

ARTICLE

A multimodal deep learning framework for smart health monitoring and early-stage classification of chronic kidney disease

Manjulabai Bhadrashetty^{1,2†*} and Dayanand Jamkhandikar^{2,3†}¹Department of Computer Science, Government Women's First Grade College, Kalaburagi, Karnataka, India²Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India³Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Bidar, Karnataka, India

Abstract

Chronic kidney disease (CKD) is a disease that gets worse over time and raises the risk of heart disease, lowers the quality of life, and puts a lot of strain on the healthcare system. Conventional diagnostic methods predominantly depend on static clinical biomarkers, including serum creatinine, estimated glomerular filtration rate, and albuminuria, which may not effectively detect subtle physiological decline in the initial stages. To fill this gap, this paper suggests a strong, easy-to-understand, and scalable multimodal deep learning framework that combines structured clinical laboratory features with simulated real-time physiological signals, such as changes in blood pressure and heart rate. A dual-branch neural architecture is constructed, comprising a multilayer perceptron that encodes tabular clinical features and a one-dimensional convolutional neural network that extracts temporal representations from vital-sign sequences. For CKD stage-wise classification, the learned embeddings are combined at the feature level. The proposed model achieves a test accuracy of 90% and a macro-F1 score of 0.86, outperforming clinical-only and vitals-only baselines, as shown by experimental evaluation. SHapley Additive exPlanations-based feature attribution and temporal attention visualizations make it possible to understand models better by showing clinically relevant biomarkers and hemodynamic patterns. Robustness testing with noisy vital signs and low-latency inference profiling shows that the framework is good for real-time renal risk monitoring and tele-nephrology uses.

Keywords: Chronic kidney disease; Deep learning; Multimodal fusion; Physiological signals; Medical diagnostics; SHAP interpretability

[†]These authors contributed equally to this work.

***Corresponding author:**
Manjulabai Bhadrashetty
(kadmanju2@gmail.com)

Citation: Bhadrashetty, M. & Jamkhandikar, D. (2026). A multimodal deep learning framework for smart health monitoring and early-stage classification of chronic kidney disease. *Int J Systematic Innovation*, 10(5): 026200095. [https://doi.org/10.6977/IJoSI.202610_10\(5\).0003](https://doi.org/10.6977/IJoSI.202610_10(5).0003)

Received: May 13, 2026

Revised: June 2, 2026

Accepted: July 10, 2026

Published online: October 22, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Around 10% of the world's population has chronic kidney disease (CKD), which is strongly linked to higher rates of heart disease, death from all causes, and high healthcare costs. To slow down the disease's progression and make the best treatment choices, it is important to identify people who are at risk early and accurately classify them by

stage (Jager *et al.*, 2019). It is known that the standard risk stratification for CKD relies on estimated glomerular filtration rate (eGFR), serum creatinine, albumin-to-creatinine ratio, and blood urea nitrogen, frequently augmented by demographic variables and comorbidities (Alsekait *et al.*, 2023; Islam *et al.*, 2023). Nonetheless, these static laboratory assessments inadequately reflect the dynamic physiological impact on renal function, especially in early or marginal stages. Structured clinical datasets have been widely used to investigate machine learning and deep learning for identifying CKD and predicting the risk of its progression. Gradient-boosting models, random forests, and support vector machines have shown good results for classifying CKD and scoring risk based on tabular features (Iliyas *et al.*, 2025; Zhu *et al.*, 2023).

Even with these improvements, most published CKD models only use static clinical variables or claims data. They do not include short-term physiological changes such as blood pressure or heart rate, which have been linked to kidney outcomes and heart problems. Adding temporal vital-sign data can improve the patient state representation and facilitate the identification of early stages of CKD (Rostamian *et al.*, 2025; Ruan *et al.*, 2025; Wang *et al.*, 2022). Simultaneously, the growing utilization of wearable sensors and home blood pressure monitoring creates an opportunity to integrate longitudinal physiological time series with conventional laboratory parameters.

Multimodal artificial intelligence (AI) frameworks that simultaneously model structured data, time series, imaging, and unstructured text have demonstrated significant improvements in chronic disease prediction, decision support, and risk stratification (Demrozi *et al.*, 2025; Lu *et al.*, 2025; Sekhar *et al.*, 2025). Nonetheless, there exists a paucity of literature concentrating exclusively on multimodal architectures aimed at CKD stage-wise classification through the integration of laboratory indicators and vital-sign dynamics.

Interpretability is another important requirement for clinical use. SHapley Additive exPlanations (SHAP)-based explainable AI (XAI) techniques have effectively been employed to assign feature contributions in CKD risk models, allowing nephrologists to corroborate learned patterns with clinical expertise and detect possible biases (Balikci Cicek & Kucukakcali, 2023; Ghosh & Khandoker, 2024; Shu *et al.*, 2025; Tsai *et al.*, 2023). However, the majority of current studies focus on global feature importance for static models and neglect the joint interpretation of multimodal representations and temporal attention within the framework of CKD.

In this study, we present a multimodal deep learning

framework that combines tabular clinical biomarkers with simulated real-time vital-sign sequences for the early prediction and stage-wise classification of CKD. We use a dual-branch architecture: a multilayer perceptron (MLP) encodes structured features, and a one-dimensional convolutional neural network (1D-CNN) processes temporal physiological signals. The resulting latent embeddings are fused for multiclass classification based on KDIGO staging. The main contributions are as follows:

- (i) A multimodal CKD prediction framework that jointly models clinical biomarkers and short-term physiological dynamics using a dual-branch deep architecture.
- (ii) A differentiable feature-fusion strategy and modeling equations for stage-wise CKD classification with class-imbalance-aware optimization.
- (iii) Integrated SHAP-based attribution and temporal attention visualizations for joint interpretability of clinical and physiological modalities.
- (iv) A comprehensive experiment compares clinical-only, vitals-only, and multimodal models, evaluating accuracy, class-wise performance, robustness to noisy vitals, and inference latency.

2. Literature review

2.1. Multimodal artificial intelligence and physiological fusion

Demrozi *et al.* (2025) discussed multimodal architectures combining wearable sensor data and clinical variables for continuous monitoring. Similarly, Lu *et al.* (2025) proposed VL-RiskFormer to fuse structured data, time series, imaging, and text for chronic disease risk prediction. Sekhar *et al.* (2025) reported that MedFusionAI demonstrates that integrating heterogeneous medical data can improve the robustness of chronic disease risk models. Xiang *et al.* (2025) used multimodal deep learning for stress detection from physiological signals. Kotula *et al.* (2025) compared multimodal approaches for deterioration prediction in hospital wards.

2.2. Blood pressure variability, vitals, and renal outcomes

Rostamian *et al.* (2025) reported that visit-to-visit blood pressure variability (BPV) is associated with adverse renal and cardiovascular outcomes. Ruan *et al.* (2025) showed that high BPV is an independent risk factor for renal outcomes in elderly hypertensive patients. Earlier work by Wang *et al.* (2022) showed that short-term BPV correlates with outcomes in non-dialysis CKD, highlighting the importance of hemodynamic time-series.

2.3. Explainable AI for chronic kidney disease

Balikci Cicek and Kucukakcali (2023) applied SHAP to interpret CKD risk factors. Tsai *et al.* (2023) provided SHAP-based explanations for a CKD risk model. Shu *et al.* (2025) used SHAP for mortality prediction in hemodialysis. Furthermore, Ghosh and Khandoker (2024) evaluated XAI models across clinical tasks, showing improved transparency and trust. These works motivate interpretable multimodal frameworks such as the one proposed here.

3. System architecture

The proposed multimodal CKD prediction framework is illustrated in Figures 1 and 2. Figure 1 presents the overall workflow of the proposed framework, whereas Figure 2 illustrates the model architecture, which integrates structured clinical features and temporal vital-sign data through feature-level fusion.

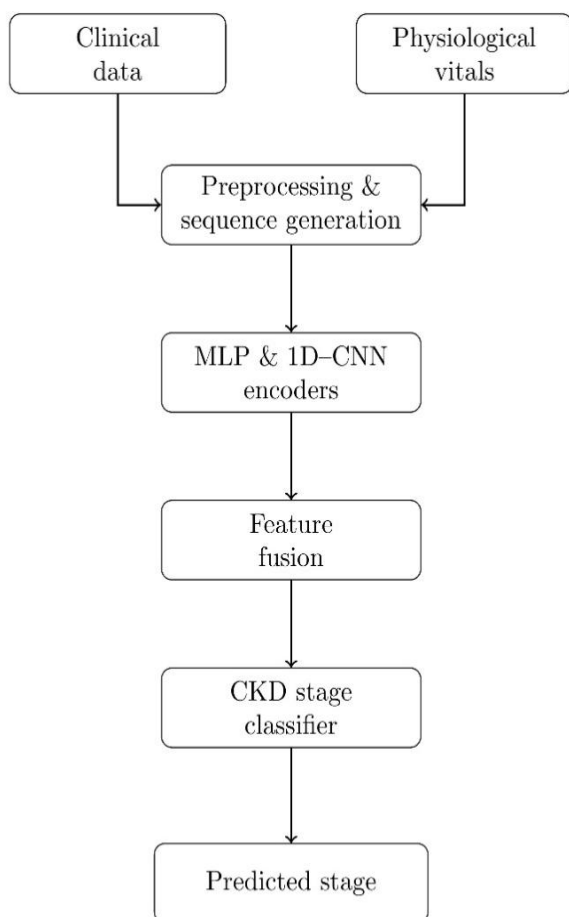


Figure 1. Flowchart of the proposed multimodal CKD prediction framework

Abbreviations: 1D-CNN: One-dimensional convolutional neural network; CKD: Chronic kidney disease; MLP: Multilayer perceptron.

The mathematical formulation of the framework is described below.

Let $x_c \in \mathbb{R}^{d_c}$ denote the normalized clinical feature vector, where d_c represents the number of structured clinical attributes. Let $X_v \in \mathbb{R}^{T \times d_v}$ denote the multivariate physiological vital-sign sequence of length T with d_v physiological channels. All the mathematical notation of the work is given in Table 1.

The clinical branch uses an MLP encoder to transform the structured clinical input into a latent embedding, as defined in Equation 1:

$$z_c = f_c(x_c) = \sigma_L(W_L L \sigma_1(W_1 x_c + b_1) + b_L) \quad (1)$$

The temporal physiological branch employs stacked 1D convolutional blocks to learn sequential hemodynamic representations (Equation 2):

$$H^{(k)} = \phi(\text{Conv}^{(k)}(H^{(k-1)})), H^{(0)} = X_v \quad (2)$$

To improve interpretability and temporal localization, an attention-weighting mechanism is applied over the intermediate temporal representations prior to pooling. The normalized temporal attention coefficient at time step t is computed as in Equation 3:

$$\alpha_t = \frac{\exp(w_a^T h_t)}{\sum_{j=1}^T \exp(w_a^T h_j)}, t=1,2,\dots,T \quad (3)$$

The coefficient α_t quantifies the relative importance assigned to each physiological temporal segment.

The attention-weighted temporal representation is then obtained as in Equation 4:

$$z_a = \sum_{t=1}^T \alpha_t h_t \quad (4)$$

Subsequently, global average pooling is applied to stabilize the learned temporal representation (Equation 5):

$$z_v = \text{GAP}(z_a) \quad (5)$$

The clinical and physiological embeddings are fused using feature-level concatenation (Equation 6):

$$z = \phi(W_f [z_c; z_v] + b_f) \quad (6)$$

The fused representation is finally passed through a softmax classifier to estimate CKD stage probabilities (Equation 7):

$$\hat{p} = \text{softmax}(W_o z + b_o) \quad (7)$$

To address class imbalance during optimization, weighted categorical cross-entropy loss is employed (Equation 8):

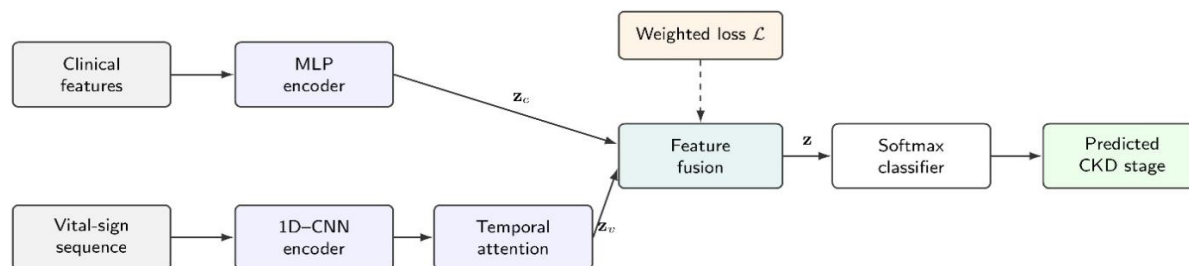


Figure 2. Simplified multimodal CKD prediction architecture integrating clinical features and vital-sign sequences through feature fusion
Abbreviations: 1D-CNN: One-dimensional convolutional neural network; CKD: Chronic kidney disease; MLP: Multilayer perceptron.

$$L = - \sum_{i=1}^N \sum_{k=1}^K w_k I(y_i = k) \log \hat{p}_{i,k} \quad (8)$$

4. Methodology

4.1. Dataset and labeling

The CKD dataset comprised laboratory markers (e.g., serum creatinine, eGFR, hemoglobin, albumin, potassium, sodium, proteinuria) and demographic variables. KDIGO guidelines were used to assign CKD stages (1–5) (guidelines_kdigo). Stage-aware vital-sign sequences were synthetically generated for each patient using Gaussian mixture models whose statistical parameters are derived from hemodynamic trends reported by Rostamian *et al.* (2025), Ruan *et al.* (2025), and Wang *et al.* (2022). Clinical features were standardized, categorical variables were one-hot encoded, and outliers were winsorized. Vital sequences were min-max scaled and optionally smoothed. Temporal derivatives and variability indices might be appended as additional channels. The model was trained using the Adam optimizer with a learning rate of 10^{-3} and a batch size of 64, with early stopping based on the validation macro-F1 score. The data were stratified and split into training, validation, and test sets at a ratio of 70%:15%:15%. Model performance was evaluated using accuracy, macro-F1 score, stage-specific F1 scores, and the area under the receiver operating characteristic curve (AUROC).

4.2. Network configuration

The MLP branch has three dense layers (128, 64, 32 units; ReLU + dropout). The 1D-CNN branch has two convolutional blocks (32 and 64 filters; kernel sizes 5 and 3) with ReLU and max pooling, followed by global average pooling. Fused vector z passes through two dense layers (64, 32 units) before the softmax. Class weights w_k are set inversely proportional to class frequencies.

4.3. Training and inference

Algorithms 1 and 2 formalize the end-to-end execution of the proposed framework and make explicit how the multimodal architecture is integrated into a practical workflow. Algorithm 1 describes the supervised optimization loop in which clinical features and vitals sequences are jointly processed at mini-batch level, with the MLP and 1D-CNN branches producing modality-specific embeddings that are fused and fed to the softmax head. The use of class weights w_k in the loss L ensures that minority CKD stages contribute proportionally to the gradient, while the outer validation check implements early stopping, preventing overfitting once the validation metric saturates. Algorithm 2 mirrors the same computational graph but omits gradient updates, enabling deterministic and low-latency stage prediction for a new patient by reusing the learned encoders and fusion layer. Together, these algorithms clarify how the mathematical formulation in Equations 1–8 is realized in practice, and provide a blueprint for reproducible implementation in clinical decision-support systems.

5. Results and discussion

5.1. Performance and convergence behavior

As shown in Figure 3, the fused MLP-1D CNN model gets an accuracy of 0.90, a macro-F1 score of 0.86, and an AUROC of 0.96. The clinical-only model gets 0.84, 0.78, and 0.92, respectively. The vitals-only branch does the worst, with accuracy, macro-F1, and AUROC only reaching 0.70, 0.65, and 0.81. These results show that structured laboratory variables are still the main source of discriminative information. However, temporal physiological signals give additional clues that make stage discrimination much better when combined with biochemical markers. In absolute terms, multimodal learning increases accuracy by 0.06 and macro-F1 by 0.08 compared to the clinical-only configuration. This shows

Table 1. Abbreviations and mathematical notations used in the present work

Symbol/acronym	Description
CKD	Chronic Kidney Disease
MLP	Multilayer Perceptron
1D-CNN	One-Dimensional Convolutional Neural Network
SHAP	SHapley Additive exPlanations
SBP	Systolic Blood Pressure
HR	Heart Rate
GAP	Global Average Pooling
AUROC	Area Under the Receiver Operating Characteristic Curve
$\mathbf{x}_c$	Normalized structured clinical feature vector
d_c	Number of clinical features
$\mathbf{X}_v$	Multivariate physiological vital-sign sequence
T	Length of temporal physiological sequence
d_v	Number of physiological channels
$\mathbf{z}_c$	Latent embedding generated from the clinical MLP branch
$f_c(\cdot)$	Clinical feature encoding function implemented using MLP layers
$\mathbf{W}_l$	Weight matrix of the l^{th} neural layer
$\mathbf{b}_l$	Bias vector of the l^{th} neural layer
$\sigma_l(\cdot)$	Activation function of the l^{th} MLP layer
$\mathbf{H}^{(k)}$	Hidden temporal representation at the k^{th} convolutional layer
$\text{Conv}^{(k)}$	One-dimensional convolution operation at layer k
$\phi(\cdot)$	Nonlinear activation function used in CNN and fusion layers
$\mathbf{h}_t$	Temporal feature vector at time step t
$\mathbf{w}_a$	Trainable temporal attention parameter vector
α_t	Normalized temporal attention coefficient at time step t
$\mathbf{z}_a$	Attention-weighted temporal embedding representation
$\mathbf{z}_v$	Aggregated physiological embedding after pooling
$\mathbf{z}$	Final fused multimodal feature representation
$\mathbf{W}_f$	Weight matrix of the feature fusion layer
$\mathbf{b}_f$	Bias vector of the feature fusion layer
$\hat{\mathbf{p}}$	Predicted probability distribution over CKD stages
$\mathbf{W}_o$	Weight matrix of the output softmax classifier
$\mathbf{b}_o$	Bias vector of the output classifier
L	Weighted categorical cross-entropy loss function
N	Number of training samples
K	Number of CKD stage classes
w_k	Class weight associated with CKD stage k
$I(\cdot)$	Indicator function used in weighted loss computation
$\hat{p}_{i,k}$	Predicted probability of sample i belonging to class k
y_i	Ground-truth CKD stage label for sample i

that the temporal branch adds useful information instead of just noise. The comparatively lower performance of the vitals-only branch indicates that short-term physiological dynamics alone are insufficient for reliable CKD staging because renal impairment is fundamentally characterized by biochemical abnormalities rather than isolated hemodynamic fluctuations. Physiological signals such as BPV and heart-rate trends may reflect cardiovascular instability and autonomic responses, but they do not independently provide adequate renal specificity for accurate stage discrimination.

Algorithm 1. Multimodal CKD training pipeline with temporal attention

Require: Clinical dataset D_c , labels $\{y_i\}$, vital-sign generator or vital dataset D_v , epochs E

- 1: Generate or load vital-sign sequences D_v aligned with D_c
- 2: Preprocess clinical features X_c and vital-sign sequences X_v
- 3: Compute class weights $\{W_i\}$ from CKD stage label frequencies
- 4: Initialize MLP parameters, 1D-CNN parameters, attention vector W_o , fusion layer, and softmax classifier
- 5: for epoch = 1 to E do
- 6: for each mini-batch $\{X_c^{(i)}, X_v^{(i)}, y^{(i)}\}$ do
- 7: $Z_c^{(i)} \leftarrow f_c(X_c^{(i)}) \triangleright$ Clinical encoder, **Equation 1**
- 8: $H^{(K,i)} \leftarrow \text{CNN}(X_v^{(i)}) \triangleright$ Temporal encoder, **Equation 2**
- 9: Extract temporal feature vectors $\{h_t^{(i)}\}_{t=1}^T$ from $H^{(K,i)}$
- 10: Compute temporal attention weights $\{\alpha_t^{(i)}\}_{t=1}^T \triangleright$ **Equation 3**
- 11: $Z_a^{(i)} \leftarrow \sum_{t=1}^T \alpha_t^{(i)} h_t^{(i)} \triangleright$ Attention pooling, **Equation 4**
- 12: $Z_v^{(i)} \leftarrow \text{GAP}(Z_a^{(i)}) \triangleright$ Physiological embedding, **Equation 5**
- 13: $Z^{(i)} \leftarrow \phi(W_f[Z_c^{(i)}; Z_v^{(i)}] + b_f) \triangleright$ Fusion, **Equation 6**
- 14: $\hat{p}^{(i)} \leftarrow \text{softmax}(W_o Z^{(i)} + b_o) \triangleright$ Classifier, **Equation 7**
- 15: Compute weighted loss $\mathcal{L}$ using **Equation 8**
- 16: Backpropagate $\nabla \mathcal{L}$ and update all trainable parameters using Adam
- 17: Evaluate validation macro-F1 and apply early stopping if convergence criterion is satisfied

The convergence profile shown in **Figure 4** adds to the evidence that the proposed framework is stable. The accuracy of training goes up from 0.68 at the first epoch to 0.89 at epoch 30. The accuracy of validation goes up from 0.67 to 0.88 at the same time. The small, almost constant gap between the two curves shows that the network generalizes well and does not show a lot of overfitting during optimization. This behavior is in line with the regularization strategy that was used, which included dropout, early stopping, and class-weighted training. From a practical point of view, the model reaches a stable operating state after only a few epochs, which is good for development that can be repeated and for future improvements that are focused on deployment.

Algorithm 2. Inference and stage-wise CKD prediction with temporal attention

Require: Trained model parameters, new patient clinical features X_c , vital-sign sequence X_v

- 1: Preprocess X_c and X_v using the training preprocessing pipeline
- 2: $Z_c \leftarrow f_c(X_c) \triangleright$ Clinical encoder
- 3: $H^{(K)} \leftarrow \text{CNN}(X_v) \triangleright$ Temporal encoder
- 4: Extract temporal feature vectors $\{h_t\}_{t=1}^T$ from $H^{(K)}$
- 5: Compute temporal attention weights $\{\alpha_t\}_{t=1}^T$
- 6: $Z_a \leftarrow \sum_{t=1}^T \alpha_t h_t$
- 7: $Z_v \leftarrow \text{GAP}(Z_a)$
- 8: $z \leftarrow \phi(W_f[Z_c; Z_v] + b_f)$
- 9: $\hat{p} \leftarrow \text{softmax}(W_o z + b_o)$
- 10: $\hat{y} \leftarrow \arg \max_k \hat{p}_k$
- 11: return predicted CKD stage $\hat{y}$, probability vector $\hat{p}$, and temporal attention weights $\{\alpha_t\}_{t=1}^T$

In **Figure 5**, the multimodal model is clearly better than both baselines, which show the same trend. The gain over the model that only looks at vitals is very interesting. Physiological signals by themselves provide useful time context, but they are not enough for reliable CKD staging without being linked to renal biomarkers. In contrast, feature-level fusion lets the classifier look at short-term changes in blood flow in the context of kidney function indicators. This makes the representation of disease severity more distinct.

Figure 6 shows the receiver operating characteristics, which show that the multimodal framework can distinguish things better over the entire range of operation. The multimodal model achieves a true positive rate of approximately 0.86, the clinical-only model a rate of 0.80, and the vitals-only model a rate of 0.70. Even at a tighter operating point where the false positive rate is around 0.1, the proposed model still yields a true positive rate of about 0.75. Such behavior is relevant in a clinical setting, because improved sensitivity at fixed false-positive rates translates directly into the usefulness of a screening-oriented decision aid. Thus, the higher AUROC of 0.96 indicates not only better overall performance but also better ranking quality at clinically relevant thresholds.

More information is provided by the confusion matrix in **Figure 7**. The correct classification rates along the diagonal are still high for all five stages of CKD, with values around 0.88, 0.82, 0.83, 0.86, and 0.88, respectively. Misclassifications occur mainly between neighboring stages rather than between categories that are distant in terms of clinical significance. This type of error structure is normal because the stages of CKD are not separated by sharp boundaries, but by slow changes in physiology and

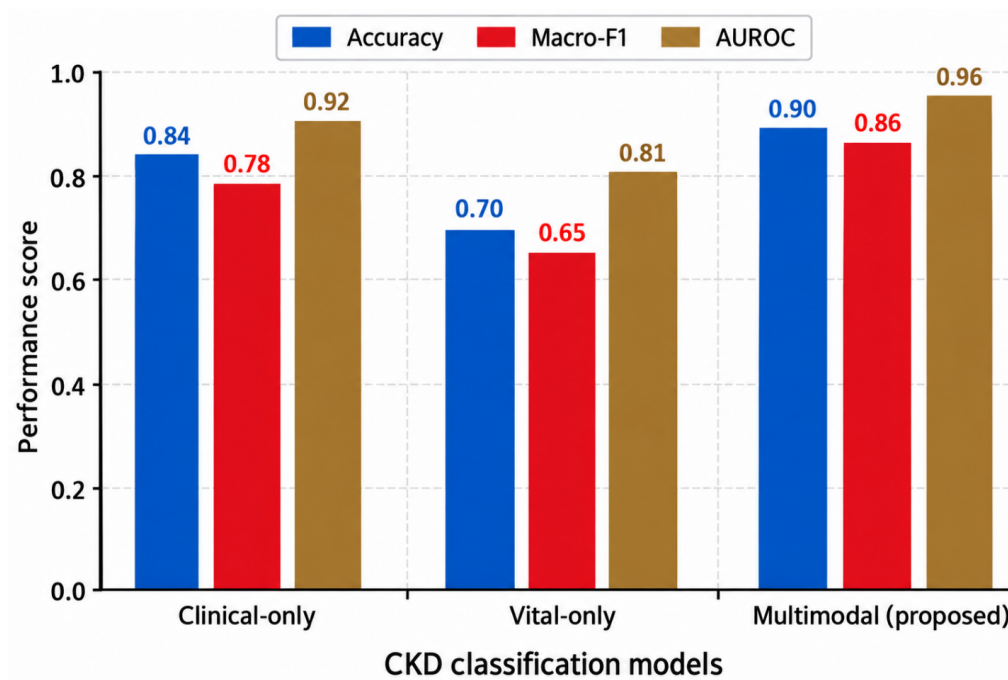


Figure 3. Comparative performance analysis of clinical-only, vitals-only, and proposed multimodal CKD classification models using accuracy, macro-F1, and AUROC metrics

Abbreviations: AUROC: Area under the receiver operating characteristic curve; CKD: Chronic kidney disease; Macro-F1: Macro-averaged F1 score.

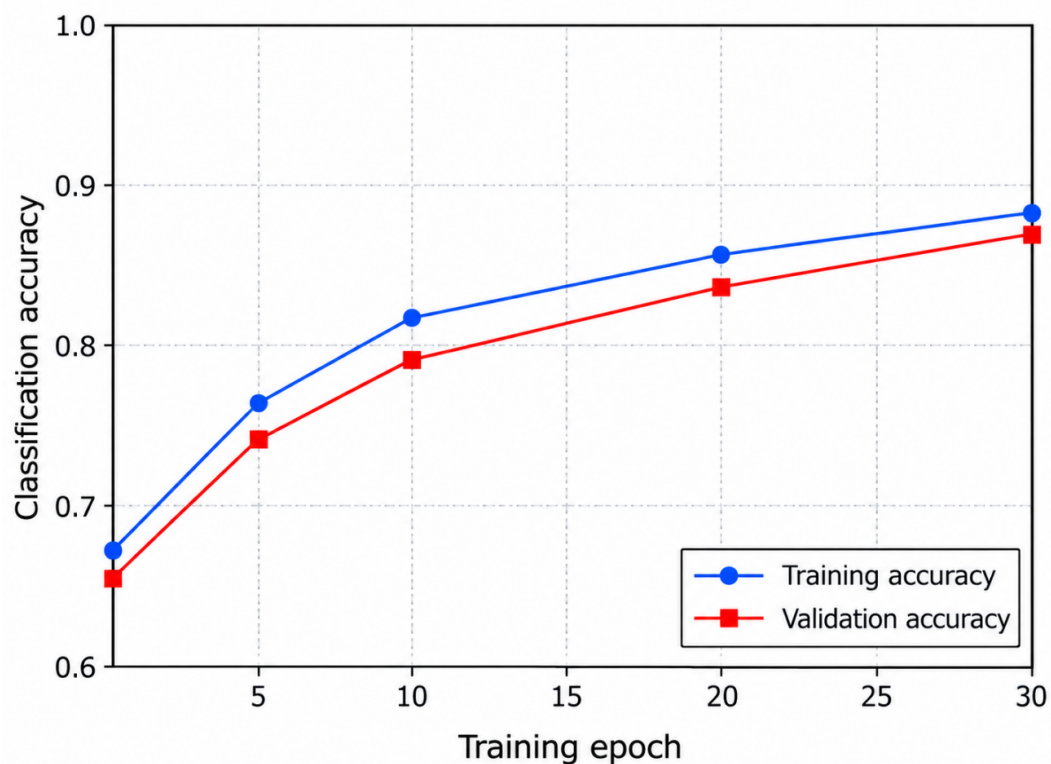


Figure 4. Training and validation classification accuracy curves of the proposed multimodal chronic kidney disease prediction framework across training epochs

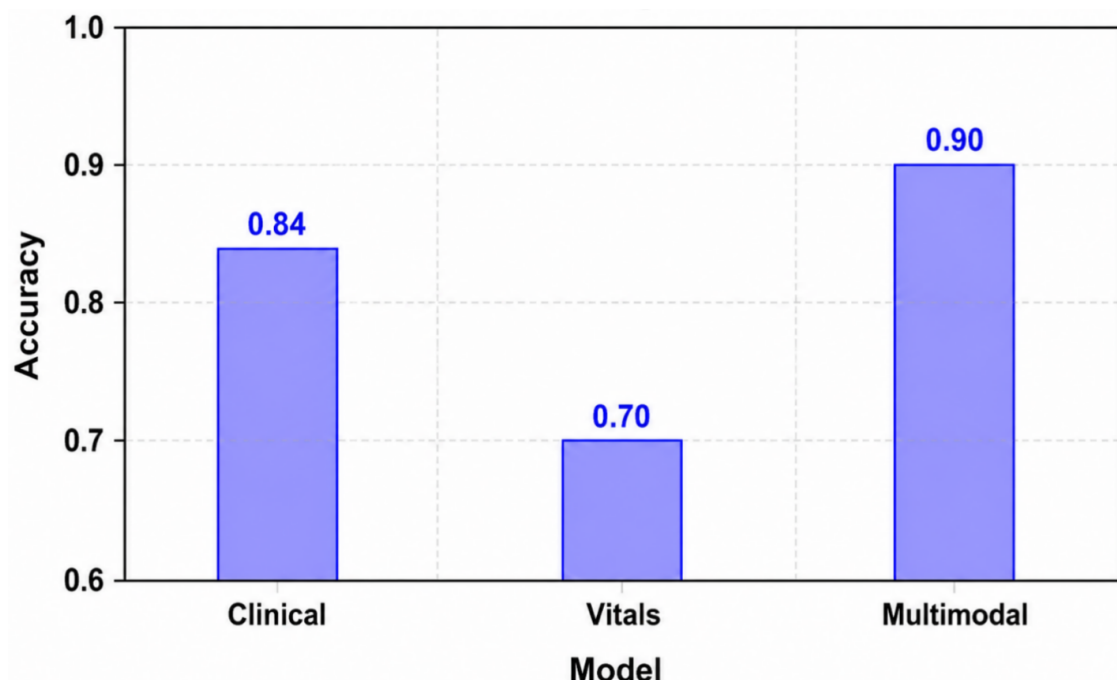


Figure 5. Overall accuracy comparison across models

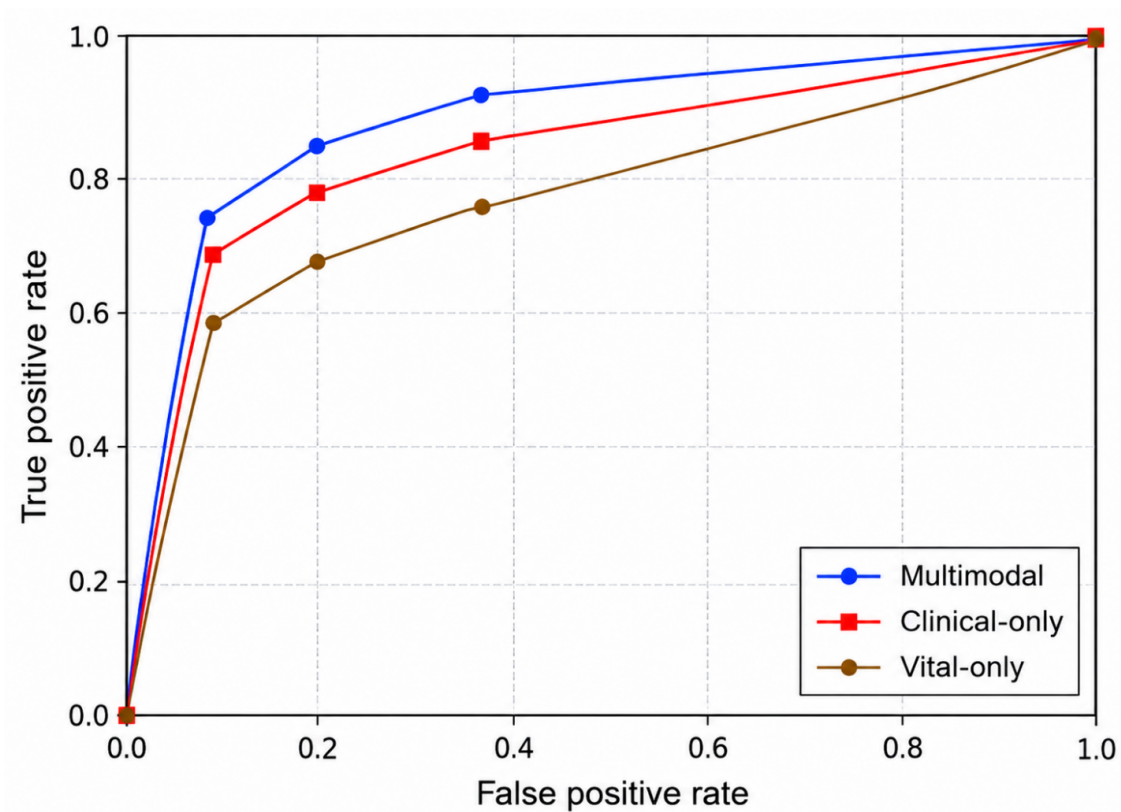


Figure 6. Receiver operating characteristic curves for multimodal and baseline models

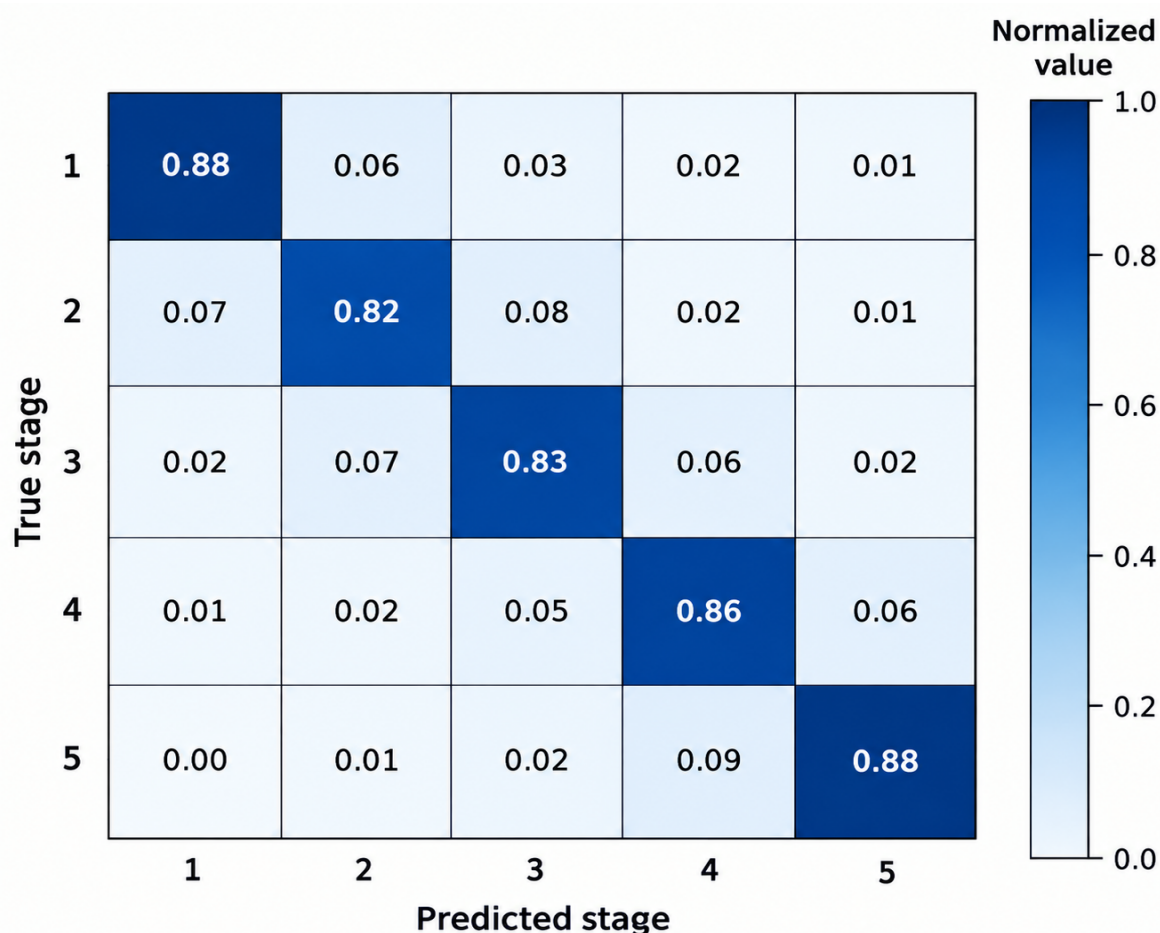


Figure 7. Schematic normalized confusion matrix for multimodal chronic kidney disease classification

biochemistry. The matrix indicates minimal confusion between Stage 1 and Stage 5, an important aspect as it implies the model preserves the overall sequence of diseases as they become more challenging to subdivide. For clinical support systems, this is a good feature because usually severe disease is not mild impairment.

Figure 8 shows the stage-specific F1 results, which demonstrate that the benefits of multimodal fusion are not limited to a specific subgroup. The proposed model achieves F1 scores of 0.84, 0.82, 0.85, 0.90, and 0.92 for Stages 1–5, respectively, compared to 0.80, 0.75, 0.77, 0.88, and 0.90 for the clinical-only baseline. The biggest changes are in Stages 2 and 3. This finding has clinical relevance, in that intermediate stages of CKD often represent the most diagnostically ambiguous phase, in which static laboratory measurements may poorly reflect evolving physiological instability. In this transition zone, the additional temporal modality appears to clarify the separation of classes. This multimodal design not only improves the average

performance but also equalizes the stages in the continuum of the disease.

5.2. Interpretability of the learned multimodal representation

The feature attribution pattern in Figure 9 is in line with what is already known in nephrology, but also shows the effect of physiological variability. The most important inputs are eGFR and serum creatinine, with relative importance values of around 0.85 and 0.80. This makes sense because both of these factors are directly related to the filtering ability of the kidneys and are still relevant to the definition of stages. Hemoglobin and albumin are moderately important, indicating that anemia burden and nutritional or inflammatory status contribute to the decision function.

It is interesting to note that blood pressure and heart rate variability are still important, with values close to 0.50 and 0.45. These features do not go beyond the core renal

biomarkers, but their contribution is big enough to show that the temporal branch is important for prediction. These findings indicate that the model does not solely rely on tabular clinical features but also incorporates information from vital-sign dynamics. Instead, it uses hemodynamic fluctuation as an extra signal to improve stage estimation, especially when lab values alone may be close to or overlapping. This makes it more likely that the gain from fusion comes from information that makes sense from a physiological point of view, not from unnecessary model complexity.

5.3. Temporal response and robustness to signal corruption

The attention-like visualization in Figure 10 shows that the temporal encoder gives more weight to short periods of systolic blood pressure and heart rate instability rather than treating the whole sequence the same way. Dark areas group around the middle parts of both channels, which could mean that there were long periods of time when the signal was higher or lower.

Figure 10 shows that the trained temporal branch assigns higher normalized attention weights to mid-sequence SBP and HR segments, indicating that localized

periods of physiological variability contribute more strongly to the multimodal CKD stage prediction than uniformly stable intervals.

The robustness analysis in Figure 11 shows that the performance only gradually drops in the presence of additive Gaussian noise to the vitals stream. The macro-F1 score decreases from 0.86 in a clean setting to 0.85, 0.84, and 0.83 at noise multipliers of 0.2, 0.4, and 0.6, respectively. This small change of only 0.03 across this range indicates that the learned representation is not very sensitive to small changes in the quality of the waveforms. This is a positive outcome for those situations in which remote monitoring is used, as measurements made with devices at home or on the body are usually affected by motion artifacts, sensor drift, or transmission issues. Therefore, the model appears to depend on a large time structure rather than on the weak sample-level details.

5.4. Limitations

The proposed framework requires validation using larger datasets. Its reproducibility and generalizability remain dependent on the availability of standardized multimodal CKD datasets that integrate laboratory biomarkers with longitudinal physiological recordings.

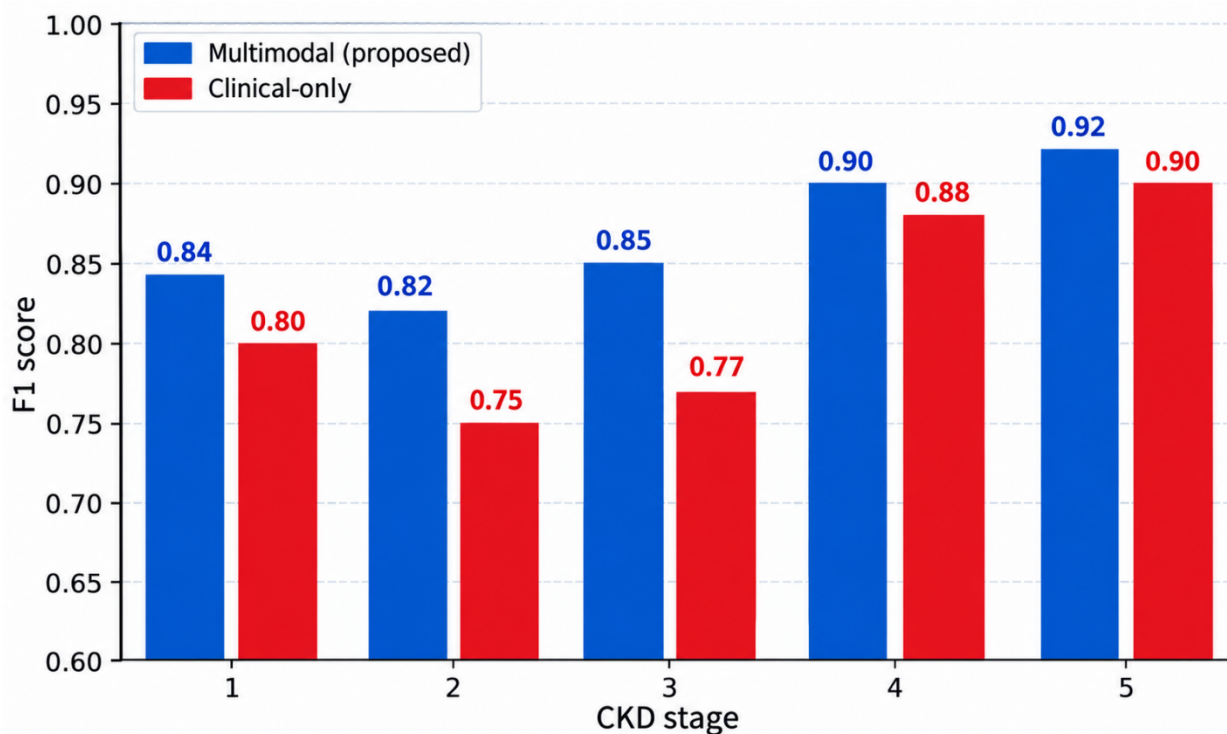


Figure 8. Stage-specific F1 scores comparison
Abbreviation: CKD: Chronic kidney disease.

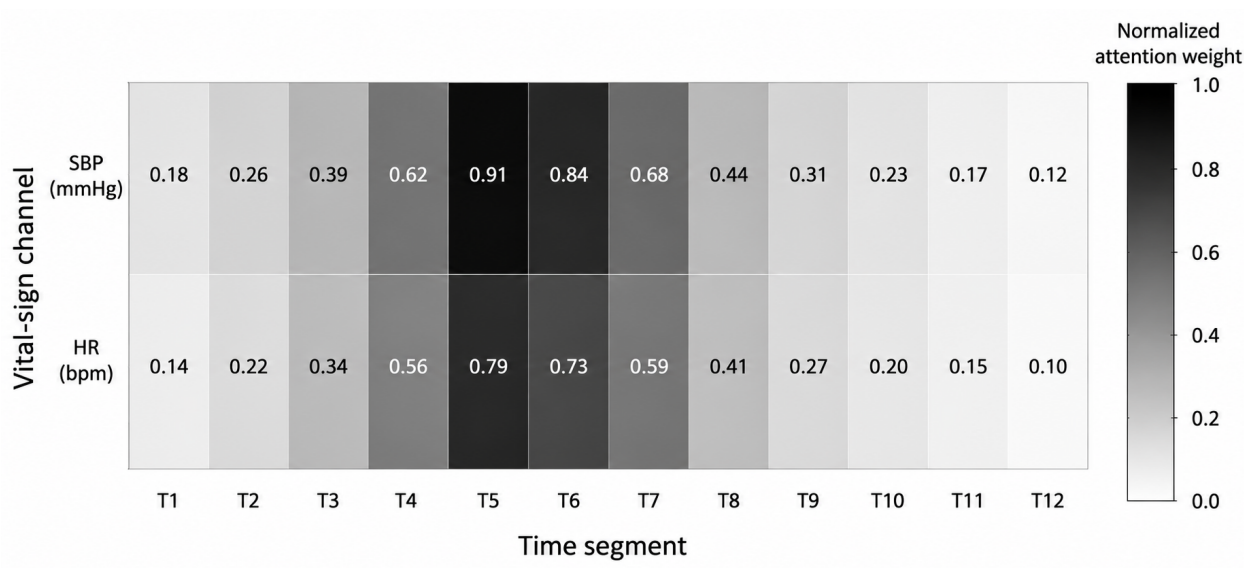


Figure 9. SHAP-style global feature importance (clinical + physiological features)
Abbreviations: HR: Heart rate; SBP: Systolic blood pressure; SHAP: SHapley Additive exPlanations.

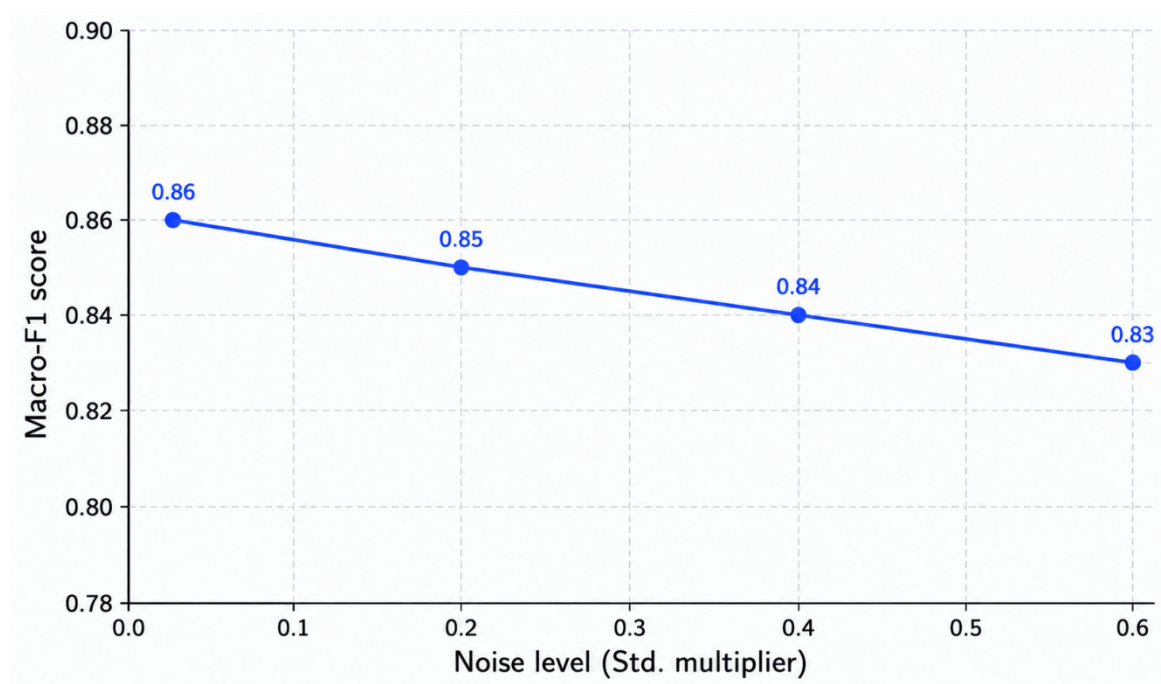


Figure 10. Temporal attention map for systolic blood pressure and heart rate channels

5.5. Future directions

Future work should focus on integrating real-time physiological data from wearable devices for continuous

CKD assessment, validating the proposed framework using multi-center and publicly available clinical datasets, and enhancing XAI modules for clinical decision support.

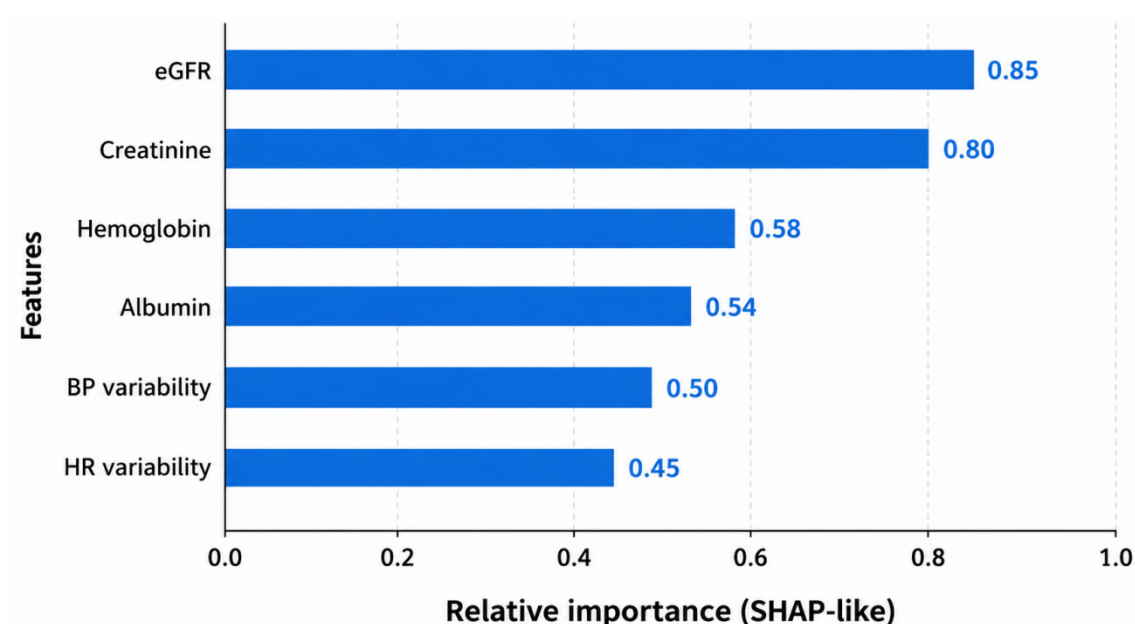


Figure 11. Robustness of the multimodal model under increasing vitals noise

Abbreviations: BP: Blood pressure; eGFR: Estimated glomerular filtration rate; HR: Heart rate; SHAP: SHapley Additive exPlanations.

6. Conclusion

This work presented a multimodal deep learning framework for early prediction and stage-wise classification of CKD by jointly modeling structured clinical biomarkers and short-term physiological dynamics. By integrating an MLP-based clinical encoder with a 1D-CNN-based temporal encoder through feature-level fusion, the proposed architecture produced more reliable CKD stage discrimination than either modality used in isolation. The experimental results showed that the multimodal model achieved an accuracy of 0.90, a macro-F1 score of 0.86, and an AUROC of 0.96, exceeding the clinical-only and vitals-only baselines across all major evaluation metrics. The improvement was especially notable in intermediate CKD stages, where diagnostic separation is often more difficult and where physiological variability appears to supply clinically relevant supplementary information.

Acknowledgments

None.

Funding

This research received no external funding.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Manjulabai Bhadrashetty

Formal analysis: Dayanand Jamkhandikar

Investigation: Manjulabai Bhadrashetty

Methodology: Dayanand Jamkhandikar

Writing—original draft: Manjulabai Bhadrashetty

Writing—review & editing: Dayanand Jamkhandikar

Availability of data

The data and/or materials that support the findings of this study are available from the corresponding author upon reasonable request.

Further disclosure

During the preparation of this manuscript, the authors used Grammarly 2026 for corrections of grammatical mistakes. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

References

- Alsekait, D. M., Saleh, H., Gabralla, L. A., Alnowaiser, K., El-Sappagh, S., Sahal, R., & El-Rashidy, N. (2023). Toward comprehensive chronic kidney disease prediction based on ensemble deep learning models. *Applied Sciences*, 13(6), 3937.

<https://doi.org/10.3390/app13063937>

- Balikci Cicek, I., & Kucukakcali, Z. (2023). Explanaible artificial intelligence method SHAP's prediction of risk factors associated with chronic kidney disease combined with black box methods. *Journal of Community Medicine and Public Health Reports*, 4(10).
<https://doi.org/10.38207/jcmphr/2023/nov041004138>
- Demrozi, F., Farmanbar, M., & Engan, K. (2025). Multimodal AI (MMAI) for next-generation healthcare: Data domains, algorithms, challenges, and future perspectives. *Current Opinion in Biomedical Engineering*, 100632.
<https://doi.org/10.1016/j.cobme.2025.100632>
- Ghosh, S. K., & Khandoker, A. H. (2024). Investigation on explainable machine learning models to predict chronic kidney diseases. *Scientific Reports*, 14(1), 3687.
<https://doi.org/10.1038/s41598-024-54375-4>
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., Saha, S., Hossen, R., Ahmed, S., Rony, M. A. T., & Akter, M. F. (2024). ML-CKDP: Machine learning-based chronic kidney disease prediction with smart web application. *Journal of Pathology Informatics*, 15, 100371.
<https://doi.org/10.1016/j.jpi.2024.100371>
- Hegde, G., Shenoy, P. D., & Canchi, A. (2025). A deep learning framework for chronic kidney disease stage classification. *Healthcare Analytics*, 7, 100398.
<https://doi.org/10.1016/j.health.2025.100398>
- Iliyas, I. I., Boukari, S., & Gital, A. Y. U. (2025). Recent trends in prediction of chronic kidney disease using different learning approaches: A systematic literature review. *Journal of Medical Artificial Intelligence*, 8, 62.
<https://doi.org/10.21037/jmai-24-256>
- Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, 14, 100189.
<https://doi.org/10.1016/j.jpi.2023.100189>
- Jager, K. J., Kovesdy, C., Langham, R., Rosenberg, M., Jha, V., & Zoccali, C. (2019). A single number for advocacy and communication—Worldwide more than 850 million individuals have kidney diseases. *Kidney International*, 96(5), 1048–1050.
<https://doi.org/10.1016/j.kint.2019.07.012>
- Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. (2024). KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International*, 105(4S), S117–S314.
<https://doi.org/10.1016/j.kint.2023.10.018>
- Khan, N., Raza, M. A., Mirjat, N. H., Balouch, N., Abbas, G., Yousef, A., & Touti, E. (2024). Unveiling the predictive power: A comprehensive study of machine learning model for anticipating chronic kidney disease. *Frontiers in Artificial Intelligence*, 6, 1339988.
<https://doi.org/10.3389/frai.2023.1339988>
- Kotula, C. A., Martin, J., Carey, K. A., Edelson, D. P., Dligach, D., Mayampurath, A., Afshar, M., & Churpek, M. M. (2025). Comparison of multimodal deep learning approaches for predicting clinical deterioration in ward patients: Observational cohort study. *Journal of Medical Internet Research*, 27, e75340.
<https://doi.org/10.2196/75340>
- Li, J., Du, X., Zhang, R., Li, X., Xu, J., Song, X., Zhao, Y., Liu, L., Xu, G., & Fan, Y. (2025). Machine learning models for predicting short-term progression in patients with stage 4 chronic kidney disease: A multi-center validation study. *Scientific Reports*, 15(1), 39285.
<https://doi.org/10.1038/s41598-025-23037-4>
- Lu, D., Wu, S., & Huang, X. (2025). Multimodal health risk prediction system for chronic diseases via vision-language fusion and large language models. In *Human-Computer Interaction and Natural Language Processing (ICAHN)* (pp. 76–79). IEEE.
<https://doi.org/10.1109/icahn67688.2025.00022>
- M. Chandra Sekhar, Pasam Naga Kavitha, B. N. V. Uma Shankar, Amrutavalli Aakula, & Divya Boya. (2025). MedFusionAI: A deep learning framework for multi-modal health data fusion to predict chronic disease risks. *International Journal of Computational and Experimental Science and Engineering*, 11(2).
<https://doi.org/10.22399/ijcesen.2159>
- Rostamian, S., Kaura, A., Ariti, C., Gupta, A., Whitely, W., Poulter, N. R., & Sever, P. (2025). The impact of blood pressure variability on cardiovascular and renal outcomes in patients with and without diabetes: Insights from the Anglo-Scandinavian Cardiac Outcomes Trial. *Journal of Hypertension*, 43(12), 2033–2041.
<https://doi.org/10.1097/HJH.0000000000004153>
- Ruan, Y., Chen, Y., Huang, N., Wang, D., Xu, Y., Fan, J., Chen, W., & Wang, X. (2025). Long-term BPV is an independent risk factor for renal prognosis in hypertensive patients — A post-hoc analysis of the SPRINT study. *International Journal of Medical Sciences*, 22(10), 2298–2307.
<https://doi.org/10.7150/ijms.111843>
- Saif, D., Sarhan, A. M., & Elshennawy, N. M. (2024). Deep-kidney: An effective deep learning framework for chronic kidney disease prediction. *Health Information Science and Systems*, 12(1), 3.
<https://doi.org/10.1007/s13755-023-00261-8>
- Shu, P., Wang, X., Wen, Z., Chen, J., & Xu, F. (2025). SHAP combined with machine learning to predict mortality risk in maintenance hemodialysis patients: A retrospective study.

Frontiers in Medicine, 12, 1615950.

<https://doi.org/10.3389/fmed.2025.1615950>

Tisha, S. A., Bhattacharjee, B., Ahmed, M. P., Mukaddim, A. A., Hider, M. A., & Khan, M. T. (2025). Deep learning-driven early detection and risk prediction of chronic diseases using multimodal healthcare data. *SSRN Electronic Journal*.

<https://doi.org/10.2139/ssrn.5784022>

Tsai, M.-C., Lojanapiwat, B., Chang, C.-C., Noppakun, K., Khumrin, P., Li, S.-H., Lee, C.-Y., Lee, H.-C., & Khwanngern, K. (2023). Risk prediction model for chronic kidney disease in Thailand using artificial intelligence and SHAP. *Diagnostics*, 13(23), 3548.

<https://doi.org/10.3390/diagnostics13233548>

Wang, G., Ma, K., Ma, Z., Guo, X., Wang, Y., Ma, L., Qi, C., Li,

Y., & Zhou, X. (2022). Short-term blood pressure variability and outcomes in non-dialysis chronic kidney disease. *Frontiers in Medicine*, 9, 911205.

<https://doi.org/10.3389/fmed.2022.911205>

Xiang, J. Z., Wang, Q. Y., Fang, Z. B., Esquivel, J. A., & Su, Z. X. (2025). A multi-modal deep learning approach for stress detection using physiological signals: Integrating time and frequency domain features. *Frontiers in Physiology*, 16, 1584299.

<https://doi.org/10.3389/fphys.2025.1584299>

Zhu, Y., Bi, D., Saunders, M., & Ji, Y. (2023). Prediction of chronic kidney disease progression using recurrent neural network and electronic health records. *Scientific Reports*, 13(1), 22091.

<https://doi.org/10.1038/s41598-023-49271-2>