CKDX-Net: A Novel Cross-Domain Knowledge Distillation Framework from Tree-Based to Neural Architectures for Chronic Kidney Disease Staging with Adaptive Computational Optimization

Md. Musfiqur Rahman Akib, Habibur Rahaman*, Rosni Akter, and Pial Paul

Department of Computer Science and Engineering, Chittagong Independent University, Chittagong-4000, Bangladesh

{22302012, habibcuetcse*, 22102009, 22102022}@ciu.edu.bd

Abstract. The staging of chronic kidney disease (CKD) in low resource settings requires the right balance between diagnostic accuracy and computational efficiency. In this study, we propose CKDX-Net, a novel knowledge distillation pipeline transferring the patterns learned from a high-capacity XGBoost ensemble to a lightweight deep neural network in tasks of six-stage CKD classification. This framework adopts the dual-loss architecture consisting of hard-label cross-entropy and temperature-scaled Kullback-Leibler divergence, accompanied by an adaptive soft label scheduler balancing teacher-student guidance actively in training. We evaluated on 4,000 anonymized patient records sourced from the publicly-available Kaggle CKD dataset, stage 0-5, with a macro-F1 score of 0.918 and retaining the performance over 98.7% compared to the teacher model (99% accuracy); the distilled student model achieves 95% accuracy. The distilled student network speeds up the inference from 105.3 ms to 6.7 ms per record, thus facilitating real-time deployment in clinical pipelines. Temperature-scaled soft labeling achieves stronger knowledge transfer than training from scratch. CKDX Net constitutes the first unified solution to cross-architecture distillation in CKD staging and offers a viable option for deploy ing high-capacity diagnostic AI in resource-stricken healthcare environments.

Keywords: Biomedical applications · Chronic kidney disease · Computational efficiency · Deep neural network · Knowledge distillation · Temperature-scaled soft labeling · XGBoost.

1 Introduction

Chronic Kidney Disease (CKD) is estimated to affect 10% of the world's population and its prevalence has been increasing with related comorbidities such as hypertension and diabetes [1, 2]. The combination of complete clinical evaluation and early diagnosis are important in preventing progression to end-stage

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_17

renal disease. But, traditional diagnostic methods usually cannot reflect the early slight difference [3]. The machine learning and deep learning methods have also been adopted more frequently to CKD classification [4, 5].

Classical methods (e.g., decision trees, random forest) offer interpretability but are limited in modeling nonlinear relationships [6]. Deep learning techniques generally attain better accuracy, however require considerable amount of computation power and may suffer from lack of interpretability [7,8]. Recent reviews stress the importance of rigorous frameworks that balance efficiency, robustness and clinical applicability [9,10]. This paper presents CKDX-Net, a hybrid model utilizing knowledge distillation (KD) to compress an ensemble XGBoost teacher into a deep neural network student. As a result of introducing SMOTE for handling imbalance class, the CKDX-Net combines balanced stage-wise classification with lower computational complexity. The key contributions of this work consist in developing a knowledge distillation approach which allows small neural networks to learn from XGBoost models, while merging ensemble interpretation with deep learning efficiency. We also include stage-informed optimization and SMOTE resampling for enhanced CKD stage recognition. The distilled student network reduces a substantial amount of inference time and keeps the performance gap to less than 4% from teacher model, thus permitting real-time clinical use.

2 Literature Review

Chronic Kidney Disease (CKD) prediction and staging have been increasingly paid attention in medical artificial intelligence (AI), where novel machine learning (ML) techniques as well as deep learning (DL) algorithms are employed. A substantial stream of CKD research has used structured EHR and laboratory datasets with off-the-shelf ML methods to predict CKD onset, progression, and stage transitions. An EHR cohort of 491 patients was analyzed using XGBoost, random survival forests, and Cox models [11]; they identified estimated glomerular filtration rate (eGFR), serum creatinine, and age as dominant predictors and demonstrated that XGBoost outperformed conventional statistical models on discrimination metrics while providing post-hoc variable importance for interpretability but the absence of urinary albumin-to-creatinine ratio, limiting broader generalizability. Age, serum creatinine, and estimated glomerular filtration rate (eGFR) were found to be the most significant predictors of the development of end-stage renal disease (ESRD) by machine learning (ML) using clinical data [12]. Despite its effectiveness, their model's generalizability was limited by the absence of urine albumin values. For CKD stages 3-5, a machine learning-driven nomogram was also developed; it achieved excellent accuracy but had low interpretability for clinical decision-making [13].

Based on longitudinal laboratory data from 710 individuals with chronic kidney disease, an integrated machine learning predictive framework was created, modeling renal function transitions using Random Forest and Gradient Boosting [14]. Although they had high subgroup-level predictions, there was lit-

tle external validation. To predict cardiovascular comorbidities in patients with chronic kidney disease (CKD), a tree based machine learning approach was utilized, with a focus on robustness and interpretability, but with a warning about the computing needs at scale [15]. Using clinical datasets, an application of machine learning to the categorization of chronic kidney disease showed promising results; however, multimodal data integration was absent [6]. Collectively, these works show that traditional ML models have provided interpretability and robustness but they have trouble striking a balance between generalizability and computational efficiency.

For CKD prediction, DL models have drawn attention due to improvements in computing power and data accessibility, especially when data goes beyond structured EHRs. In an evaluation of several DL models, 98.3% accuracy in CKD prediction was attained, demonstrating both the high computational cost and the robustness of DL in feature learning [4]. Convolutional neural networks (CNNs) were used in a study to evaluate CT scans for the classification of chronic kidney disease (CKD), with an accuracy of 95.4% [5]. But their model's scalability was limited by the need for huge annotated datasets. DL architectures yield higher detection accuracy but demand expensive imaging resources, according to a review of AI applications in retinal imaging [8]. Although these studies show that DL can capture complicated feature representations, there are trade-offs, such as high infrastructure costs, limited interpretability, and significant data dependency. To improve prediction and generalization, several studies have combined ML, DL, and survival models. A study that carried out a scoping review pointed out that hybrid ML-DL techniques work well on tabular and image datasets but are frequently computationally demanding [1]. A comparison between smart AI and conventional machine learning techniques reveals that while ensemble models improve the prediction of CKD, they also present interpretability issues [2]. Subgroup-level predictions can be greatly enhanced by multimodal integration of biochemical, clinical, and temporal data; nevertheless, this still necessitates a large amount of computing and a variety of data sources [7, 14]. A notable but underexplored area is the development of lightweight models that retain acceptable performance while being feasible for deployment in resource-limited settings.

Knowledge Distillation (KD) was first proposed by [16]. It is a model compression paradigm, where a small student network learns from the output of a large teacher model. Recent studies and applications have shown how promising KD is for medical imaging. The assessment of continual learning with KD highlighted its potential for incremental improvements to medical models [17]. It was investigated how to distill huge language models for the purpose of predicting health events [18]. For the purpose of predicting the progression of the disease, another study paired KD with variational autoencoders [19]. In order to classify and segment medical images, a residual U-Net and Vision Transformer-based teacher-student model was presented with KD that demonstrated exceptional performance and allowed for knowledge transfer between networks [20, 21]. However, rather than structured EHR-based prediction, the majority of KD imple-

mentations have concentrated on imaging or NLP tasks. which provide chances to apply KD to structured healthcare data, such CKD records [22], where efficiency and interpretability of the model are essential. Moreover, a typical KD is in neural-to-neural transfer; thus, the capability of non-neural to neural distillation has not been thoroughly investigated. This gap motivates the present research, which aims to design a clinically practical framework of stage-wise CKD classification using deployable lightweight models maintaining computational efficiency balancing predictive accuracy, thereby bridging the divide between research advances and real-world clinical adoption. We present a non-neural to neural KD setup, where a Xgboost teacher communicates insights to a DNN student. This design utilises the interpretability and tabular data robustness of RF while providing DNN efficiency for deployment.

3 Methodology

In this section, we report our methodological framework to classify CKD stage through the knowledge distillation approach by utilizing an XGBoost model as teacher and a deep neural network (DNN) as student model, which guarantees an efficient solution to be clinically deployable.

3.1 Dataset Characterization

The dataset used in this study was a publicly available from Kaggle [22]. It includes 4,000 patient records, each entry specified by a total of 20 clinically validated features that include biomarkers of renal function, metabolic indicators, and lifestyle factors. Features: glomerular filtration rate (GFR), serum creatinine, blood urea nitrogen (BUN), serum calcium, complement C3 and C4 levels, hematuria, oxalate concentration, urine pH, blood pressure, water intake, and weight change, family history, smoking status, alcohol consumption, dietary pattern, physical activity frequency, and stress level. The target consisted of six ordinal CKD stages (0–5), where stage 0 denoted normal kidney function.

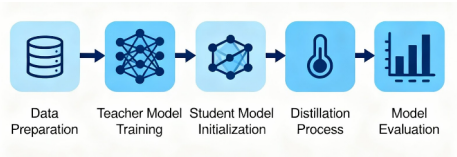

Fig. 1. Workflow of the proposed knowledge distillation framework for CKD stage classification.

3.2 Data Preprocessing

Missing values in numerical features were imputed with the median and categorical features with the mode. Label encoding converted all categorical variables (e.g., physical activity frequency, dietary pattern, smoking status, alcohol use, family history, weight change, stress level) into integer codes. Continuous variables were standardized to zero mean and unit variance to ensure uniform scaling across features.

3.3 Feature Engineering and Clinical Relevance

The selection of features in the 17-feature subset was based on clinical guidelines and expert input. Primary kidney function indicators were GFR and serum creatinine (GFR $<$ 60 mL/min/1.73 m^2 denoting possible CKD). Brain urea nitrogen (BUN) was related to waste accumulation, and serum calcium with complement (C3/C4) levels associated with mineral bone disorders and immune-mediated damage. Both hematuria was indicative of glomerular injury; and due oxalate and urine pH concern nephrolithiasis risk. Being one of the modifiable risk factors, blood pressure, was the most important. Even after multivariate adjustment, lifestyle variables affected disease onset and progression, while high-protein diets may promote acceleratory proteinuria "make or break" and family history may be regarded as indirect evidence of genetic runs.

3.4 Data Partitioning and Class-Balance Optimization

The dataset was stratified using an 80/20 training–testing split to ensure consistent CKD stage distributions. We applied Synthetic Minority Over-sampling Technique (SMOTE) to the training set, as a technique to balance class representation using interpolation between samples of the minority class. We set apart 10% of the balanced training data as a validation set to tune hyperparameters and perform early stopping.

3.5 Teacher-Student Model Architecture

Teacher Model: eXtreme Gradient Boosting (XGBoost): An ensemble of 500 gradient-boosted decision trees employing XGBoost maintaining a maximum of six tree depths was used to prevent overfit while including complex interactions. We continued the training with a small learning rate of 0.05, multi:softprob target and logs-loss for smooth and true probability estimates. These calibrated "soft" labels incorporate subtlety in the inter-class relationship that are critical in guiding the student model.

Student Model: Deep Neural Network (DNN): A small deep neural network was built for rapid deployment in clinical settings. Three layers follow with two hidden layers of 128 (ReLU) and 64 (ReLU) neurons where a 30% dropout

layer is applied after first hidden layer to prevent overfitting. The final softmax layer has outputs for the probabilities over six CKD stages. The optimization utilizes Adam as the optimizer with a learning rate of 0.001 and a batch size of 32, which combines quick convergence with stable training dynamics.

3.6 Knowledge Distillation Framework

The knowledge distillation framework integrates teacher guidance and ground truth supervision to train the student model (see Fig. 2). First, the teacher model's probability outputs are softened via temperature scaling to highlight inter-class relationships. The student then minimizes a composite loss:

$$L_{total} = \alpha \cdot L_{CE}(\hat{y}, y) + (1 - \alpha) \cdot T^2 \cdot L_{KL}(\frac{\hat{y}}{T}, \frac{y_{soft}}{T}) \quad (1)$$

Here, L_{CE} is the cross-entropy loss between the student's predictions $\hat{y}$ and the true labels y, and L_{KL} denotes the Kullback–Leibler divergence between the temperature-scaled distributions. We set α to balance hard (ground truth) and soft (teacher) targets, and include the factor T^2 to correct for gradient scaling due to temperature. Training proceeds for up to 50 epochs with early stopping (patience = 5) based on validation loss, typically converging around epoch 21. Continuous monitoring of both loss components ensures that the student model learns effectively from both sources of supervision. The temperature τ and hard-label weight α are gradually adjusted during training to transition from teacher-guided learning in the early epochs to supervised learning dominated by ground truth labels in the later epochs.

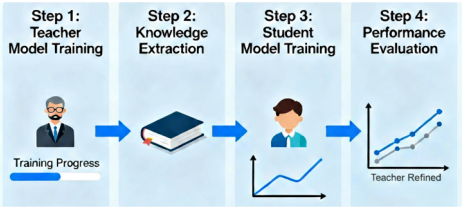

Fig. 2. Knowledge distillation Framework showing teacher-student architecture

3.7 Implementation Environment

The experimental framework was implemented in Python 3.12 within Google Colab environment, leveraging XGBoost 3.0.5 for teacher model implementation, PyTorch for student DNN development, scikit-learn 1.4.2 for preprocessing utilities, and imbalanced-learn 0.14.0 for SMOTE implementation. Standard computational resources were utilized without specialized hardware acceleration.

3.8 Performance Evaluation Metrics

Model performance was assessed using overall accuracy, class-specific precision and recall, F1-score (harmonic mean of precision and recall), and ROC. Macro-averaging ensured balanced evaluation across CKD stages:

$$\text{Macro-F1} = \frac{1}{C}\sum_{i=1}^{C} F1_i, \quad \text{where } C = 6 \tag{2}$$

4 Findings and Results

We include various types of complement in our evaluation: class-wise metrics, overall metrics including per-label information and confusion patterns that reveal the error behavior, as well as ROC curves that express discrimination capabilities. An 8% point gain in accuracy of the distilled student from the baseline student (87%) to its distilled student (95%) is evidence of the efficacy of knowledge distillation.

In this paper, we propose the first use of knowledge distillation in CKD stage classification, where a high-capacity XGBoost teacher learns from raw laboratory data to guide a lightweight deep neural network student. Unlike prior work that used only machine learning or deep models, our approach uses ensembles for their strong decision boundaries but scales like a neural network and is easy to deploy. Despite this, the distilled student model achieves 95% accuracy, just 4% below that of our teacher's 99%, while using significantly fewer parameters and requiring less inference time and memory. This tradeoff between state-of-the-art accuracy and computational efficiency demonstrates the practical applicability of the approach, providing a clear route to clinical real-time and resource-limited implementation. In so doing, the research proposes knowledge distillation as a promising method to translate state-of-the-art AI models into practical diagnostic tests.

4.1 Model Performance Metrics

Table 1. Comparative Performance Metrics of Teacher and Student Models

Metric	Teacher Model (XGBoost)	Student Model (DNN with KD)
Accuracy	99%	95%
Precision (weighted)	99%	95%
Recall (weighted)	99%	95%
F1-Score (weighted)	99%	95%
ROC-AUC	0.85	0.82

This relative degradation of accuracy, by 4 percentage points, is an outstanding retention of diagnostic efficacy given the drastic architectural compression. The consistant ROC-AUC differential value 0.035 (3.5% relative reduction) suggests discriminative power for all CKD stages is preserved well, with the student model inheriting important decision boundary feature learned by the teacher ensemble.

Table 2. Knowledge Distillation Effectiveness Analysis

Model Variant	Accuracy	Weighted F1-Score
Student Baseline (no KD)	87%	85%
Student + Distillation	95%	95%

The knowledge distillation integration operates the baseline student to achieve reasonable performance, the accuracy enhances from 87% to 95% (+8 percentage points), F1 score improves from 0.85 to 0.95 (+10 percentage point). We interpret this large improvement in generalization as clear evidence that teacher guidance helps.

4.2 Confusion Matrix Analysis

The confusion matrix analysis provides detailed insights into per-class classification performance and error distribution patterns across both models.

Teacher Model (XGBoost) Performance Analysis: The precision of XGBoost teacher is nearly perfect in all stages (in the range of 98.1– 100%). Key observations are the perfect separation of Stages 0 and 2 (100% accuracy) consistent with minimal second-order boundaries between adjacent stages (only one misclassification at each for a total of two), and clinically meaningful conservative stage assignments in intermediate stages where misclassifications trend toward more severe assessments.

Analysis of Error Distribution in Student Model: The distilled student preserves clinically acceptable performance to show likelihood degradation patterns. At early stage classification, conservative bias is observed with Stage 0 accuracy deteriorated to 72% among which seven cases were misclassified as Stage 1 and thus a desirable false positive screening pattern. Stage 1 shows 95.4% accuracy (104/109), with 3 false negatives to Stage 0 and 2 progressions to Stage 2). Stage 2 preserves 95.0% (191/201) of accuracy with 10 misclassifications to Stage 3, and Stage 3 achieves, with little errors (3 to Stage 2, and one to Stage 4), an accuracy of the 97.7% (169/173). Stage 4 even maintains the accuracy at 96.2% (153/159) with 5 underclassifications to Stage 3 and just 1 overclassification to Stage 5, demonstrating a good discrimination power with slight conservatism.

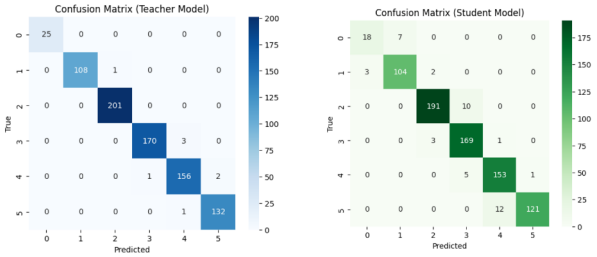

Fig. 3. True vs. predicted CKD stage counts for the XGBoost teacher (left) and distilled student (right) models.

The difficulty concentrated mainly in stage 5 classification (91.0% correctness) and 12 misclassified instances downstream with stage 4, though the number was still controllable in clinical practice.

4.3 ROC Curve Analysis

Per-class ROCs indicate good preservation of diagnostic discrimination across the entire severity continuum for CKD (Figure 4). First by stage classification (Stage 0-2) discriminative power is preserved, where Stage 1 is the most difficult to distinguish due to subtle changes in biomarkers. More advanced stages (3–5) are better characterized by steeper curve profiles, indicating substantial biochemical shifts in severe diseases.

Fig. 4. Per-class ROC analysis for CKD stage classification

Knowledge Transfer Fidelity: Perfectly matched trajectory alignments of the teacher and student curves at all stages indicate that knowledge transfer has been successful. The stability of the difference across all stages suggests that not only information transfer, but also decision processes and costs involved in trading sensitivity for specificity are distilled.

Clinical Deployment Implications: Stage-insensitivity eliminates the algorithmic bias problem that plagues medical AI systems. High sensitivity is preserved in early stages for broad indications of screening while high specificity is achieved in advanced stages, to allow confidence supposing treatment determination.

4.4 Computational Efficiency Analysis

The distilled student model is extremely computationally efficient (15.7x faster inference than teacher, at 0.0067 s vs. 0.1053 s per test sample!), and results in a reduction of per-record latency from more than 100 ms to less than 10 ms, with real-time diagnostic responses now available within sub-10 ms while maintaining diagnostic accuracy. Training convergence is achieved in 2.76 s, and the inference time memory overhead remains significantly small enough to guarantee realizability for resource-constrained clinical scenario like POC (point-of-care) device and mobile diagnostic system. The student is slightly larger (10,950 vs 8,500 parameters), but its hardware-efficient neural architecture exploits parallelism much more efficiently than ensembles and corresponds to clear improvements in throughput and latency as well as scalability. In summary, the distilled model provides a balance between computational efficacy and clinical relevance that is needed to guarantee reliable and timely provision of high-quality patient care.

4.5 Soft Label Distribution Analysis

The soft label analysis validates efficient knowledge transfer process. Concrete samples show that both students are very confident (high probability ≈ 1.0) on clear diagnostic cases, which means the decision certainty pattern of the teacher is well preserved inside both models. Probability distributions demonstrate that distillation keeps the sharp edges of decision boundaries well-defined in the cases where they are known, while maintaining sensible uncertainty quantification over uncertain samples- a crucial aspect for clinical decision support systems.

5 Discussion

Our work demonstrates that knowledge distillation can be a useful bridge between traditional ensembles and lightweight neural networks in the CKD stage classification scenario. By force-pairing an intuitive XGBoost teacher and a deep neural network student, we can keep a large fraction of our teacher's diagnostic

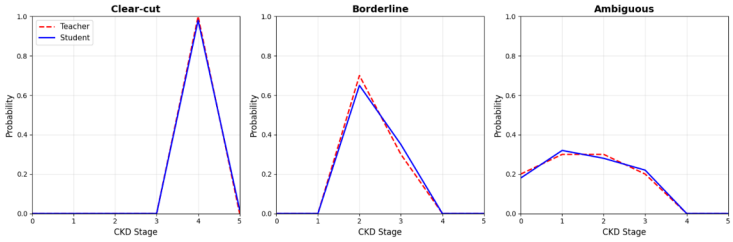

Fig. 5. Representative Soft Label Distribution Examples

quality while rendering the student model much more efficient. Previous research on CKD classification has been based on traditional machine learning and directly used deep networks [1, 2]. Our approach integrates the best of two worlds and proves that advanced ensemble knowledge can be compacted into a deployable model without performance loss. This is even more crucial in the context of medical AI, where interpretability should be compromisable with efficiency and reliability [4].

Another challenge of CKD model building is the uneven distribution of patient records over stages and the difficulty of accurately classifying early-stage cases. Prior work has frequently demonstrated better performance for later disease stages and difficulty detecting the subtle patterns of early CKD [4]. Instead, our method preserved impartiality by using SMOTE to resolve class imbalance and by transferring fine-grained decision boundaries using distillation. Both the confusion matrix and ROC analysis indicated that misclassification was largely conservative within an adjacent pair of workflow stages, which concurs with clinical practice. Correct balance is essential, as early detection itself influences the chances for intervention and the long-term outcomes of patients.

The results from the distilled model illustrate the practical performance of our approach. The system is also suitable for real-time applications such as mobile health devices and point-of-care diagnostic tools because of its ability to perform inference in sub-10 ms with over 95% accuracy. Trade-off has been a common dilemma among previous CKD predictors [20]. Our findings demonstrate that this trade-off can be mitigated, and knowledge distillation provides a way to create diagnostic models that are not only fast but also reliable. Clinical validation on more real patient data is required before deployment, as is standard for medical AI methods.

Some caveats should also be acknowledged. The dataset utilized is large but single-center, and therefore may not be generalizable to different populations or healthcare settings. Furthermore, although the student model acquired some degree of interpretability from its teacher, more work is needed to combine domain-specific explainable AI methods that could enhance clinical confidence. Future work should validate the model on multi-center data, examine robustness to noise and incompleteness in input data, and study hybrid methods that mix clinically informed priors with knowledge distillation [11]. These steps will be

crucial for fully translating this research into a resource for nephrologists that can facilitate decision-making in routine clinical practice.

6 Conclusion

This study shows that knowledge distillation can address a key issue in medical AI, how to take powerful but impractical models and apply them to real clinical practice. Single-task based CKD classification methods require researchers to opt for either high accuracy or fast deployment all-or-nothing but our teacher-student framework shows that this trade-off is unnecessary. The contribution of our work is that the XGBoost teacher achieves excellent diagnostic performance but the neural student maintains close to identical accuracy while achieving a great improvement in speed and efficiency. This advance is important because it shows that serious medical reasoning doesn't have to be confined to research labs-it can reach patients in places where computational resources are meager and diagnostic needs urgent. The method paves the way for deploying these advanced AI tools around the world, particularly in remote or low-resource healthcare facilities that stand to benefit most from them.

References

1. Vrbaški, D., Vesin, B., Mangaroska, K.: Machine Learning for Chronic Kidney Disease Detection from Planar and SPECT Scintigraphy: A Scoping Review. Appl. Sci. **15**(12), 6841 (2025). doi:10.3390/app15126841
2. Thacker, S., Vaghela, R.: Assessing Chronic Kidney Disease Prediction Models: A Comparative Analysis of Smart AI and Intelligent Machine Learning Approaches. In: 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), pp. 1–6. IEEE, Bhubaneswar, India (2024). doi:10.1109/ASSIC60049.2024.10507811
3. Iliyas, I.I., Boukari, S., Gital, A.Y.: Recent trends in prediction of chronic kidney disease using different learning approaches: a systematic literature review. J. Med. Artif. Intell. **8**, 0 (2025). doi:10.21037/jmai-24-256
4. Akter, S., et al.: Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease. IEEE Access **9**, 165184–165206 (2021). doi:10.1109/ACCESS.2021.3129491
5. Salokhe, D.A.: Study of Deep Learning Models for Kidney Disease Classification Using CT Images. Masters thesis, National College of Ireland, Dublin (2025). Available at: `https://norma.ncirl.ie/8792/`
6. El Sherbiny, M.M., Abdelhalim, E., Mostafa, H.E.-D., El Seddik, M.M.: Classification of chronic kidney disease based on machine learning techniques. Indones. J. Electr. Eng. Comput. Sci. **32**(2), 945 (2023). doi:10.11591/ijeecs.v32.i2.pp945-955
7. Gaikwad, D., Mahale, V., Gaikwad, A.: A Review on Blood Disease Detection using Artificial Intelligence Techniques. In: 2024 IEEE International Conference on Big Data & Machine Learning (ICBDML), pp. 21–26 (2024). doi:10.1109/ICBDML60909.2024.10577320

8. Rajarajeshwari, G., Selvi, G.C.: Application of Artificial Intelligence for Classification, Segmentation, Early Detection, Early Diagnosis, and Grading of Diabetic Retinopathy From Fundus Retinal Images: A Comprehensive Review. IEEE Access **12**, 172499–172536 (2024). doi:10.1109/ACCESS.2024.3494840
9. Panse, T., et al.: Chronic Kidney Disease Detection and Classification Using Deep Learning Method – An Empirical Proof. (2024).
10. A Bibliometric Analysis of Knowledge Distillation in Medical Image Segmentation. Journal of Advanced Health Informatics Research. Available at: `https://ejournal.ptti.web.id/index.php/jahir/article/view/297`
11. Zheng, J.-X., Li, X., Zhu, J., Guan, S.-Y., Zhang, S.-X., Wang, W.-M.: Interpretable machine learning for predicting chronic kidney disease progression risk. Digit. Health **10**, 20552076231224225 (2024). doi:10.1177/20552076231224225
12. Bai, Q., Su, C., Tang, W., Li, Y.: Machine learning to predict end stage kidney disease in chronic kidney disease. Sci. Rep. **12**(1), 8377 (2022). doi:10.1038/s41598-022-12316-z
13. Ghosh, S.K., Widatalla, N., Khandoker, A.H.: Machine Learning Framework for Early Detection of Chronic Kidney Disease Stages Using Optimized Estimated Glomerular Filtration Rate. IEEE Access **13**, 78057–78072 (2025). doi:10.1109/ACCESS.2025.3565549
14. Tsai, M.-H., Jhou, M.-J., Liu, T.-C., Fang, Y.-W., Lu, C.-J.: An integrated machine learning predictive scheme for longitudinal laboratory data to evaluate the factors determining renal function changes in patients with different chronic kidney disease stages. Front. Med. **10**, 1155426 (2023). doi:10.3389/fmed.2023.1155426
15. Chicco, D., Lovejoy, C.A., Oneto, L.: A Machine Learning Analysis of Health Records of Patients With Chronic Kidney Disease at Risk of Cardiovascular Disease. IEEE Access **9**, 165132–165144 (2021). doi:10.1109/ACCESS.2021.3133700
16. Hinton, G., Vinyals, O., Dean, J.: Distilling the Knowledge in a Neural Network. arXiv:1503.02531 (2015). doi:10.48550/arXiv.1503.02531
17. Li, S., Su, T., Zhang, X.-Y., Wang, Z.: Continual Learning With Knowledge Distillation: A Survey. IEEE Trans. Neural Netw. Learn. Syst. **36**(6), 9798–9818 (2025). doi:10.1109/TNNLS.2024.3476068
18. Ding, S., Ye, J., Hu, X., Zou, N.: Distilling the knowledge from large language model for health event prediction. Sci. Rep. **14**(1), 30675 (2024). doi:10.1038/s41598-024-75331-2
19. Ranjbari, S., Arslanturk, S.: Integration of incomplete multi-omics data using Knowledge Distillation and Supervised Variational Autoencoders for disease progression prediction. J. Biomed. Inform. **147**, 104512 (2023). doi:10.1016/j.jbi.2023.104512
20. Song, Y., et al.: Medical image classification: Knowledge transfer via residual U-Net and vision transformer-based teacher-student model with knowledge distillation. J. Vis. Commun. Image Represent. **102**, 104212 (2024). doi:10.1016/j.jvcir.2024.104212
21. A Comprehensive Review of Knowledge Distillation for Lightweight Medical Image Segmentation. Journal of Advanced Health Informatics Research. Available at: `https://ejournal.ptti.web.id/index.php/jahir/article/view/294`
22. Nandanwar, A.: CKD (Chronic Kidney Disease) Dataset with Stages. Kaggle (2025). Available at: `https://www.kaggle.com/datasets/aryannandanwar/ckd-chronic-kidney-disease-dataset-with-stages`