Cover Letter to the Editor-in-Chief

Date: [Current Date]

To: Prof. Dimitrios I. Fotiadis

Editor-in-Chief, IEEE Journal of Biomedical and Health Informatics (J-BHI)

Manuscript ID: JBHI-05124-2025

Title: DualSeg: Unified multi-scale framework with dual-stage encoder for glomerular segmentation

Dear Prof. Fotiadis and the Associate Editor,

We would like to express our sincere gratitude to you and the reviewers for the comprehensive and constructive evaluation of our manuscript. We have found the comments extremely valuable in improving the quality, rigor, and clarity of our work.

We have carefully revised the manuscript in accordance with the "Major Revision" decision. In this revised version, we have significantly expanded our experimental validation by including new state-of-the-art baselines (e.g., InceptionNeXt, U-Mamba), adding a new external validation dataset (KPMP), and visualizing Effective Receptive Fields (ERF) to clarify our architectural novelty. Furthermore, we have refined the mathematical formulations and included statistical significance testing to bolster our claims.

Below, we provide a point-by-point response detailing how each comment has been addressed. Changes in the revised manuscript have been highlighted for ease of review.

We hope that these revisions satisfactorily address the reviewers' concerns and that the manuscript is now suitable for publication in J-BHI.

Sincerely,

Ling He, Ph.D. (Corresponding Author)

College of Biomedical Engineering, Sichuan University

On behalf of all authors

Point-by-Point Response to Reviewer 1

Major Concerns

1. Experimental Validation and Efficiency Claims (FLOPs/Memory/Latency):

Comment: The reviewer noted that efficiency claims were unsubstantiated without benchmarks like FLOPs, memory, and latency.

Response: We appreciate this critical feedback. To substantiate our efficiency claims, we have performed a comprehensive complexity analysis.

- Action: We have added a "Performance vs. FLOPs" comparison in Fig. 1 (Bottom) and integrated these metrics into our analysis. As shown in the figure, DualSeg achieves an optimal trade-off, securing the highest Dice score while maintaining significantly lower FLOPs (Floating Point Operations) compared to Transformer-based counterparts like UNETR and Swin UNETR.
- Clarification on Latency/Memory: While we acknowledge the value of latency and memory metrics, these are highly hardware-dependent (varying significantly between different GPU architectures and optimization libraries). We focused on FLOPs as it provides a hardware-neutral, theoretical measure of algorithmic complexity, which is the primary focus of this methodological paper. We believe the FLOPs analysis sufficiently demonstrates the architectural efficiency of the proposed Wave-Swin and VRWKV integration.

2. Dataset Limitations (H&E Staining):

Comment: The reviewer pointed out that limiting evaluation to PAS staining limits clinical relevance, as H&E is also common.

Response: We agree that cross-stain generalization is a vital aspect of clinical utility.

- Constraint: Unfortunately, publicly available, high-quality datasets with pixel-level glomerular annotations for H&E stained images are currently scarce, preventing us from performing a direct supervised training comparison on H&E data comparable to the PAS benchmarks.
- Action: To rigorously address the core concern of *generalizability* and *robustness*, we introduced a new, independent human dataset: the KPMP dataset (referenced in Section V.A). Even though this dataset is PAS-stained, it represents a significant "cross-center" and "cross-species" challenge. We applied models trained *solely* on murine data (KPIs) directly to this human dataset without fine-tuning. The results (Table and Fig. 8)

demonstrate that DualSeg significantly outperforms baselines in this challenging domain-shift scenario, implying robust feature learning that we believe will translate to other staining protocols in future work.

3. Incomplete Architectural Comparisons (Missing Baselines):

Comment: The reviewer noted the omission of modern efficient architectures like InceptionNeXt and Mamba-based models.

Response: We accept this valid criticism. We have significantly expanded our comparative experiments.

- Action: We have added comparisons against InceptionNeXt, VM-UNet-V2, and U-Mamba across all datasets.
- Results: As shown in Table I (KPIs) and Table II (HuBMAP), as well as the scatter plot in Fig. 1, DualSeg consistently outperforms these modern architectures. Specifically, while InceptionNeXt offers efficiency, it struggles with the morphological variance in the 5/6Nx subset. DualSeg achieves a superior balance of global context (via VRWKV) and local precision (via Wave-Swin) compared to the pure Mamba or CNN-Next approaches.

4. Clinical Deployment Analysis:

Comment: The reviewer requested a deployment feasibility analysis and clinical context for the mDSC improvements.

4. Clinical Deployment Analysis

Comment: The reviewer requested a deployment feasibility analysis and clinical context for the mDSC improvements.

Response: We appreciate this critical feedback regarding the translational gap between metrics and practice.

- Feasibility (Hardware): We acknowledge that a full-scale deployment study on edge devices (e.g., with limited GPU memory) is outside the current scope, which focuses on methodological architecture. We have explicitly added this constraint to our Limitations (Section VI.D) and flagged optimization for CPU-only workstations as a priority for our future work.
- Clinical Context (Diagnostic Confidence): To address the concern about whether the 2.73% improvement translates to diagnostic confidence, we argue that reliability is as crucial as accuracy in clinical settings:

    - Boundary Precision: As shown in the new Error Maps, our method significantly reduces HD95. Clinically, this precise boundary delineation is vital for quantifying *interstitial fibrosis*, where pixel-level errors can lead to incorrect CKD staging.
    - Statistical Stability: Furthermore, to ensure these gains support robust clinical decision-making, we performed statistical significance testing (t-test) (details provided in response to Reviewer 2). The results confirm that DualSeg's performance is statistically superior ($p<0.05$) with lower variance, offering the consistency pathologists require to trust automated tools.

5. Architectural Design Justification:

Comment: The reviewer questioned the "local-to-global" sequence and the empirical selection of window sizes.

Response: We have strengthened the justification for our design choices.

- Action (Sequencing): We conducted an ablation study on module ordering (referenced in Table III). The results show that the "Attention-Wave" (global-to-local) configuration yields significantly lower performance (mDSC 85.78%) compared to our proposed "Wave-Attention" (local-to-global) design (mDSC 92.98%). This confirms our hypothesis that refining local features *before* modeling global dependencies is crucial for histological images.
- Action (Window Size): In Section III.A.2, we have clarified that our dynamic window sizes ($7, 11, 15$) are not arbitrary but are grounded in biological priors. The maximum window size was selected to cover the average glomerular diameter (approx. 154px) after downsampling, ensuring the network captures the full structural context without integrating excessive background noise.

Minor Concerns:

Comment: Comparison figures lack error maps; component analysis ignores overhead.

Response:

- Action: We have updated Fig. 6, Fig. 7, and Fig. 8 to include Error Maps. These maps visually highlight over-segmentation (green) and under-segmentation (yellow), providing a clear, qualitative assessment of where baseline models fail compared to DualSeg.

Point-by-Point Response to Reviewer 2

1. Novelty and Visual Evidence:

Comment: The reviewer felt the novelty compared to TransUNet/H2Former was not convincing and requested visual analysis (e.g., feature maps).

Response: We have clarified the distinction between DualSeg and previous hybrids.

- Action: We added a visualization of Effective Receptive Fields (ERF) in Fig. 1 (Top).
- Explanation: This visualization demonstrates that while CNNs focus locally and standard Transformers produce noisy global patterns, DualSeg (Fig. 1-IV) achieves a "clean, global ERF." This visually proves that our specific combination of Wave-Swin (dynamic local) and VRWKV (linear global with Z-shift) yields a structural perception capability distinct from and superior to TransUNet or H2Former.

2. Generalization and Citations:

Comment: The reviewer requested cross-stain/center validation and suggested specific citations regarding stain augmentation.

Response:

- Action (Validation): As mentioned in the response to Reviewer 1, we introduced the KPMP dataset to perform cross-center and cross-species validation (training on Mouse/KPIs, testing on Human/KPMP). This is a rigorous test of generalization.
- Action (Citations): We have cited the suggested works (References [46] and [47] in the revised bibliography) and incorporated them into our discussion on stain invariance and future directions for handling multistained images in Section II.B.

3. Clinical Relevance and Interpretability:

Comment: The reviewer asked for statistical support and a connection to diagnostic benefits.

Response:

- Action: We performed t-tests to validate the statistical significance of our improvements. Fig. 10 now includes error bars and significance asterisks ($p<0.05$ to $p<0.001$), confirming that our performance gains are statistically robust and not due to random variance.
- Action: We expanded Section VI.C to explain that the lower HD95 score implies higher boundary fidelity, which is clinically essential for distinguishing between the glomerular capsule and surrounding interstitial fibrosis—a key marker in diagnosing diabetic nephropathy.

4. Methodological Presentation:

Comment: The reviewer found the math dense and requested simplification.

Response:

- Action: We have significantly streamlined Section III (Methodology). We simplified the notation, clearly defined all symbols (including inputs/outputs for the Wave and VRWKV blocks), and removed redundant derivations. The focus is now on the conceptual flow: how the Wave component handles texture and how the VRWKV component manages global spatial heterogeneity.

5. Experimental Analysis (Inference Time, Failures):

Comment: The reviewer requested inference metrics and qualitative examples of failure cases.

Response:

- Action (Metrics): As detailed in Fig. 1, we have included FLOPs comparisons to address efficiency.
- Action (Failure Cases): We have added a new "Failure Case Analysis" in Section VI.B and Fig. 9. We candidly discuss an instance where DualSeg failed to segment a globally sclerotic glomerulus due to truncation at the image edge. This provides a balanced view of the model's current limitations.

Minor Issues:

Comment: Fix font sizes, typos, and figure alignment.

Response:

- Action: We have proofread the manuscript extensively.

- **Figure fonts have been enlarged for readability.**
- **The typo "uqualitative" and hyphenation issues have been corrected.**
- **Dataset names (HuBMAP, KPIs, KPMP) are now capitalized consistently.**
- **References have been checked for formatting compliance.**