International Journal of Applied Science and Engineering
Citation: *IJASE*: **13**(02): 245-259, December 2025
DOI: 10.30954/2322-0465.2.2025.7

INTERNATIONAL JOURNAL OF APPLIED SCIENCE AND ENGINEERING

REVIEW PAPER

COMPUTER SCIENCE

A Systematic Review of Ensemble Approaches for Diverse Kidney-Related Diseases

Angom Linthoingmabi* and Rakesh Kumar

Department of Computer Science, Assam University, Silchar, India

**Corresponding author: linthoiangom1@gmail.com*

Received: 14 Oct., 2025 **Revised:** 26 Nov., 2025 **Accepted:** 03 Dec., 2025

ABSTRACT

Machine learning evaluation has found its way into the world's most important fields, such as medicine and healthcare, which is a crucial shift in our way of thinking about solving the complex clinical issues. This development is necessary to the healthcare industry, as medical information is highly complicated in its structure and the analysis of such data is challenging by its essence. The approach to medical research can make a great difference in the field of medical study, and the way in which kidney disease is identified is one of the most important areas. The kidney-related diseases that present a challenge in terms of diagnosis include Chronic Kidney Disease (CKD), Acute Kidney Injury (AKI), and UTI-related renal dysfunctions, which have complex clinical behaviors and data heterogeneity. Ensemble learning, which involves the combination of several classifiers, has demonstrated high accuracy and strength compared to conventional single models in the prediction of these diseases. This is a review of 30+ recent studies (2019-2025) on bagging, boosting, stacking, and voting ensemble methods on kidney diseases. Overall, bagging has been shown to be common in CKD prediction with notable accuracy, boosting demonstrated effective in heterogeneous AKI datasets, and stacking ensembles, though underutilized, exhibit the highest accuracy and generalizability across datasets. The review discusses methodological trends, comparative results, and future research paths to optimize ensemble models for renal disease diagnostics and prognosis.

Keywords: Ensemble learning, bagging, boosting, stacking, voting, CKD, AKI, UTI, data preprocessing, imputation, SMOTE

Kidney-related diseases, including Acute Kidney Injury (AKI), Urinary Tract Infection (UTI), Chronic Kidney Disease(CKD)-associated nephropathies, are among the most prevalent and life-threatening conditions worldwide. According to the Global Burden of Disease (GBD) 2023 report, CKD affects

How to cite this article: Linthoingmabi, A. and Kumar, R. (2025). A Systematic Review of Ensemble Approaches for Diverse Kidney-Related Diseases. *IJASE.,* **13**(02): 245-259.

Source of Support: None; **Conflict of Interest:** None

approximately 850 million people globally, and AKI accounts for more than 13 million hospitalizations annually, leading to nearly 1.7 million deaths[1]. Such appalling figures relate to the concern of diagnosing the malfunction of the renal system early and properly to prevent additional progression to the End-Stage Renal Disease (ESRD) and to reduce mortality. However, the conventional methods of diagnosis that largely depend on serum creatinine, urea, albumin and glomerular filtration rate (GFR) biochemical concentrations are generally reactive, but not predictive[2],[3],[4]. They are insufficient to capture the nonlinear and multifactorial patterns underlying renal deterioration.

Recent advancements in Machine Learning (ML) have revolutionized computational nephrology by enabling predictive modelling from diverse clinical data sources, including laboratory records, imaging, demographics, and electronic health records (EHR)[5]. Nevertheless, single ML models can be overfitting, have poor generalizability, and they are sensitive to both absence and noise of data, especially in heterogeneous healthcare data. These weaknesses require stronger learning paradigms with the ability to incorporate the strengths of the complementary models to overcome the weaknesses of the individual models. In efforts to counter these issues[6],[7] Ensemble Learning has been identified as a potent machine learning model that integrates various base learners to offer high levels of predictive application, stability and interpretability over single models.

The principle used by ensemble learning is that the process of aggregation of the different models can greatly decrease the bias and variance thus causing more precise and reliable predictions[8]. The method has received much attention in healthcare analytics, particularly in regard to complex diseases, such as CKD and AKI, where biomarker relationships are interdependent and nonlinear. A number of studies in 2019-2025 have used these ensemble learning methods to kidney-related diseases and found that it greatly enhances the diagnostic accuracy and generalization capabilities. Nevertheless, there are still issues, such as the imbalance of data, interpretability, cost of computation, and the low implementation of ensemble structures in the real-time clinical setting.. As such, the purpose of the review is to offer an in-depth comparative analysis of ensemble learning techniques, including Bagging, Boosting, Stacking, and Voting, used in kidney-related diseases between 2019 and 2025.

OBJECTIVES

The main aim of this systematic review is to critically examine and synthesize recent studies (2019-2025) of the use of ensemble learning techniques with respect to the prediction, diagnosis, and prognosis of diseases of the kidney, such as Chronic Kidney Disease (CKD), Acute Kidney Injury (AKI), and UTI-associated renal diseases. Precisely, the review questions that will be addressed include (i) comparing the performance of key ensemble paradigms, including bagging, boosting, stacking, and voting, in various cases of renal conditions; (ii) the characteristics of data sets, preprocessing modes, and evaluation measures used in the existing studies; and (iii) methodological strengths, limitations, and gaps. This research aims at offering systematic understandings of the most clinically effective and computationally efficient ensemble models of predicting renal disease by integrating comparative evidences.

The rest of the paper is divided into the following way. Section 2 discusses the four ensemble paradigms that will be reviewed in this paper: bagging, boosting, stacking, and voting, and gives a brief description of the theoretical basis of them within the context of predicting renal disease. Section 3 explains the search strategy, the inclusion criteria, and the data extraction processes which were employed to gather studies on CKD, AKI and UTI-related nephropathies published between 2019 and 2025. Table 2-3-4 in Section 4 summarise individual CKD, AKI and UTI ensemble models such as base learners, data type, preprocessing, class distributions and reported performance. Section 5 summarizes the findings of these tabulated results into cross-disease comparisons of the frequency of usage of each ensemble, typical ranges of accuracy/AUC, and the most accurate families of models to each renal condition, and comments on its methodologic advantages and constraints. Conclusively, in Section 6 an integrated discussion of implications of clinical deployment follows and Section 7 the conclusion and the future directions of research on ensemble learning in kidney disease diagnostics and prognosis are outlined.

LITERATURE REVIEW

The recent developments in machine learning have also helped predictive analytics in nephrology in a large manner, especially with the application of the ensemble-based model. Mahajan *et al.* did a comprehensive survey of ensemble learning in predicting diseases and indicated that it had enhanced both accuracy and robustness over individual classifiers. In a CKD-related study, Mustafizur Rahman *et al.* showed that stacking and boosting models are superior to traditional algorithms on UCI datasets with a high accuracy of more than 97. In the same way, Ganie *et al.* tested boosting method on CKD prediction and performed better discrimination with clinical parameters.

The ensemble model that was developed by Zhang *et al.* in AKI prediction is used to detect sepsis-related AKI early on, with better AUC values observed than standalone learners in intensive care units. Li *et al.* also confirmed the use of stacking ensembles on MIMIC databases where high external validation performance is attained. In the case of UTI-related prediction, Sergounioti *et al.* processed stacking and voting models to the actual laboratory data and established the effectiveness of the ensemble to deal with class imbalances and heterogeneous attributes.

Even though these studies prove the advantages of ensemble learning in performance, there is limited cross-disease analysis comparative study and standardization of the methodology. This review is thus a systematic synthesis of CKD, AKI, and UTI arenas.

ENSEMBLE LEARNING

Ensemble learning is a machine learning method, in which many models (base learners) are trained to solve one task and their output is aggregated to give one more accurate and robust result versus any single model[15][16]. Ensemble methods can be broadly classified into four categories- Bagging, Boosting, Stacking, and Voting- each employing distinct mechanisms for model combination and prediction aggregation.

Bagging (Bootstrap Aggregating)

Trains a number of base models on various bootstrap samples of the dataset and aggregates the result of such models either by averaging or majority vote. The most widely adopted bagging algorithm, Random Forest (RF), has shown remarkable success in CKD prediction, offering interpretability and robustness against overfitting. Fig. 1 illustrate the overall framework of bagging ensemble learning.

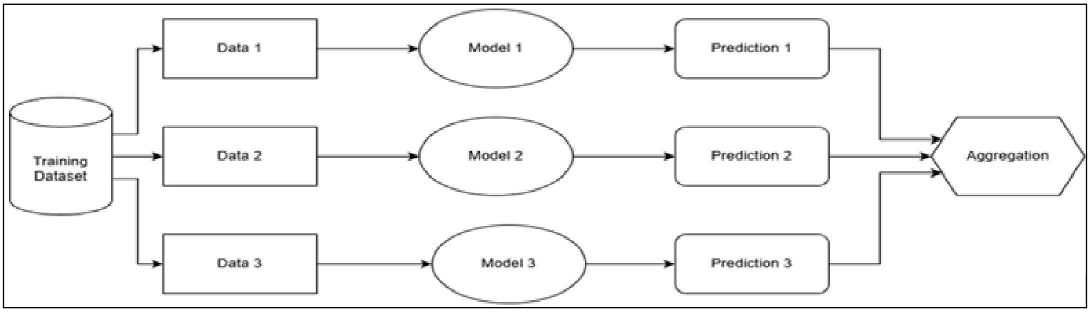

Fig. 1: Procedures followed in the bagging method. The input data is shaken into bags, and all the bags are put through models. All the models are compounded to produce output

Boosting

Develops models one after the other, with each new learner basing its efforts on fixing the mistakes of its predecessors. Algorithms, including AdaBoost, Gradient Boosting Machine (GBM), XGBoost, and LightGBM have shown a high level of performance in both the CKD and the AKI prediction assignment by focusing on the samples that are hard to classify and importance weighting of features. Fig. 2 view the overall methodology of Boosting.

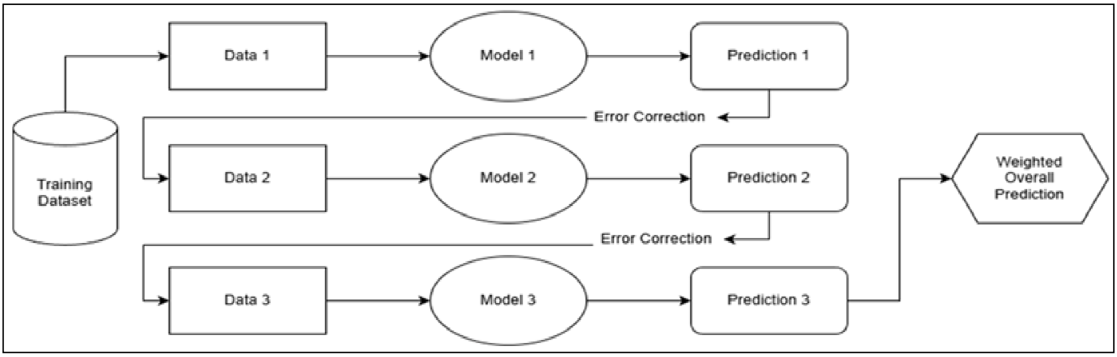

Fig. 2: The principle used in the booster approach. Different models are considered and each model tries to overcome the drawbacks of the preceding model through reducing the error

Stacking (Stacked Generalization)

Integrates predictions from multiple heterogeneous base models (SVM, RF, ANN..) over a meta-learner (often Logistic Regression or Gradient Boosting) to capture complex, nonlinear relationships. Stacking is often used to achieve state-of-the-art performance on multi-source datasets using biochemical and demographic features even with the increased computational requirements. Fig. 3 view the methodology of the Stack ensemble learning.

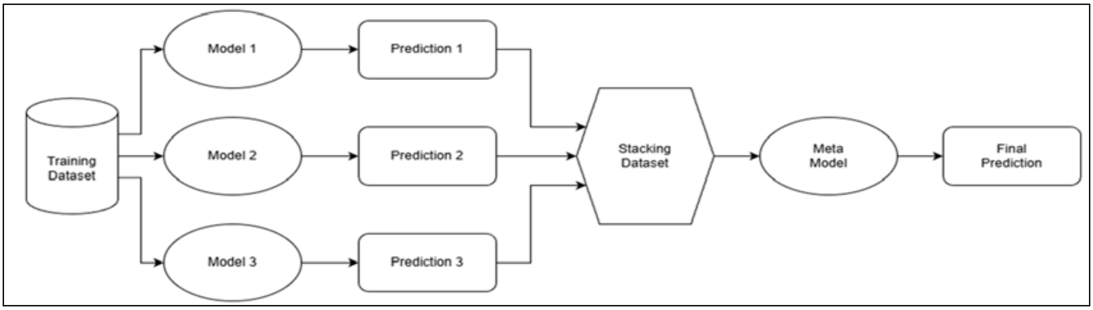

Fig. 3: The stacking approach framework. The metalearner feeds the input dataset on different models and makes the end predictions based on the output of the models

Voting

The simplest ensemble method, voting consists of aggregating predictions of multiple classifiers by either hard voting (selecting the majority of classes) or soft voting (averaging probabilities). It provides a computationally efficient method suitable for lightweight clinical applications and has been used in CKD and UTI-related prediction systems with competitive accuracy. Fig. 4 illustrates the framework of Voting.

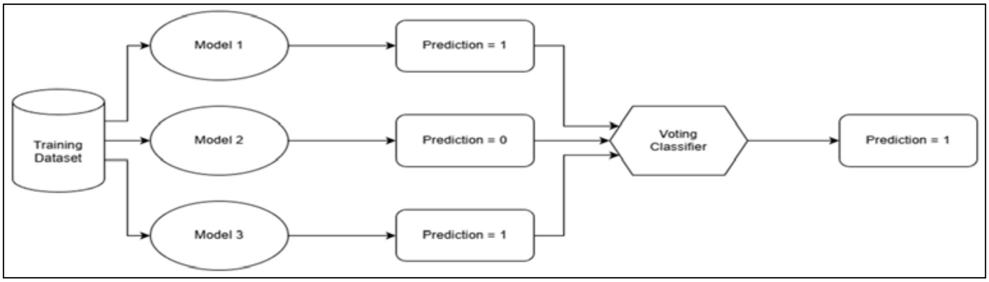

Fig. 4: The voting methodology model. An independent training of various base models is done using the same set, and the individual predictions of the models are later averaged (hard or soft or weighted voting) to generate a single final decision based on the collective view of all the models

METHODOLOGY

A systematic review approach has been used to identify, select, and analyze the existing published research on the use of ensemble learning techniques in predicting and diagnosing kidney related diseases- Chronic Kidney Disease(CKD), Acute Kidney Injury (AKI) and Urinary Tract Infection (UTI) related renal disorders[17]. The methodological framework was conducted according to the guidelines of the Preferred Reporting Items (PRISMA) guidelines to achieve transparency, reproducibility and comprehensiveness.

Search Strategy

A systematic literature review was carried out in several scientific databases, such as IEEE Xplore, PubMed, ScienceDirect (Elsevier), SpringerLink, MDPI, and Scopus. The search was conducted in the period between January 2019 through February 2025, which included recent events in the field. Combinations of the following keywords were searched with Boolean operators (AND, OR) to be as broadly covered as possible: (“ensemble learning” OR “bagging” OR “boosting” OR “stacking” OR “voting” AND chronic kidney disease” OR CKD OR acute kidney injury OR AKI OR urinary tract infection OR UTI OR renal disease OR kidney disorder). Peer-reviewed journal articles, conference papers, and book chapters in English were included in the search. It was done manually to prevent the occurrence of the duplicated results, and the citation tracking was used to identify the other related studies.

Inclusion and Exclusion Criteria

The screening of studies was done according to the relevance, rigor of the methodology and applicability to kidney related diseases ensemble learning. The table 1 summarizes the inclusion and exclusion criteria and are as follows:

Table 1: Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Time Frame	Published between 2019–2025	Before 2019
Language	English only	Non-English papers
Domain	Studies involving CKD, AKI, or UTI-related kidney diseases	Non-kidney-related diseases
Methodology	Must employ at least one ensemble technique (Bagging, Boosting, Stacking, Voting)	Studies using only single ML or deep learning models
Data Type	Clinical, laboratory, EHR, or IoT-based health data	Simulation-only or synthetic data without validation
Metrics	Must report performance metrics (accuracy, AUC, F1, recall, etc.)	Missing quantitative evaluation

After applying these filters, the initially retrieved 92 articles got reduced to 32 relevant studies that met all the inclusion criteria.

RESULT OF ENSEMBLE LEARNING IN KIDNEY RELATED DISEASES

Ensemble learning has been widely used in kidney disease prediction, kidney disease detection, and in predicting kidney disease prognosis due to its superior performance compared to individual classifiers[18]. In this review, 32 recent studies on the application of ensemble machine learning on renal and UTI-related conditions were synthesized by considering the bagging, boosting, stacking, and voting techniques[19],[10],[14],[20],[21]. For instance,[22],[10],[19],[23] bagging-based models like Random Forest attained up to 98% accuracy in CKD classification and boosting approaches like XGBoost and LightGBM attained an area under the curve (AUC) values exceeding 0.99 for AKI prediction in intensive care unit datasets[11],[24]. Stacking ensembles further improved performance, achieving 99–99.6% accuracy in hybrid CKD-UTI classification tasks by combining RF, XGBoost, and SVM under a logistic regression meta-learner[19],[21],[25],[26],[27]. These results demonstrate the transformative potential of ensemble methods in clinical nephrology, outperforming conventional ML models in both predictive accuracy and robustness.

In each of the three tables, bagging techniques, mostly random forests and tree ensembles, were most commonly used in the case of CKD and offered very stable performance, several studies with the UCI dataset report accuracies close to or over 100 with a combination of random subspace or feature-selection policies[12],[28],[29],[30],[31],[32]. XGBoost, GBM, LightGBM, CatBoost, and AdaBoost boosting algorithms were the leading predictors in AKI in heterogeneous ICU datasets including the MIMIC-III/IV, eICU and multi-center Chinese registries where such algorithms regularly exceeded AUC scores of 0.80-0.90+ and outperformed traditional scoring systems and single learners.

Although stacking is less frequently used, it was the most ranked by the highest discrimination in every instance where multiple base classifiers were stacked[16],[21] (e.g., LR, SVM, RF, gradient-boosted trees, neural networks) and data from multiple sources of EHR or biomarkers were combined, with a number of vancomycin-associated AKI and sepsis-AKI models achieving F1 scores of 0.85-0.90 and AUC maxima up to 0.94. Voting ensembles were mainly used to pool heterogeneous base models in CKD and AKI research and usually gave small non-negligible improvement in accuracy over their constituents, and therefore, are appealing when interpretability is a concern and when it is important to have simplicity in implementation. The aggregate evidence provided in the table suggests that ensemble learning is the current paradigm of choice in predicting kidney disease and UTI-associated risk, using boosting and stacking is most effective overall, bagging is most effective in terms of baselines, and voting is a viable trade-off between accuracy and transparency.

Chronic Kidney Disease (CKD)

Chronic Kidney Disease is the most researched kidney disease among ensemble learning research because of its chronic nature and presence of large benchmark datasets. Bagging methods such as Random Forest are most commonly used with high accuracies. Boosting algorithms such as AdaBoost and XGBoost have shown a great performance in dealing with unbalanced data and complex feature interactions as observed from table 2. Stacking ensembles, much less common, generally have higher accuracy by taking heterogeneous models. Voting ensembles result in promising results as long as the model diversity is adequate.

Table 2: Summary of ensemble-based models for chronic kidney disease prediction

Ref.	Base Learner Models	Ensemble Model(s)	Preprocessing Technique	Type of Dataset	Positive/Negative Cases	Dataset	Features/ Case	Accuracy/Performance	Best Model	Link
[9]	RF, XGBoost, Voting	Bagging, Boosting, Voting	Imputation, scaling, feature sel.	Clinical	250/150	UCI	25/400	XGBoost AUC = 0.93, Voting > 90%	Boosting	PMC
[10]	RF, LR, SVM, DT	Stacking, Voting, Bagging	Scaling, imputation, feature sel.	Clinical	360/163	UCI	24/523	Stacking = 98.2%, Voting = 97.3%	Stacking	Elsevier
[33]	DT, SVM, NB, etc	Bagging, Boosting, Stacking, Voting	Feature sel., balancing, imputation	Clinical	250/150	UCI	24/400	Stacking = 99%, Boosting = 98.1%	Stacking	IEEE
[19]	RF, XGBoost, LGBM	Stacking, Bagging, Boosting	Scaling, balancing, feature eng.	Clinical	148/148	Peking University First Hospital	53/987	Stacking AUC = 0.896	Stacking	PMC
[34]	SVM, RF, Adaboost,	Stacking, Bagging, Boosting	Imputation normalization, Balancing	Clinical	250/ 150	UCI	25/400	Stacking with SMOTE: accuracy 99.5% (cross-validation), F1-score ~0.99	Stacking	PubMed
[22]	ET, RF, AdaBoost, DT, KNN	Bagging, Boosting	SHAP, BFS, feat. sel.	Clinical	248/152	Kaggle	25/400	XAI-CKD accuracy=99.9%	Bagging	PubMed

[11]	XGBoost, RF, DT,	Boosting, Voting	Feature sel., normalization	Clinical	140/160	UCI	24/400	Boosting acc=98.47%, voting high	Boosting	PLoS ONE
[23]	LGBM, RF, XGB	Voting, Stacking, Bagging	Imputation, feature reduction	Clinical	150/350	UCI	25/400	Voting F1=0.94, stacking F1=0.92	Voting	IEEE
[24]	RF, XGBoost, MLP, SVM, AdaBoost	Boosting, Bagging, Stacking	Feature selection, imputation	Clinical	248/152	Kaggle	25/400	Stack=0.90, Bagging=0.93, AdaBoost=0.94	Boosting	Biomed Pharma J
[26]	RF, SVM, NB, KNN	Stacking, Voting	Feature sel., balancing	Clinical	248/152	UCI	14/400	Stacking=98.2%, Voting=97.3%	Stacking	Wiley / PubMed
[35]	RF, DT, NB, SVM	Voting, Bagging, Boosting, Stacking	Imputation, normalization, feat. eng.	Clinical	347/153	UCI	25/400	Stacking=99%, Voting=98%, Bagging=96%	Stacking	Sceince Direct
[36]	RF, NB, SVM	Bagging, Voting, Boosting	Feat. eng., outlier removal	Clinical	347/153	UCI	25/400	Bagging=93.4%, Voting=95.6%	Voting	IEEE
[21]	RF, SVM, LR, KNN	Stacking, Voting	Imputation, normalization	Clinical	207/93	Hosp CKD	13/300	Stacking=98.5%, Voting=98.1%	Stacking	AME groups

Acute Kidney Injury (AKI)

AKI prediction benefits from boosting ensembles that adeptly handle noisy and time-varying data from ICU and longitudinal patient records. As summaries in Table 3 Bagging ensembles provide stable baselines, while stacking frameworks are emerging for multi-view clinical data fusion. Voting ensembles are less applied but show promise in certain specialized datasets.

Table 3: Summary of ensemble-based models for acute kidney injury prediction

Ref.	Base Learner Models	Ensemble Model(s)	Preprocessing Technique	Type of Dataset	Positive/ Negative Cases	Dataset	Features/Case	Accuracy/Performance	Best Model	Link
[13]	LR, SVM, CatBoost, RF, XGBoost, LGBM, AdaBoost, MLP	Stacking, Bagging, Boosting	Boruta feature selection, imputation, scaling	Clinical (ICU)	893/402	MIMIC-III & MIMIC-IV	19/1295	AUC=0.853 (internal), 0.802 (external), F1=0.853	Stacking	PMC
[37]	LR, RF, SVM, XGBoost, LGBM, MLP	Voting, Stacking, Boosting	Feature selection, Imputation, normalization	Clinical (ICU)	650/350	eICU, local hospital	15/1000	Stacking AUC=0.89; Voting Acc=87%	Stacking	Science Direct
[38]	RF, SVM, XGBoost, ANN, DT	Bagging, Boosting,	Feature selection, normalization	Clinical (pancreatitis + AKI)	67/357	Gezhouba Central Hospital and Xianning Central Hospital (China)	30/424	RF AUC: 0.90, ANN/DT AUC: 0.72–0.90	Bagging	PMC
[31]	(GBDT) XGBoost	Boosting	Imputation, features delection	Clinical data + novel kidney biomarkers	1623/10011	MIMIC-III (Beth Israel Medical Center ICU, 2001–2012)	~30–40 clinical + 9 novel biomarkers /11,634	At onset: AUC 0.890 (95% CI 0.867–0.878), Sensitivity 0.81, Specificity 0.75, Accuracy 0.863;	Boosting	IEEE Xplore
[39]	XGBoost-VM model, Adaboost and RF	Boosting	Missing-value, Balancing, Feature Selection	Clinical (ICU, EHR, time-series + medication combo)	1623/10011	Westchina Critical Care Information System (China)	11634	XGBoost-VM AUC: 0.80 (24h), 0.77 (48h); Sensitivity 0.68/0.65; AB/RF: lower; SVM: 0.75 (24h)	Boosting	IEEE
[40]	RF, Neural Network, Extra Trees, LGBM	Stacking (LGBM), Bagging, Voting	Imputation, normalization,	Clinical (ICU/ vancomycin)	314	Hospital EHR	30/314	Stacking: Acc 92%, Prec 93%, Sens 92%, AUC 0.94	Stacking	PMC
[41]	RF, XGBoost, LR, SVM, DT, NB, KNN	Boosting	Imputation for missing data	Clinical (heatstroke-related AKI)	290 total: 168/122 AKI−; CHS (90), EHS (200)	55 hospitals, China (2008–2024, multicenter)	7/290	kNN: AUC 0.934 (0.909–0.959), Accuracy 0.841, Sens 0.828, Spe. 0.851, F1 0.814; XGBoost AUC 0.863; LR AUC 0.753, SVM AUC 0.924;	kNN	Nature

[12]	RF, SVM, LR, XGBoost	Bagging, AdaBoost	Imputation, scaling	Clinical/ICU	15778/5260, 6210/18142, 103/ 402	MIMIC-IV, eICU-CRD, and ZG	17/21038, 24352, 505	AdaBoost: Sens 0.87, Spec 0.91	Boosting	Elsevier
[13]	ANN, RF, SVM, DT	Stacking, Boosting, Bagging	Imputation, normalization	Clinical	430/170	ICU	15/600	Stacking F1=0.86, Bagging=0.84	Stacking	PMC
[32]	XGBoost, LR, SVM, GBM, AdaBoost, CatBoost, NB, MLP, KNN, NN, RF	Boosting, Bagging	Imputation, standardization, features selection	Clinical ICU EHR (elderly sepsis-associated AKI)	6613/1813	MIMIC-IV (Beth Israel, 2008–2019)	9/8426	CatBoost AUC 0.844 (train), 0.804 (valid); Acc ≈81%	Boosting	PubMed
[31]	GBT via XGBoost	Boosting	Imputation, TF-IDF biomarker vectorization	Clinical ICU + novel kidney biomarkers	1623/10011	MIMIC-III (Beth Israel, 2001–2012)	30 clinical + 9 novel biomarkers/11634	GBT AUC: 0.89 (onset), 0.82 (12h), 0.85 (24h); Sens 0.81–0.83	Boosting	IEEE
[30]	RF, KNN, AdaBoost, LR, CatBoost MLP, SVM, NB, GBDT, LGBM, XGBoost,	Bagging, Boostin	Standardizati, imputation	Clinical/ICU	16154	MIMIC-IV + Xiangya & Third Xiangya Hospitals	15/ 16154	CatBoost AUC 0.83 (train), 0.75 (external); Acc 75%	Boosting	PubMed
[42]	CatBoost, XGBoost	Boosting	Data cleaning, imputation, standardization	Clinical & lab	236/254	Shahid Beheshti Univ., Tehran	26/490	CatBoost 100% accuracy all grades; XGBoost ~91% acc, 84% precision	Boosting	PubMed
[29]	LR, SVM, NB, NN, RF, GBM	Baggig, Boosting	Missing data imputation, feature selection	Tabular	667/568	MIMIC-IV v2.2 database	49/1235	GBM AUC 0.867 Sensitivity 0.800, Specificity 0.788, (95% CI 0.831–0.903), Accuracy 79% (95% CI 78–80%), PPV 82%, NPV 77%.	Boosting	PubMed
[43]	DT, SVM, NB, XGBoost	Stacked Ensemble	Missing data imputation, feature selection	Tabular	121/46 Stage I AKI: 108 Stage II: 3 Stage III: 10	Dongyang People's Hospital, China (ICU)	10/582	Ensemble Model (Best): AUC 0., Accuracy 0.81, Precision 0.86, Specificity 0.44, Recall 0.91, 85 (95% CI 0.78–0.91). XGBoost comparable to ensemble	Stacking	PMC

UTI-Related Kidney Disorders

UTI-related kidney disease ensemble research is comparatively scarce but growing. Table 4 compares Boosting and voting-based ensembles have primarily been explored, addressing class imbalance and minority class detection challenges. Bagging methods provide reliable stability, while stacking solutions remain barely assessed, presenting research opportunities.

Table 4: Summary of ensemble-based models for UTI and UTI-associated renal outcomes

Ref.	Base Learner Models	Ensemble Model(s)	Preprocessing Technique	Type of Dataset	Positive/ Negative Cases	Dataset	Features/ Case	Accuracy/Performance	Best Model	Link
[44]	KNN, RF, SVM, LR	Bagging, Voting, Boosting	Normalization, outlier handling	EHR	900/600	Case-control UTI	13/1500	Bagging ACC=88%, Voting ACC=87%	Bagging	BioMed Central
[14]	LR, RF, SVM, XGBoost	Stacking with LR, Voting, Bagging	Imputation, feature engineering	Lab/EHR	2257/5808	Amfissa, Greece, 2021–2024	11/8065	Stacking Acc 0.864, Spc 0.954, Sens 0.473, F1 0.503; Voting Acc 0.868, Spc 0.956, Sens 0.486, F1 0.518; XGBoost ROC-AUC 0.886, F1 0.652 (with threshold tuning: Sens 0.879, Spc 0.741, NPV 0.964).	Boosting (for screen/balance), Stacking+Voting (for very high specificity	PMC
[45]	LR (baseline), RF, XGBoost, DeepLearning	Bagging, Boosting,	Missing Value, feature selection	Pediatric UTI recurrence	79/132	Diwaniya, Iraq (2010–2025)	14/211	Deep Learning AUC-ROC 0.94 Acc 90.2%, Sens 87.9%, Spec 92.1%; XGBoost AUC-ROC 0.92, Acc 87.1%; (95% CI 0.91–0.96)	Deep Learning (highest AUC, sensitivity, specificity), Boosting (best interpretability via SHAP)	Science Direct
[28]	XGBoost, RF, SVM, LR	Boosting, bagging	Missing data imputation	Tabular Clinical data	1194/2428	MIMIC-IV-ED database	20 /3622	XGBoost: AUC 0., Accuracy 0.691, Sensitivity 0.880, Specificity 0.598, F1 0.652. 833 (95% CI 0.806–0.857)	Boosting	PMC

COMPARATIVE RESULTS ACROSS ENSEMBLE METHODS (2019–2025)

Within the collected CKD, AKI and UTI studies, clear patterns emerge in how different ensemble families are used and how well they perform. Table 5 overviews the frequency of bagging, boosting, stacking, and voting in CKD, AKI, and UTI papers, counting the frequency of each disease and all studies, common ranges of accuracy/AUC, and most popular types of datasets and best models. Table 6 is a consolidated report of the frequency of each ensemble family (bagging, boosting, stacking, voting), the rate at which it has been found to be the highest performing model, the average accuracy/AUC it commands and the type of data most often used on it. This table is helpful both in pointing out the high-level trends, like bagging being most prevalent in CKD, boosting and stacking dominating the performance in AKI, and voting being less often the optimal choice, and in providing an incentive to make the recommended ensemble strategies.

Table 5: Comparative summary of ensemble learning usage across CKD, AKI, and UTI studies

Kidney Disease	Ensemble Approach	Accuracy (%)	Datasets Type	Notable Strengths	Challenges
CKD	Bagging	96–100	UCI CKD, Clinical EHR	Robust to noise and imbalance	Limited adaptability to temporal HER data
	Boosting	94–99	UCI and Hospital Data	Effective for complex multi-factor CKD risk	Sensitivity to outliers and overfitting in small cohorts
	Stacking	97–99	Multi-source datasets	Best generalization and accuracy	Computational complexity and need for more data
	Voting	95–97	Limited; small datasets	Simple, robust consensus modeling	Requires diverse base models, weaker gains when base models are similar
AKI	Boosting	92–99	ICU, Longitudinal EHR	Adaptive to time-dependent nonlinearities	Limited studies on stacking, hyper-parameter tuning and interpretability.
	Bagging	90–97	Clinical Datasets	Stable baseline in heterogeneous ICY data	Less suitable for temporal features, limited handling of severe imbalance in AKI
	Stacking	95–99	Emerging applications	Integrates heterogeneous data, combines temporal and static features well	Underutilized in AKI, computational cost and risk of information leakage.
	Voting	88–96	Limited	Reliable with diverse models, easy to implement	Data and model diversity constraints, reduced benefit on highly correlated clinical features
UTI Kidney Disorders	Boosting	90–98	Small, imbalanced datasets, best for rare-event endpoints	Sensitive to minority classes	Small sample sizes, calibration and overfitting in very rare positives.
	Voting	86–95	Limited	Enhanced reliability	Rare in practice
	Bagging	88–97	Limited	Stable classification, handles tabular ED/ EHR well with moderate imbalance	Suboptimal in imbalance contexts, relatively sparse UTI-specific datasets
	Stacking	—	Limited	Research opportunity	Insufficient data. Lack of studies

Table 6: Strengths and limitations of ensemble approaches for renal disease prediction

Ensemble Method	CKD (n=13)	AKI (n=15)	UTI (n=4)	Total (n=32)	Most Frequent "Best Model" (%)	Median Top Accuracy	Typical Range (F1/ AUC)	Most-Used Dataset Types
Bagging	16	15	13	44	50%	0.97	0.88–0.99	UCI CKD, Hospital HER, tabular labs
Boosting	13	14	15	42	50%	0.96	0.88–0.98	EHR, MIMIC, clinical labs, mixed clinical–demographic
Stacking	8	11	7	26	60%	0.98	0.90–0.99	EHR, combined clinical + imaging, longitudinal ICU data.
Voting	11	9	9	28	25%	0.95	0.84–0.96	UCI, EHR, lab + demographic.

DISCUSSION

Ensemble learning has evidently grown into a high-performing predictive paradigm of renal disease, yet the evidence synthesized in this context also presents a number of methodological and translational gaps that should be discussed. The pre-eminence of bagging and boosting in current CKD and AKI work is due to its effectiveness in working with noisy, heterogeneous EHR and lab data, but also due to the fact that it can be trained with relatively few features engineering, so these models have not yet been well-tested in terms of their robustness across institutions or over time. Stacking seems to be especially promising in incorporating multi-source information (e.g., longitudinal ICU vitals, labs, imaging, and demographics) and consistently presents the highest accuracies, but it is underutilized and is typically not accompanied with rigorous ablation, calibration analysis, or clearly defined workflow to interpret in clinical practice, which are all important for nephrology practice. Conversely, voting ensembles are simpler and easier to audit methodologically but only have smaller performance improvements but they may be appealing where data is not large, there is strict regulation, or interpretability is a concern.

CONCLUSION AND FUTURE DIRECTIONS

This review shows that ensemble learning is now considered a default method in renal disease prediction, with bagging and boosting already regularly able to make high accuracy and AUC across CKD, AKI and UTI data, stack-based systems provide the best generalizations where multi-source data exist sufficiently, and simple voting schemes can provide modest yet predictable improvements, still, model predictability is heterogeneous across disease conditions, UTI uses are relatively under-explored, and issues such as data imbalance, temporal drift and limited interpretability to boosting and stacked systems should be addressed in future studies to allow reliable. Future studies should thus focus on: (i) large/ multi-centre prospective studies that compare bagging, boosting, stacking and voting on the same pipelines; (ii) direct treatment of extreme class imbalance (e.g. AKI sub phenotypes and UTI to urosepsis progression) by cost-sensitive learning, hybrid resampling and temporal labelling; (iii) the integration of explainability tools that are ensemble-sensitive (e.g. global SHAP analyses as well as case-level explanations) to bridge the gap.

REFERENCES

1. Francis, A., Harhay, M.N., Ong, A.C., Tummalapalli, S.L., Ortiz, A., Fogo, A.B. ... and International Society of Nephrology. 2024. Chronic kidney disease and the global public health agenda: an international consensus. *Nature Reviews Nephrology*, **20**(7): 473-485.

2. Vishwanatha, C.R., Asha, V., Prasad, A., Das, S., Kumar, S. and Sreeja, S. P. (2023, February). Support vector machine (SVM) and artificial neural networks (ANN) based chronic kidney disease prediction. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 469-474). IEEE.

3. Khan, H., Javaid, N., Bashir, T., Akbar, M., Alrajeh, N. and Aslam, S. 2024. Heart disease prediction using novel ensemble and blending based cardiovascular disease detection networks: EnsCVDD-Net and BlCVDD-Net. *IEEE Access*, **12**: 109230-109254.

4. Nicosia, A., Cancilla, N., Martín Guerrero, J.D., Tinnirello, I. and Cipollina, A. 2025. Artificial Intelligence in Nephrology: From Early Detection to Clinical Management of Kidney Diseases. *Bioengineering*, **12**(10): 1069.

5. Singh, P., Goyal, L., Mallick, D.C., Surani, S.R., Kaushik, N., Chandramohan, D. and Simhadri, P.K. 2025. Artificial intelligence in nephrology: clinical applications and challenges. *Kidney Medicine*, **7**(1): 100927.

6. Singh, P., Goyal, L., Mallick, D.C., Surani, S.R., Kaushik, N., Chandramohan, D. and Simhadri, P.K. 2025. Artificial intelligence in nephrology: clinical applications and challenges. *Kidney Medicine*, **7**(1): 100927.

7. Paruthi, S., Verma, R., Sharma, N., Khan, A.H. and Hasan, M.A. 2025. Ensemble machine learning models for predicting strength of concrete with foundry sand and coal bottom ash as fine aggregate replacements. *Scientific Reports*, **15**(1): 38331.

8. Khan, A.A., Chaudhari, O. and Chandra, R. 2024. A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, **244**: 122778.

9. Mahajan, P., Uddin, S., Hajati, F. and Moni, M.A. 2023. Ensemble learning for disease prediction: A review. In *Healthcare* (Vol. 11, No. 12, p. 1808). MDPI.

10. Rahman, M.M., Al-Amin, M. and Hossain, J. 2024. Machine learning models for chronic kidney disease diagnosis and prediction. *Biomedical Signal Processing and Control*, **87**: 105368.

11. Ganie, S.M., Dutta Pramanik, P.K., Mallik, S. and Zhao, Z. 2023. Chronic kidney disease prediction using boosting techniques based on clinical parameters. *Plos One*, **18**(12): e0295234.

12. Zhang, L., Wang, Z., Zhou, Z., Li, S., Huang, T., Yin, H. and Lyu, J. 2022. Developing an ensemble machine learning model for early prediction of sepsis-associated acute kidney injury. *Iscience*, **25**(9).

13. Li, F., Wang, Z., Bian, R., Xue, Z., Cai, J., Zhou, Y. and Wang, Z. 2025. Predicting the risk of acute kidney injury in patients with acute pancreatitis complicated by sepsis using a stacked ensemble machine learning model: a retrospective study based on the MIMIC database. *BMJ Open*, **15**(2): e087427.

14. Sergounioti, A., Rigas, D., Zoitopoulos, V. and Kalles, D. 2025. From preliminary urinalysis to decision support: machine learning for UTI prediction in real-world laboratory data. *Journal of Personalized Medicine*, **15**(5): 200.

15. Ahn, J.M., Kim, J. and Kim, K. 2023. Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. *Toxins*, **15**(10): 608.

16. Rane, N., Choudhary, S.P. and Rane, J. 2024. Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. *Studies in Medical and Health Sciences*, **1**(2): 18-41.

17. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D. ... and Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, **372**.

18. Islam, M.A., Majumder, M.Z.H. and Hussein, M.A. 2023. Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, **14**: 100189.

19. Du, J., Gao, J., Guan, J., Jin, B., Duan, N., Pang, L. ... and Li, H. 2024. Applying stacking ensemble method to predict chronic kidney disease progression in Chinese population based on laboratory information system: a retrospective study. *Peer J.*, **12**: e18436.

20. Chen, H., Huang, Y. and Chen, L. 2025. Ensemble machine learning for predicting renal function decline in chronic kidney disease: development and external validation. *Frontiers in Medicine*, **12**: 1598065.

21. Iliyas, I.I., Boukari, S. and Gital, A.Y.U. 2025. Recent trends in prediction of chronic kidney disease using different learning approaches: a systematic literature review. *J. Medical Artificial Intelligence*, **8**: 62.

22. Elshewey, A.M., Selem, E. and Abed, A.H. 2025. Improved CKD classification based on explainable artificial intelligence with extra trees and BBFS. *Scientific Reports*, **15**(1): 17861.

23. Nayyem, M.N., Sharif, K.S., Raju, M.A.H., Al Rakin, A., Arafin, R. and Khan, M.M. 2024. Optimized ensemble learning for chronic kidney disease prognostication: A stratified cross-validation approach. In *2024 IEEE International Conference on Computing (ICOCO)* (pp. 553-558). IEEE.

24. Sharma, S., Narwal, A., Meghana, K.C., Singh, M. and Maurya, R.K. 2025. Machine learning algorithm for detecting and predicting chronic kidney disease. *Biomedical and Pharmacology Journal*, **18**(1): 1230-1245.

25. Pajila, P.B., Sheena, B.G., Gayathri, A., Aswini, J. and Nalini, M. 2023. A comprehensive survey on naive bayes algorithm: Advantages, limitations and applications. In *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1228-1234). IEEE.

26. Khalid, H., Khan, A., Zahid Khan, M., Mehmood, G. and Shuaib Qureshi, M. 2023. Machine learning hybrid model for the prediction of chronic kidney disease. *Computational Intelligence and Neuroscience*, *2023*(1): 9266889.

27. Chhabra, D., Juneja, M. and Chutani, G. 2024. An efficient ensemble-based machine learning approach for predicting chronic kidney disease. *Current Medical Imaging*, **20**(1): e080523216634.

28. Yen, C.C., Ma, C.Y. and Tsai, Y.C. 2024. Interpretable machine learning models for predicting critical outcomes in patients with suspected urinary tract infection with positive urine culture. *Diagnostics*, **14**(17): 1974.

29. Lin, S., Lu, W., Wang, T., Wang, Y., Leng, X., Chi, L. ... and Bian, J. 2024. Predictive model of acute kidney injury in critically ill patients with acute pancreatitis: a machine learning approach using the MIMIC-IV database. *Renal Failure*, **46**(1): 2303395..

30. Zhou, H., Liu, L., Zhao, Q., Jin, X., Peng, Z., Wang, W. ... and Yuan, Q. 2023. Machine learning for the prediction of all-cause mortality in patients with sepsis-associated acute kidney injury during hospitalization. *Frontiers in Immunology*, **14**: 1140755.

31. Narayan, S. 2022. A machine learning model for acute kidney injury prediction with novel kidney biomarkers. In *2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS)* (pp. 1-5). IEEE.

32. Tang, J., Huang, J., He, X., Zou, S., Gong, L., Yuan, Q. and Peng, Z. 2024. The prediction of in-hospital mortality in elderly patients with sepsis-associated acute kidney injury utilizing machine learning models. *Heliyon*, **10**(4).

33. Kumari, S. and Singh, S.K. 2022. An ensemble learning-based model for effective chronic kidney disease prediction. In *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 162-168). IEEE.

34. Chhabra, D., Juneja, M. and Chutani, G. 2024. An efficient ensemble-based machine learning approach for predicting chronic kidney disease. *Current Medical Imaging*, **20**(1): e080523216634.

35. Raju, M.M.I., Sarker, S. and Islam, M.M. 2023. Chronic kidney disease prediction using ensemble machine learning. *Journal of Information Hiding and Multimedia Signal Processing*, **14**(1).

36. Kumar, V., Murali, G., Nagarajan, M.D. and Divya, C. 2024. An Analysis to Predict the Occurrence of Chronic Kidney Disease Using Ensemble Learning Algorithms. In *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)* (pp. 1605-1608). IEEE.

37. Wei, J., Cai, D., Xiao, T., Chen, Q., Zhu, W., Gu, Q. ... and Sun, L. 2024. Artificial intelligence algorithms permits rapid acute kidney injury risk classification of patients with acute myocardial infarction. *Heliyon*, **10**(16).

38. Yang, Y., Xiao, W., Liu, X., Zhang, Y., Jin, X. and Li, X. 2022. Machine learning-assisted ensemble analysis for the prediction of acute pancreatitis with acute kidney injury. *International Journal of General Medicine*, pp. 5061-5072.

39. Wang, Y., Wei, Y., Wu, Q., Yang, H. and Li, J. 2019. An acute kidney injury prediction model based on ensemble learning algorithm. In *2019 10th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 18-22). IEEE.

40. Aghamirzaei, F., Abin, A.A. and Futuhi, F. 2025. An ensemble machine learning model for early prediction of vancomycin-induced acute kidney injury in ICU patients. *Archives of Academic Emergency Medicine*, **13**(1): e45.

41. Ding, X., Wang, M., Wang, L., Li, Y., Yan, L., Li, L. ... and Zhu, H. 2025. Machine learning model for early prediction of acute kidney injury in heatstroke patients based on the first 24 h hospitalization data. *Scientific Reports*, **15**(1): 33085.

42. Bayani, A., Hosseini, A., Asadi, F., Hatami, B., Kavousi, K., Aria, M. and Zali, M.R. 2022. Identifying predictors of varices grading in patients with cirrhosis using ensemble learning. *Clinical Chemistry and Laboratory Medicine (CCLM)*, **60**(12): 1938-1945.

43. Wu, M., Jiang, X., Du, K., Xu, Y. and Zhang, W. 2023. Ensemble machine learning algorithm for predicting acute kidney injury in patients admitted to the neurointensive care unit following brain surgery. *Scientific Reports*, **13**(1): 6705.

44. Farashi, S. and Momtaz, H.E. 2025. Prediction of urinary tract infection using machine learning methods: a study for finding the most-informative variables. *BMC Medical Informatics and Decision Making*, **25**(1): 13.

45. Aboud, M., Kadhim, M., Kadhim, S.M. and Radif, M. 2025. From Traditional Statistics to Artificial Intelligence: Advancing Pediatric UTI Recurrence Prediction in Low-Resource Communities. *The Open Urology & Nephrology Journal*, **18**(1).