IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

NephroSense: A Multimodal Framework for Early Chronic Kidney Disease Risk Screening

Neha Barde[1], Janhavi Gholve[2], Sanika Naik[3], Monica Charate[4]
Usha Mittal Institute of Technology, Santacruz, Mumbai, India[1-4]
[1] nehabarde0605@gmail.com, [2] jgholve2020@gmail.com
[3] sanikanaik1129@gmail.com, [4] monica.charate1983@gmail.com

Abstract : *Chronic kidney disease (CKD) is often detected late because laboratory screening is inaccessible or costly in many regions. This paper presents NephroSense, a web-based multimodal screening aid that estimates CKD risk from two complementary sources of evidence: a health questionnaire and an optional smartphone eye photograph. The questionnaire is scored by a soft-voting ensemble of XGBoost, Random Forest and Logistic Regression trained on three public datasets, while the conjunctival photograph is scored by a fine-tuned EfficientNet-B3 network that acts as an anaemia proxy. The two scores are combined by a fixed 75:25 weighted average and mapped to a Low, Moderate or High risk category. SHAP, LIME and Grad-CAM explanations and a nearby-hospital finder accompany each result. On internal held-out data, the ensemble reached 85% cross-validated accuracy and the image model reached 96.5% validation accuracy. These figures are internal validation results, not clinical validation, and NephroSense is positioned as a pre-clinical screening aid rather than a diagnostic instrument.*

Keywords: Chronic Kidney Disease, Multimodal Fusion, Ensemble Learning, EfficientNet-B3, Conjunctival Pallor, Explainable AI, Health Screening

I. INTRODUCTION

Chronic kidney disease (CKD) progresses silently for years before symptoms appear. By the time a patient reports fatigue, swelling or reduced urine output, glomerular function has often already fallen well below normal. Population screening based on serum creatinine and urinalysis is effective, but it depends on access to a laboratory, a clinician to interpret the report and, in many regions, an out-of-pocket cost that discourages routine testing in asymptomatic individuals. This gap motivates software tools that combine information a person can supply without a laboratory visit, such as demographic details, self-reported comorbidities and whatever laboratory values they already hold, with inputs that an ordinary smartphone can capture.

NephroSense is one such tool. It is a web-based screening aid that estimates CKD risk from two complementary sources of evidence: a health questionnaire and an optional smartphone photograph of the eye. It was built around three design principles. First, every predictive component must degrade gracefully: a missing eye photograph or an unreadable one should reduce the information available rather than crash the pipeline or silently substitute a guess. Second, the prediction must be explainable, so that both the questionnaire model and the image model expose which inputs drove the result. Third, the output must be framed as a screening estimate, with a recommendation to consult a physician in every risk band, and never as a diagnosis.

The main contributions of this work are:

- A questionnaire module based on a soft-voting ensemble of XGBoost, Random Forest and Logistic Regression, trained on three combined public datasets;
- An image module that uses a fine-tuned EfficientNet-B3 network to estimate an anaemia-risk score from conjunctival photographs;

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

- A fixed 75:25 weighted-average fusion of the two scores into a Low, Moderate or High CKD risk category;
- An explainability layer (SHAP, LIME and Grad-CAM) and a nearby-hospital finder, delivered through a single web application.

The remainder of this paper is organised as follows. Section II reviews related work. Section III states the problem and objectives. Section IV describes the system architecture and methodology, Section V the multimodal fusion, and Section VI the explainability and hospital-lookup modules. Section VII reports the results, Section VIII discusses limitations, and Section IX concludes.

II. LITERATURE REVIEW

Machine-learning approaches to CKD prediction have made extensive use of the University of California, Irvine (UCI) Chronic Kidney Disease dataset, a 400-record collection of 24 clinical attributes [1]. Gradient-boosted trees such as XGBoost have become a standard baseline for tabular clinical prediction because of their robustness to mixed categorical and continuous features and their low sensitivity to feature scaling [2]. Class imbalance, a frequent property of clinical datasets, is commonly addressed by synthetic oversampling of the minority class, first proposed as the Synthetic Minority Over-sampling Technique (SMOTE) [3].

Separately, a body of work has used the conjunctiva, the richly vascularised membrane lining the inner eyelid, as a non-invasive proxy for haemoglobin concentration, because pallor of this tissue is a long-established bedside sign of anaemia [4]. Convolutional networks pretrained on large natural-image corpora and fine-tuned on conjunctival photographs have been used for this purpose, with EfficientNet-family backbones favoured for their accuracy-to-parameter efficiency [5]. Because anaemia is a common consequence of reduced erythropoietin production in CKD, an anaemia signal is a plausible but indirect contributor to a kidney-risk estimate; it is not a measurement of kidney function.

For interpretability, SHapley Additive exPlanations (SHAP) [6] and Local Interpretable Model-agnostic Explanations (LIME) [7] are the two dominant post-hoc frameworks for tabular classifiers, and both are used in this work to give two independent views of feature attribution. Recent studies have also explored multimodal CKD screening that combines clinical and image-based information. Retinal imaging has shown promising results, but several approaches rely on private, hospital-based or biopsy-linked datasets. Table I summarises representative studies.

TABLE I: Literature Survey of Existing CKD Prediction and Screening Approaches

Sr.	Paper and Author	Methodology	Key Observation
1	Early CKD Prediction Using Lifestyle and Demographic Indicators, Valencia et al. (2026)	AdaBoost with SHAP/LIME on NHANES data	72.6% accuracy; BUN, ACR, creatinine and age were key predictors.
2	Explainable AI for CKD Prediction in Medical IoT, Rezk et al. (2025)	GAN imputation, prototypical networks, MAML, SHAP/LIME	Up to 99.99% accuracy; handles missing data and limited samples.
3	Explainable Machine Learning Model for CKD Prediction, Arif et al. (2024)	MLP, KNN imputation, SelectKBest, LIME	100% accuracy on UCI CKD dataset; no SHAP component.
4	Explainable Disease Surveillance for Multiple Chronic Diseases, Khan et al. (2025)	Random Forest with SHAP rule extraction on EHR	CKD F1 of about 0.77 at 12 months; predicts multiple diseases.
5	Deep Learning-Based CKD Prediction, Chellam et al. (2024)	DNN with OLPP feature reduction	98.89% accuracy; no XAI component.
6	Noninvasive CKD Screening from Retinal Images, Wu et al. (2025)	EfficientNet-B5 with retinal and clinical data	AUC 0.839–0.993; private biopsy-linked dataset from 9 hospitals.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, November 2026 Impact Factor: 8.2

Sr.	Paper and Author	Methodology	Key Observation
7	CKD Prediction Using Retinal Imaging and Urine Dipstick Data, Bhak et al. (2025)	Multimodal deep learning with retinal images and urine dipstick data	Better than retinal-only screening; lower performance for age ≥ 65.

Research gap: existing studies mainly focus on machine learning, deep learning or explainable AI separately, and retinal-CKD studies often depend on private clinical datasets. A multimodal platform that combines CKD risk prediction, explainability and accessible user-oriented screening, built on public data, remains limited. NephroSense addresses this gap.

III. PROBLEM STATEMENT AND OBJECTIVES

Problem Statement

Early CKD detection requires laboratory investigations that are not always available for routine screening, particularly in primary-care and resource-limited settings. Questionnaire-only tools ignore complementary visual evidence, and image-based tools in the literature depend on specialised retinal imaging and private datasets. The research problem addressed here is: how can a questionnaire-based model and a smartphone eye-image model be combined, using public datasets and an interpretable fusion rule, to give an accessible and explainable CKD risk estimate that remains usable when one input is missing?

Objectives

The objectives of this work are:

- to build a questionnaire-based CKD risk model, using a soft-voting ensemble of XGBoost, Random Forest and Logistic Regression, on three combined public datasets;
- to build an image-based anaemia-proxy model by fine-tuning EfficientNet-B3 on conjunctival photographs;
- to combine the two scores with a fixed 75:25 weighted average and report a Low, Moderate or High risk category;
- to explain each prediction using SHAP, LIME and Grad-CAM, and to help users with elevated risk find nearby nephrology-capable facilities;
- to evaluate both learned components on held-out data and state clearly the limits of that evaluation.

IV. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Overview

NephroSense is implemented as a Streamlit web application backed by a local SQLite database, with the machine-learning components written in Python. The tabular ensemble is built on XGBoost and scikit-learn, with SMOTE provided by imbalanced-learn and explanations provided by the SHAP and LIME packages; the image pipeline uses PyTorch, torchvision and OpenCV. Fig. 1 shows the overall workflow. A registered user completes the health questionnaire and, optionally, uploads a conjunctival eye photograph. The two inputs are scored independently, fused into a single risk percentage and category, and shown together with explanations and a hospital finder. Every optional dependency is imported defensively, so that a missing package disables only the affected feature, and each module falls back to a documented, less precise behaviour or an explicit refusal to score rather than failing silently.

B. Datasets and Preprocessing

Three independent public sources are combined for the questionnaire model: the UCI Chronic Kidney Disease ARFF dataset [1], a second independently collected Bangladeshi CKD cohort distributed as comma-separated values, and a subset of the U.S. National Health and Nutrition Examination Survey (NHANES) 2017–2018 cycle [9]. Because NHANES does not report a clinical CKD label, a target label is derived by applying the 2021 CKD-EPI creatinine equation [8] to each respondent's age, sex and serum creatinine, and thresholding the estimated GFR at 60 mL/min/1.73 m² in line with the KDIGO definition of reduced kidney function [10]. Categorical fields recorded as free text are

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

mapped to a consistent binary encoding across the three sources, and rows that are exact duplicates across the numeric feature columns are removed so that no record is counted twice.

Missing numeric values are imputed with the column median computed on the combined training data, and the same medians are stored with the trained model so that identical imputation is applied at inference time to any field a user leaves blank. Three clinically motivated features are engineered after imputation: the blood-urea-to-creatinine ratio, an anaemia-severity term equal to the shortfall of haemoglobin below a 14 g/dL reference point (floored at zero), and an interaction term between age and systolic blood pressure. For the image module, 859 usable conjunctival photographs were assembled from three sources, as summarised in Table II.

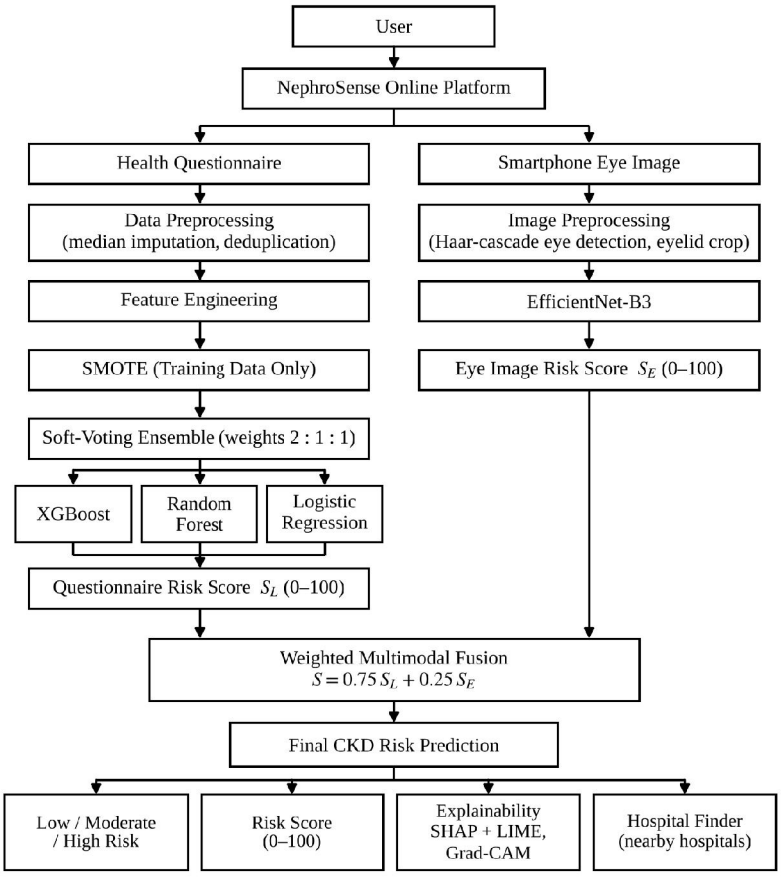

Fig. 1 NephroSense system architecture: the health questionnaire and the optional eye image are scored independently, combined by a fixed 75:25 weighted fusion, and presented with explanations and a hospital finder.

TABLE II: Data Sources Used by the Two Learned Modules

Module	Source	Description
Questionnaire	UCI CKD dataset [1]	400 records, 24 clinical attributes (ARFF)
Questionnaire	Bangladeshi CKD cohort	Independently collected cohort (CSV)
Questionnaire	NHANES 2017–2018 [9]	Demographics, blood pressure, biochemistry, blood count; CKD label derived from eGFR < 60
Eye image	India and Italy	Conjunctival photographs paired with laboratory haemoglobin; labelled with WHO thresholds

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

Module	Source	Description
Eye image	Ghana	Conjunctival photographs labelled anaemic or non-anaemic; 859 usable images in total

C. Questionnaire Model: Soft-Voting Ensemble

The tabular classifier is a soft-voting ensemble of three base learners: an XGBoost gradient-boosted tree classifier [2], a random forest with balanced class weighting, and a logistic-regression classifier preceded by feature standardisation. Their predicted probabilities are combined with weights of 2, 1 and 1, which gives the gradient-boosted model, the strongest individual learner on this data, twice the influence of either of the others:

$$P_{ens}(x) = [2 \cdot P_{XGB}(x) + P_{RF}(x) + P_{LR}(x)] / 4 \quad (1)$$

where PXGB, PRF and PLR are the CKD probabilities produced by the three learners for input x. Training uses an 80/20 stratified train–test split. SMOTE [3] is applied to the training partition only, so that the held-out partition keeps its natural class distribution. The XGBoost hyperparameters (number of trees, maximum depth and learning rate) are selected by a five-fold grid search scored on recall, reflecting the design choice that in screening a missed positive case is costlier than a false alarm.

Instead of the conventional 0.5 cut-off, the decision threshold is tuned on the precision–recall curve of the held-out partition: the lowest threshold at which precision remains at or above 0.85 is selected, which maximises recall under that precision floor. If the raw probability places a case in the lowest risk band but the tuned threshold would classify it as CKD-positive, the reported level is escalated from Low to Moderate, so that the label shown to the user is never more reassuring than the tuned decision boundary implies. Generalisation is additionally estimated with ten-fold stratified cross-validation on the original, non-resampled data.

D. Eye-Image Module: EfficientNet-B3

The optional visual-screening module estimates an anaemia-risk score from a photograph of the eye, on the premise that conjunctival pallor is a recognised, if imprecise, bedside indicator of low haemoglobin [4]. World Health Organization haemoglobin thresholds (13.0 g/dL for men and 12.0 g/dL for women) are used to derive a binary label for the two haemoglobin-labelled sources, so that all three sources share a common target definition.

The training photographs are tight crops of the lower, inner eyelid, so supplying an uncropped phone photograph would create a mismatch between training and inference. The inference pipeline therefore first locates the eye with a Haar-cascade detector, crops the lower-inner third of the detected region and only then passes the crop to the network. If no eye region can be located, the module declines to produce a score rather than proceeding on an unreliable crop.

The classifier is an EfficientNet-B3 backbone [5], pretrained on ImageNet and fine-tuned with its final layer replaced by a single linear output trained with a binary cross-entropy loss weighted inversely to class frequency. Augmentation (horizontal flipping and mild colour jitter) is applied during training only. Model selection uses an 80/20 random split with a fixed seed, and the checkpoint with the highest validation accuracy is retained.

V. MULTIMODAL FUSION

The questionnaire score SL and the eye-image score SE, both expressed on a 0–100 scale, are combined by a fixed weighted average:

$$S = wL \cdot SL + wE \cdot SE, \quad wL = 0.75, \quad wE = 0.25 \quad (2)$$

The higher weight on the questionnaire reflects the stronger evidence carried by laboratory and demographic predictors, while the eye score is an indirect proxy. If only one modality is available, it receives a weight of 100% instead of the missing score being treated as zero. The fused score is then mapped to a risk category using fixed boundaries: Low for $S < 34$, Moderate for $34 \leq S < 67$, and High for $S \geq 67$ (Fig. 2). The weights and boundaries are design choices, not empirically optimised parameters.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

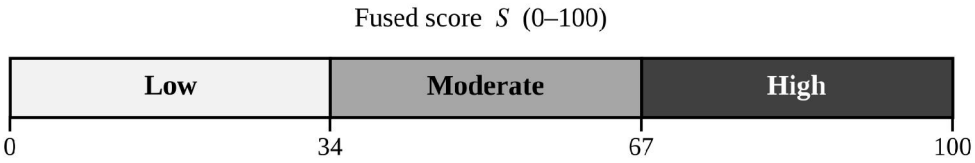

Fig. 2 Mapping of the fused score S to the three risk categories.

VI. EXPLAINABILITY AND HOSPITAL LOOKUP

A. Explainability

Because the ensemble score is the least directly interpretable output, two independent post-hoc explanations are computed for every scored assessment. A SHAP TreeExplainer [6] is applied to the XGBoost component of the ensemble, which is extracted from the voting ensemble because the efficient tree-based algorithm requires a single tree model, and returns the five features contributing most strongly to that patient's prediction. A LIME tabular explainer [7], which perturbs the input locally and fits an interpretable surrogate, provides a methodologically independent second view. For the image module, a Grad-CAM map is overlaid on the cropped eye region to show which part of the crop the network weighted most heavily. This states where the model looked, not whether that region is clinically informative in a given case.

B. Nearby Hospital Finder

Given a city, town or village name, the system searches for nearby hospitals and dialysis-capable facilities using the Geoapify Places API, with OpenStreetMap services as a fallback that needs no API key. Facilities are heuristically tagged as government or private by keyword matching, and facilities whose names suggest nephrology, dialysis or transplant services are flagged and prioritised, so that users with elevated risk can find centres equipped for renal care more easily.

VII. RESULTS AND DISCUSSION

A. Reported Metrics

Table III summarises the headline metrics produced by the training and evaluation pipelines of the two learned components. The questionnaire ensemble reached 85% accuracy under ten-fold stratified cross-validation on the combined, deduplicated dataset, with the decision threshold tuned for a precision of at least 0.85. The EfficientNet-B3 model reached a best validation accuracy of 96.5% on its held-out 80/20 split of 859 images.

TABLE III: Reported Model Metrics

Component	Metric	Value
Questionnaire ensemble	Cross-validated accuracy	85%
Questionnaire ensemble	Decision threshold	Tuned at precision ≥ 0.85
Eye model (EfficientNet-B3)	Validation accuracy	96.5%
Eye model (EfficientNet-B3)	Training images	859 (India, Italy, Ghana)
Fusion layer	Questionnaire : eye weighting	75% : 25%
Fusion layer	Risk categories (fused score)	Low < 34; Moderate 34–66; High ≥ 67

B. Discussion

Three properties of these results are worth noting when judging whether the numbers are meaningful. First, the cross-validated accuracy is computed on the combined dataset rather than on any single source, so it reflects the harder task of generalising across three populations collected under different conditions. Second, the decision threshold is tuned for precision rather than raw accuracy; it is chosen to keep false alarms bounded while maximising the fraction of true

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

CKD cases still caught, rather than to maximise the headline figure. Third, the image model estimates anaemia, not kidney function, so its accuracy describes how well it separates anaemic from non-anaemic conjunctivae in the available datasets and should not be read as CKD-detection accuracy.

The fusion design keeps the contribution of each modality transparent: the questionnaire carries three quarters of the weight, the eye image one quarter, and either one alone is used at full weight when the other is absent. The reported figures are internal validation results, and the limitations in the next section apply to all of them.

VIII. LIMITATIONS

Several limitations should be considered when interpreting the reported metrics. First, both learned components are evaluated on held-out partitions of the same public datasets used for training, and external prospective validation on other populations, imaging equipment and disease prevalence has not been carried out. Second, the image checkpoint is selected by validation accuracy and the questionnaire threshold is tuned on the held-out partition, so both reported figures may be optimistic. Third, the conjunctival anaemia proxy is indirect, and the 75:25 weighting and the risk boundaries are design choices rather than optimised values. Fourth, the NHANES CKD labels are derived from an eGFR threshold rather than from clinician diagnosis. Finally, the facility finder classifies facilities from name keywords rather than verified registries, and its completeness depends on third-party mapping providers, so users should confirm services directly.

The interface displays an unconditional emergency-symptom banner directing users to immediate in-person care regardless of the computed risk. All results are labelled as screening estimates rather than diagnoses and recommend physician consultation for every risk band, which positions NephroSense as a triage aid rather than a substitute for clinical care.

IX. CONCLUSION AND FUTURE WORK

This paper described NephroSense, a multimodal CKD screening aid that combines a soft-voting ensemble trained on three public clinical datasets with a fine-tuned EfficientNet-B3 conjunctival-image anaemia proxy through a fixed 75:25 weighted fusion, and reports Low, Moderate or High risk with SHAP, LIME and Grad-CAM explanations and a nearby-hospital finder. On internal held-out data the ensemble reached 85% cross-validated accuracy and the image model 96.5% validation accuracy; these are not clinical validation.

Future work includes prospective validation of the fused risk score against clinically confirmed outcomes in an independent cohort; extending the questionnaire model with the urinary albumin-to-creatinine ratio, a KDIGO-recommended marker not consistently available in the current training sources; evaluating whether the fusion weights can be set empirically on an independent cohort; and testing the image module on a wider range of smartphone cameras and populations.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Technology, Usha Mittal Institute of Technology, for its support throughout this work.

REFERENCES

1. L. Rubini, P. Soundarapandian, and P. Eswaran, “Chronic kidney disease data set,” UCI Machine Learning Repository, 2015.
2. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2016, pp. 785–794.
3. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.
4. S. Suner, G. Crawford, J. McMurdy, and G. Jay, “Non-invasive determination of hemoglobin by digital

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, November 2026
Impact Factor: 8.2

photography of palpebral conjunctiva," J. Emerg. Med., vol. 33, no. 2, pp. 105–111, 2007.
5. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn., 2019, pp. 6105–6114.
6. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 4765–4774.
7. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2016, pp. 1135–1144.
8. L. A. Inker et al., "New creatinine- and cystatin C-based equations to estimate GFR without race," N. Engl. J. Med., vol. 385, no. 19, pp. 1737–1749, 2021.
9. Centers for Disease Control and Prevention, "National health and nutrition examination survey, 2017–2018," National Center for Health Statistics, 2020.
10. KDIGO CKD Work Group, "KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease," Kidney Int. Suppl., vol. 3, no. 1, pp. 1–150, 2013.