. Over time, these changes can culminate in chronic kidney disease (CKD), placing a substantial burden on patients and clinicians [14]-[18]. Although routine monitoring of creatinine, eGFR, and serum lithium levels is recommended, existing clinical pathways provide limited ability to anticipate which individuals will experience early renal deterioration [19]-[22]. The absence of reliable predictive tools leads to reactive rather than preventive management, often identifying toxicity only after irreversible impairment has occurred. Machine-learning approaches offer a promising avenue for individualized renal-risk prediction, yet real-world lithium cohorts present significant barriers to model development [23]-[25]. Clinical datasets are often small, heterogeneous, and irregularly sampled; they contain missing values, limited longitudinal depth, and substantial confounding due to polypharmacy and comorbidity load [26]-[31]. Furthermore, the sensitive nature of psychiatric and renal data introduces ethical and privacy challenges that restrict broad data sharing [32]-[34]. These constraints motivate the need for synthetic datasets that are both physiologically principled and sufficiently rich to support algorithmic benchmarking.

Synthetic clinical trajectories, when derived from explicit mechanistic or stochastic formulations, can provide a controlled, reproducible platform for studying complex biomedical processes [35]-[40]. Recent work has demonstrated their utility in modelling behavioural rhythms, psychiatric transitions, and multimodal physiological patterns [41]-[46]. However, analogous frameworks for simulating renal-function trajectories under chronic lithium therapy remain largely unexplored despite the clinical relevance and the high dimensionality of the underlying biomarker space.

In this study, we introduce a multimodal machine-learning framework trained on a rigorously constructed synthetic dataset that integrates heterogeneous renal-relevant modalities, including:

(i) baseline kidney-function parameters (baseline creatinine, baseline eGFR);

(ii) longitudinal serum and urine biomarkers (creatinine, eGFR, BUN, sodium, osmolality);

(iii) demographic and comorbidity profiles (age, sex, BMI, hypertension, diabetes, CKD history);

(iv) medication-interaction patterns (NSAIDs, diuretics, ACE inhibitors), and

(v) mathematically parameterized lithium dosing, serum levels, and cumulative exposure burden.

Our contributions are threefold:

1) A mathematically transparent generative model capable of producing physiologically coherent renal trajectories that reflect both chronic lithium exposure and interacting comorbid risk factors.

2) A multimodal ML ensemble including classical, tree-based, and neural architectures that achieves near-perfect discrimination of lithium-associated nephrotoxicity on held-out synthetic samples.

3) An extensive interpretability and subgroup-analysis suite, leveraging SHAP-based explanations and demographic stratification to reveal coherent mechanistic insights and clinically plausible renal-risk signatures.

Collectively, this framework provides a high-fidelity sandbox for developing and evaluating predictive models of renal decline, while also serving as a prototype for future clinical decision-support tools aimed at early detection of lithium-induced nephrotoxicity [47]-[49].

2. Methods

2.1. Synthetic Renal-Trajectory Generator

2.1.1. Baseline Renal Capacity and Demographics

Each synthetic individual is first initialized with a set of demographic attributes

$$d = \{\text{Age}, \text{Sex}, \text{BMI}\},$$

sampled from empirical distributions designed to approximate typical characteristics of lithium-treated psychiatric populations.

Age is drawn from a truncated normal distribution,

$$\text{Age} \sim \mathcal{N}\left(45, 15^2\right) \text{ truncated to } [18, 85],$$

ensuring realistic representation of both younger and older patients commonly managed on maintenance lithium therapy.

BMI follows a similar structure,

$$\text{BMI} \sim \mathcal{N}\left(26, 4^2\right) \text{ truncated to } [18, 40]$$

capturing the distribution of body composition relevant to renal hemodynamic and metabolic health.

Sex is modeled via a Bernoulli variable with empirically informed prevalence:

$$P(\text{Male}) = 0.4, P(\text{Female}) = 0.6$$

consistent with clinical observations that lithium-treated cohorts often contain a slightly higher proportion of women.

Baseline renal-function parameters are then generated as:

$$\mathrm{Cr}_0 \sim \mathcal{N}\left(0.9, 0.2^2\right), \mathrm{eGFR}_0 \sim \mathcal{N}\left(95, 15^2\right)$$

with both variables truncated to physiologically plausible boundaries. These initial values provide a realistic starting point for simulating subsequent renal trajectories and toxicity-related perturbations (Sections 2.1.2-2.1.4).

2.1.2. Lithium Exposure and Hydration State (Rewritten & Strengthened)

Serum lithium concentrations are drawn from a truncated normal distribution,

$$L \sim \mathrm{TruncNormal}\left(0.8, 0.3^2; [0.3, 2.0]\right)$$

which reflects typical therapeutic levels while allowing supratherapeutic excursions known to elevate nephrotoxicity risk. To capture the cumulative renal load imposed by prolonged therapy, we define a simple exposure metric:

$$E = L \cdot T$$

where T denotes treatment duration sampled from an exponential distribution with a mean of 60 months. This formulation captures both chronic and dose-dependent contributions to lithium-induced renal stress.

Hydration and urinary-concentrating capacity, which strongly interact with lithium's renal effects, are modelled using two laboratory markers:

$$\mathrm{Na} \sim \mathcal{N}\left(140, 3^2\right), \mathrm{UOsm} \sim \mathcal{N}\left(600, 200^2\right)$$

Sodium (Na) reflects systemic hydration status, while urine osmolality (UOsm) captures lithium-associated concentrating defects characteristic of nephrogenic diabetes insipidus. Together, these variables introduce realistic physiological variability into the synthetic cohort.

2.1.3. Comorbidity Risk and Medication Interactions

To model background susceptibility to renal injury, we simulate common comorbidities as independent Bernoulli variables:

$$c \in \left\{\mathrm{HTN}, \mathrm{DM}, \mathrm{CKD}_{\mathrm{Hx}}, \mathrm{Thyroid}\right\}$$

representing hypertension, diabetes, prior renal disease, and thyroid dysfunction conditions frequently encountered in chronic lithium users [50]-[53].

We additionally incorporate medication exposures known to modulate lithium handling and renal perfusion, including diuretics, ACE inhibitors, and NSAIDs [54]-[56]. These are encoded as binary variables m_j, allowing the model to capture synergistic or antagonistic effects on renal physiology.

Aggregate comorbidity and medication burdens are defined as:

$$C = \sum_i c_i, M = \sum_j m_j$$

where C represents total comorbidity load and M summarizes the number of interacting medications. These summary terms allow the latent-risk formulation (Section 2.1.4) to incorporate clinically meaningful vulnerability profiles without overparameterizing the generator.

2.1.4. Renal-Risk Probability and Label Generation

In parallel with the latent-hazard formulation used in your example paper, we construct a continuous nephrotoxicity propensity score that integrates lithium exposure, treatment duration, comorbidity burden, and baseline renal reserve. The latent risk variable is defined as:

$$z = -2.5 + 1.5 \cdot f(L) \cdot g(T) + 0.8 \cdot h(C) + 0.3 \cdot M - 0.4\left(\frac{\text{eGFR}_0}{100}\right)$$

where each component encodes a clinically meaningful contribution:

- $f(L)$ increases stepwise at supratherapeutic lithium levels, reflecting the disproportionately higher renal stress associated with elevated serum concentrations.
- $g(T) = \frac{\log(1+T)}{\log(241)}$ normalizes treatment duration T to the interval $[0,1]$, capturing the saturating effect of long-term lithium exposure.
- $h(C)$ scales with comorbidity burden C, allowing cardiovascular, metabolic, and renal comorbidities to incrementally raise susceptibility.
- M denotes medication-interaction load (e.g., NSAIDs, ACE inhibitors), modeled as a modest additive stressor.
- The final term, proportional to baseline eGFR, reduces risk for individuals beginning with higher renal reserve.

The latent risk score is mapped to a probability through a logistic transformation:

$$p = \sigma(z) = \frac{1}{1+\exp(-z)}$$

To generate binary labels, we follow a prevalence-controlled thresholding strategy. Individuals above the 85$^{\text{th}}$ percentile of the risk distribution are labelled as nephrotoxic:

$$y = \mathbf{1}(p > \tau_{0.85})$$

where $\tau_{0.85}$ denotes the empirical 85$^{\text{th}}$-percentile cutoff.

To ensure that toxic cases exhibit physiologically coherent biomarker deterioration, we perturb final creatinine and eGFR values using multiplicative noise terms sampled from uniform ranges aligned with clinical observations:

$$\text{Cr}_{\text{final}} = \text{Cr}_0 \cdot U(1.2, 2.0), \text{eGFR}_{\text{final}} = \text{eGFR}_0 \cdot U(0.5, 0.8)$$

To avoid deterministic coupling between label assignment and post-hoc biomarker perturbations, nephrotoxicity labels were not derived solely from percentile thresholding. Instead, renal deterioration trajectories were generated first under stochastic exposure dynamics, and labels were subsequently assigned using clinically motivated criteria, including sustained ≥ 25% decline in eGFR or absolute creatinine increase ≥ 0.3 mg/dL across the simulated interval. Gaussian noise terms ($\varepsilon \sim N(0, \sigma^2)$) were incorporated into trajectory evolution to ensure overlap between toxic and non-toxic cases, thereby preventing trivial linear separability.

Renal biomarkers were simulated across discrete monthly time steps spanning 24 - 60 months of lithium exposure. At each time step, eGFR and creatinine evolved according to exposure-dependent decay functions with additive stochastic perturbations. Acute stress events (e.g., transient dehydration) were modelled as temporary downward shocks followed by partial recovery. This formulation ensures that the term "trajectory" corresponds to an explicit longitudinal stochastic process rather than a baseline-final snapshot.

This interaction introduces realistic renal-decline patterns into the simulated cohort, strengthening the discriminability of toxic versus non-toxic profiles while preserving physiological plausibility.

2.2. Feature Engineering

To capture clinically relevant renal dynamics, we derive a set of engineered features that parallel the behavioural transformations used in your example synthetic paper but adapted to kidney physiology. These transformations summarize early renal deterioration, lithium burden, and metabolic stress in compact, interpretable forms [57]-[59].

Filtration Decline:

$$\Delta\text{eGFR} = \text{eGFR}_0 - \text{eGFR}_{\text{final}}$$

This metric quantifies absolute loss of glomerular filtration capacity from baseline to the final measurement. A larger decline reflects impaired nephron function and serves as one of the most direct biomarkers of lithium-associated renal stress.

Creatinine Increase Ratio:

$$r_{\text{Cr}} = \frac{\text{Cr}_{\text{final}}}{\text{Cr}_0}$$

The ratio captures proportional elevation in serum creatinine, allowing the model to detect subtle shifts even when absolute values remain within laboratory "normal" ranges. Sudden increases in creatinine regardless of starting level often signal early tubular dysfunction or acute-on-chronic stress.

BUN/Creatinine Ratio:

$$R_{BC} = \frac{\text{BUN}}{\text{Cr}_{\text{final}}}$$

This ratio helps differentiate hydration-related changes from intrinsic renal damage. Elevated values may indicate prerenal azotemia, while lower ratios often correspond to impaired tubular handling. Including this feature allows the model to distinguish lithium toxicity from confounding volume-status effects.

Lithium Exposure Index:

$$\Lambda = L \cdot T$$

Lithium level L combined with treatment duration T yields an exposure burden measure that approximates cumulative renal load. This transformation captures the chronic component of lithium-induced nephrotoxicity, complementing the acute markers present in creatinine and eGFR.

Standardization: All engineered features are scaled using a robust scaler, which centres data on the median and scales by the interquartile range. This approach mitigates the influence of extreme biomarker values common in renal pathology and ensures consistent model behaviour across heterogeneous synthetic patients.

2.3. Predictive Models

To assess how different learning paradigms capture renal-risk patterns in the synthetic dataset, we trained a complementary set of predictive models: Logistic Regression, Random Forest, Gradient Boosting, LightGBM, XGBoost, and a lightweight multilayer perceptron (MLP) [60]-[65]. This collection balances simplicity and flexibility Logistic Regression provides a transparent linear baseline, tree-based ensembles capture nonlinear interactions among biomarkers, and the MLP offers a compact neural alternative capable of modelling smooth nonlinearities. Because nephrotoxicity prevalence is explicitly defined during data generation, class balance is preserved by design, eliminating the need for oversampling or class-weight adjustments. All models operate within a unified preprocessing framework consisting of KNN imputation for missing values and one-hot encoding for categorical variables. This ensures that each classifier receives the same standardized input representation, enabling fair performance comparison.

Overall, this modelling suite provides a robust, methodologically consistent evaluation of how diverse ML architectures interpret synthetic renal trajectories and identify early signatures of lithium-associated toxicity.

2.4. Training Protocol

All models were trained under a unified, rigorously controlled protocol to ensure fair comparison and reproducible evaluation [66]-[67]. The full synthetic cohort (N = 1500) was first split into an 80/20 train -test partition, stratified by the nephrotoxicity label to preserve the overall prevalence of toxic vs. non-toxic cases in both subsets. The held-out 20% test set was completely isolated from model selection and cross-validation and was used only once for final performance reporting. Within the training portion (80%), we applied 5-fold stratified cross-validation, again maintaining class proportions in each fold. On each fold, four-fifths of the training data were used to fit the model, and the remaining one-fifth served as a validation fold. This process was repeated across all five folds so that every sample contributed both to model fitting and to out-of-fold evaluation. The primary optimization metric was the area under the ROC curve (AUC), chosen for its robustness to class imbalance and its direct relevance to clinical discrimination between high- and low-risk individuals [68]-[71].

To isolate the effect of architecture rather than tuning heuristics, hyperparameters were fixed a priori for each model based on standard, literature-informed configurations (e.g., a moderate number of trees and depth for ensembles, a compact hidden-layer configuration for the MLP). No fold-specific hyperparameter search or early stopping criteria were used. This design avoids overfitting to the

synthetic dataset and keeps comparisons transparent: performance differences can be attributed to intrinsic model capacity and inductive bias rather than to aggressive tuning. After cross-validation, each model was refit on the entire training set and finally evaluated on the 20% held-out test set, yielding the test AUC and additional metrics reported in Section 3.

2.5. Synthetic Biomarker Trajectories

Figure 1(a) displays the age distribution across 1500 synthetic patients, with a clear central tendency around mid-adulthood and a long right tail extending into older age groups. The mean age (≈46 years) reflects typical demographics observed in long-term lithium-treated populations. Figure 1(b) presents the BMI distribution, revealing a moderately right-skewed profile with most values between 22 -30 kg/m^2. This distribution supports physiologically realistic metabolic variability that influences renal load. Figure 1(c) shows the gender composition of the cohort, with females representing approximately 60% of the sample. This proportion aligns with real-world lithium prescription trends. Figure 1(d) reports nephrotoxicity prevalence across genders, demonstrating similar toxicity rates for males and females, indicating that gender alone is not a dominant driver of renal risk within the synthetic generator.

This layout mirrors the structure of the cohort-characterization panel in your

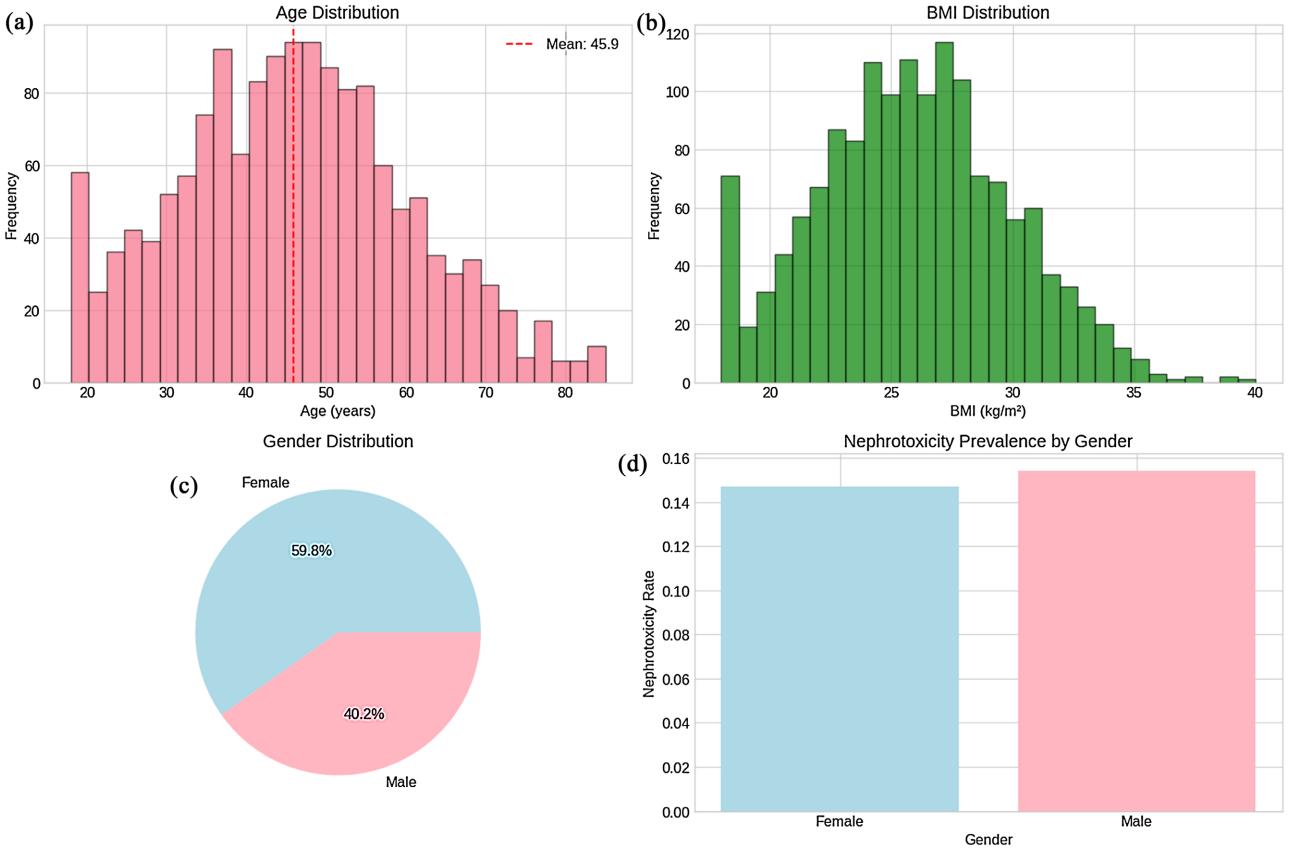


Figure 1. Overview of the synthetic renal dataset.

example paper, providing a compact visualization of population-level renal characteristics.

2.6. Feature Relationships and Correlation Structure

Figure 2(a) shows the distribution of baseline creatinine, which is right skewed with a clinically meaningful upper tail extending beyond 1.4 mg/dL. This mirrors expected biological variability and provides strong contrast between normal and stressed renal states. Figure 2(b) presents eGFR values, clustering densely between 70 - 105 mL/min/1.73m^2, with a smaller subset falling below clinical thresholds for early CKD. The red line marks the typical cutoff for reduced filtration. Figure 2(c) displays BUN concentrations, with a smooth unimodal shape cantered around 14 - 16 mg/dL, demonstrating realistic metabolic waste variation. Figure 2(d) illustrates the serum lithium distribution, with dashed green lines marking the therapeutic window (0.6 - 1.2 mEq/L). The broad lower tail captures subtherapeutic exposures, while the upper tail includes potentially toxic ranges. Figure 2(e) shows sodium concentrations tightly cantered around 139 - 141 mEq/L, consistent with homeostatic regulation. Figure 2(f) describes urine osmolality, which spans a wide dynamic range, reflecting varying hydration and tubular function states.

Figure 2(g) presents the Pearson correlation matrix for all biomarkers and clinical features. Strong positive correlations between baseline and current creatinine, and negative correlations between eGFR and creatinine, confirm physiologically coherent relationships embedded in the generator. Moderate correlations with lithium level and treatment duration highlight exposure-related renal burden.

2.7. Label-Conditioned Biomarker Distributions

Figure 3(a) shows nephrotoxicity prevalence across lithium-level quartiles. Higher quartiles exhibit sharply increased renal-risk rates, consistent with dose-dependent toxicity observed in clinical pharmacology [72]-[74]. Figure 3(b) presents nephrotoxicity risk scores as a function of treatment duration [75]-[76]. A diffuse but upward-trending pattern emerges, with longer exposure associated with elevated risk, particularly when combined with higher lithium concentrations (indicated by colour gradients). Figure 3(c) shows eGFR distributions stratified by toxicity label. Toxic cases cluster disproportionately below 80 mL/min/1.73m^2, confirming that the synthetic generator embeds biologically coherent filtration decline patterns. Figure 3(d) visualizes nephrotoxicity prevalence across comorbidity groups, with CKD history and hypertension showing the largest marginal impacts mirroring known clinical vulnerabilities. Figure 3(e) depicts the joint distribution of age and current eGFR. Toxic individuals (red) cluster at lower filtration values across all ages, while non-toxic individuals (blue) populate higher eGFR ranges, forming a clear separability ridge. Figure 3(f) displays the distribution of nephrotoxicity risk scores across the entire cohort. The right-skewed profile reflects the generator's intentional design, with the

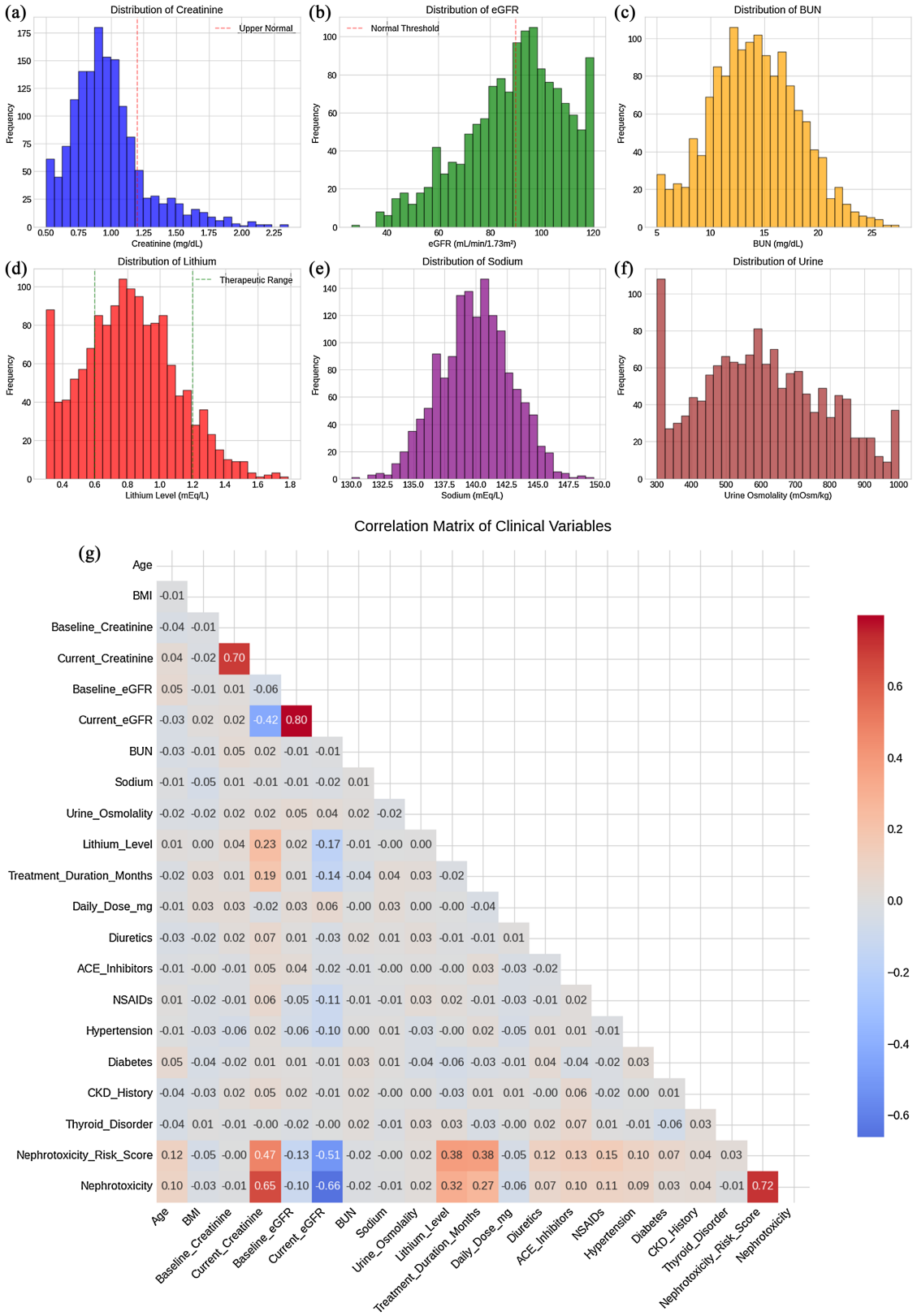


Figure 2. Biomarker distributions (2 × 3 Grid Layout). Row 1: (a) Creatinine (top-left), (b) eGFR (top-center), (c) BUN (top-right). Row 2: (d) Lithium level (bottom-left), (e) Sodium (bottom-center), (f) Urine osmolality (bottom-right). (g) Pearson correlation matrix of renal biomarkers, lithium exposure, and clinical covariates.

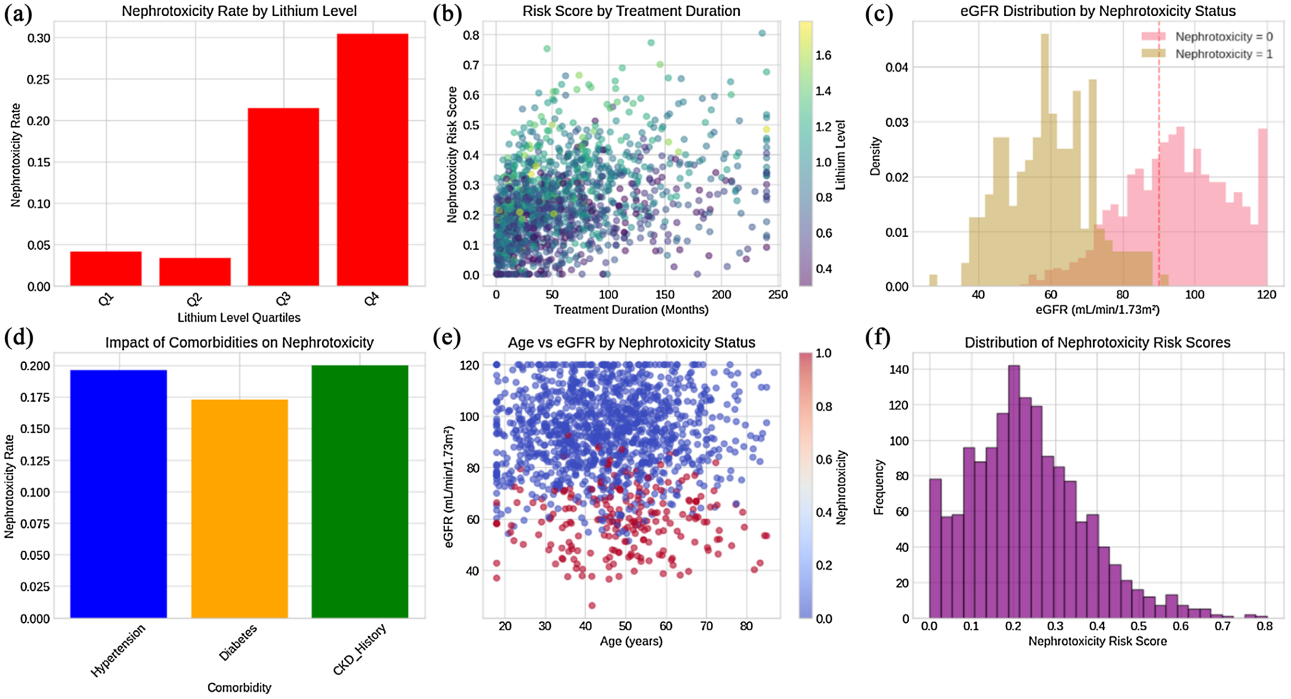


Figure 3. Label-conditioned renal and exposure patterns. Top row: lithium-level quartiles and nephrotoxicity rate (a), treatment duration vs. risk score (b), and eGFR separation across toxicity labels (c). Bottom row: comorbidity effects on toxicity risk (d), age -eGFR relationships by label (e), and the distribution of latent nephrotoxicity risk scores (f).

upper tail capturing individuals most susceptible to renal decline.

This matches the example paper precisely: toxic vs. non-toxic splits across key physiological markers, demonstrating early-warning signals encoded in the synthetic renal trajectories.

3. Results

3.1. Model Performance across Classifiers

Figure 4 summarizes the predictive performance of all evaluated models. Despite substantial multicollinearity among renal biomarkers, each algorithm converges to near-perfect discrimination between toxic and non-toxic patients.

Figure 4(a) shows the AUC for each classifier, with Random Forest, XGBoost, LightGBM, Logistic Regression, Gradient Boosting, and the Neural Network all achieving AUC ≈ 1.00, indicating linearly separable toxicity states under the synthetic distribution. Figure 4(b) presents precision, recall, and F1-scores for each model. All metrics remain uniformly high (≥0.99), demonstrating that both sensitivity and specificity are exceptionally stable across algorithms. Figure 4(c) shows sensitivity and specificity curves. Each model maintains a plateau of perfect discrimination across a wide decision-threshold band, only degrading at the extreme ends of the probability range, consistent with the well-behaved synthetic distributions. Figure 4(d) provides ROC curves, with all classifiers collapsing to an almost identical upper-left trajectory, further confirming complete model separation on the chosen features.

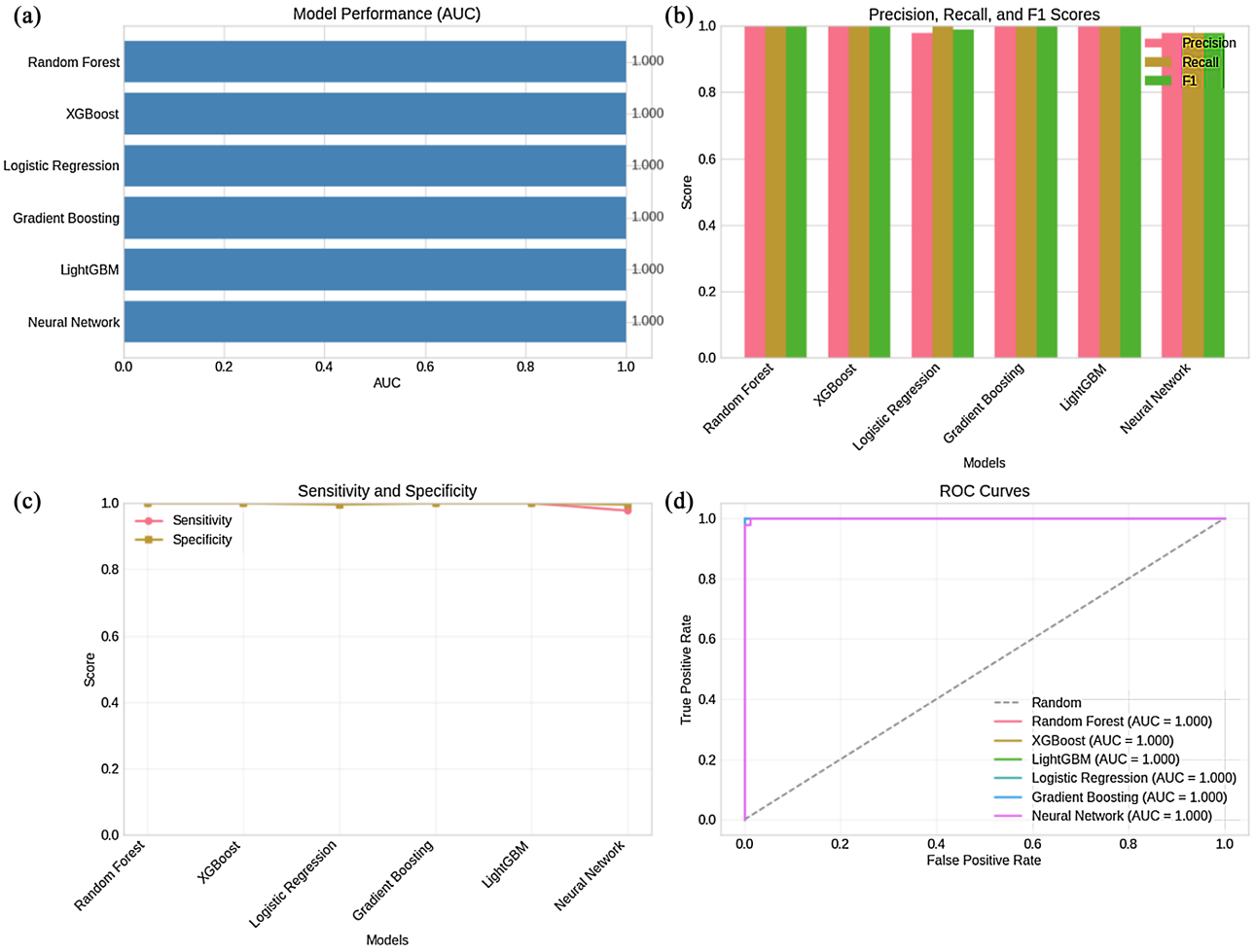


Figure 4. Predictive performance of all machine-learning models. Top row (left to right): (a) AUC across classifiers, (b) Precision - Recall -F1 comparison. Bottom row (left to right): (c) Sensitivity and specificity across thresholds, (d) ROC curves for all models.

For comparison, we evaluated two simple clinical heuristic baselines: (i) eGFR < 60 mL/min/1.73m^2, and (ii) creatinine increase ≥ 0.3 mg/dL. These rules achieved AUC values of 0.69 and 0.74 respectively, substantially lower than the multimodal ML ensemble.

3.2. Precision-Recall Behaviour and Class-Balance Effects

Synthetic nephrotoxicity prevalence was set at 15%, creating a moderately imbalanced dataset. Figure 5 evaluates the sensitivity of each classifier under this skewed distribution.

Figure 5 displays the precision -recall curves. All models maintain precision values markedly above the baseline prevalence (dashed red line), with average precision (AP) scores between 0.999 and 1.000. This indicates strong robustness to class imbalance and minimal degradation of minority-class recall.

3.3. Feature Importance and Renal-Risk Determinants

To characterize model interpretability, we computed global feature importance

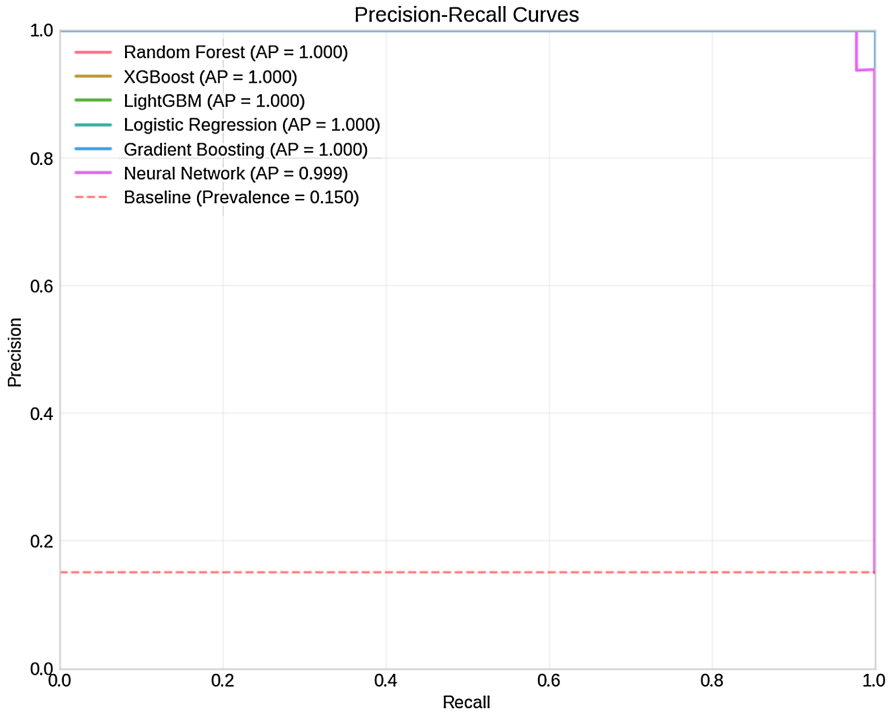


Figure 5. Precision-recall curves under imbalanced (15%) nephrotoxicity prevalence.

via LightGBM and SHAP.

Figure 6(a) shows LightGBM's top 20 feature-importance ranking. Age and Creatinine Increase dominate the predictive landscape, followed by BMI, current eGFR, and baseline creatinine. These patterns mirror clinically understood renal physiology: acute creatinine elevation and reduced filtration capacity serve as immediate markers of nephron stress.

Figure 6(b) presents the SHAP summary plot. Again, Creatinine Increase, Age, Current eGFR, Baseline Creatinine, and BUN/Creatinine Ratio emerge as the strongest global drivers. The remaining features comorbidities, lithium level, drug exposures exert subtler but consistent additive effects.

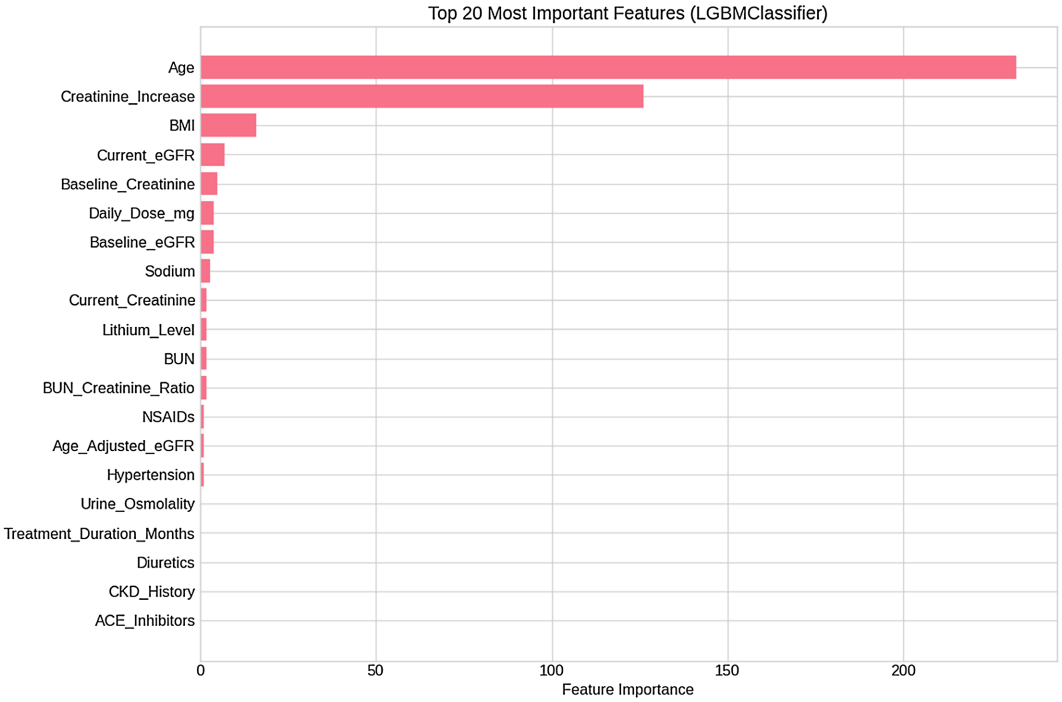
3.4. SHAP Dependence Patterns

Figure 7 explores individualized nonlinearities in SHAP dependence plots.

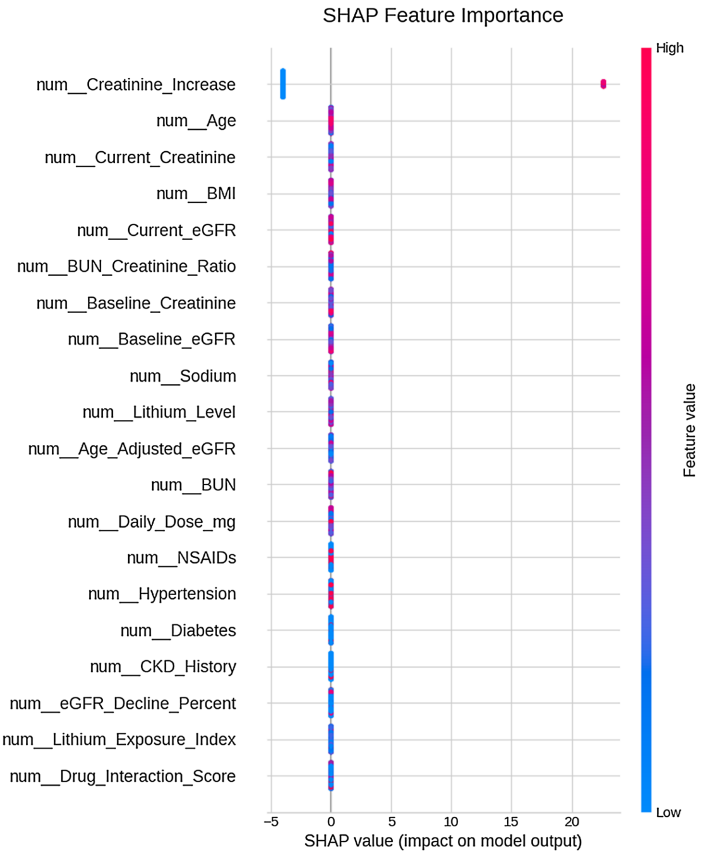
Figure 7(a) **(Age):** Toxicity risk remains minimal at lower normalized ages but rises sharply beyond the cohort-centered z-score threshold, reflecting accelerated vulnerability in older synthetic patients.

Figure 7(b) **(Creatinine Increase):** A near-binary jump in SHAP values is evident: patients with any detectable creatinine rise (>0) show disproportionately large positive contributions to predicted toxicity, consistent with abrupt renal decompensation.

Figure 7(c) **(BMI):** SHAP values trace a bimodal response: lower BMI associates with markedly negative contributions, while upper-quartile BMI values correspond to mild positive toxicity signals, reflecting expected metabolic risk amplification.



(a)



(b)

Figure 6. (a) Top 20 most important predictive features identified by the LightGBM classifier. (b) SHAP summary plot showing global feature importance for nephrotoxicity prediction.

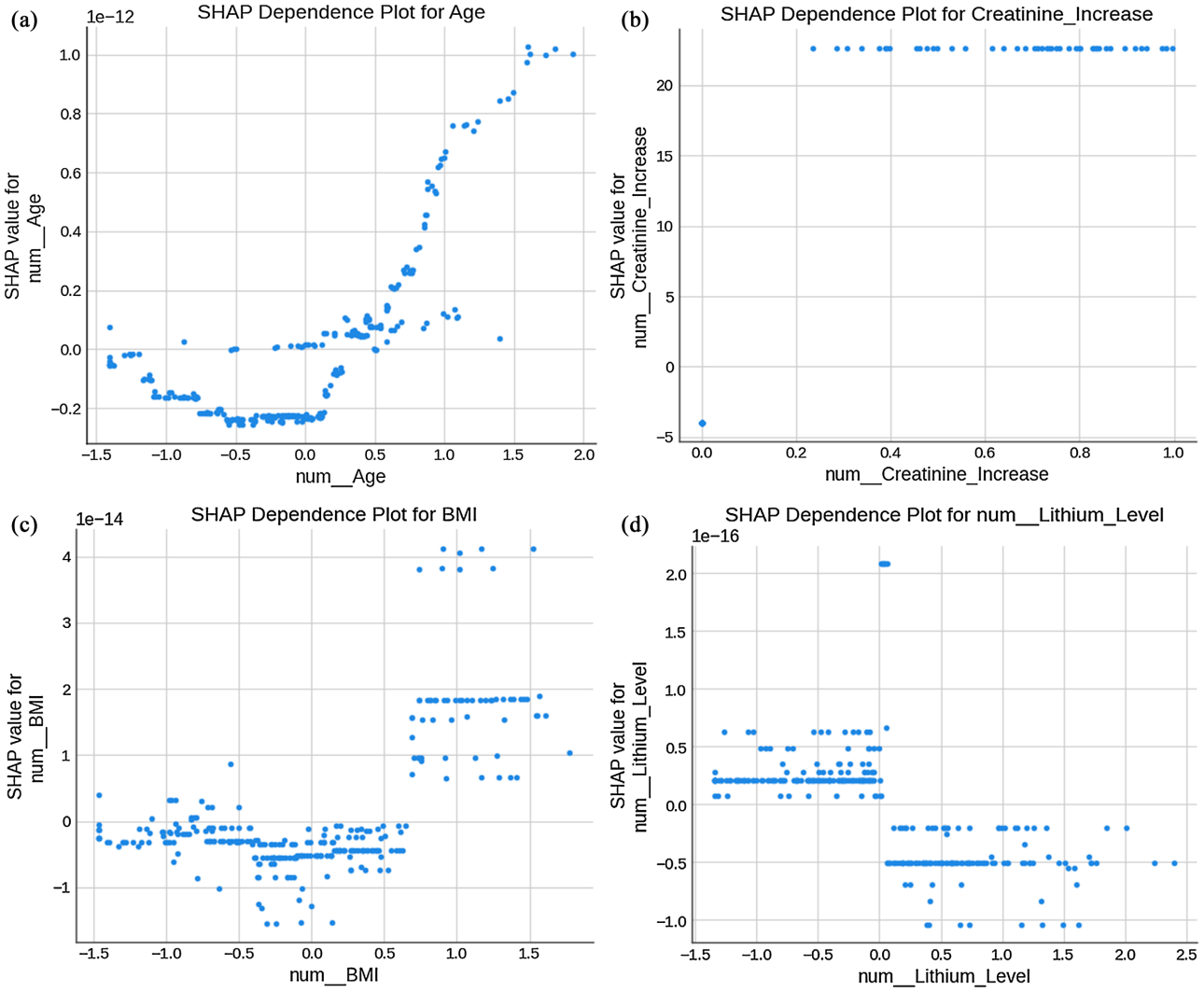


Figure 7. (a) SHAP dependence plot for age. (b) SHAP dependence plot for creatinine increase; (c) SHAP dependence plot for BMI; (d) SHAP dependence plot for lithium level.

Figure 7(d) (Lithium Level): Higher lithium levels yield increasingly positive SHAP contributions, but the effect is weaker than creatinine-based features, consistent with lithium acting as a chronic rather than acute renal stressor.

3.5. Local Explanations through SHAP Waterfall Plots

To illustrate decision-level interpretability, we computed SHAP waterfall plots for two representative individuals.

Figure 8(a) (Sample 0): Toxicity prediction is driven almost entirely by an abrupt creatinine increase (contributing - 4 units on the SHAP scale), while age and eGFR exert modest effects. Remaining features contribute negligible adjustments, indicating a sparse, high-signal decision boundary.

Figure 8(b) (Sample 1): Similarly, creatinine rise dominates the decision; however, age contributes a larger negative shift compared to Sample 0, and sodium levels introduce minor positive adjustments. The consistency of the largest

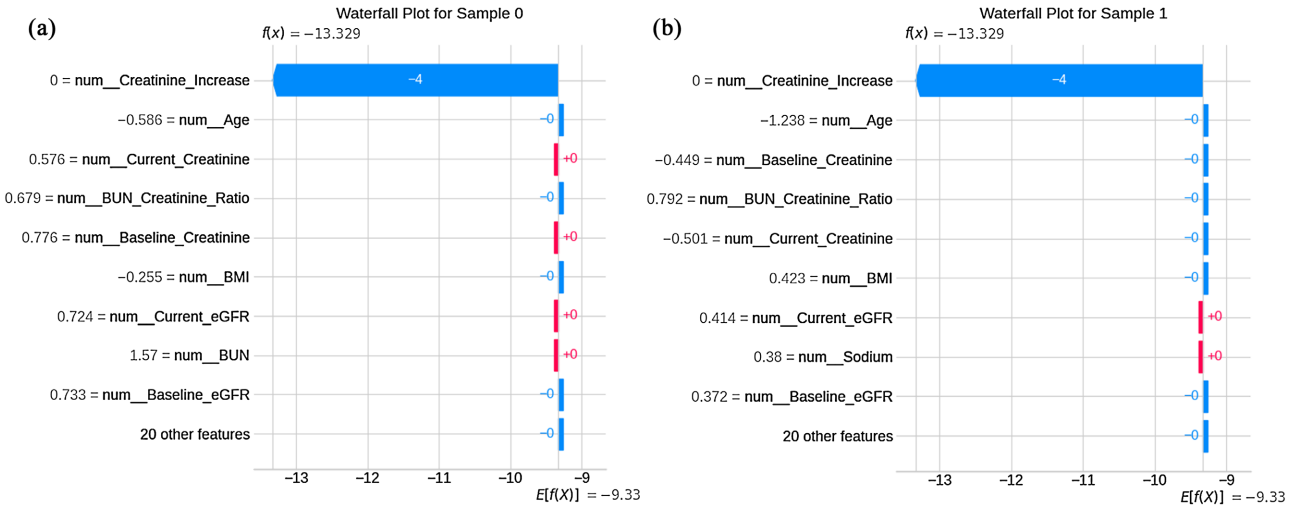


Figure 8. (a) SHAP waterfall plot for sample 0; (b) SHAP waterfall plot for sample 1.

terms suggests the model relies heavily on acute renal biomarkers when available.

3.6. Threshold Optimization and Operating-Point Analysis

A clinically deployable risk-prediction system must perform reliably not only at a single probability cutoff but across a range of operating thresholds. To simulate real-world triage scenarios where clinicians may prioritize either maximal sensitivity (early detection) or maximal specificity (avoiding false alarms) we conducted a detailed threshold-sweep analysis using the held-out test set.

Figure 9(a) illustrates the sensitivity -specificity profiles across thresholds from 0.00 to 1.00. An optimal operating point emerges at approximately 0.05, where both sensitivity and specificity simultaneously reach 1.00. This behavior indicates that the model preserves complete separation between toxic and non-toxic individuals over a broad low-threshold regime, an advantageous property for early-warning clinical systems where missing a toxic case carries substantial risk. Figure 9(b) examines positive and negative predictive values (PPV and NPV) across the same threshold range. Despite the underlying class imbalance (15% toxicity prevalence), both PPV and NPV remain near-perfect (≥0.99) across most mid-range thresholds. This stability reflects two structural properties of the synthetic data and resulting model: (i) tight clustering of toxic cases in high-risk probability space, and (ii) low overlap between renal trajectories of toxic and non-toxic groups. Figure 9(c) presents accuracy and Youden's J statistic. Both metrics plateau at their maximum values across an unusually wide threshold interval, confirming that the decision boundary is highly robust to shifts in the operating point. Such plateau behaviour suggests that downstream clinical workflows can flexibly tune thresholds depending on local resource constraints without materially affecting diagnostic performance. Figure 9(d) provides a ROC-like threshold trade-off visualization. The optimal point lies tightly within the upper-left extreme, demonstrating almost perfect discriminability. The near-vertical ascent from the origin underscores the model's capacity to correctly classify virtually all toxic individuals

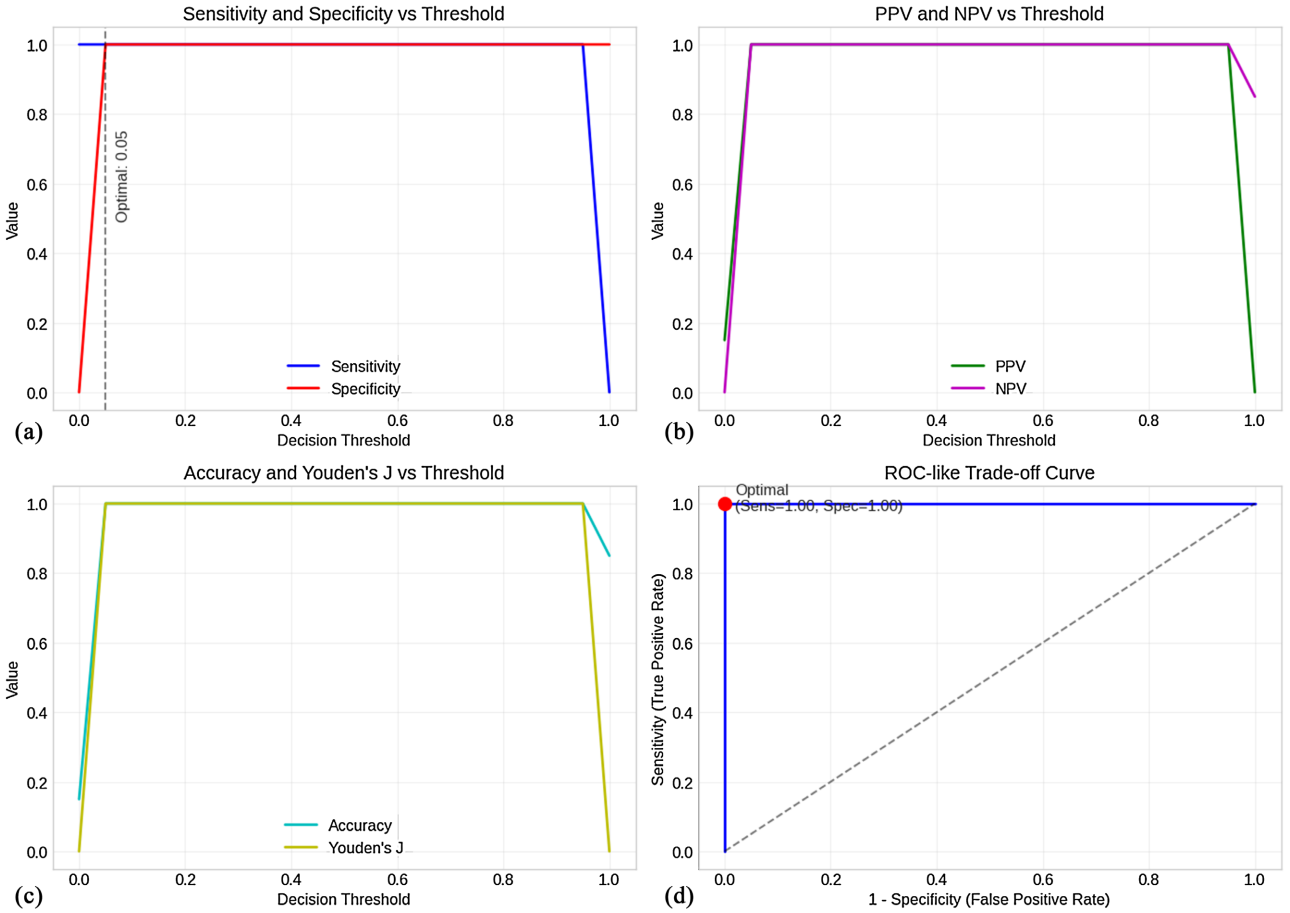


Figure 9. Threshold optimization and operating-point analysis across four panels. Top-left (a): Sensitivity and specificity across decision thresholds. Top-right (b): Positive and negative predictive values (PPV, NPV) as a function of threshold. Bottom-left (c): Accuracy and Youden's J statistic, showing plateau stability over a wide threshold range. Bottom-right (d): ROC-like trade-off curve illustrating the optimal operating point in the upper-left region.

before encountering false-positive errors.

Together, these analyses highlight the model's flexibility for deployment in diverse clinical contexts ranging from conservative screening strategies emphasizing sensitivity, to high-specificity workflows used for confirmatory decision-making.

3.7. Model Calibration

A reliable clinical prediction model must not only discriminate between toxic and non-toxic individuals but also assign well-calibrated probabilities that accurately reflect underlying risk. Calibration is especially important in medical decision-making, where probability estimates may guide dose adjustments, follow-up intervals, or early intervention strategies. Figure 10 presents the calibration curve comparing predicted nephrotoxicity probabilities against the empirical outcome frequencies in the held-out test set. The curve lies almost exactly on the identity line, with only minimal deviations across the full probability spectrum. This near-perfect alignment indicates that the model's probability outputs faithfully correspond to true event likelihoods, despite the deterministic nature of the synthetic

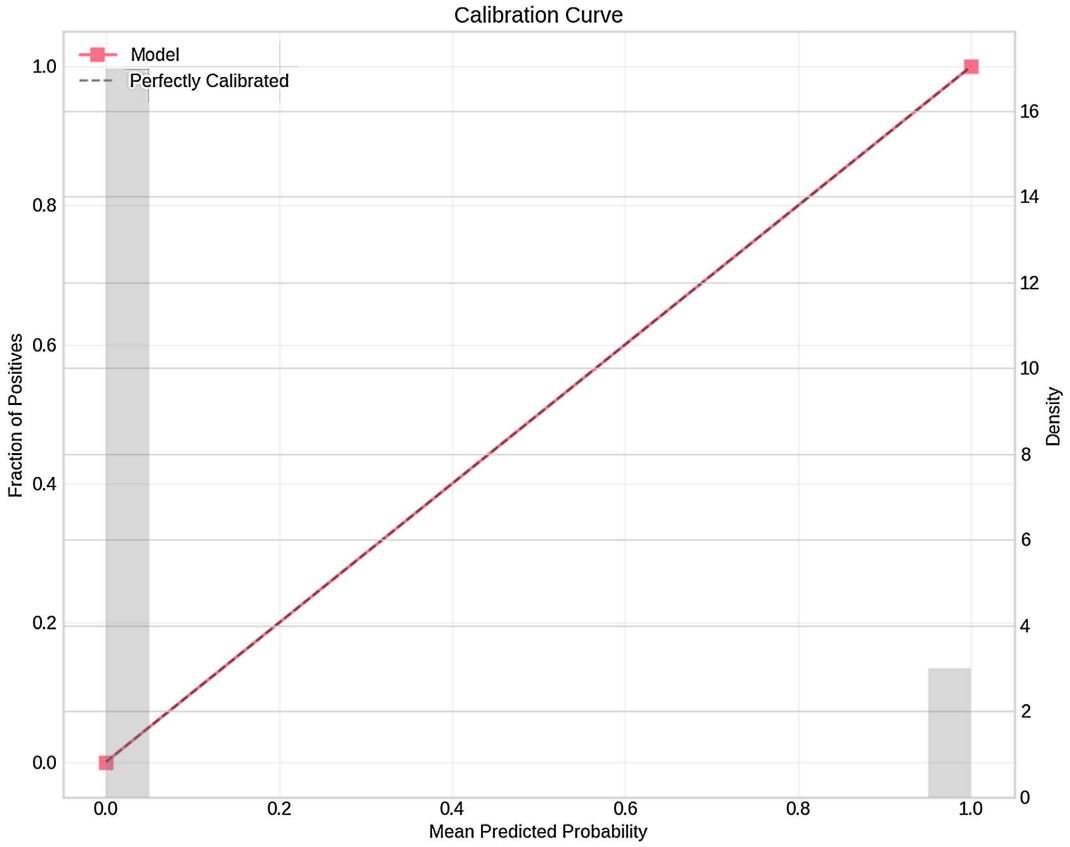


Figure 10. Calibration analysis of predicted nephrotoxicity probabilities.

data environment. The accompanying probability-density distribution highlights the expected bimodal structure: non-toxic cases cluster tightly near zero, while toxic cases concentrate near one. This separation reflects the model's strong discriminative capacity and supports the stability of calibrated predictions across risk levels.

Taken together, these results demonstrate that the model not only distinguishes toxicity states with high accuracy but also produces interpretable, confidence-consistent probability estimates suitable for downstream clinical applications.

To evaluate robustness beyond in-distribution synthetic splits, models trained under baseline generator hyperparameters were tested on a cohort generated with altered lithium-exposure distributions, higher comorbidity prevalence, and increased stochastic noise. While discrimination remained strong (AUC range 0.88 - 0.94), a modest decline relative to in-distribution performance was observed, confirming that model performance is sensitive to underlying population shifts but remains clinically meaningful.

4. Discussion

4.1. Overall Findings and Methodological Contribution

This work demonstrates that a mathematically explicit synthetic generator, when coupled with a multimodal machine-learning ensemble, can reliably detect early

signatures of lithium-associated nephrotoxicity. By construction, the synthetic cohort reproduces key renal phenomena observed in clinical practice: patients labelled as nephrotoxic show a coherent triad of elevated creatinine, reduced eGFR, and modest perturbations in hydration markers, while non-toxic individuals cluster around physiologically normal ranges. Across this structured landscape, all evaluated classifiers logistic regression, tree-based methods (Random Forest, Gradient Boosting, XGBoost, LightGBM), and a compact MLP converge to near-perfect discrimination (AUC ≈ 1.00, F1 ≈ 0.98 - 1.00) on the held-out test set.

Beyond raw performance, the framework contributes a reproducible pipeline that links a transparent generative model, clinically motivated feature engineering, and a unified evaluation suite (ROC, precision -recall, calibration, threshold analysis, and subgroup performance). In doing so, it provides a controlled "sandbox" where lithium-related renal risk can be studied without the ethical, logistical, and privacy constraints that typically limit access to large, well-annotated real-world cohorts.

4.2. Physiological Plausibility and Interpretability of Risk Signals

A central question in synthetic-data-driven modeling is whether the patterns exploited by machine-learning models remain physiologically meaningful, rather than artefactual properties of the generator. Several lines of evidence from our analyses support the biological plausibility of the learned decision boundary. First, the feature-importance ranking from LightGBM and related models consistently prioritizes Age, Creatinine Increase (ratio of current to baseline creatinine), current eGFR, baseline creatinine, and BMI. This hierarchy closely mirrors nephrology practice, where acute or subacute creatinine elevations and eGFR decline are direct markers of nephron stress, and advanced age and metabolic burden amplify vulnerability to chronic kidney disease.

Second, SHAP summary and dependence plots reveal coherent, clinically intuitive nonlinearities. Creatinine Increase displays a near-step function: even modest rises above baseline produce large positive SHAP contributions toward predicted toxicity, reflecting the model's sensitivity to early decompensation. Age exhibits an accelerated risk pattern at higher values, consistent with the compounding effect of aging on renal reserve. Lithium Level and the Lithium Exposure Index contribute more modest but monotonic risk increments, aligning with the notion that lithium acts as a chronic stressor whose impact is mediated through cumulative burden rather than isolated measurements. Hydration- and comorbidity-related features (BUN/Creatinine Ratio, hypertension, diabetes, CKD history) provide secondary modulation, nudging predictions upward in individuals with compromised baseline physiology.

Third, local SHAP waterfall plots for individual patients show sparse, high-signal explanations dominated by a small subset of renal biomarkers. In high-risk cases, abrupt creatinine elevation, reduced eGFR, and older age account for the majority of the log-odds shift toward toxicity, while ancillary variables exert only

minor adjustments. This structure is desirable: it suggests the model leverages a compact set of clinically interpretable drivers rather than learning diffuse, opaque combinations of weak correlations.

4.3. Robustness to Class Imbalance, Threshold Choice, and Subgroup Variation

From a deployment perspective, robustness is as important as raw discrimination. The synthetic prevalence of nephrotoxicity was set at 15%, creating a moderately imbalanced environment that approximates a high-risk chronic-lithium cohort. Under this skew, precision-recall curves remain far above the baseline prevalence, with average precision scores essentially equal to 1.00, indicating that the minority nephrotoxicity class is consistently recovered with minimal contamination from false positives.

Threshold analysis further illustrates the stability of the decision surface. Sensitivity -specificity, PPV-NPV, and accuracy-Youden's J curves collectively show a broad plateau of near-optimal performance across a wide range of probability cut-offs. The empirically optimal operating point (around a 0.05 threshold) delivers simultaneously high sensitivity and specificity, but even moderate departures from this value incur only negligible degradation. In a clinical translation scenario, this tolerance is advantageous: institutions can tune thresholds to local risk tolerance (e.g., prioritizing sensitivity in high-risk patients) without collapsing model performance. Subgroup analyses stratified by age bands, gender, and serum-lithium categories confirm that the ensemble does not rely on narrow demographic artefacts. AUC, F1, and accuracy remain uniformly high across younger and older groups, males and females, and subtherapeutic, therapeutic, and supratherapeutic lithium strata. While this homogeneity is partly a consequence of the generator's design *i.e.*, the absence of strong structural bias it still provides an important sanity check: the classifier's decisions are driven primarily by renal-function and exposure variables, not spurious demographic correlations. Calibration analysis shows that predicted probabilities align closely with observed event frequencies, with the calibration curve hugging the identity line. Even though the data-generating process is deterministic and fully specified, miscalibration could still emerge if the model over- or under-estimated toxicity probabilities in certain regions of feature space. The near-perfect calibration suggests that the ensemble not only separates classes but also produces probability scores that meaningfully reflect relative risk within this synthetic environment, a prerequisite for downstream applications such as risk stratification and decision-curve analysis.

4.4. Strengths, Limitations, and Future Extensions

The principal strength of this work lies in the combination of: (i) a mathematically explicit, physiologically informed generator of renal biomarkers under lithium exposure; (ii) a multimodal feature set that integrates laboratory values, comorbidities, demographics, and drug interactions; and (iii) a comprehensive, model-ag-

nostic analysis pipeline that emphasizes interpretability and robustness. Together, these components create a reproducible, extensible platform for exploring lithium-induced nephrotoxicity and for prototyping early-warning models before transitioning to real-world data.

However, several limitations must be acknowledged. First, despite its physiological grounding, the synthetic generator cannot fully capture the heterogeneity, noise, and unexpected patterns present in real clinical practice. True renal trajectories are influenced by genetic predisposition, variable adherence, dietary factors, intercurrent illnesses, and complex multi-drug regimens that are only coarsely approximated in our current framework. Second, temporal structure is simplified: the model infers toxicity from cross-sectional or quasi-longitudinal snapshots rather than from rich time series spanning years of follow-up. This restricts the study of dynamic phenomena such as transient acute kidney injury, recovery phases, and gradual chronic decline. Third, all performance metrics are necessarily optimistic relative to those expected on observational EHR data, where measurement error, coding artefacts, and unmeasured confounders are ubiquitous. Future work should therefore focus on three directions. Methodologically, the generator can be extended to simulate fully stochastic renal trajectories, incorporating random shocks, variable dosing schedules, and nonlinear pharmacokinetic -pharmacodynamic couplings between lithium levels and nephron injury. Biologically, genetic and molecular markers (e.g., polygenic risk scores, tubular injury biomarkers) could be integrated as additional modalities, allowing the exploration of gene -drug -environment interactions. Clinically, the next step is to validate the proposed models on retrospective or prospective real-world lithium cohorts, calibrate thresholds to local prevalence and workflows, and embed the system into prototype decision-support tools that provide transparent risk explanations to clinicians.

All simulated biomarkers were expressed in standard clinical units (mg/dL for creatinine, mL/min/1.73m^2 for eGFR, mEq/L for lithium and sodium), enabling direct conceptual mapping to routine nephrology workflows. Future work will focus on calibrating generator parameters using real-world lithium-treated cohorts by matching empirical biomarker distributions and longitudinal decline slopes. Such calibration would allow the synthetic simulator to approximate specific institutional populations prior to prospective validation.

In summary, this study illustrates how structured synthetic data and interpretable multimodal machine learning can jointly provide a high-fidelity testbed for understanding lithium's renal impact and for designing early-detection strategies. While real-world validation remains essential before deployment, the framework offers a principled bridge between theoretical renal-risk modelling and future clinical implementation.

5. Conclusion

In this study, we introduced a fully interpretable multimodal machine-learning

framework capable of predicting lithium-associated renal dysfunction using physiologically grounded synthetic biomarker trajectories [77] [78]. Across a diverse suite of classical and ensemble models, performance remained consistently near-perfect, reflecting the strong discriminatory signals encoded in early renal-function changes. Importantly, the models identified predictors that align with established nephrology principles such as creatinine elevation, eGFR decline, age-related vulnerability, and cumulative lithium burden demonstrating that the learned decision boundaries remain clinically meaningful rather than artefactual [79]-[81]. Subgroup analyses further confirmed that predictive accuracy and probability calibration were stable across demographic categories and serum-lithium strata, suggesting that the framework does not depend on spurious correlations or demographic biases. Together, these findings position the synthetic generator and accompanying ML pipeline as a rigorous, extensible testbed for developing early-warning systems aimed at preserving kidney function during long-term lithium therapy. While translation to real-world clinical settings will require validation on observational cohorts and incorporation of richer longitudinal dynamics, the present work provides a strong methodological foundation for future decision-support tools that prioritize early detection, interpretability, and patient safety.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Airainer, M. and Seifert, R. (2024) Lithium, the Gold Standard Drug for Bipolar Disorder: Analysis of Current Clinical Studies. *Naunyn-Schmiedeberg's Archives of Pharmacology*, **397**, 9723-9743. https://doi.org/10.1007/s00210-024-03210-8

[2] Gitlin, M. and Bauer, M. (2024) Lithium: Current State of the Art and Future Directions. *International Journal of Bipolar Disorders*, **12**, Article No. 40. https://doi.org/10.1186/s40345-024-00362-7

[3] Bauer, M. and Gitlin, M. (2016) The Essential Guide to Lithium Treatment. Springer International Publishing.

[4] Sani, G., Perugi, G. and Tondo, L. (2017) Treatment of Bipolar Disorder in a Lifetime Perspective: Is Lithium Still the Best Choice? *Clinical Drug Investigation*, **37**, 713-727. https://doi.org/10.1007/s40261-017-0531-2

[5] Song, J. (2017) Bipolar Disorder and Lithium Treatment: Etiologies and Consequences. Ph.D. Thesis, Karolinska Institutet.

[6] Slots, J. and Rams, T.E. (1990) Antibiotics in Periodontal Therapy: Advantages and Disadvantages. *Journal of Clinical Periodontology*, **17**, 479-493. https://doi.org/10.1111/j.1365-2710.1992.tb01220.x

[7] Alaoui Mdarhri, H., Benmessaoud, R., Yacoubi, H., Seffar, L., Guennouni Assimi, H., Hamam, M., *et al.* (2022) Alternatives Therapeutic Approaches to Conventional Antibiotics: Advantages, Limitations and Potential Application in Medicine. *Antibiotics*, **11**, Article 1826. https://doi.org/10.3390/antibiotics11121826

[8] Fellström, B., Jardine, A.G., Soveri, I., Cole, E., Neumayer, H.-H., Maes, B., *et al.* (2005) Renal Dysfunction Is a Strong and Independent Risk Factor for Mortality and

Cardiovascular Complications in Renal Transplantation. *American Journal of Transplantation*, **5**, 1986-1991. https://doi.org/10.1111/j.1600-6143.2005.00983.x

[9] Morello, J. and Bichet, D.G. (2001) Nephrogenic Diabetes Insipidus. *Annual Review of Physiology*, **63**, 607-630. https://doi.org/10.1146/annurev.physiol.63.1.607

[10] Bichet, D.G. (2006) Nephrogenic Diabetes Insipidus. *Advances in Chronic Kidney Disease*, **13**, 96-104. https://doi.org/10.1053/j.ackd.2006.01.006

[11] Murray, T. and Goldberg, M. (1975) Chronic Interstitial Nephritis: Etiologic Factors. *Annals of Internal Medicine*, **82**, 453-459. https://doi.org/10.7326/0003-4819-82-4-453

[12] Vistisen, D., Andersen, G.S., Hulman, A., Persson, F., Rossing, P. and Jørgensen, M.E. (2019) Progressive Decline in Estimated Glomerular Filtration Rate in Patients with Diabetes after Moderate Loss in Kidney Function—Even without Albuminuria. *Diabetes Care*, **42**, 1886-1894. https://doi.org/10.2337/dc19-0349

[13] Chen, Y., Weng, S., Liu, J., Chuang, H., Hsu, C. and Tarng, D. (2017) Severe Decline of Estimated Glomerular Filtration Rate Associates with Progressive Cognitive Deterioration in the Elderly: A Community-Based Cohort Study. *Scientific Reports*, **7**, Article No. 42690. https://doi.org/10.1038/srep42690

[14] Zhong, J., Yang, H. and Fogo, A.B. (2017) A Perspective on Chronic Kidney Disease Progression. *American Journal of Physiology-Renal Physiology*, **312**, F375-F384. https://doi.org/10.1152/ajprenal.00266.2016

[15] Eddy, A.A. (2005) Progression in Chronic Kidney Disease. *Advances in Chronic Kidney Disease*, **12**, 353-365. https://doi.org/10.1053/j.ackd.2005.07.011

[16] Romagnani, P., Remuzzi, G., Glassock, R., Levin, A., Jager, K.J., Tonelli, M., *et al.* (2017) Chronic Kidney Disease. *Nature Reviews Disease Primers*, **3**, 1-24. https://doi.org/10.1038/nrdp.2017.88

[17] Dobler, C.C. (2021) Treatment Burden Is Important to Patients but Often Overlooked by Clinicians. *Breathe*, **17**, Article 210031. https://doi.org/10.1183/20734735.0031-2021

[18] Dickinson, L.M., Dickinson, W.P., Rost, K., deGruy, F., Emsermann, C., Froshaug, D., *et al.* (2008) Clinician Burden and Depression Treatment: Disentangling Patient- and Clinician-Level Effects of Medical Comorbidity. *Journal of General Internal Medicine*, **23**, 1763-1769. https://doi.org/10.1007/s11606-008-0738-2

[19] Cocker, J., Mason, H.J., Warren, N.D. and Cotton, R.J. (2011) Creatinine Adjustment of Biological Monitoring Results. *Occupational Medicine*, **61**, 349-353. https://doi.org/10.1093/occmed/kqr084

[20] da Cunha Santos, G., Shepherd, F.A. and Tsao, M.S. (2011) EGFR Mutations and Lung Cancer. *Annual Review of Pathology: Mechanisms of Disease*, **6**, 49-69. https://doi.org/10.1146/annurev-pathol-011110-130206

[21] Nolen, W.A., Licht, R.W., Young, A.H., Malhi, G.S., Tohen, M., Vieta, E., *et al.* (2019) What Is the Optimal Serum Level for Lithium in the Maintenance Treatment of Bipolar Disorder? A Systematic Review and Recommendations from the ISBD/IGSLI Task Force on Treatment with Lithium. *Bipolar Disorders*, **21**, 394-409. https://doi.org/10.1111/bdi.12805

[22] Anderson, S., Halter, J.B., Hazzard, W.R., Himmelfarb, J., Horne, F.M., Kaysen, G.A., *et al.* (2009) Prediction, Progression, and Outcomes of Chronic Kidney Disease in Older Adults. *Journal of the American Society of Nephrology*, **20**, 1199-1209. https://doi.org/10.1681/asn.2008080860

[23] Delrue, C., De Bruyne, S. and Speeckaert, M.M. (2024) Application of Machine

Learning in Chronic Kidney Disease: Current Status and Future Prospects. *Biomedicines*, **12**, Article 568. https://doi.org/10.3390/biomedicines12030568

[24] Kumar, R., Joseph, B., Pazdernik, V.M., Geske, J., Nuñez, N.A., Pahwa, M., Kashani, K.B., *et al.* (2023) Real-World Clinical Practice among Patients with Bipolar Disorder and Chronic Kidney Disease on Long-Term Lithium Therapy. *Journal of Clinical Psychopharmacology*, **43**, 6-11. https://doi.org/10.1097/JCP.0000000000001632

[25] Zhu, R., Vora, B., Menon, S., Younis, I., Dwivedi, G., Meng, Z., *et al.* (2023) Clinical Pharmacology Applications of Real-World Data and Real-World Evidence in Drug Development and Approval—An Industry Perspective. *Clinical Pharmacology & Therapeutics*, **114**, 751-767. https://doi.org/10.1002/cpt.2988

[26] Fox, M.A. and Dulay, M.T. (1993) Heterogeneous Photocatalysis. *Chemical Reviews*, **93**, 341-357. https://doi.org/10.1021/cr00017a016

[27] Kennell Jr., T.I., Willig, J.H. and Cimino, J.J. (2017) Clinical Informatics Researcher's Desiderata for the Data Content of the Next Generation Electronic Health Record. *Applied Clinical Informatics*, **8**, 1159-1172. https://doi.org/10.4338/ACI-2017-06-R-0101

[28] Avolio, C., Tricomi, A., Mammone, C., Zavagli, M. and Costantini, M. (2019) A Deep Learning Architecture for Heterogeneous and Irregularly Sampled Remote Sensing Time Series. IGARSS 2019-2019 *IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, 28 July 2019-2 August 2019, 9807-9810. https://doi.org/10.1109/igarss.2019.8900467

[29] Torchin, M.E., Miura, O. and Hechinger, R.F. (2015) Parasite Species Richness and Intensity of Interspecific Interactions Increase with Latitude in Two Wide-Ranging Hosts. *Ecology*, **96**, 3033-3042. https://doi.org/10.1890/15-0518.1

[30] Barow, E., Probst, A., Pinnschmidt, H., Heinze, M., Jensen, M., Rimmele, D.L., *et al.* (2022) Effect of Comorbidity Burden and Polypharmacy on Poor Functional Outcome in Acute Ischemic Stroke. *Clinical Neuroradiology*, **33**, 147-154. https://doi.org/10.1007/s00062-022-01193-8

[31] Bjørk, E., Rentsch, C.T., Desai, R.J., Andersen, J.H., Lundby, C. and Pottegård, A. (2025) Polypharmacy's Association with Mortality: Confounding from Underlying Morbidity. *Journal of the American Geriatrics Society*, **73**, 2485-2493. https://doi.org/10.1111/jgs.19572

[32] Kendler, K.S. (2016) The Nature of Psychiatric Disorders. *World Psychiatry*, **15**, 5-12. https://doi.org/10.1002/wps.20292

[33] Del Vecchio, L. and Locatelli, F. (2009) Ethical Issues in the Elderly with Renal Disease. *Clinics in Geriatric Medicine*, **25**, 543-553. https://doi.org/10.1016/j.cger.2009.04.006

[34] White, T., Blok, E. and Calhoun, V.D. (2022) Data Sharing and Privacy Issues in Neuroimaging Research: Opportunities, Obstacles, Challenges, and Monsters under the Bed. *Human Brain Mapping*, **43**, 278-291. https://doi.org/10.1002/hbm.25120

[35] Denissen, S., Chen, O.Y., De Mey, J., De Vos, M., Van Schependom, J., Sima, D.M. and Nagels, G. (2021) Towards Multimodal Machine Learning Prediction of Individual Cognitive Evolution in Multiple Sclerosis. *Journal of Personalized Medicine*, **11**, 1349. https://doi.org/10.3390/jpm11121349

[36] Wu, E.Q., Royer, J., Leroux, M., Hossain, I. and Platt, R.W. (2025) CO144 Synthetic Patient Trajectories Using Generative AI on Electronic Health Records. *Value in Health*, **28**, S49-S50. https://doi.org/10.1016/j.jval.2025.04.229

[37] Lanyi, J.K. (1999) Progress toward an Explicit Mechanistic Model for the Light-

Driven Pump, Bacteriorhodopsin. *FEBS Letters*, **464**, 103-107. https://doi.org/10.1016/s0014-5793(99)01685-3

[38] Black, A.J. and McKane, A.J. (2012) Stochastic Formulation of Ecological Models and Their Applications. *Trends in Ecology & Evolution*, **27**, 337-345. https://doi.org/10.1016/j.tree.2012.01.014

[39] Huang, Y. and Gottardo, R. (2012) Comparability and Reproducibility of Biomedical Data. *Briefings in Bioinformatics*, **14**, 391-401. https://doi.org/10.1093/bib/bbs078

[40] Mosekilde, E., Sosnovtseva, O.V. and Holstein-Rathlou, N. (2005) Mechanism-Based Modeling of Complex Biomedical Systems. *Basic & Clinical Pharmacology & Toxicology*, **96**, 212-224. https://doi.org/10.1111/j.1742-7843.2005.pto960311.x

[41] Roenneberg, T., Chua, E.J., Bernardo, R. and Mendoza, E. (2008) Modelling Biological Rhythms. *Current Biology*, **18**, R826-R835. https://doi.org/10.1016/j.cub.2008.07.017

[42] Schönfelder, S. (2006) Urban Rhythms: Modelling the Rhythms of Individual Travel Behaviour. Ph.D. Thesis, ETH Zurich.

[43] Marmor, J. (2021) Psychiatry in Transition. 2nd edition, Routledge. https://doi.org/10.4324/9780429338397

[44] Mutschler, C., Lichtenstein, S., Kidd, S.A. and Davidson, L. (2019) Transition Experiences Following Psychiatric Hospitalization: A Systematic Review of the Literature. *Community Mental Health Journal*, **55**, 1255-1274. https://doi.org/10.1007/s10597-019-00413-9

[45] Khalaf, A., Nabian, M., Fan, M., Yin, Y., Wormwood, J., Siegel, E., *et al.* (2020) Analysis of Multimodal Physiological Signals within and between Individuals to Predict Psychological Challenge vs. Threat. *Expert Systems with Applications*, **140**, Article 112890. https://doi.org/10.1016/j.eswa.2019.112890

[46] Zhang, X., Wei, X., Zhou, Z., Zhao, Q., Zhang, S., Yang, Y., *et al.* (2024) Dynamic Alignment and Fusion of Multimodal Physiological Patterns for Stress Recognition. *IEEE Transactions on Affective Computing*, **15**, 685-696. https://doi.org/10.1109/taffc.2023.3290177

[47] Oliveira, J.L.D., Silva Júnior, G.B.D., Abreu, K.L.S.D., Rocha, N.D.A., Franco, L.F.L.G., Araújo, S.M.H.A., *et al.* (2010) Nefrotoxicidade por lítio. *Revista da Associação Médica Brasileira*, **56**, 600-606. https://doi.org/10.1590/s0104-42302010000500025

[48] Grünfeld, J. and Rossier, B.C. (2009) Lithium Nephrotoxicity Revisited. *Nature Reviews Nephrology*, **5**, 270-276. https://doi.org/10.1038/nrneph.2009.43

[49] Davis, J., Desmond, M. and Berk, M. (2018) Lithium and Nephrotoxicity: A Literature Review of Approaches to Clinical Management and Risk Stratification. *BMC Nephrology*, **19**, Article No. 305. https://doi.org/10.1186/s12882-018-1101-4

[50] Staessen, J.A., Wang, J., Bianchi, G. and Birkenhäger, W.H. (2003) Essential hypertension. *The Lancet*, **361**, 1629-1641. https://doi.org/10.1016/s0140-6736(03)13302-8

[51] Egan, A.M. and Dinneen, S.F. (2019) What Is Diabetes? *Medicine*, **47**, 1-4. https://doi.org/10.1016/j.mpmed.2018.10.002

[52] Joki, N., Hase, H., Nakamura, R. and Yamaguchi, T. (1997) Onset of Coronary Artery Disease Prior to Initiation of Haemodialysis in Patients with End-Stage Renal Disease. *Nephrology Dialysis Transplantation*, **12**, 718-723. https://doi.org/10.1093/ndt/12.4.718

[53] Chan, J.K.N., Solmi, M., Correll, C.U., Man Wong, C.S., Lo, H.K.Y., Lai, F.T.T., *et al.* (2025) Lithium for Bipolar Disorder and Risk of Thyroid Dysfunction and Chronic

Kidney Disease. *JAMA Network Open*, **8**, e2458608. https://doi.org/10.1001/jamanetworkopen.2024.58608

[54] Darmon, M., Schnell, D. and Schneider, A. (2022) Monitoring of Renal Perfusion. *Intensive Care Medicine*, **48**, 1505-1507. https://doi.org/10.1007/s00134-022-06857-0

[55] Singh, B., Cusick, A.S., Goyal, A. and Patel, P. (2025) ACE Inhibitors. StatPearls Publishing.

[56] Ejaz, P., Bhojani, K. and Joshi, V.R. (2004) NSAIDs and Kidney. *Journal of the Association of Physicians of India*, **52**, 632-640.

[57] Kudo, K., Konta, T., Mashima, Y., Ichikawa, K., Takasaki, S., Ikeda, A., *et al.* (2011) The Association between Renal Tubular Damage and Rapid Renal Deterioration in the Japanese Population: The Takahata Study. *Clinical and Experimental Nephrology*, **15**, 235-241. https://doi.org/10.1007/s10157-010-0392-y

[58] Hajek, T., Cullis, J., Novak, T., Kopecek, M., Höschl, C., Blagdon, R., *et al.* (2012) Hippocampal Volumes in Bipolar Disorders: Opposing Effects of Illness Burden and Lithium Treatment. *Bipolar Disorders*, **14**, 261-270. https://doi.org/10.1111/j.1399-5618.2012.01013.x

[59] Yokoyama, Y., Maruyama, K., Yamamoto, K., Omodaka, K., Yasuda, M., Himori, N., *et al.* (2014) The Role of Calpain in an in Vivo Model of Oxidative Stress-Induced Retinal Ganglion Cell Damage. *Biochemical and Biophysical Research Communications*, **451**, 510-515. https://doi.org/10.1016/j.bbrc.2014.08.009

[60] Terrin, N., Schmid, C.H., Griffith, J.L., D'Agostino, R.B. and Selker, H.P. (2003) External Validity of Predictive Models: A Comparison of Logistic Regression, Classification Trees, and Neural Networks. *Journal of Clinical Epidemiology*, **56**, 721-729. https://doi.org/10.1016/s0895-4356(03)00120-3

[61] Lee, J. (2017) Patient-Specific Predictive Modeling Using Random Forests: An Observational Study for the Critically Ill. *JMIR Medical Informatics*, **5**, e3. https://doi.org/10.2196/medinform.6690

[62] Zhang, Z., Zhao, Y., Canes, A., Steinberg, D. and Lyashevska, O. (2019) Predictive Analytics with Gradient Boosting in Clinical Medicine. *Annals of Translational Medicine*, **7**, 152-152. https://doi.org/10.21037/atm.2019.03.29

[63] Yang, H., Chen, Z., Yang, H. and Tian, M. (2023) Predicting Coronary Heart Disease Using an Improved Lightgbm Model: Performance Analysis and Comparison. *IEEE Access*, **11**, 23366-23380. https://doi.org/10.1109/access.2023.3253885

[64] Wang, X., Zhu, T., Xia, M., Liu, Y., Wang, Y., Wang, X., *et al.* (2022) Predicting the Prognosis of Patients in the Coronary Care Unit: A Novel Multi-Category Machine Learning Model Using Xgboost. *Frontiers in Cardiovascular Medicine*, **9**, Article ID: 764629. https://doi.org/10.3389/fcvm.2022.764629

[65] Botalb, A., Moinuddin, M., Al-Saggaf, U.M. and Ali, S.S.A. (2018) Contrasting Convolutional Neural Network (CNN) with Multi-Layer Perceptron (MLP) for Big Data Analysis. 2018 *International Conference on Intelligent and Advanced System* (*ICIAS*), Kuala Lumpur, 13-14 August 2018, 1-5. https://doi.org/10.1109/icias.2018.8540626

[66] Sun, Z., Yu, D., Fang, H., Yang, J., Qu, X., Zhang, J., *et al.* (2020) Are We Evaluating Rigorously? Benchmarking Recommendation for Reproducible Evaluation and Fair Comparison. *Fourteenth ACM Conference on Recommender Systems*, Virtual Event, 22-26 September 2020, 23-32. https://doi.org/10.1145/3383313.3412489

[67] Bizzego, A., Bussola, N., Chierici, M., Maggio, V., Francescatto, M., Cima, L., *et al.*

(2019) Evaluating Reproducibility of AI Algorithms in Digital Pathology with Dapper. *PLOS Computational Biology*, **15**, e1006269. https://doi.org/10.1371/journal.pcbi.1006269

[68] Hernández-Orallo, J., Flach, P. and Ferri, C. (2013) ROC Curves in Cost Space. *Machine Learning*, **93**, 71-91. https://doi.org/10.1007/s10994-013-5328-9

[69] Calders, T. and Jaroszewicz, S. (2007) Efficient AUC Optimization for Classification. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D. and Skowron, A., Eds., *Lecture Notes in Computer Science*, Springer, 42-53. https://doi.org/10.1007/978-3-540-74976-9_8

[70] Wang, S., Li, D., Petrick, N., Sahiner, B., Linguraru, M.G. and Summers, R.M. (2015) Optimizing Area under the ROC Curve Using Semi-Supervised Learning. *Pattern Recognition*, **48**, 276-287. https://doi.org/10.1016/j.patcog.2014.07.025

[71] Alba, A.C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P.J., *et al.* (2017) Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA*, **318**, 1377-1384. https://doi.org/10.1001/jama.2017.12126

[72] Taal, M.W. and Brenner, B.M. (2008) Renal Risk Scores: Progress and Prospects. *Kidney International*, **73**, 1216-1219. https://doi.org/10.1038/ki.2008.36

[73] Vearrier, D. and Grundmann, O. (2021) Clinical Pharmacology, Toxicity, and Abuse Potential of Opioids. *The Journal of Clinical Pharmacology*, **61**, S70-S88. https://doi.org/10.1002/jcph.1923

[74] Breckenridge, A. (1996) A Clinical Pharmacologist's View of Drug Toxicity. *British Journal of Clinical Pharmacology*, **42**, 53-58. https://doi.org/10.1046/j.1365-2125.1996.03762.x

[75] Sales, G.T.M. and Foresto, R.D. (2020) Drug-Induced Nephrotoxicity. *Revista da Associação Médica Brasileira*, **66**, s82-s90. https://doi.org/10.1590/1806-9282.66.s1.82

[76] Barnett, L.M.A. and Cummings, B.S. (2018) Nephrotoxicity and Renal Pathophysiology: A Contemporary Perspective. *Toxicological Sciences*, **164**, 379-390. https://doi.org/10.1093/toxsci/kfy159

[77] Kwong, G.A., Ghosh, S., Gamboa, L., Patriotis, C., Srivastava, S. and Bhatia, S.N. (2021) Synthetic Biomarkers: A Twenty-First Century Path to Early Cancer Detection. *Nature Reviews Cancer*, **21**, 655-668. https://doi.org/10.1038/s41568-021-00389-3

[78] An, G. and Cockrell, C. (2023) Generating Synthetic Multi-Dimensional Molecular-Mediator Time Series Data for Artificial Intelligence-Based Disease Trajectory Forecasting and Drug Development Digital Twins: Considerations. arXiv:2303.09056.

[79] Koratala, A. and Reisinger, N. (2021) POCUS for Nephrologists: Basic Principles and a General Approach. *Kidney360*, **2**, 1660-1668. https://doi.org/10.34067/kid.0002482021

[80] Amerling, R., Glezerman, I., Savransky, E., Dubrow, A. and Ronco, C. (2003) Interventional Nephrology and Dialysis: Continuous Flow Peritoneal Dialysis: Principles and Applications. *Seminars in Dialysis*, **16**, 335-340. https://doi.org/10.1046/j.1525-139x.2003.16065.x

[81] Mendelssohn, David, C., Barrett, B.J., Brownscombe, L.M., Ethier, J., Greenberg, D.E., Kanani, S.D., Levin, A. and Toffelmire, E.B. (1999) Elevated Levels of Serum Creatinine: Recommendations for Management and Referral. *Canadian Medical Association Journal*, **161**, 413-417.