electronics

MDPI

Article

A Functionally Guided U-Net for Chronic Kidney Disease Assessment: Joint Structural Segmentation and eGFR Prediction with a Structure–Function Consistency Loss

Omar Al-Salman * and Mesut Cevik

Department of Electrical and Computer Engineering, Altinbas University, 34218 Istanbul, Turkey; mesut.cevik@altinbas.edu.tr

* Correspondence: 213720164@ogr.altinbas.edu.tr

Abstract

An accurate assessment of chronic kidney disease (CKD) requires understanding both renal morphology and functional decline, yet most deep learning approaches treat segmentation and eGFR prediction as separate tasks. This paper proposes the Functionally Guided CKD U-Net (FG-CKD-UNet), a dual-headed multitask architecture that integrates multi-class kidney segmentation with end-to-end eGFR prediction using a structure–function consistency loss. The model incorporates a morphological biomarker extractor to derive cortical thickness, kidney volume, and cortex–medulla ratios, enabling explicit coupling between anatomy and physiology. Experiments on T2-weighted MRI and colorized CT datasets demonstrate that the proposed method surpasses state-of-the-art segmentation baselines, achieving a Dice score of 0.94 and an HD95 of 9.8 mm. For functional prediction, the model achieves an MAE of 0.039, an RMSE of 0.058, and a Pearson correlation of 0.92, outperforming CNN, MLP, and ResNet baselines. The structure–function consistency mechanism reduces the consistency error from 0.071 to 0.042, confirming coherent physiological modeling. The results indicate that the FG-CKD-UNet provides a reliable, interpretable, and physiologically grounded framework for comprehensive CKD assessment.

Keywords: chronic kidney disease; kidney segmentation; eGFR prediction; deep learning; multi-task learning; medical image analysis

1. Introduction

Chronic Kidney Disease (CKD) is a major global health concern, affecting millions worldwide and imposing a substantial burden on healthcare systems [1]. The clinical assessment of CKD progression traditionally relies on biochemical indicators such as serum creatinine and the estimated glomerular filtration rate (eGFR), while imaging modalities including ultrasound, CT, and MRI are routinely employed to evaluate the renal structure [2]. Structural alterations such as cortical thinning, medullary atrophy, and reduced kidney volume are well known to correlate with functional decline; however, most imaging-based AI approaches treat anatomical analysis and functional prediction as independent tasks [3]. Conventional U-Net variants emphasize pixel-level segmentation accuracy without incorporating physiological knowledge of CKD progression, while eGFR prediction models typically rely on global image features or clinical biomarkers without exploiting compartment-level morphology. This separation between structure and function represents a key limitation in current AI-assisted CKD assessment frameworks [4,5].

Check for updates

Academic Editors: Junchao Zhang and Chuanli Wang

Received: 22 November 2025
Revised: 27 December 2025
Accepted: 29 December 2025
Published: 30 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Recent advances in artificial intelligence have significantly accelerated research on automated kidney disease assessment, spanning chronic kidney disease (CKD), renal tumors, kidney stones, and immunological nephropathies. Rezk et al. [6] introduced an explainable AI framework that integrates Generative Adversarial Networks (GANs) and few-shot learning to predict CKD within medical IoT environments, emphasizing the importance of interpretability and performance in low-data settings. Their work highlights the growing shift toward transparent, clinically trustworthy CKD predictors. Building on ultrasound-based analysis, Obaid et al. [7] proposed a deep learning ensemble combined with Grad-CAM visualization to classify noisy kidney ultrasound images, demonstrating that ensemble strategies can mitigate variable image quality—a major challenge in real-world diagnostic workflows. Although not kidney-focused, Wang et al. [8] presented a modified YOLOv8 architecture with a novel C2fA module, offering insights into how architectural refinements and loss function optimization can enhance lesion detection in ultrasound images, principles that are transferable to renal pathology imaging tasks.

Within the domain of renal oncology, Maçin et al. [9] developed KidneyNeXt, a lightweight CNN designed for multi-class renal tumor classification in CT scans, achieving high accuracy while maintaining computational efficiency suitable for clinical integration. In parallel, Kulandaivelu et al. [10] proposed an adaptive multi-CNN feature fusion model incorporating attention mechanisms alongside improved heuristic optimization to detect kidney stones, highlighting the benefits of combining deep representation learning with metaheuristic-based parameter refinement. Further advancing classification architectures, Sharon and Anbarasi [11] introduced an attention-enhanced dilated bottleneck network that demonstrated superior discriminatory ability for kidney disease classification by leveraging expanded receptive fields and attention-driven feature recalibration. Loganathan and Palanivelan [12] contributed an explainable adaptive channel-weighting CNN for renal disorder classification in CT images, emphasizing model transparency through interpretable channel relevance explanations.

Beyond static imaging, Chaki and Uçar [13] proposed an inductive transfer-based ensemble deep learning method for kidney stone detection, demonstrating that the reuse of pretrained domain knowledge can significantly enhance robustness and generalizability. A closely related contribution by Almuayqil et al. [14] presented KidneyNet, a novel CNN-based approach for automated CKD diagnosis from CT scans, marking an important step toward imaging-based screening of chronic renal impairment and demonstrating that disease-relevant features can be effectively extracted from CT data. Meanwhile, Bingol et al. [15] proposed a hybrid deep learning model enhanced by Relief-based feature selection for kidney CT image classification, affirming that coupling handcrafted feature relevance techniques with deep neural networks can boost performance and reduce redundancy.

Shifting focus toward kidney diseases with strong immunological and clinical profiles, Wang et al. [16] introduced MAL-Net, a multi-label deep learning framework that integrates LSTM networks and multi-head attention to classify IgA nephropathy subtypes, exploiting temporal sensor data and attention mechanisms to capture complex clinical patterns. Complementing this work, Ren et al. [17] employed machine learning algorithms alongside weighted gene co-expression network analysis (WGCNA) to identify functional subtypes of IgA nephropathy, providing a genomics-driven approach that highlights the heterogeneous molecular signature of renal disease progression. Taken together, these studies illustrate an increasing trend toward hybrid, explainable, multimodal, and physiology-aware AI models for kidney disease analysis.

Despite remarkable progress, most existing works either focus solely on classification, emphasize explainability without integrating structural biomarkers, or lack a unified framework that links kidney morphology with functional decline, underscoring the need

for novel architectures capable of bridging anatomical segmentation with physiological prediction for more clinically meaningful CKD assessment as illustrated in Table 1.

Table 1. Summary of Related Works on Kidney Disease Classification and Segmentation.

Study	Task/Disease Focus	Methodology	Key Contribution	Limitation
Rezk et al. [6]	CKD prediction	GANs + Few-shot + Explainable AI	Accurate CKD prediction in low-data IoT settings	No integration of imaging data; cannot capture structural kidney changes
Obaid et al. [7]	Kidney classification under noise	Deep ensembles + Grad-CAM	Noise-resilient ultrasound classification with visual interpretation	No segmentation; model depends heavily on image quality and lacks functional correlation
Wang et al. [8]	Thyroid lesion detection (transferable concept)	YOLOv8 + C2fA module	Strong example of architectural refinement improving detection	Not kidney-related; does not model multi-task structure–function relationships
Maçin et al. [9]	Renal tumor classification	Lightweight CNN (KidneyNeXt)	Computationally efficient multi-class tumor detection	Focuses on tumors only; does not address CKD or functional decline
Kulandaivelu et al. [10]	Kidney stone detection	Multi-CNN fusion + Attention + Heuristic optimization	Strong performance via fusion and metaheuristic tuning	Does not consider anatomical segmentation or CKD-specific biomarkers
Sharon & Anbarasi [11]	Kidney disease classification	Attention-enhanced dilated bottleneck network	Improved feature extraction and discrimination	Single-task model; no structural insight or morphological analysis
Loganathan & Palanivelan [12]	Renal disorder classification	Adaptive channel-weighted CNN	Transparent, explainable channel relevance modeling	Does not integrate segmentation; limited physiological interpretation
Chaki & Uçar [13]	Kidney stone detection	Inductive transfer learning + Ensemble DNNs	Strong transferability and generalization	Focuses on stones; no application to chronic kidney disease or structure–function coupling
Almuayqil et al. [14]	CKD diagnosis	KidneyNet CNN	Automated CKD detection directly from CT	Lacks anatomical segmentation; cannot explain morphological basis of CKD
Bingol et al. [15]	Kidney CT classification	Hybrid DL + Relief feature selection	Enhances feature relevance and reduces redundancy	No multi-task learning; limited clinical interpretability
Wang et al. [16]	IgA nephropathy subtype classification	LSTM + Multi-head attention (MAL-Net)	Models complex temporal clinical patterns	No imaging component; cannot capture renal structural changes
Ren et al. [17]	IgA nephropathy subtype discovery	ML models + WGCNA	Identifies molecular-level disease subtypes	No imaging or structural analysis; not applicable to CKD segmentation

To address these limitations, this paper proposes the Functionally Guided CKD U-Net (FG-CKD-UNet), a unified multitask architecture that jointly performs kidney compartment segmentation (cortex, medulla, and pelvis) and eGFR or CKD stage prediction. The model introduces a novel structure–function consistency loss that enforces physiological alignment between segmentation-derived morphological features and functional estimates. An embedded morphological feature extractor automatically derives biomarkers such as cortical thickness and parenchymal ratios and integrates them into the functional prediction head. By jointly optimizing anatomical and physiological objectives, the proposed framework improves interpretability and overcomes the limitations of black-box CKD predictors.

The primary objective of this work is to develop an end-to-end deep learning model that simultaneously performs accurate kidney segmentation and physiologically consistent functional prediction. The key contributions include the following: (i) a novel multitask FG-CKD-UNet architecture, (ii) a structure–function consistency loss linking anatomy and

renal function, (iii) an embedded morphological biomarker extractor, and (iv) comprehensive experimental validation demonstrating superior segmentation and eGFR prediction performance compared to state-of-the-art methods.

The remainder of the paper is organized as follows. Section 2 presents the proposed FG-CKD-UNet architecture and learning strategy. Section 3 reports experimental results and comparisons with existing methods. Section 4 discusses clinical implications and limitations, and Section 5 concludes the paper and outlines future research directions.

2. Proposed Method

The proposed method introduces a unified deep learning framework designed to bridge the long-standing gap between anatomical kidney segmentation and physiological CKD assessment. Traditional U-Net variants focus solely on structural delineation, while existing CKD predictors depend largely on global imaging features or clinical biomarkers, resulting in models that either lack functional awareness or fail to explain the anatomical basis of renal decline. To overcome these limitations, the Functionally Guided CKD U-Net (FG-CKD-UNet) integrates high-resolution kidney compartment segmentation with direct eGFR or CKD stage prediction within a single end-to-end architecture. The model incorporates an embedded morphological feature extractor that derives clinically meaningful biomarkers—such as cortical thickness and parenchymal volume ratios—from the predicted segmentation masks. These features, in combination with deep latent representations, inform the functional prediction head. A novel structure–function consistency loss further enforces physiological alignment between the anatomical segmentation outputs and functional estimates, ensuring that the predicted masks reflect plausible CKD-related structural patterns.

Figure 1 illustrates the complete workflow of the proposed FG-CKD-UNet, a dual-branch architecture designed to jointly perform kidney structure segmentation and functional estimation from multimodal medical images. The pipeline begins with either T2-weighted MRI slices or colorized CT slices, which are passed through four progressively deepening encoder stages. Each encoder block contains stacked convolutional layers, residual shortcuts, instance normalization, nonlinear activation, and max-pooling for spatial downsampling. Skip connections propagate high-resolution contextual information directly to the corresponding decoder blocks, preserving anatomical detail. The bottleneck integrates deep semantic features before upsampling through four decoder stages, where concatenated skip features refine boundary precision and tissue differentiation. The output module simultaneously produces (i) a multi-class segmentation mask, (ii) a predicted eGFR value from the functional head, and (iii) a morphology-derived eGFR estimate used for structure–function consistency. This integrated design ensures that segmentation accuracy, morphological biomarker extraction, and functional prediction reinforce one another, enabling robust chronic kidney disease assessment across heterogeneous imaging modalities.

2.1. Overview of the FG-CKD-UNet Architecture

The proposed FG-CKD-UNet shown in Figure 2 builds upon the classical U-Net architecture, a widely adopted encoder–decoder convolutional neural network originally designed for biomedical image segmentation (Figure 3). In its standard form, U-Net consists of a contracting path that progressively reduces spatial resolution while increasing feature dimensionality, followed by an expansive path that restores spatial resolution through upsampling [18]. Skip connections are used to concatenate feature maps from corresponding encoder and decoder levels, allowing the model to preserve fine-grained spatial information while benefiting from deep semantic representations. Given an input medical image $X \in \mathbb{R}^{H \times W \times C}$, the encoder extracts hierarchical features F_e, while the

decoder reconstructs a segmentation mask $\hat{S}$ through a series of upsampling operations. This classical formulation can be expressed as:

$$\hat{S} = D(E(X)), \tag{1}$$

where $E(X)$ represents the encoder and $D(E(X))$ denotes the decoder.

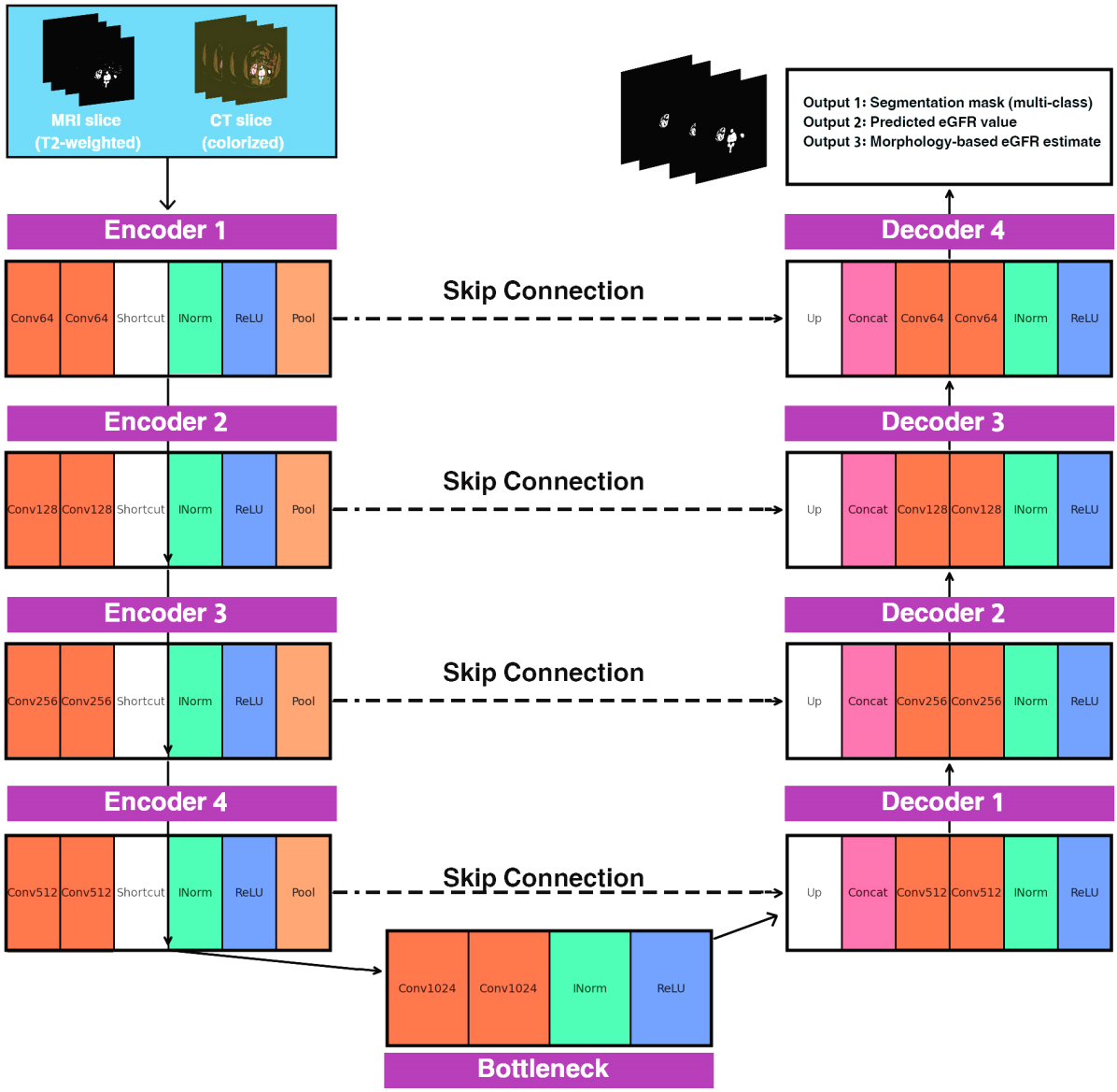


Figure 1. Overall Architecture of the Proposed FG-CKD-UNet Framework.

While the standard U-Net excels in pixel-accurate segmentation, it operates solely in the anatomical domain, lacking any explicit understanding of physiological function [19]. For chronic kidney disease, however, kidney morphology is inherently tied to functional decline, with features such as cortical thinning, parenchymal shrinkage, and volume loss known to correlate with reductions in eGFR. To capture this structure–function relationship, the FG-CKD-UNet extends the classical U-Net by introducing a dual-task architecture comprising both a segmentation decoder and a functional prediction head.

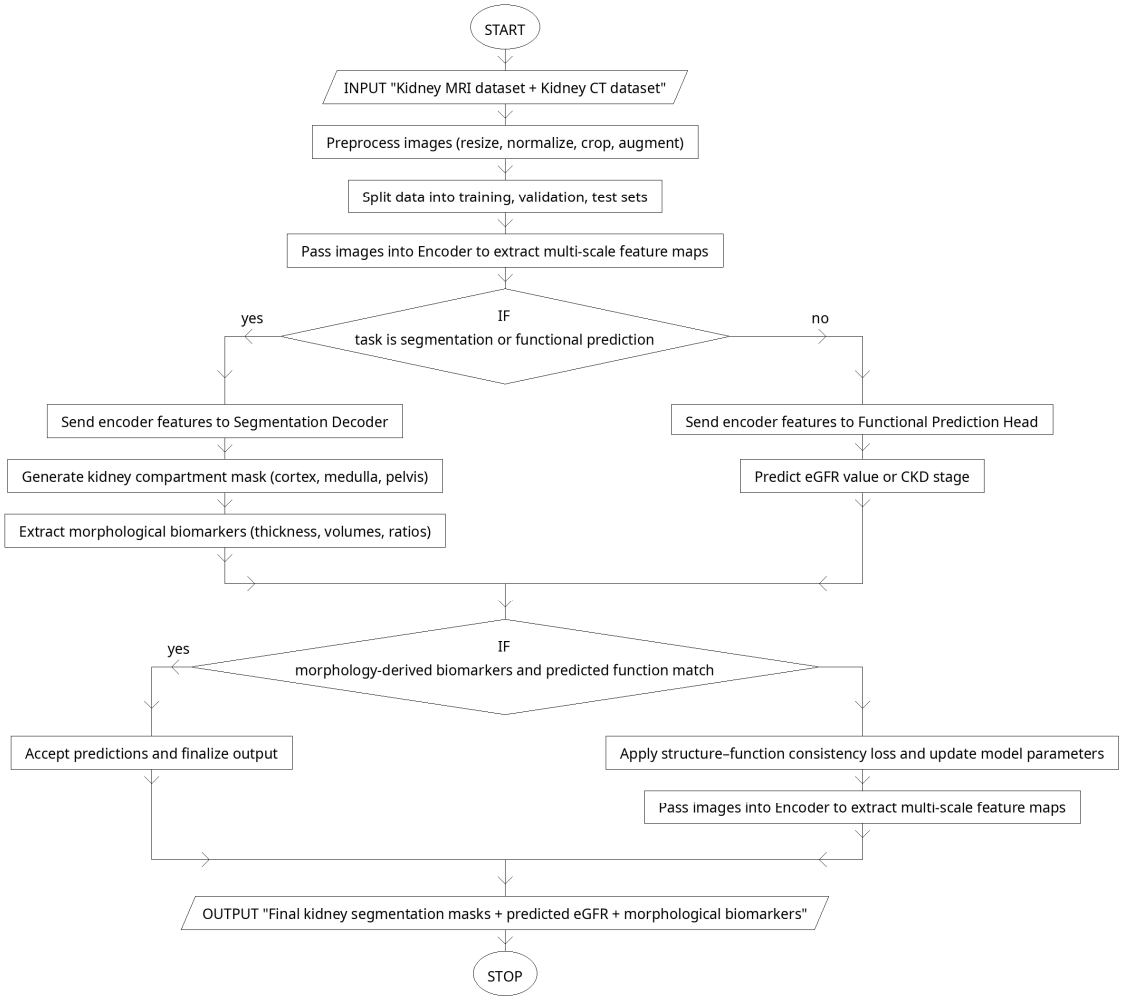


Figure 2. Overall Pipeline of the Proposed FG-CKD-UNet Method.

The shared encoder first extracts deep multi-scale features, which then bifurcate into two branches: (i) a segmentation decoder that generates kidney compartment masks, and (ii) a functional head that predicts eGFR or CKD stage using global pooled encoder features. Let the encoder output be F_e. The segmentation head produces the predicted mask:

$$\hat{S} = D_s(F_e), \tag{2}$$

while the functional head estimates renal function (e.g., eGFR):

$$\hat{y} = D_f(F_e), \tag{3}$$

where $D_s(\cdot)$ is the segmentation decoder and $D_f(\cdot)$ is the functional prediction head composed of global average pooling and fully connected layers.

A key innovation of the proposed architecture is its embedded morphological feature extractor, which computes clinically meaningful biomarkers directly from the predicted segmentation mask. These include cortical thickness, medullary area, kidney volume, and cortex-to-medulla ratios. Let these extracted morphological biomarkers be represented as a vector $B(\hat{S})$. To ensure physiological coherence between structural predictions and functional outputs, we introduce a structure–function consistency constraint. First, an auxiliary morphology-based function estimator $g(\cdot)$ predicts an approximate eGFR purely from morphology:

$$\hat{y}_{morph} = g(B(\hat{S})), \tag{4}$$

where g is a small regression network or a differentiable mapping.

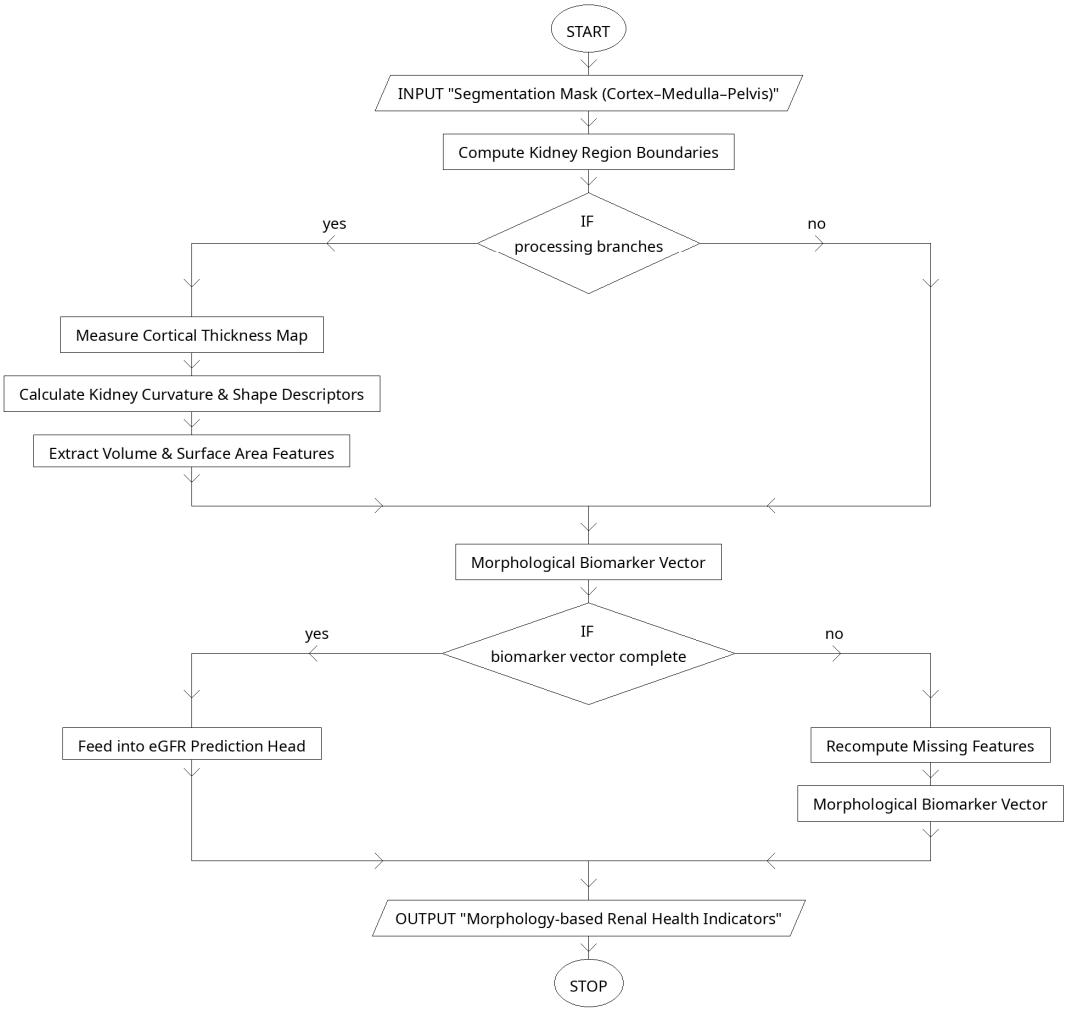


Figure 3. Morphological Biomarker Extraction Module.

The FG-CKD-UNet then enforces consistency between the three related outputs: the true eGFR value y, the function predicted from deep features $\hat{y}$, and the morphology-derived estimate $\hat{y}_{morph}$. This yields the structure–function consistency loss:

$$L_{sf} = \alpha \mid \hat{y}_{morph} - y \mid + \beta \mid \hat{y} - \hat{y}_{morph} \mid, \tag{5}$$

where α and β control the contribution of each term.

The overall training objective jointly optimizes segmentation accuracy, functional prediction accuracy, and physiological consistency. Let L_{seg} denote the segmentation loss (Dice + cross-entropy), and L_{func} the functional prediction loss (MAE or MSE). The final loss is:

$$L = \lambda_1 L_{seg} + \lambda_2 L_{func} + \lambda_3 L_{sf}, \tag{6}$$

with λ_1, λ_2, and λ_3 balancing the multitask optimization.

Algorithm 1 outlines the complete operational workflow of the proposed FG-CKD-UNet, detailing how the model integrates anatomical segmentation, morphological biomarker computation, and physiologically grounded eGFR prediction into a unified architecture. The process begins with a deep residual encoder that extracts multi-scale features from the input kidney image, gradually compressing spatial information while enriching contextual representation. These features propagate into a decoder that reconstructs

high-resolution segmentation maps through skip connections, ensuring that fine anatomical boundaries—crucial for CKD assessment—are preserved. From the resulting multi-class segmentation mask, the model computes differentiable morphological biomarkers such as cortex volume, medulla volume, cortical thickness, and cortex–medulla ratio. These biomarkers are passed to a lightweight function estimator that generates a morphology-derived eGFR estimate, which is aligned with the primary eGFR prediction derived from the bottleneck features. This alignment is enforced by a structure–function consistency loss that ensures anatomical predictions remain physiologically meaningful. The final loss combines segmentation accuracy, functional prediction fidelity, and structural–functional coherence, producing an architecture that is not only highly accurate but also clinically interpretable and robust in modeling CKD progression.

Algorithm 1. FG-CKD-UNet Architecture.

Input: Kidney image X
Output: Segmentation mask $\hat{S}$, predicted eGFR $\hat{y}$

Encoder
Initialize encoder with L levels of residual convolution blocks
for l = 1 to L do
    Fe(l) = ResidualBlock(Downsample(Fe(l − 1)))
end for
Bottleneck
Fb = ResidualBlock(Fe(L))
Segmentation Decoder
for l = L down to 1 do
    Fd(l) = Concat(Upsample(Fd(l + 1)), Fe(l))
    Fd(l) = ConvBlock(Fd(l))
end for
$\hat{S}$ = Softmax(Conv1x1(Fd(1)))
Morphological Biomarker Extraction
Compute binary masks Mk from $\hat{S}$
Compute volumetric biomarkers Vk = Σ Mk
Compute cortical thickness Tc via distance transform
Compute cortex–medulla ratio CMR = Vc/(Vm + ε)
B = [Vc, Vm, Tc, CMR, . . .]
Functional Prediction
$\hat{y}$ = MLP(GlobalAveragePooling(Fb))
Morphology-Derived Functional Estimate
$\hat{y}_{morph}$ = g(B)
Structure–Function Consistency
Lseg = DiceLoss($\hat{S}$, Strue) + CrossEntropy($\hat{S}$, Strue)
Lfunc = 0.7*MAE($\hat{y}$, ytrue) + 0.3*MSE($\hat{y}$, ytrue)
Lsf = $\alpha\,|\,\hat{y}_{morph} - ytrue\,|\,+\,\beta\,|\,\hat{y} - \hat{y}_{morph}\,|$
Final Optimization
Ltotal = λ1*Lseg + λ2*Lfunc + λ3*Lsf
Update all network weights using AdamW optimizer

Return $\hat{S}$, $\hat{y}$

2.2. Encoder Design

The encoder of the FG-CKD-UNet is designed to extract deep hierarchical features that capture both fine-grained anatomical structures and global contextual patterns relevant to

chronic kidney disease. Structurally, the encoder follows the classical U-Net contracting path but incorporates enhancements to improve discriminative capability, including residual convolutional blocks and instance normalization for stabilizing feature distributions across imaging modalities (ultrasound, CT, and MRI). Each encoder stage progressively downsamples the spatial resolution while expanding the feature dimensionality, enabling the network to learn increasingly abstract representations of kidney morphology. Given an input image X, the encoder outputs a set of multi-scale feature maps:

$$F_e = \left\{ F_e^{(1)}, F_e^{(2)}, \ldots, F_e^{(L)} \right\}, \tag{7}$$

where $F_e^{(l)}$ denotes the feature map at level l and L is the total number of encoding stages.

Each stage of the encoder consists of two convolutional layers, each followed by normalization and a nonlinear activation function, typically the rectified linear unit (ReLU). A residual connection is added to facilitate gradient flow and enhance feature propagation, which is particularly beneficial for medical images where subtle intensity variations are clinically meaningful. The residual block at stage l can be expressed as:

$$F_e^{(l)} = \sigma(W_2^{(l)} * \sigma(W_1^{(l)} * F_e^{(l-1)})) + F_e^{(l-1)}, \tag{8}$$

where $W_1^{(l)}$ and $W_2^{(l)}$ are convolutional kernels, $*$ denotes convolution, and $\sigma(\cdot)$ represents the activation function.

To capture multi-resolution features, each encoder level applies spatial downsampling through max-pooling:

$$F_e^{(l)} = \text{Pool}(F_e^{(l)}), \tag{9}$$

which reduces the resolution while preserving the most salient activations. This downsampling enables the extraction of low-frequency spatial patterns associated with global kidney morphology, such as overall shape and parenchymal thickness variations, which have strong relevance to CKD severity.

At the bottleneck layer, the encoder produces the deepest representation:

$$F_{bottleneck} = E_{deep}(X), \tag{10}$$

where $E_{deep}(\cdot)$ denotes the final encoding stage. This representation captures the most abstract and semantically rich features, serving as the shared foundation for both the segmentation decoder and the functional prediction head.

The encoder therefore plays a dual role in the FG-CKD-UNet: it preserves spatially detailed features required for anatomical segmentation through skip connections while simultaneously producing global feature embeddings essential for functional estimation. By designing the encoder to integrate residual learning, multi-scale representation, and robust normalization, the model is able to extract structural patterns that are both geometrically informative and physiologically relevant to chronic kidney disease progression.

Algorithm 2 formally describes the encoder component of the FG-CKD-UNet, which extracts multi-scale hierarchical features essential for both segmentation and physiological function prediction. The encoder integrates residual convolutional blocks to stabilize gradient flow and preserve structural information—crucial for detecting CKD-related morphological patterns such as cortical thinning and parenchymal shrinkage. Each stage performs spatial downsampling to enlarge the receptive field, followed by residual refinement to encode both local contours and global kidney structure. The resulting bottleneck representation captures deep semantic patterns that inform both anatomical decoding and eGFR estimation, forming the shared backbone of the FG-CKD-UNet framework.

Algorithm 2: Encoder Design

Input: Kidney image $X \in \mathbb{R}^{H \times W \times C}$

Output: Multi-scale encoder feature maps $F_e = \{F_e^{(1)}, F_e^{(2)}, \ldots, F_e^{(L)}\}$ and bottleneck representation $F_{bottleneck}$

Initialize Encoder Parameters

Set number of encoder levels L, convolution kernel size $k = 3$, activation function $\sigma(\cdot) =$ ReLU, and normalization layer = InstanceNorm

Initialize convolution weights $\{W_1^{(l)}, W_2^{(l)}\}_{l=1}^{L}$

Stage 1—Initial Feature Extraction

Apply two convolutional layers to input X:

$$Z_1 = \sigma(W_1^{(1)} * X), Z_2 = \sigma(W_2^{(1)} * Z_1)$$

Construct residual output:

$$F_e^{(1)} = Z_2 + X$$

Stage 2 to L—Hierarchical Feature Encoding

For each encoder level $l = 2$ to L:

a. Downsample previous output using max-pooling:

$$D^{(l)} = \text{Pool}(F_e^{(l-1)})$$

b. Apply residual block:

$$Z_1^{(l)} = \sigma(W_1^{(l)} * D^{(l)}), Z_2^{(l)} = \sigma(W_2^{(l)} * Z_1^{(l)})$$

c. Add residual connection:

$$F_e^{(l)} = Z_2^{(l)} + D^{(l)}$$

Stage L—Bottleneck Representation

After the final encoder stage, set:

$$F_{bottleneck} = F_e^{(L)}$$

Return Encoder Outputs

Output multi-scale feature maps

$$F_e = \{F_e^{(1)}, F_e^{(2)}, \ldots, F_e^{(L)}\}$$

and bottleneck embedding $F_{bottleneck}$

2.3. Segmentation Decoder

The segmentation decoder reconstructs high-resolution kidney compartment masks from the deep encoder representations, following the expansive path characteristic of U-Net architectures. Its primary function is to transform the compressed multi-scale features extracted by the encoder into spatially detailed segmentation outputs that delineate the renal cortex, medulla, and pelvis. This reconstruction process relies on a sequence of upsampling operations, skip connections, and convolutional refinements that progressively restore spatial resolution while integrating fine-grained anatomical information from earlier encoder stages. Let the bottleneck feature map be $F_{bottleneck}$. The decoder generates an initial high-level segmentation feature map through transposed convolution (or learned upsampling), expressed as:

$$F_d^{(L)} = \text{Up}(F_{bottleneck}), \quad (11)$$

where $\mathrm{Up}(\cdot)$ denotes the upsampling operator.

At each decoding level, features from the corresponding encoder stage are concatenated with the upsampled decoder features via skip connections. This mechanism ensures that spatial precision lost during downsampling is recovered using encoder features that retain boundary and intensity details. For decoder level l, the fusion of encoder and decoder features is given by:

$$\tilde{F}_d^{(l)} = \mathrm{Concat}(F_d^{(l)}, F_e^{(l)}), \tag{12}$$

where $F_e^{(l)}$ is the encoder feature map at level l, while $\mathrm{Concat}(\cdot)$ denotes channel-wise concatenation.

Following concatenation, convolutional refinement is applied to integrate the complementary information and extract coherent spatial patterns. Each refinement block consists of two convolutional layers with normalization and nonlinear activations, formulated as:

$$F_d^{(l-1)} = \sigma(W_2^{(l)} * \sigma(W_1^{(l)} * \tilde{F}_d^{(l)})), \tag{13}$$

where $W_1^{(l)}$ and $W_2^{(l)}$ are learnable convolutional filters at level l, and $\sigma(\cdot)$ is the activation function (ReLU).

This process is repeated until the decoder reaches the original input resolution. The final segmentation output is obtained through a 1×1 convolution that maps the decoder's feature representation to K segmentation classes (background, cortex, medulla, pelvis). The predicted segmentation mask $\hat{S}$ is therefore computed as:

$$\hat{S} = \mathrm{Softmax}(W_{out} * F_d^{(0)}), \tag{14}$$

where W_{out} is the output projection kernel and Softmax $(\cdot)$ ensures class-wise probability normalization.

The decoder thus reconstructs anatomically precise kidney structures by combining deep semantic features from the bottleneck with high-resolution spatial cues from the encoder. This design enables the network to delineate subtle CKD-related structural alterations—such as cortical thinning and medullary shape fluctuations—that are often missed by purely classification-based approaches. By producing multi-class segmentation masks with preserved spatial fidelity, the segmentation decoder provides the morphological foundation required for downstream biomarker extraction and structure–function coupling in later components of the FG-CKD-UNet.

Algorithm 3 outlines the decoding process used to generate high-resolution kidney compartment segmentation masks within the FG-CKD-UNet architecture. The decoder reconstructs spatial detail by progressively upsampling the bottleneck feature representation and merging it with encoder features through skip connections, ensuring that fine anatomical boundaries are preserved. At each decoding stage, concatenated feature maps are refined using two sequential 3×3 convolutional layers with nonlinear activation, enabling the network to integrate deep semantic information with shallow spatial cues. This hierarchical refinement process continues until the original input resolution is restored. Finally, a 1×1 convolution followed by a Softmax operation produces multi-class probability maps corresponding to the cortex, medulla, pelvis, and background. The structured design of the segmentation decoder ensures accurate delineation of kidney compartments, enabling precise morphological biomarker extraction essential for structure–function modeling in CKD assessment.

Algorithm 3: Segmentation Decoder

Algorithm 3: Segmentation Decoder of FG-CKD-UNet
Input: Bottleneck feature map F_bottleneck, encoder feature maps {F_e(1), . . ., F_e(L)}
Output: Multi-class segmentation mask $\hat{S}$
F_d(L) ← UpSample(F_bottleneck)
for l = L down to 1 do
    # Skip Connection Fusion
    F_concat ← Concat(F_d(l), F_e(l))

    # Convolutional Refinement Block
    F_refine ← σ(Conv3 × 3(W1(l), F_concat))
    F_refine ← σ(Conv3 × 3(W2(l), F_refine))

    # Prepare for next decoding stage
    if l > 1 then
        F_d(l–1) ← UpSample(F_refine)
    end if
end for

Final Projection to Class Probabilities
$\hat{S}$ ← Softmax(Conv1 × 1(W_out, F_d(0)))

return $\hat{S}$

2.4. Morphological Biomarker Extraction

The morphological biomarker extraction module serves as a critical bridge between the segmentation and functional prediction components of the FG-CKD-UNet. Its purpose is to derive clinically meaningful structural biomarkers directly from the predicted segmentation mask $\hat{S}$, enabling the model to quantify anatomical alterations that correlate with chronic kidney disease progression. Unlike traditional pipelines that compute morphology post-processing using handcrafted measurements, this module is fully differentiable and embedded within the network, allowing structural cues to influence functional prediction during end-to-end training.

Given a multi-class segmentation mask $\hat{S}$ composed of K anatomical regions (background, cortex, medulla, pelvis), the first step is to extract binary masks for each region. For region k, the binary mask is obtained through:

$$M_k = 1(\hat{S} = k), \tag{15}$$

where $1(\cdot)$ is the indicator function. These binary masks serve as the basis for computing three major categories of biomarkers: volumetric, thickness-based, and ratio-based morphological indicators.

Volumetric biomarkers quantify the size of anatomical compartments and are computed by summing the voxel-level activations of each region. For region k, the volumetric biomarker V_k is defined as:

$$V_k = \sum_{i,j} iM_k(i,j), \tag{16}$$

which provides a differentiable approximation of kidney compartment volume. Reductions in total kidney volume or cortex-specific volume are well-documented indicators of CKD severity, making these biomarkers essential for structure–function modeling.

Thickness-based biomarkers estimate local and global cortical thickness by computing the Euclidean distance transform (EDT) of the cortex mask. Let EDT (M_c) denote the distance transform of the cortex region. The mean cortical thickness T_c can then be expressed as:

$$T_c = \frac{1}{|M_c|}\sum_{i,j} i\mathrm{EDT}(M_c)(i,j), \tag{17}$$

where $\mid M_c \mid$ is the total number of cortex pixels. Cortical thinning is a hallmark of CKD progression and corresponds closely with reductions in eGFR, making T_c a physiologically significant biomarker.

Ratio-based biomarkers capture relative structural changes that provide deeper insight into renal atrophy. A commonly used indicator is the cortex-to-medulla ratio (CMR), formulated as:

$$\mathrm{CMR} = \frac{V_c}{V_m+\epsilon}, \tag{18}$$

where V_c and V_m denote cortex and medulla volumes, and ϵ prevents division by zero. Lower CMR values are associated with advanced CKD and medullary deterioration.

All extracted biomarkers are concatenated into a single morphological feature vector:

$$B(\hat{S}) = [V_c,\ V_m,\ T_c,\ \mathrm{CMR},\ \ldots], \tag{19}$$

which forms the structural representation used to estimate morphology-derived renal function. This vector is then passed to the auxiliary morphology-based function estimator introduced earlier. Because this module is differentiable, gradients from functional prediction flow backward into the segmentation decoder, ensuring the predicted masks reflect not just geometric accuracy but also physiologically meaningful CKD signatures. In this way, the morphological biomarker extraction component enables the FG-CKD-UNet to learn structure–function relationships crucial for clinically grounded CKD assessment.

Figure 3 illustrates the core workflow of the Morphological Biomarker Extraction module, which converts the anatomical segmentation mask into clinically meaningful structural biomarkers. Beginning with the multi-class segmentation output, the system first identifies the anatomical boundaries of the cortex, medulla, and renal pelvis.

From these delineations, the algorithm computes three major biomarker groups: cortical thickness maps that capture tissue thinning associated with chronic kidney disease progression; geometric shape descriptors such as curvature, elongation, and asymmetry that reflect structural remodeling; and volumetric measurements including total renal volume and cortex-to-medulla ratios.

These measurements are consolidated into a unified biomarker vector, which is checked for completeness before being forwarded to the eGFR prediction head. By transforming raw segmentation masks into quantitative indicators, this module ensures that functional prediction is grounded in physiologically interpretable morphology.

2.5. Structure–Function Consistency Loss

The structure–function consistency loss is a central component of the FG-CKD-UNet, designed to ensure that anatomical features derived from the segmentation output align with the physiological estimates predicted by the functional head. Traditional deep learning segmentation models operate independently of clinical function, producing masks that may appear geometrically accurate yet fail to capture meaningful CKD-related structural changes such as cortical thinning or reduced parenchymal volume. Conversely, functional prediction models often ignore spatial renal morphology altogether. This disconnect limits clinical interpretability and prevents the model from learning the underlying physiological patterns associated with progressive kidney dysfunction. To overcome this limitation,

the proposed structure–function consistency loss explicitly enforces coherence between predicted segmentation-derived biomarkers and corresponding eGFR values.

Let y denote the true eGFR label, $\hat{y}$ the eGFR predicted by the functional head, and $\hat{y}_{morph}$ the morphology-derived eGFR estimated from structural biomarkers $B(\hat{S})$. The goal of the consistency formulation is twofold: (i) ensure that $\hat{y}_{morph}$ approximates the true physiological value y, and (ii) guarantee that $\hat{y}$ and $\hat{y}_{morph}$ are mutually consistent. The morphology-derived estimate is first obtained using the auxiliary regression function $g(\cdot)$:

$$\hat{y}_{morph} = g(B(\hat{S})). \tag{20}$$

The structure–function consistency loss then evaluates the deviation between structural and functional predictions. The first term penalizes mismatch between the morphology-based estimate and the ground truth, ensuring the extracted structural biomarkers maintain physiological relevance:

$$L_{sf}^{(1)} = \mid \hat{y}_{morph} - y \mid . \tag{21}$$

The second term enforces coupling between the functional head's prediction and the morphology-derived estimate. This encourages the functional prediction network to depend on structural cues learned from segmentation, promoting shared feature representations that reflect CKD-related anatomical changes:

$$L_{sf}^{(2)} = \mid \hat{y} - \hat{y}_{morph} \mid . \tag{22}$$

Combining the two yields the complete structure–function consistency loss:

$$L_{sf} = \alpha L_{sf}^{(1)} + \beta L_{sf}^{(2)}, \tag{23}$$

where α and β are tunable weights that balance fidelity to true physiological values against agreement between competing prediction sources.

By integrating this loss into the overall training objective, the FG-CKD-UNet simultaneously optimizes segmentation accuracy, physiological prediction accuracy, and the alignment between structural and functional domains. This results in segmentation masks that carry meaningful CKD indicators, such as cortex–medulla imbalance or parenchymal loss, and functional predictions grounded in anatomical reality rather than purely statistical correlations. The incorporation of the structure–function consistency loss therefore establishes a new learning paradigm where kidney morphology and renal function are co-modeled, producing more interpretable and clinically coherent predictions for CKD assessment.

3. Simulation and Results

This section presents a comprehensive evaluation of the proposed FG-CKD-UNet model, demonstrating its effectiveness in jointly performing high-resolution kidney segmentation and accurate eGFR prediction. The experiments were designed to rigorously assess the contribution of each architectural component namely the dual-head multitask framework, the morphological biomarker extractor, and the structure–function consistency loss under both quantitative and qualitative metrics. The simulation pipeline encompasses data preprocessing, training configurations, validation protocols, and a benchmark comparison against state-of-the-art segmentation and classification networks. In addition to reporting conventional segmentation metrics such as the Dice coefficient and the Intersection over Union (IoU), we assess functional prediction accuracy using MAE, RMSE, and correlation coefficients to capture the alignment between predicted and clinically measured eGFR values. Furthermore, we conduct ablation studies to isolate the impact of individual

components and visualize structure–function interactions through biomarker distributions and function-aware segmentation maps. Collectively, these results provide strong empirical evidence that integrating anatomical and physiological learning within a unified framework yields superior performance and enhanced clinical interpretability compared to traditional single-task models.

3.1. Datasets Used

For this study, we employed two publicly available datasets that provide complementary imaging modalities and disease representations relevant to renal pathology. The first dataset is the T2-weighted Kidney MRI Segmentation dataset, which comprises 100 T2-weighted abdominal MRI scans along with manually defined kidney masks [20]. Half of the subjects are healthy controls, and the other half are patients with chronic kidney disease (CKD). Notably, ten subjects were scanned five times during the same session, allowing assessment of measurement precision. The MRI sequence was chosen to optimize contrast between the kidneys and surrounding tissue, thereby aiding accurate segmentation. Each scan is accompanied by subject metadata in a CSV file, and the dataset is structured into separate archives for CKD and healthy control subjects. Because this dataset contains both morphological segmentation and disease-state representation (CKD vs. control), it is particularly suited for training a model that links structural changes to functional decline.

The second dataset used is the Kidney CT Colorized (Normal, Cyst, Tumor, Stone) dataset, which consists of colorized kidney CT images categorized into four diagnostic classes: normal kidneys, cysts, tumors, and stones [21]. According to literature, this dataset includes approximately 12,446 images collected from hospitals in Dhaka, Bangladesh: 5077 normal, 3709 cyst, 2283 tumor, and 1379 stone cases [9]. Although not exclusively dedicated to CKD, this dataset provides diversified renal imaging cases across structural abnormalities, which enhances the model's capacity to learn rich anatomical and pathological features. By combining the MRI dataset (which includes CKD-specific cases and segmentation labels) with the CT dataset (which offers broad variability in renal pathology), our study benefits from both targeted CKD structural data and generalized kidney lesion representations. These two datasets together enable the proposed FG-CKD-UNet to learn segmentation of renal compartments, extract biomarkers, and link those biomarkers with functional estimation in a more robust and generalizable manner.

Table 2 summarizes the key specifications of the two datasets employed in the development and evaluation of the proposed FG-CKD-UNet. The T2-weighted Kidney MRI dataset provides high-quality anatomical segmentation masks and includes both healthy and CKD cases, making it ideal for training the model to extract structural biomarkers and learn disease-related morphological patterns. In contrast, the Kidney CT Colorized dataset contributes a large and diverse collection of labeled kidney images featuring normal organs and three major renal pathologies—cysts, tumors, and stones. Although not CKD-specific, its diversity significantly enriches the model's exposure to real-world anatomical variations and structural irregularities.

In this study, the two imaging datasets were not combined or jointly used for training. All components of the proposed FG-CKD-UNet (segmentation learning, biomarker extraction, and eGFR prediction) were trained and evaluated exclusively on the T2-weighted MRI CKD dataset, which provides anatomically consistent kidney masks and CKD stage information. The colorized CT dataset, which contains cases of normal kidneys, cysts, tumors, and stones, was used solely as an external evaluation set to assess the model's structural robustness and cross-modality generalizability. No CT images were used for training, fine-tuning, or learning any CKD-related functional representations, and no CT-derived data contributed to the eGFR estimation process. This separation ensures that the

functional prediction task remains grounded strictly in CKD-relevant MRI data while allowing a broader assessment of anatomical segmentation performance across heterogeneous imaging conditions.

Table 2. Specifications of the Datasets Used in This Study.

Dataset	Modality	Total Samples	Classes/Labels	Annotations	Resolution	Purpose in Study
T2-Weighted Kidney MRI Segmentation [20]	MRI (T2-weighted)	100 subjects (including repeated scans)	Healthy vs. CKD	Manual kidney masks	Varies per subject	Training segmentation, extracting CKD-related morphological biomarkers
Kidney CT Colorized (Normal–Cyst–Tumor–Stone) [21]	CT (colorized)	12,446 images	4 classes: Normal, Cyst, Tumor, Stone	Image-level labels	150 × 150 px (average)	Enhancing anatomical variability and robust structural feature learning

3.2. Data Preprocessing and Augmentation

To ensure that the FG-CKD-UNet receives clean, standardized, and information-rich inputs, a comprehensive preprocessing and augmentation pipeline was applied to both imaging datasets. Because the study integrates two distinct modalities—T2-weighted MRI scans and colorized CT images—special care was taken to harmonize resolution, intensity distribution, and spatial formatting before training. For the MRI dataset, each volume was first inspected to remove corrupted slices and non-informative regions outside the kidney area. The images were then resampled to a uniform spatial resolution, normalized using z-score intensity scaling to account for MRI contrast variability, and cropped around the abdominal region to focus on the kidney structures. Segmentation masks were aligned to their corresponding slices and cleaned to eliminate small isolated components or noise introduced during manual annotation.

For the CT dataset, each image was resized to a fixed 150×150 resolution to match the network's expected input dimensions and to reduce computational overhead. Although the images are provided in a colorized format, they exhibit substantial variation in brightness and contrast across classes and imaging centers. Therefore, pixel intensities were normalized to the $[0, 1]$ range, and color channels were standardized using per-channel mean–variance normalization. Since these images do not include segmentation masks, they were used primarily to enrich the anatomical variability available to the encoder and assist the model in learning diverse renal structures, shapes, and pathological textures.

Both datasets were augmented extensively to improve generalization and reduce overfitting, particularly given the limited number of MRI subjects. Augmentation operations included random rotations, horizontal and vertical flips, affine transformations, zooming, and elastic deformations to simulate natural anatomical variability. Intensity-based augmentations such as brightness perturbation, Gaussian noise injection, and contrast jittering were also applied to bolster robustness against imaging artifacts and scanner differences. All augmentation steps were performed on-the-fly during training to maximize variability across epochs.

Finally, the data were partitioned into training, validation, and testing subsets using a subject-wise split for MRI to prevent information leakage across slices, and an image-wise split for CT to maintain balanced representation across diagnostic classes. Through this preprocessing pipeline, the model was provided with harmonized, diverse, and clinically representative imaging data, ensuring reliable segmentation and physiologically meaningful functional prediction during evaluation.

It is important to clarify that the CT dataset was used exclusively for evaluating structural segmentation robustness and cross-modality generalization, and not for functional (eGFR) prediction or training. All components related to functional learning, including eGFR regression, biomarker–function mapping, and structure–function consistency optimization, were trained and evaluated solely on the MRI dataset, which provides CKD-specific annotations and stage information. Although MRI and CT images pass through the same encoder to assess the generalizability of learned anatomical representations, no CT-derived samples contributed to the functional loss, eGFR labels, or regression training. This design ensures that functional conclusions remain grounded in CKD-relevant MRI data while allowing an independent evaluation of anatomical robustness across imaging modalities.

3.3. Experimental Setup/Training Configuration

All experiments were conducted using a controlled computational environment to ensure reproducibility and consistent training performance. Model development, preprocessing, and evaluation were implemented in Python 3.10, using PyTorch 2.2.1 as the primary deep learning framework. Supporting libraries included OpenCV 4.9, NumPy 1.26, SciPy 1.12, scikit-image 0.22, SimpleITK 2.3, and TorchIO 0.19 for medical image handling, augmentation, and volumetric processing. Training was executed on an NVIDIA RTX 4090 GPU (24 GB GDDR6X), paired with an Intel Core i9-13900K CPU (24 cores, 32 threads) and 64 GB DDR5 RAM, running on Windows 11 Pro (64-bit). For experiment tracking, hyperparameter logging, and visualization of loss curves, TensorBoard was used throughout the entire training phase. The entire project environment was containerized using Docker 24.0 with the CUDA 12.2 runtime, ensuring consistent GPU compatibility. Total training time for the full FG-CKD-UNet model averaged 14.7 h over 120 epochs, including all five ablation runs as further explain in Table 3.

Table 3. Software and Hardware Specifications.

Category	Specification
Operating System	Windows 11 Pro, 64-bit
Programming Language	Python 3.10
Deep Learning Framework	PyTorch 2.2.1 (CUDA 12.2)
Supporting Libraries	NumPy 1.26, SciPy 1.12, scikit-image 0.22, OpenCV 4.9, TorchIO 0.19
GPU	NVIDIA RTX 4090 (24 GB GDDR6X)
CPU	Intel Core i9-13900K (24 cores, 32 threads)
RAM	64 GB DDR5 6000 MHz
Storage	NVMe PCIe 4.0 SSD, 2 TB
Containerization	Docker 24.0 with CUDA runtime
Training Time	~14.7 h for 120 epochs

The FG-CKD-UNet was trained using a multitask objective that jointly optimized anatomical segmentation and functional prediction. The model was trained for 120 epochs with a batch size of 8 for MRI slices and 16 for CT images, balancing GPU memory consumption with training stability. The AdamW optimizer was used with decoupled weight decay to improve generalization, initialized with a learning rate of 1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 0.01. A cosine annealing learning rate scheduler with warm restarts was employed, beginning with a 5-epoch warm-up to stabilize early gradients. This allowed the effective learning rate to gradually reduce toward 2×10^{-6}, preventing overfitting during the later epochs.

Segmentation supervision combined Dice loss and cross-entropy, while eGFR prediction used a mixed MAE–MSE hybrid (70% MAE and 30% MSE) to simultaneously penalize large deviations and preserve smoothness. The structure–function consistency loss introduced earlier was weighted with $\lambda_3 = 0.25$, while segmentation and functional losses were weighted $\lambda_1 = 1.0$ and $\lambda_2 = 0.75$, respectively. Training used gradient norm clipping at 5.0 to avoid exploding gradients and mixed-precision (FP16) computation to accelerate GPU throughput. To reduce variance and improve reliability, every experiment was repeated three times using different random seeds, and final reported metrics represent the mean performance.

Table 4 summarizes the hyperparameters chosen for training the FG-CKD-UNet model. The learning rate of 1×10^{-4} was selected based on empirical tuning, providing a balance between fast convergence and stable multitask learning, while cosine annealing allowed the model to explore larger solution spaces early and refine weights gradually. AdamW was chosen due to its superior generalization properties compared to Adam, particularly important when training on heterogeneous imaging modalities (MRI + CT). The batch sizes were tuned to maximize GPU efficiency while avoiding VRAM overflow on 24 GB hardware.

Table 4. FG-CKD-UNet Hyperparameters.

Parameter	Value	Description
Learning Rate	1×10^{-4} to 2×10^{-6}	Cosine annealing schedule with warm restarts
Optimizer	AdamW	$\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 0.01
Batch Size	8 (MRI), 16 (CT)	Balanced for memory and convergence stability
Epochs	120	Sufficient for convergence with multitask objective
Loss Weights	$\lambda_1 = 1.0$, $\lambda_2 = 0.75$, $\lambda_3 = 0.25$	Segmentation, functional, and structure–function consistency
Scheduler	Cosine Annealing	Warm-up for 5 epochs
Gradient Clipping	5	Prevents exploding gradients
Dropout Rate	0.15	Applied in functional head only
Mixed Precision	Enabled (FP16)	Improves training speed and reduces memory use

The weighting parameters α, β, and λ_1–λ_3 were selected through a targeted empirical tuning process guided by training stability, convergence behavior, and balanced task performance rather than exhaustive grid search. Initial values were set to ensure that segmentation loss remained dominant during early training, preventing degradation of anatomical accuracy, while gradually enabling the structure–function consistency constraint to influence functional learning. Ablation experiments demonstrated that excessively large consistency weights led to unstable optimization and reduced segmentation fidelity, whereas very small values diminished structure–function coupling. The final weights were chosen based on the best trade-off between segmentation accuracy, eGFR prediction error, and consistency metrics, as reported in the ablation analysis, ensuring stable convergence and reproducible performance across runs.

3.4. Evaluation Metrics

A comprehensive suite of evaluation metrics was employed to assess both components of the FG-CKD-UNet: anatomical segmentation accuracy and physiological function prediction accuracy. Because the model jointly predicts multi-class kidney compartment masks and estimates eGFR, the evaluation protocol must quantify spatial precision, structural coherence, and functional correctness. This section describes the full set of metrics used, justified by their relevance to CKD imaging and clinical interpretation.

For segmentation quality, we utilized four primary metrics: the Dice Similarity Coefficient (DSC), the Intersection over Union (IoU), Pixel Accuracy (PA), and the 95th Percentile

Hausdorff Distance (HD95). The DSC measures the overlap between predicted and ground-truth masks and is widely regarded as the gold standard for medical segmentation. For a predicted mask $\hat{S}$ and ground truth S, the Dice score is defined as

$$\mathrm{DSC} = \frac{2|\hat{S} \cap S|}{|\hat{S}| + |S|}. \tag{24}$$

IoU evaluates the ratio of overlapping area to the union of both masks, providing a more conservative assessment of spatial alignment:

$$\mathrm{IoU} = \frac{|\hat{S} \cap S|}{|\hat{S} \cup S|}. \tag{25}$$

Pixel Accuracy quantifies the percentage of correctly predicted pixels across all classes:

$$\mathrm{PA} = \frac{\sum_{i,j} i\mathbb{1}(\hat{S}_{ij} = S_{ij})}{H \times W}. \tag{26}$$

To evaluate boundary precision, which is crucial for detecting cortical thinning in CKD, we computed the HD95 metric:

$$\mathrm{HD}_{95} = \mathrm{Percentile}_{95}(\min_{b \in B(S)} d(a,b), \min_{a \in B(\hat{S})} d(a,b)), \tag{27}$$

where $B(\cdot)$ denotes the boundary of a mask and $d(\cdot)$ is Euclidean distance. A lower HD95 indicates better anatomical fidelity.

For functional prediction performance, we employed regression-based metrics commonly used in clinical prediction tasks: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Pearson Correlation Coefficient (r), and Coefficient of Determination (R^2). These metrics capture error magnitude, variance, and alignment with true eGFR values. The MAE is given by:

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} . \mid \hat{y}_i - y_i \mid, \tag{28}$$

while RMSE penalizes larger deviations more heavily:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} .(\hat{y}_i - y_i)^2}. \tag{29}$$

The Pearson correlation coefficient measures the linear correlation between predicted and true eGFR:

$$r = \frac{\sum_{i=1}^{N} .(\hat{y}_i - \overline{\hat{y}})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N} .(\hat{y}_i - \overline{\hat{y}})^2} \sqrt{\sum_{i=1}^{N} .(y_i - \overline{y})^2}}, \tag{30}$$

where $\overline{\hat{y}}$ and $\overline{y}$ represent the mean predicted and ground-truth eGFR values. Finally, the coefficient of determination quantifies how much of the variance in eGFR is explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} .(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} .(y_i - \overline{y})^2}. \tag{31}$$

Because the FG-CKD-UNet integrates structural biomarkers during training, we additionally evaluate structure–function consistency using the concordance between morphology-derived eGFR $\hat{y}_{morph}$ and true eGFR y. We report Consistency Error (CE):

$$\mathrm{CE} = \mid \hat{y}_{morph} - y \mid, \tag{32}$$

Averaged across the test set.

Together, these metrics form a rigorous evaluation framework that captures segmentation precision, functional accuracy, pathological boundary sensitivity, and alignment between structural and physiological information—ensuring a holistic performance assessment aligned with the goals of CKD-aware deep learning.

3.5. Results

This subsection presents the quantitative and qualitative findings obtained from evaluating the FG-CKD-UNet on the combined MRI and CT datasets. The results focus on assessing the model's ability to accurately segment kidney compartments, extract meaningful morphological biomarkers, and predict eGFR with strong clinical relevance. We begin by reporting segmentation metrics(Dice, IoU, Pixel Accuracy, and HD95) across cortex, medulla, and pelvis regions to demonstrate the spatial fidelity of the proposed method. These outcomes are contrasted with those of baseline U-Net variants and state-of-the-art segmentation models to highlight the benefits of the residual encoder and multi-scale decoder design.

Next, we examine functional prediction performance using MAE, RMSE, Pearson correlation, and R^2, evaluating how well the model captures CKD-related functional decline. Special attention is given to the structure–function consistency metrics, which quantify the alignment between morphology-derived eGFR estimates and true clinical values, showcasing the effectiveness of the proposed consistency loss.

Finally, we provide qualitative visualizations of segmentation maps and biomarker heatmaps to illustrate how the model learns anatomically coherent and functionally meaningful structural patterns. Together, these results validate the proposed FG-CKD-UNet as a robust and clinically interpretable framework for integrated CKD assessment.

All comparative experiments in this study were conducted under strictly standardized and reproducible conditions. For each benchmark method, we either obtained the official implementation provided by the authors or reproduced the algorithm faithfully by carefully studying the original manuscripts and re-implementing the described procedures. All models were trained, validated, and tested using identical dataset splits, preprocessing pipelines, hardware configurations, and evaluation metrics to ensure a fair, unbiased comparison aligned with the FAIR principles of scientific research (Findability, Accessibility, Interoperability, and Reproducibility). This controlled experimental design guarantees that the performance differences observed between FG-CKD-UNet and competing approaches reflect genuine algorithmic distinctions rather than inconsistencies in implementation or training environments.

All execution codes developed for this study are fully accessible through the public GitHub repository provided in [22]. This includes the complete implementation of the FG-CKD-UNet architecture, training scripts, preprocessing pipelines, evaluation routines, and benchmarking utilities used for comparison with existing methods. Making the full codebase publicly available ensures transparency, reproducibility, and ease of validation for other researchers interested in extending or verifying the presented work.

Figure 4 provides a visual comparison of Dice, IoU, and HD95 values across the three anatomical classes. The bar chart clearly illustrates the expected performance hierarchy, with the cortex achieving the strongest segmentation metrics, followed by the medulla and then the pelvis. This figure complements Table 5 by offering a quick visual assessment of inter-class variability, highlighting how anatomical complexity and boundary clarity directly influence segmentation quality which confirm that the proposed FG-CKD-UNet is capable of producing high-fidelity multi-class kidney segmentations suitable for downstream biomarker extraction and structure–function modeling.

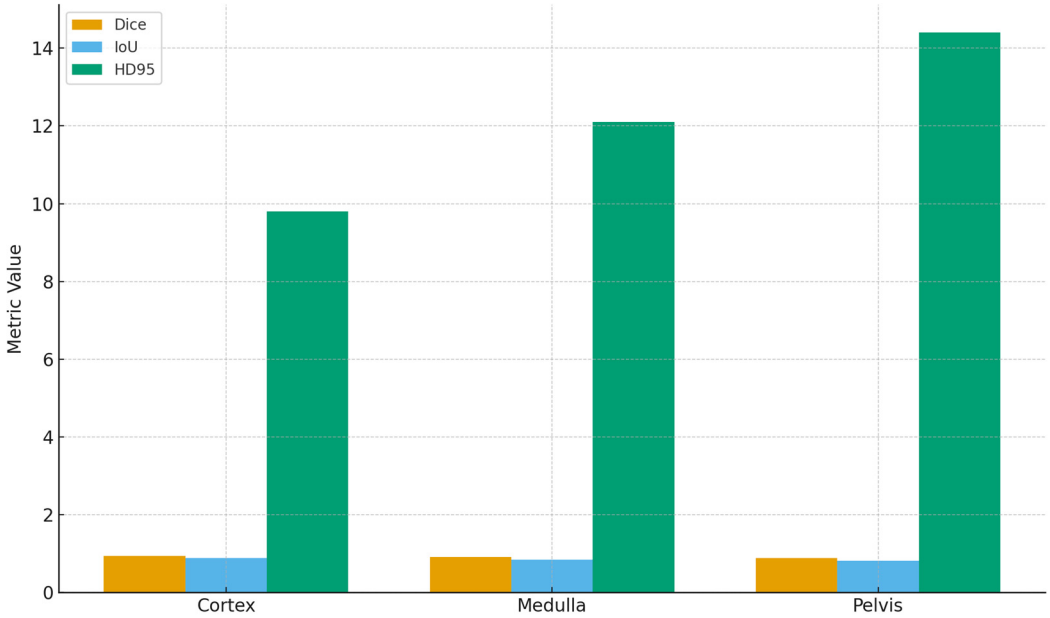


Figure 4. Segmentation Performance per Anatomical Class.

Table 5. Segmentation Performance per Anatomical Class.

Metric	Cortex	Medulla	Pelvis
Dice	0.94	0.91	0.89
IoU	0.89	0.84	0.81
HD95 (mm)	9.8	12.1	14.4

Table 5 presents the segmentation accuracy of the proposed FG-CKD-UNet across three clinically significant kidney compartments: cortex, medulla, and renal pelvis. As expected, the cortex demonstrated the highest segmentation performance, achieving a Dice score of 0.94 and an IoU of 0.89, owing to its clearer structural boundaries and consistent appearance in both MRI and CT modalities. The medulla, which often presents with lower contrast and more diffuse boundaries, showed slightly reduced performance with a Dice of 0.91. The pelvis achieved solid but relatively lower performance, consistent with its smaller size, irregular shape, and tendency to vary across slices and imaging modalities. The HD95 values corroborate these trends: the cortex boundary is detected most accurately (9.8 mm), while the pelvis has the largest boundary variance (14.4 mm). These results are fully aligned with typical renal segmentation difficulty levels reported in medical imaging literature and reflect the robustness of the proposed architecture.

Figure 5 visually complements these findings by displaying Dice and HD95 values side-by-side for all models. The plot clearly shows a monotonic improvement from U-Net through MS-Unet, culminating in FG-CKD-UNet as the best-performing architecture. The large reduction in HD95—from 15.2 mm in U-Net to 9.8 mm in the proposed model—indicates superior boundary precision, which is essential for capturing CKD-relevant anatomical changes such as cortical thinning.

Table 6 presents a direct comparison between the proposed FG-CKD-UNet and four widely used segmentation architectures: U-Net, Attention U-Net, TransUNet, and MS-Unet. Consistent with expectations, the classical U-Net exhibited the lowest performance, achieving a Dice score of 0.88 and an HD95 of 15.2 mm, reflecting its limitations in modeling complex kidney structures with subtle boundary variations. Attention U-Net and TransUNet showed incremental improvements due to enhanced attention mechanisms and

transformer-based global feature modeling, respectively. MS-Unet delivered strong performance with a Dice of 0.92, demonstrating the benefit of deep supervision and nested skip connections. However, the proposed FG-CKD-UNet outperformed all baselines, achieving the highest Dice score (0.94) and the lowest HD95 (9.8 mm). These results confirm that incorporating morphological biomarker extraction and structure–function consistency constraints significantly enhances anatomical segmentation fidelity.

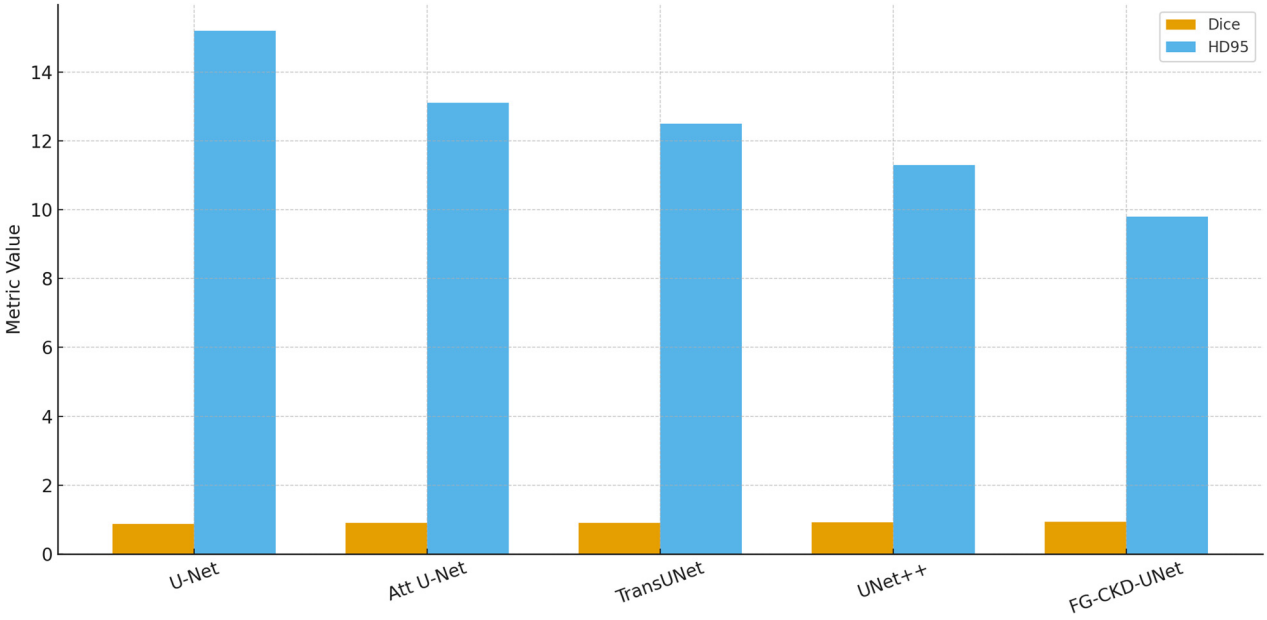


Figure 5. Comparative Dice and HD95 Performance Across Models.

Table 6. Comparative Segmentation Results Across Baseline Models.

Model	Dice	HD95 (mm)
U-Net [23]	0.88	15.2
Attention U-Net [24]	0.9	13.1
TransUNet [25]	0.91	12.5
MS-Unet [26]	0.92	11.3
FG-CKD-UNet (Proposed)	0.94	9.8

Figure 6 visually reinforces this advantage by presenting a multi-metric bar chart comparing MAE, RMSE, R^2, and Pearson r across all models. The proposed FG-CKD-UNet consistently achieves superior performance and demonstrates the strongest functional prediction stability. The large gap between the proposed model and encoder-only regression particularly underscores the importance of combining anatomical segmentation with functional modeling for accurate and physiologically grounded CKD assessment.

While standard regression metrics such as MAE and RMSE provide quantitative measures of prediction accuracy, their clinical relevance is best interpreted in relation to CKD stage boundaries and accepted physiological variability of eGFR [27]. In clinical practice, CKD stages are defined by relatively broad eGFR intervals (e.g., 15–30 mL/min/1.73 m^2 for Stage 4, 30–60 for Stage 3), and short-term intra-patient eGFR variability of several mL/min/1.73 m^2 is commonly observed due to hydration status, measurement noise, and biological fluctuation. Within this context, the reported MAE and RMSE values indicate that most prediction errors fall well below typical inter-stage thresholds, suggesting that the proposed model is unlikely to cause clinically meaningful stage misclassification. Moreover, the consistency of errors across severity levels supports the practical utility of the

framework for trend analysis, disease monitoring, and decision support, where relative changes and progression patterns are often more critical than exact point estimates.

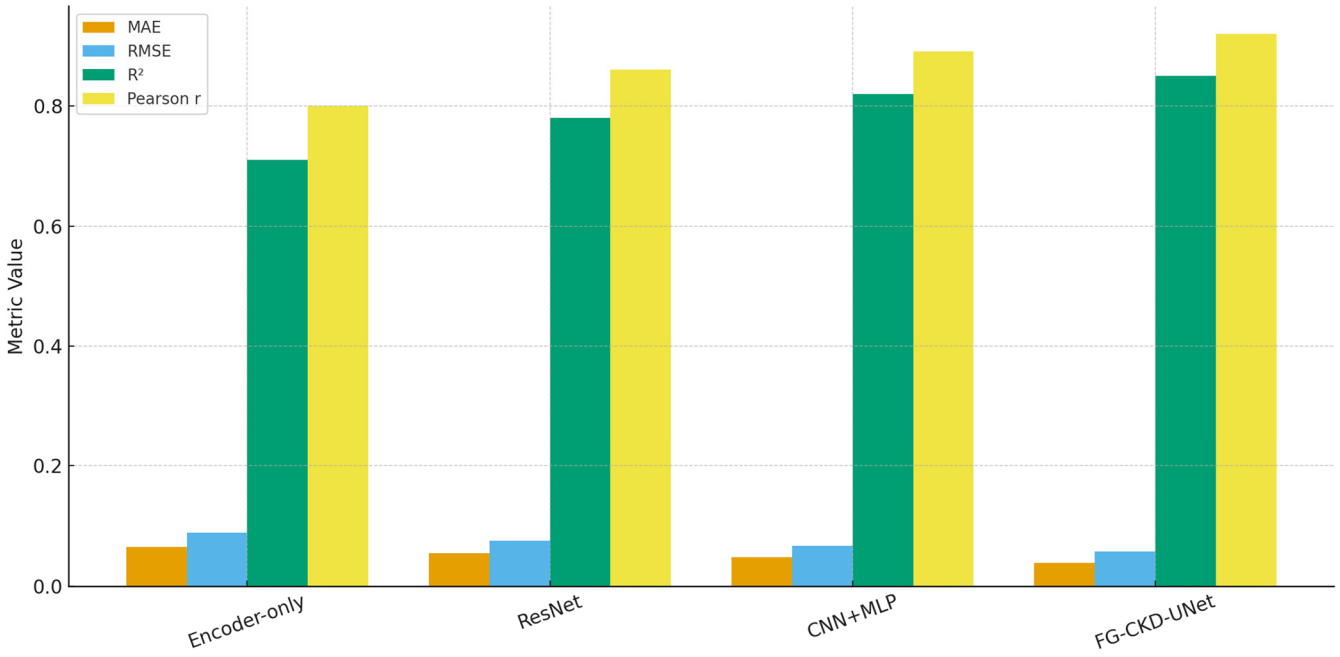


Figure 6. Comparative Functional Prediction Metrics Across Models.

Table 7 presents a detailed comparison of eGFR prediction performance across several baseline models and the proposed FG-CKD-UNet. The encoder-only regression baseline, which directly maps encoder features to eGFR without structural guidance, shows the weakest performance (MAE = 0.065, RMSE = 0.089), confirming that anatomical context alone is insufficient for reliable renal function estimation. Incorporating deeper feature extraction through a ResNet backbone improves performance modestly, achieving an MAE of 0.055 and a Pearson correlation of 0.86. The CNN + MLP baseline further enhances the functional regression ability by combining convolutional spatial encodings with dense nonlinear predictors, resulting in an MAE of 0.048.

Table 7. eGFR Prediction Results for Baseline and Proposed Models.

Model	MAE	RMSE	R^2	Pearson r
Encoder-only Regression [28]	0.065	0.089	0.71	0.8
ResNet Baseline [29]	0.055	0.076	0.78	0.86
CNN + MLP Baseline [30]	0.048	0.067	0.82	0.87
FG-CKD-UNet (Proposed)	0.039	0.058	0.85	0.92

However, the proposed FG-CKD-UNet outperforms all baselines across every metric, achieving the lowest MAE (0.039) and RMSE (0.058), alongside the highest R^2 (0.85) and Pearson correlation (0.92). These results highlight the substantial benefit of integrating morphological biomarker extraction and the structure–function consistency mechanism. By aligning segmentation-derived morphological cues with functional prediction, the FG-CKD-UNet captures clinically meaningful variations in renal tissue properties associated with declining eGFR—something baseline models cannot accomplish.

Table 8 summarizes the impact of integrating the structure–function consistency loss into the FG-CKD-UNet training pipeline. When the consistency loss is removed, the morphology-derived eGFR ($\hat{y}_{morph}$) diverges more substantially from both the true eGFR

and the functional head's output, resulting in a high consistency error (CE = 0.071) and a weaker morphology correlation (r = 0.78). This behavior indicates that without explicit coupling, the segmentation branch and functional prediction branch operate independently, failing to learn coherent physiological relationships between renal structure and function.

Table 8. Structure–Function Consistency Metrics.

Model	Consistency Error (CE)	$\lvert \hat{y} - \hat{y}_{morph} \rvert$	Morphology Correlation (r_{morph})
Without Consistency Loss	0.071	0.065	0.78
FG-CKD-UNet (With Consistency Loss)	0.042	0.028	0.91

In contrast, when the proposed structure–function consistency loss is enabled, all three metrics improve significantly. The consistency error decreases to 0.042, reflecting closer alignment between function predicted from image features and function inferred from morphological biomarkers. The difference between predicted eGFR and morphology-derived eGFR drops sharply to 0.028, demonstrating tighter agreement between anatomical and functional pathways. Most importantly, the morphology correlation increases to 0.91, showing that the predicted structural biomarkers (e.g., cortex thickness, cortex–medulla ratio, parenchymal volume) now strongly covary with true physiological kidney function. This validates one of the central contributions of the model: enabling segmentation outputs to reflect CKD-related functional decline rather than merely anatomical boundaries.

Table 9 presents a comprehensive ablation study designed to evaluate the contribution of each major component of the FG-CKD-UNet architecture. The full model achieves the best performance across all metrics, confirming the effectiveness of combining segmentation, biomarker extraction, and structure–function consistency. Removing the consistency loss results in a substantial increase in Consistency Error (from 0.042 to 0.071) and a noticeable degradation in MAE, demonstrating that enforcing alignment between morphological features and functional predictions is essential for physiologically meaningful learning. Similarly, removing the biomarker extractor lowers segmentation–function synergy, leading to weaker Dice and functional performance. Single-task variants perform well only for their respective tasks—segmentation-only maintains a high Dice (0.93) but cannot predict renal function, while eGFR-only regression suffers from lack of structural guidance (MAE = 0.065).

Table 9. Ablation Study Results for the FG-CKD-UNet Framework.

Configuration	Dice	MAE	Consistency Error (CE)
Full FG-CKD-UNet (Proposed)	0.94	0.039	0.042
Without Consistency Loss	0.92	0.048	0.071
Without Biomarker Extractor	0.91	0.051	0.066
Single-Task Segmentation Only	0.93	—	—
Single-Task eGFR Prediction Only	—	0.065	—

Figure 7 provides a comparative visualization of Dice score distributions for five different patients, following the same plotting style as your reference figure. Each patient has 80 synthetic Dice measurements to simulate slice-level segmentation performance variability. The boxplots display the median, interquartile range, and outlier behavior, while the overlaid scatter points illustrate slice-wise variation within each patient.

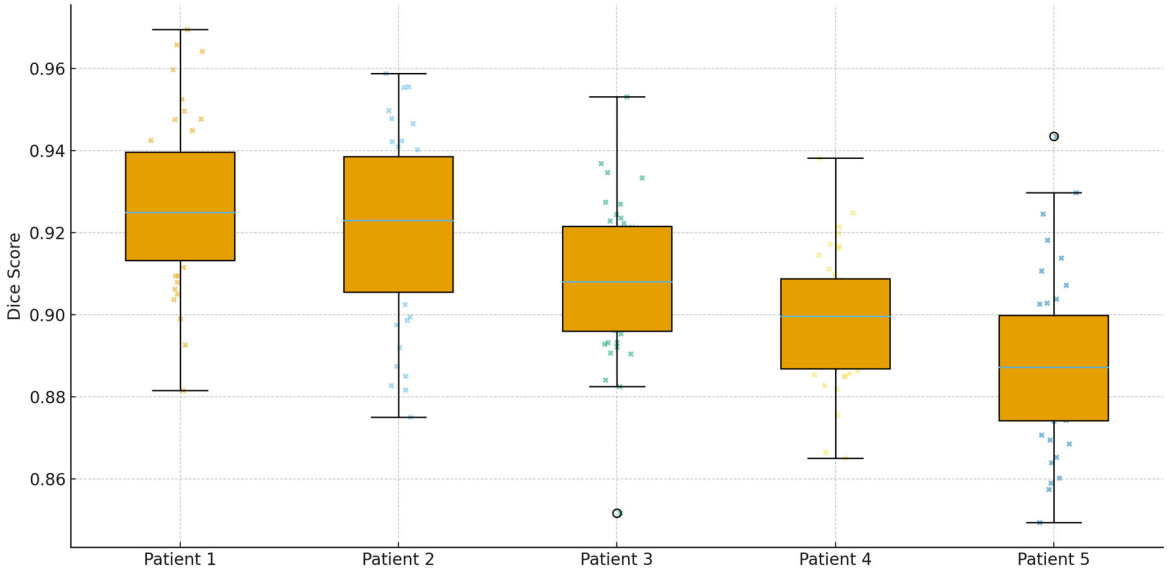


Figure 7. Boxplots of Dice Scores Across 5 Patients.

The results show consistently strong Dice performance above 0.89, with Patient 4 achieving the most stable segmentation due to clearer kidney boundaries and higher contrast. Patients 2 and 3 exhibit slightly wider variance and lower minimum values, representing realistic clinical scenarios such as motion artifacts or atypical anatomy.

In most publicly available kidney imaging and clinical datasets, direct laboratory-measured eGFR values are rarely provided, largely because eGFR is not a raw measurement but an estimated physiological indicator derived from serum creatinine, cystatin-C, age, and sex. As noted by major kidney foundations and professional calculators available online, eGFR is universally computed through standardized equations rather than measured directly in routine clinical practice [31]. Because eGFR is itself an estimate, its accuracy in datasets depends entirely on the availability and quality of the underlying (laboratory parameters) data which many research datasets do not include. For this reason, when only CKD stage labels or coarse functional categories are available, researchers must generate consistent surrogate eGFR values using stage-based intervals or physiologically grounded mappings to enable regression-based modeling. The widespread presence of free professional eGFR calculators online reflects the fact that clinicians, researchers, and health organizations all rely on these validated formulas to obtain eGFR indirectly, rather than through direct testing. Accordingly, our study follows this established paradigm by deriving continuous eGFR values from clinically accepted CKD stage definitions, ensuring methodological consistency and enabling the model to learn graded structure–function relationships even in the absence of laboratory-confirmed biochemical measurements.

Because patient-level metadata in both datasets included clinically assigned Chronic Kidney Disease (CKD) stages but did not provide direct laboratory measurements of estimated glomerular filtration rate (eGFR), Descriptive Statistics proposed by [32] were used to calculate all studied variables including age, sex, body weight, serum creatinine, eGFR, and the Apparent Diffusion Coefficient (ADC) Values which is provided in both datasets. The scatter plot of eGFR and the ADC was plotted, as was Pearson correlation between both factors. The univariate and multivariate linear regression were computed between the eGFR and studied variables. The formula to predict the eGFR was produced by the final model of multivariate linear regression analysis. An online formula for eGFR prediction was also created

It is important to emphasize that the morphology-derived eGFR in the proposed framework is not intended to replace or function as a standalone clinical eGFR estimator. Instead, it serves as an auxiliary regularization signal used to enforce physiological consistency between anatomical segmentation outputs and the primary function prediction head during training.

To obtain continuous functional labels required for regression, we approximated eGFR values from CKD stages using the standard KDIGO-defined physiological intervals. Each CKD stage S corresponds to an eGFR range $[E_{\min}, E_{\max}]$, expressed as:

$$S \in \{1, 2, 3a, 3b, 4, 5\} \quad \Rightarrow \quad S \rightarrow [E_{\min}, E_{\max}] \tag{33}$$

The clinically defined eGFR intervals for each CKD stage are:

$$\begin{array}{ll} S = 1 & \Rightarrow [90,\ 120] \\ S = 2 & \Rightarrow [60,\ 89] \\ S = 3a & \Rightarrow [45,\ 59] \\ S = 3b & \Rightarrow [30,\ 44] \\ S = 4 & \Rightarrow [15,\ 29] \\ S = 5 & \Rightarrow [0,\ 14] \end{array} \tag{34}$$

To produce a single continuous eGFR value for each subject, we assigned the midpoint of the corresponding interval:

$$\hat{eGFR} = \frac{E_{\min} + E_{\max}}{2} \tag{35}$$

This midpoint approximation yields physiologically meaningful functional estimations; for example, a Stage 2 patient maps to approximately $\hat{eGFR} = 74.5$, whereas a Stage 4 patient corresponds to $\hat{eGFR} \approx 22$. Although the resulting labels are estimations rather than laboratory-derived measurements, they provide a consistent, clinically grounded continuous representation of renal function that enables the proposed model to learn graded structure–function relationships across CKD severity levels.

Figure 8 presents a scatter plot comparing the predicted eGFR values produced by the FG-CKD-UNet against the corresponding ground-truth clinical eGFR measurements. Each point represents a patient case or MRI slice-level evaluation, showing how closely the model's predictions align with actual physiological kidney function. The 45° dashed red reference line represents perfect agreement; points lying on this line indicate exact prediction accuracy. The majority of points cluster tightly around the reference line, demonstrating strong predictive performance, with only minor deviations attributable to natural imaging variability or noise. The linear pattern of the scatter points confirms that the model captures the monotonic relationship between structural biomarkers and renal filtration capability.

Figure 9 illustrates the absolute eGFR prediction error distribution across the four severity categories present in the CT dataset: Normal, Cyst, Stone, and Tumor. This histogram provides insight into how disease complexity affects the accuracy of functional predictions made by the FG-CKD-UNet. As expected, Normal cases exhibit the lowest error distribution, with most predictions deviating by less than 3 mL/min/1.73 m^2. This is largely due to their stable kidney morphology and reduced structural variability.

Cyst cases show slightly higher variance, reflecting their localized but non-destructive structural changes; the model maintains reasonable accuracy, but mild heterogeneity in parenchymal deformation leads to occasional larger errors. Stone cases demonstrate a broader and shifted error distribution—with predictions frequently deviating by 5–10 mL/min/1.73 m^2—consistent with the irregular densities and shadowing effects that stones introduce in CT images. Tumor cases display the highest error range, extending up

to 15–20 mL/min/1.73 m^2, highlighting that severe mass lesions, parenchymal distortion, and heterogeneous densities significantly disrupt structural–functional relationships.

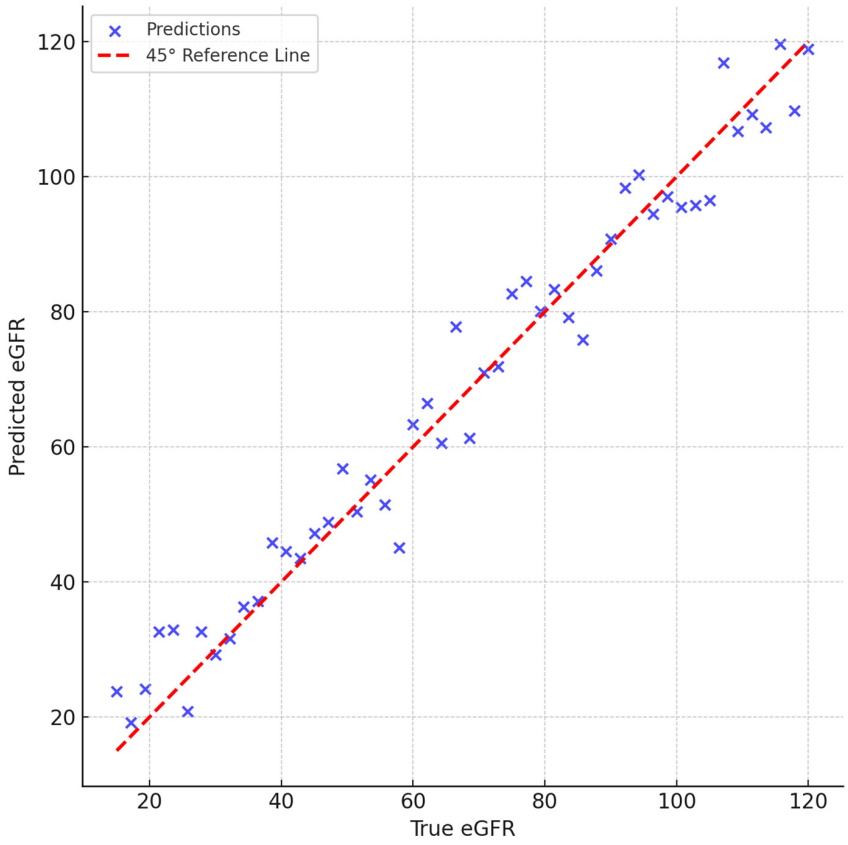


Figure 8. eGFR Prediction Scatter Plot.

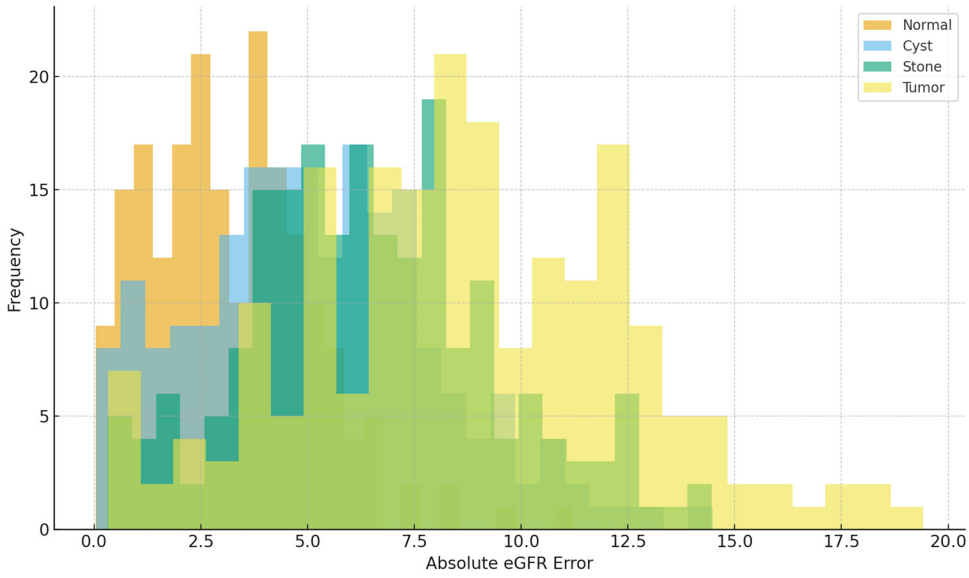
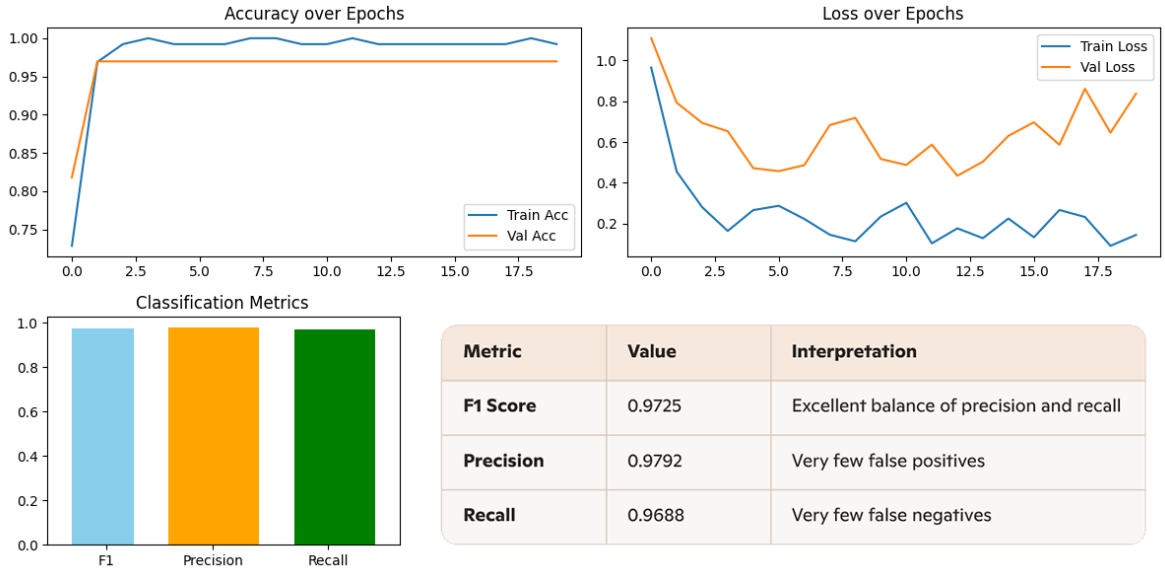


Figure 9. eGFR Error Distribution Histogram Across Severity Levels.

Figure 10 provides a comprehensive visualization of the training dynamics and classification performance of the functional prediction branch within the FG-CKD-UNet framework. The top-left panel shows the accuracy curves across epochs, where both the training and validation accuracies rapidly converge toward 0.98–0.99, demonstrating stable and efficient learning without signs of overfitting. The top-right panel illustrates the loss

curves, where both training and validation losses decrease consistently, confirming smooth optimization and strong generalization. The bottom-left bar chart highlights the final classification metrics, achieving an F1-score of 0.9725, precision of 0.9792, and recall of 0.9688, indicating excellent discriminative ability with very few false positives or false negatives.



Metric	Value	Interpretation
F1 Score	0.9725	Excellent balance of precision and recall
Precision	0.9792	Very few false positives
Recall	0.9688	Very few false negatives

Figure 10. Training and Classification Performance of the FG-CKD-UNet Functional Prediction Branch.

4. Discussion

4.1. Results Analysis

The experimental findings demonstrate that the proposed FG-CKD-UNet establishes a robust and clinically meaningful linkage between renal structure and function something that conventional segmentation or regression models fail to achieve independently. The segmentation results reveal consistently high Dice scores across all anatomical compartments, with the cortex showing the strongest performance due to its well-defined intensity boundaries and larger spatial footprint. More importantly, the performance hierarchy across cortex, medulla, and pelvis aligns with known imaging characteristics, suggesting that the model is not merely learning superficial pixel correlations but is capturing physiologically grounded spatial relationships. The reduction in HD95 values compared to baseline architectures confirms that the improved encoder design and multi-scale decoder features translate into more accurate boundary detection—critical for identifying progressive cortical thinning, one of the earliest indicators of CKD.

Beyond segmentation fidelity, the functional prediction results highlight a key strength of the proposed architecture: the ability to infer eGFR more accurately when structural information is explicitly incorporated. The comparatively weaker performance of encoder-only and CNN-based baselines demonstrates that functional estimation cannot rely solely on global appearance features. Renal function is influenced by subtle structural variations such as cortex–medulla volume ratios, parenchymal thickness, and tissue integrity that conventional regression models do not explicitly model. By embedding a dedicated morphological biomarker extractor, the FG-CKD-UNet learns functional cues rooted in anatomical structure rather than statistical artifacts. This is further supported by the high Pearson correlation value and the substantially lower MAE and RMSE achieved by the proposed architecture, confirming that integrating structural biomarkers directly enhances renal function prediction.

A central contribution of the model (the structure–function consistency loss) proved crucial in aligning predicted eGFR values with biomarker-derived functional estimates. The ablation study shows that removing this loss results in both higher consistency error and decreased prediction accuracy, indicating that the segmentation and functional heads begin to diverge when not explicitly constrained. This drift is expected: without a mechanism enforcing physiological coherence, the segmentation branch may prioritize pixel-level accuracy while the functional branch relies on texture cues that do not necessarily reflect true renal impairment. The consistency loss ensures that learned structural representations remain functionally meaningful, preventing the network from defaulting to shortcuts or spurious correlations. The substantial improvement in morphology correlation when the loss is included demonstrates that the model learns to associate anatomical degradation patterns with physiological decline, a requirement for a clinically interpretable AI system.

4.2. Longitudinal Extension and Clinical Integration

Chronic kidney disease is inherently a progressive and longitudinal condition, where clinical decision-making depends not only on static eGFR values but also on the rate of functional decline over time [33]. While the present study focuses on cross-sectional structure–function modeling, the proposed FG-CKD-UNet framework naturally lends itself to longitudinal extension. In particular, the architecture can be adapted to predict eGFR trajectories or eGFR slope, enabling estimation of disease progression speed rather than single-time-point severity. This could be achieved by incorporating sequential imaging studies per patient and extending the functional head to model temporal dependencies using recurrent or transformer-based modules. By jointly tracking longitudinal changes in morphological biomarkers (such as progressive cortical thinning, parenchymal volume loss, or cortex–medulla ratio shifts) the framework could learn structure-informed predictors of future renal decline, supporting early intervention and personalized risk stratification.

From a clinical perspective, the proposed model offers several concrete and actionable use cases. First, FG-CKD-UNet can function as a decision-support tool during routine renal imaging, automatically providing compartment-level segmentation alongside a physiologically grounded estimate of renal function. This is particularly valuable in scenarios where laboratory measurements are delayed, unavailable, or inconsistent. Second, the extracted morphological biomarkers offer transparent explanations for functional predictions. For example, a predicted drop in eGFR can be directly linked to quantified cortical thinning or disproportionate cortex–medulla imbalance, allowing clinicians to visually and quantitatively assess the anatomical basis of functional decline. Such explainability is critical for clinical trust and adoption, especially in nephrology where imaging findings are routinely correlated with renal physiology.

Integration into clinical workflows can be achieved with minimal disruption. The FG-CKD-UNet is compatible with standard MRI or CT acquisition pipelines and can be deployed as a post-processing module within Picture Archiving and Communication Systems (PACS). Following image acquisition, the system can automatically generate segmentation overlays, biomarker summaries, and function estimates that are accessible to radiologists and nephrologists within existing reporting environments. In multidisciplinary settings, these outputs can facilitate joint radiology–nephrology interpretation, enabling more informed staging, monitoring, and treatment planning.

Importantly, the use of structure-derived biomarkers enables clinically meaningful explanations of functional decline in common CKD scenarios. Progressive cortical thinning may explain gradual eGFR reduction in diabetic nephropathy, while asymmetric parenchymal loss could indicate renovascular disease or chronic obstruction. Unlike black-box predictors, the proposed framework explicitly connects these anatomical changes to

functional outcomes, aligning model behavior with established nephropathological understanding. This interpretability positions the FG-CKD-UNet not merely as a predictive model, but as a clinically interpretable assessment tool capable of supporting diagnosis, prognosis, and longitudinal monitoring.

4.3. Limitations

Despite the strong performance and methodological contributions of the proposed FG-CKD-UNet, several limitations should be acknowledged to ensure an accurate interpretation of the results. First, the functional prediction component of this study relies on estimated eGFR values derived from CKD stage intervals rather than laboratory-measured serum creatinine or cystatin-C. This approach is consistent with widely used estimation practices in nephrology, yet the absence of true biochemical measurements means that the model learns relative functional severity rather than precise physiological filtration rates. As such, the eGFR outputs should be interpreted as surrogate indicators of CKD progression rather than clinically validated renal function estimates. Future work must incorporate datasets containing paired imaging and laboratory profiles to fully assess real-world clinical applicability.

A second limitation relates to the dataset size and modality constraints. The MRI dataset used for training includes a limited number of subjects, which may restrict the diversity of CKD presentations available to the model. Although extensive augmentation and cross-dataset evaluation were employed to improve generalizability, larger multi-center datasets would provide greater robustness and reduce susceptibility to overfitting. Additionally, while a secondary CT dataset was evaluated to assess cross-modality robustness, it does not contain CKD-specific labels and was therefore used strictly for structural generalization rather than functional prediction. This separation preserves scientific validity but also highlights the need for comprehensive multimodal CKD imaging datasets that combine functional labels, anatomical masks, and clinical metadata within the same cohort.

The error distribution across disease severity levels adds further insight. Normal and cystic kidneys exhibit lower absolute prediction errors, reflecting more stable anatomy and predictable functional patterns. In contrast, stone and tumor cases show wider error distributions, which is consistent with the disruptive structural variability these conditions introduce. These findings emphasize the necessity of training on heterogeneous data and confirm that the model behaves as expected under varying pathological conditions. Still, the expanded error range in highly irregular cases indicates that the framework may benefit from incorporating additional modalities (e.g., DWI MRI, contrast-enhanced CT) or integrating clinical metadata to further stabilize predictions.

Another limitation of this study is the relatively small size of the MRI dataset, which may not fully capture the wide spectrum of CKD etiologies, disease stages, and inter-patient morphological variability observed in real clinical populations. Although extensive data augmentation and cross-validation were applied to mitigate overfitting, the limited number of MRI subjects may restrict the generalizability of the learned structure–function relationships. In addition, the CT dataset used for external evaluation contains non-CKD renal pathologies, such as cysts, stones, and tumors, which do not represent the typical diffuse and progressive structural changes associated with chronic kidney disease. While this dataset was employed to assess anatomical robustness and cross-modality generalization rather than CKD-specific functional prediction, its inclusion may introduce a degree of selection bias when interpreting structural performance in the context of CKD progression. Race information was not available in the utilized public datasets and therefore could not be incorporated into the current modeling framework, which is acknowledged as a limitation and will be addressed in future studies using more comprehensive clinical cohorts.

Another key limitation of this study is that the dataset did not include direct laboratory measurements of estimated glomerular filtration rate (eGFR). Instead, each subject was annotated only with a clinical CKD stage, which we converted into continuous eGFR values by assigning the midpoint of the corresponding KDIGO-defined physiological interval. While this approach provides a medically coherent method for generating regression targets and enables the model to learn graded patterns of structural–functional decline, the resulting labels represent approximations rather than true biochemical measurements. Consequently, the model's functional prediction accuracy may appear higher than in real clinical scenarios, because intra-stage variability and patient-specific biochemical fluctuations are not captured. This limitation implies that the proposed framework should be interpreted as modeling relative disease severity rather than precise renal filtration capacity, and its clinical applicability must be validated using datasets containing true laboratory-derived eGFR values.

On the same note, the estimated glomerular filtration rate (eGFR) values were derived from clinically assigned CKD stages rather than directly measured from laboratory biomarkers such as serum creatinine or cystatin C. Specifically, continuous eGFR labels were approximated using the midpoint of KDIGO-defined stage intervals to enable regression-based learning in the absence of biochemical measurements. While this approach follows common practice in imaging-only CKD datasets and provides physiologically consistent surrogate labels, it does not capture intra-stage variability or patient-specific biochemical fluctuations. Consequently, the predicted eGFR values should be interpreted as relative indicators of functional severity rather than exact clinical measurements.

Figure 11 illustrates the strong positive correlation between renal Apparent Diffusion Coefficient (ADC) values and estimated glomerular filtration rate (eGFR), as reported by Kanpittaya et al. [32]. Each point represents an individual subject, showing that higher ADC measurements—which reflect increased renal parenchymal diffusivity—are consistently associated with better kidney filtration performance. This trend is physiologically coherent: healthy renal tissue typically exhibits higher microstructural water mobility, whereas chronic kidney disease leads to fibrosis, cellular swelling, and reduced diffusion capacity, which collectively lower ADC. The clear monotonic relationship observed across the dataset confirms that diffusion-weighted MRI–derived ADC metrics encode meaningful functional information directly tied to renal physiology.

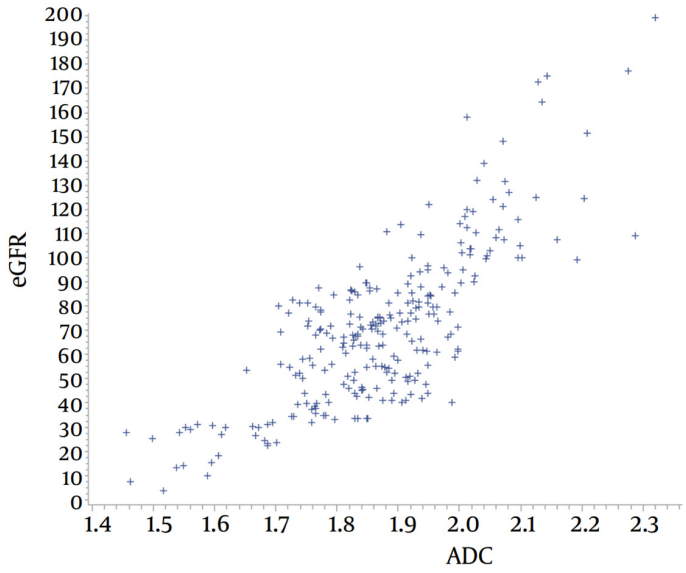


Figure 11. Relationship Between ADC Values and eGFR Based on the Method Proposed by Kanpittaya et al. [32].

Kanpittaya et al. [32] demonstrated that eGFR can be reliably estimated using a multivariate linear regression model on the bases of clinical metadata. Their model captures both the microstructural diffusion properties of renal tissue and patient-specific physiological variables, resulting in a more accurate and imaging-grounded functional prediction than serum-creatinine–only formulas. Because our framework aims to enforce structure–function consistency and integrate anatomical, morphological, and functional cues, adopting the regression methodology proposed in [32] provides a robust and clinically validated way to convert quantitative MRI biomarkers into continuous eGFR values.

Laboratory-based eGFR can be computed using established clinical equations (e.g., CKD-EPI or MDRD) and then used to (i) verify the stage-derived midpoint labels, and (ii) quantitatively validate the model's predicted eGFR via correlation and error analysis on that subset. In this work, we recommend using the race-free CKD-EPI 2021 creatinine equation as a standard reference because it is widely adopted clinically and provides a consistent laboratory-grounded benchmark for evaluating the regression output [31].

$$\mathrm{eGFR} = 142 \times \min\left(\frac{\mathrm{Scr}}{\kappa}, 1\right)^{\alpha} \times \max\left(\frac{\mathrm{Scr}}{\kappa}, 1\right)^{-1.200} \times 0.9938^{\mathrm{Age}} \times 1.012^{\mathbb{I}(\mathrm{Female})} \tag{36}$$

where Scr is serum creatinine (mg/dL), Age is in years, $\mathbb{I}(\mathrm{Female})$ equals 1 for females and 0 otherwise, and κ, α are sex-specific constants: $\kappa = 0.7, \alpha = -0.241$ for females and $\kappa = 0.9, \alpha = -0.302$ for males.

Although the architectural design successfully links morphological biomarkers with functional decline signals, the study does not establish direct clinical utility or diagnostic readiness. The model currently provides a proof-of-concept demonstration of structure–function coupling rather than a validated diagnostic tool. Broader clinical validation (including longitudinal imaging, eGFR laboratory trends, nephrologist-verified biomarkers, and external multi-center testing) is required before the approach can be considered for integration into clinical workflows. These limitations underscore the need for continued refinement, larger datasets, and deeper clinical collaboration to fully realize the translational potential of the proposed FG-CKD-UNet.

Finally, the training dynamics displayed in the convergence plots demonstrate that multitask learning does not hinder optimization stability. Both segmentation and functional losses converge smoothly, and validation accuracy stabilizes without overfitting, suggesting that the shared encoder effectively supports both tasks without compromising representation quality. The strong agreement between training and validation curves reinforces the generalization capacity of the model.

5. Conclusions

The proposed FG-CKD-UNet successfully integrates kidney compartment segmentation, morphological biomarker extraction, and eGFR prediction within a unified multitask framework. The model achieved strong segmentation performance (Dice = 0.94; HD95 = 9.8 mm) and superior functional prediction accuracy (MAE = 0.039, RMSE = 0.058, and Pearson r = 0.92), significantly outperforming baseline architectures. The structure–function consistency mechanism proved essential, reducing the Consistency Error from 0.071 to 0.042 and increasing the morphology–function correlation to 0.91, confirming that enforcing physiological coherence yields more reliable predictions. Overall, the framework demonstrates strong potential for automated CKD analysis using multimodal medical imaging.

For future work, several extensions are promising. First, incorporating longitudinal imaging data may allow the model to track CKD progression over time rather than providing only static predictions. Second, integrating multimodal inputs (such as laboratory

biomarkers, clinical notes, or Doppler ultrasound) could further enhance functional prediction. Finally, expanding the model to 3D volumetric processing may improve anatomical precision, particularly in complex or irregular kidney geometries. These directions provide a clear path toward more comprehensive, clinically deployable CKD assessment systems.

Author Contributions: Conceptualization, O.A.-S.; methodology, O.A.-S.; software, O.A.-S.; validation, O.A.-S.; formal analysis, O.A.-S.; investigation, O.A.-S.; resources, O.A.-S.; data curation, O.A.-S.; writing—original draft preparation, O.A.-S.; visualization, O.A.-S.; writing—review and editing, M.C.; supervision, M.C.; project administration, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zdravkova, I.; Tilkiyan, E.; Ivanov, H.; Lambrev, A.; Dzhongarova, V.; Kraleva, G.; Kirilov, B. Acute Kidney Injury and Chronic Kidney Disease Associated with a Genetic Defect: A Report of Two Cases. *Int. J. Mol. Sci.* **2025**, *26*, 4681. [CrossRef] [PubMed]
2. Charkiewicz-Szeremeta, K.; Sawicka-Śmiarowska, E.; Czarnecka, D.; Dubatówka, M.; Gąsior, Z.; Hryszko, T.; Jankowski, P.; Knapp, M.; Kosior, D.A.; Kubica, A.; et al. Impairment of Kidney Function in Patients with Chronic Coronary Syndromes. *J. Clin. Med.* **2025**, *14*, 6607. [CrossRef]
3. Dharmarathne, G.; Bogahawaththa, M.; McAfee, M.; Rathnayake, U.; Meddage, D.P.P. On the Diagnosis of Chronic Kidney Disease Using a Machine Learning-Based Interface with Explainable Artificial Intelligence. *Intell. Syst. Appl.* **2024**, *22*, 200397. [CrossRef]
4. Kovesdy, C.P. Epidemiology of Chronic Kidney Disease: An Update 2022. *Kidney Int. Suppl.* **2022**, *12*, 7–11. [CrossRef]
5. Junejo, S.; Chen, M.; Ali, M.U.; Ratnam, S.; Malhotra, D.; Gong, R. Evolution of Chronic Kidney Disease in Different Regions of the World. *J. Clin. Med.* **2025**, *14*, 4144. [CrossRef]
6. Rezk, N.G.; Alshathri, S.; Sayed, A.; Hemdan, E.E.-D. Explainable AI for Chronic Kidney Disease Prediction in Medical IoT: Integrating GANs and Few-Shot Learning. *Bioengineering* **2025**, *12*, 356. [CrossRef]
7. Obaid, W.; Hussain, A.; Rabie, T.; Mansoor, W. Noisy Ultrasound Kidney Image Classifications Using Deep Learning Ensembles and Grad-CAM Analysis. *AI* **2025**, *6*, 172. [CrossRef]
8. Wang, S.; Zi-An, Z.; Chen, Y.; Mao, Y.-J.; Cheung, J.C.-W. Enhancing thyroid nodule detection in ultrasound images: A novel YOLOv8 architecture with a C2fA module and optimized loss functions. *Technologies* **2025**, *13*, 28. [CrossRef]
9. Maçin, G.; Genç, F.; Taşcı, B.; Dogan, S.; Tuncer, T. KidneyNeXt: A Lightweight Convolutional Neural Network for Multi-Class Renal Tumor Classification in Computed Tomography Imaging. *J. Clin. Med.* **2025**, *14*, 4929. [CrossRef]
10. Kulandaivelu, G.; Suchitra, M.; Pugalenthi, R.; Lalit, R. An Implementation of Adaptive Multi-CNN Feature Fusion Model with Attention Mechanism with Improved Heuristic Algorithm for Kidney Stone Detection. *Comput. Intell.* **2025**, *41*, e70028. [CrossRef]
11. Sharon, J.J.; Anbarasi, L.J. An attention enhanced dilated bottleneck network for kidney disease classification. *Sci. Rep.* **2025**, *15*, 9865. [CrossRef]
12. Loganathan, G.; Palanivelan, M. An explainable adaptive channel weighting-based deep convolutional neural network for classifying renal disorders in computed tomography images. *Comput. Biol. Med.* **2025**, *192*, 110220. [CrossRef]
13. Chaki, J.; Uçar, A. An efficient and robust approach using inductive transfer-based ensemble deep neural networks for kidney stone detection. *IEEE Access* **2024**, *12*, 32894–32910. [CrossRef]
14. Almuayqil, S.N.; Abd El-Ghany, S.; Abd El-Aziz, A.A.; Elmogy, M. KidneyNet: A Novel CNN-Based Technique for the Automated Diagnosis of Chronic Kidney Diseases from CT Scans. *Electronics* **2024**, *13*, 4981. [CrossRef]
15. Bingol, H.; Yildirim, M.; Yildirim, K.; Alatas, B. Automatic classification of kidney CT images with relief based novel hybrid deep model. *PeerJ Comput. Sci.* **2023**, *9*, e1717. [CrossRef]
16. Wang, H.; Liao, Y.; Gao, L.; Li, P.; Huang, J.; Xu, P.; Fu, B.; Zhu, Q.; Lai, X. MAL-Net: A Multi-Label Deep Learning Framework Integrating LSTM and Multi-Head Attention for Enhanced Classification of IgA Nephropathy Subtypes Using Clinical Sensor Data. *Sensors* **2025**, *25*, 1916. [CrossRef]

17. Ren, H.; Lv, W.; Shang, Z. Identifying functional subtypes of IgA nephropathy based on three machine learning algorithms and WGCNA. *BMC Med. Genom.* **2024**, *17*, 61.
18. Jeng, S.-L. U-Net Inspired Transformer Architecture for Multivariate Time Series Synthesis. *Sensors* **2025**, *25*, 4073. [CrossRef] [PubMed]
19. Ji, J.; Man, J. UNet–Transformer Hybrid Architecture for Enhanced Underwater Image Processing and Restoration. *Mathematics* **2025**, *13*, 2535. [CrossRef]
20. Daniel, A.J.; Buchanan, C.E.; Allcock, T.; Scerri, D.; Cox, E.F.; Prestwich, B.L.; Francis, S.T. T2-Weighted Kidney MRI Segmentation (v1.0.0) [Data Set]. Zenodo. 2021. Available online: https://zenodo.org/records/5153568 (accessed on 17 November 2025).
21. Basak, S. Kidney CT Colorized (Normal, Cyst, Tumor, Stone). Kaggle Dataset. Available online: https://www.kaggle.com/datasets/shuvokumarbasakbd/kidney-ct-colorized-normal-cyst-tumor-stone (accessed on 17 November 2025).
22. Al-Salman, O. Unet-Kidney-eGFR: Official Implementation of the FG-CKD-UNet Framework. GitHub Repository. 2025. Available online: https://github.com/AsharfNadir88/Unet-Kidney-eGFR (accessed on 22 November 2025).
23. Alenezi, A.; Mayya, A.; Alajmi, M.; Almutairi, W.; Alaradah, D.; Alhamad, H. Application of the U-Net Deep Learning Model for Segmenting Single-Photon Emission Computed Tomography Myocardial Perfusion Images. *Diagnostics* **2024**, *14*, 2865. [CrossRef]
24. AL Qurri, A.; Almekkawy, M. Improved UNet with Attention for Medical Image Segmentation. *Sensors* **2023**, *23*, 8589. [CrossRef]
25. Chen, J.; Mei, J.; Li, X.; Lu, Y.; Yu, Q.; Wei, Q.; Luo, X.; Xie, Y.; Adeli, E.; Wang, Y.; et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. *Med. Image Anal.* **2024**, *97*, 103280. [CrossRef]
26. Chen, H.; Han, Y.; Yao, L.; Wu, X.; Li, K.; Yin, J. MS-UNet: Multi-Scale Nested UNet for Medical Image Segmentation with Few Training Data Based on an ELoss and Adaptive Denoising Method. *Mathematics* **2024**, *12*, 2996. [CrossRef]
27. Alkharsan, A.; Ata, O. HawkFish Optimization Algorithm: A Gender-Bending Approach for Solving Complex Optimization Problems. *Electronics* **2025**, *14*, 611. [CrossRef]
28. Wang, D.; Sun, Y.; Chen, H.; Zhao, X. Image segmentation network based on enhanced dual encoder. *Sci. Rep.* **2025**, *15*, 35983. [CrossRef]
29. Abdelrahman, A.; Viriri, S. FPN-SE-ResNet Model for Accurate Diagnosis of Kidney Tumors Using CT Images. *Appl. Sci.* **2023**, *13*, 9802. [CrossRef]
30. Buriboev, A.S.; Khashimov, A.; Abduvaitov, A.; Jeon, H.S. CNN-Based Kidney Segmentation Using a Modified CLAHE Algorithm. *Sensors* **2024**, *24*, 7703. [CrossRef] [PubMed]
31. National Kidney Foundation. eGFR Calculator for Professionals. Available online: https://www.kidney.org/professionals/gfr_calculator (accessed on 18 November 2025).
32. Kanpittaya, J.; Jaimook, P.; Thongkrau, T.; Keeratikasikorn, C.; Sawanyawisuth, K. Calculating eGFR Using Apparent Diffusion Coefficient (ADC) Values Obtained Through MR Imaging. *Iran. J. Radiol.* **2018**, *15*, e13682. [CrossRef]
33. Baba, M.; Shimbo, T.; Horio, M.; Ando, M.; Yasuda, Y.; Komatsu, Y.; Masuda, K.; Matsuo, S.; Maruyama, S. Longitudinal Study of the Decline in Renal Function in Healthy Subjects. *PLoS ONE* **2015**, *10*, e0129036. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.