# Enhancing Predictive Accuracy of Allograft Outcomes: Beyond Traditional Clinical and Biopsy Scoring Methods

Akshita Gupta[a], Ahmed Naglah[b], Kuang-Yu Jen[c], Avi Rosenberg[d], Luis Rodrigues[e], William L Clapp[f], Kawther Alquadan[b], John E Tomaszewski[g], Benjamin Shickel[b], and Pinaki Sarder[b]

[a]Health Outcomes and Biomedical Informatics, Department of Medicine, University of Florida, Gainesville, FL
[b]Division Of Nephrology, Hypertension, and Renal Transplantation, Department of Medicine, University of Florida, Gainesville, FL
[c]Department of Pathology and Laboratory Medicine, University of California, Davis, CA
[d]Department of Pathology, Johns Hopkins University, Baltimore, MD
[e]Department of Nephrology, University of Coimbra, Coimbra, Portugal
[f]Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL
[g]Department of Pathology & Anatomical Sciences, University at Buffalo, Buffalo, NY

## ABSTRACT

Kidney transplantation is a critical intervention for managing end-stage renal disease, with post-transplant biopsies providing essential insights into graft function. This study aimed to develop and evaluate an automated system for classifying kidney biopsies taken immediately after transplantation to predict patient outcomes, specifically changes in estimated glomerular filtration rate (eGFR). We utilized a dataset of 56 Formalin-Fixed, Paraffin-Embedded (FFPE) whole slide images of time-zero biopsies, stained with Periodic Acid-Schiff (PAS), along with time series eGFR data recorded at 3 months, 6 months and 12 motnhs post-transplant. We employed ComPRePS (an in-house multicompartment segmentation tool) to analyze biopsy images, segmenting compartments such as cortical and medullary interstitium, non-globally and globally sclerotic glomeruli, tubules, and arteries/arterioles. Focus was placed on glomeruli, from which embeddings were extracted using a pretrained Vision Transformer (ViT) model (vit-base-patch16-224-in21k). These embeddings captured critical features for training an Artificial Neural Network (ANN) model to classify patients based on the percent change in eGFR from 3 months to 12 months, categorizing them into 'eGFR decline' and 'eGFR stable' groups. The proposed classification model achieved an accuracy of 0.73 in predicting eGFR changes. We validated these results by visualizing the two patient groups in a reduced dimensional domain using UMAP (Uniform Manifold Approximation and Projection), which revealed a clear distinction between the groups. This result underscores the potential of early biopsies for predicting long-term graft outcomes, enhancing patient management by providing early insights. We also trained a baseline ANN model using glomeruli hand-crafted features from ComPRePS, which achieved an accuracy of 0.55. Our proposed model achieved higher accuracy than baseline. Further research is needed to refine the model, expand the dataset, and validate findings across diverse populations to improve prediction accuracy and clinical applicability.

**Keywords:** eGFR; Biopsy; Glomeruli; Vision Transformer; Artificial Neural Network

## 1. INTRODUCTION

Kidney Transplant is a critical treatment option for patients with end-stage renal disease (ESRD). Over 100,000 kidney transplants are performed worldwide each year.[1] In 2023, around 28,000 kidney transplants were performed in the United States. Kidney disease recurrence after transplantation occurs in approximately 10-30%

Further author information: (Send correspondence to Pinaki Sarder,PhD)
E-mail: Pinaki.sarder@medicine.ufl.edu, Telephone: 716-829-2265

Medical Imaging 2025: Digital and Computational Pathology, edited by John E. Tomaszewski
Aaron D. Ward, Proc. of SPIE Vol. 13413, 134130I · © 2025 SPIE
1605-7422 · doi: 10.1117/12.3048196

of patients, depending on the underlying condition that led to ESRD.[2] In practice, renal pathologists evaluate biopsy samples and interpret various clinical and laboratory data.Pathologists assess the health of the glomeruli, tubules, interstitium, and blood vessels. Typically, pathologists will grade and classify the types and severity of rejection to guide treatment decisions. Some of those traditional methods are Kidney Donor Profile Index (KDPI),[3] which rely primarily on donor demographics and clinical data. The most direct method for assessing transplant success is through a kidney biopsy, typically performed at different intervals post-transplant (e.g., protocol biopsies or for-cause biopsies when there's suspicion of rejection).[4] While these methods provide some predictive value, they are often only minimally to moderately effective in forecasting long-term graft outcomes. Similarly, current visual and semi-quantitative biopsy scoring systems such as the Banff[5–8] classification, the Maryland Aggregate Pathology Index (MAPI),[9] and the Remuzzi score[10] are widely used to assess transplant biopsies. The success of a transplant largely depends on the swift and accurate assessment of the graft's condition, often achieved through a zero-hour post-transplant biopsy. In this study, we aim to improve the chances of patient survival by decreasing the need for another transplant. By accurately forecasting the function of renal transplant allografts and consequently adapting treatment approaches would have a substantial impact on enhancing the survival of renal transplant allografts.
Kidney transplantation is the most effective treatment for end-stage renal disease. This study developed and evaluated an automated system to predict transplant outcomes using time-zero biopsies. Analyzing 56 Formalin-Fixed, Paraffin-Embedded (FFPE) biopsies stained with Periodic Acid-Schiff (PAS) and eGFR data at 3 months, 6 months, and 1 year, we employed ComPRePs[11] (an in-house multicompartment segmentation tool) to extract glomeruli and hand-crafted features. Embeddings from these glomeruli were generated using a pre-trained Vision Transformer (ViT)[12] model and used to train an Artificial Neural Network (ANN).[13] The model achieved an accuracy of 0.73. UMAP[14] visualization revealed a clear distinction between eGFR decline and stability. This approach underscores the potential of early biopsies in predicting long-term graft outcomes, highlighting its significance for improving patient management and guiding early interventions.

## 2. MATERIALS & METHODS

### 2.1 Dataset

The dataset consists of 56 Formalin-Fixed Paraffin-Embedded (FFPE) whole slide images (WSIs) of kidney transplant biopsies, all stained with Periodic Acid-Schiff (PAS). These biopsies were taken at time zero (T0), which corresponds to the moment right after transplantation, providing a detailed view of the kidney tissue's initial condition.

For each of these 56 patients, there is corresponding time series data of estimated glomerular filter rate (eGFR) values, which serve as indicators of kidney function over time. The eGFR data includes follow-up measurements taken at 3 months, 6 months, 1 year, and 3 years post-transplantation. This time series allows monitoring the progression of kidney function and evaluating the long-term outcomes of the transplant, such as stability, improvement, or decline in function.

The kidney transplants related to the dataset were performed in 2016 or later in UCDavis. The ages of the patients at the time of transplantation range from 16 to 75 years, with a wide distribution among different age groups.

### 2.2 Image Segmentation

ComPRePS (Computational Pathology Research and Precision Segmentation) is a cloud-based tool designed to streamline kidney FTU segmentation in whole slide images (WSI). Using Facebook AI Research's Detectron2 library, ComPRePS performs complex panoptic segmentation of six kidney-specific FTU (non-globally and globally sclerotic glomeruli, cortical and medullary interstitium, tubules, arteries/arterioles) and also supports pipelines for peritubular capillaries (PTC) and interstitial fibrosis and tubular atrophy (IFTA). ***Fig*** 1 shows the segmentation results.

The training workflow involves segmenting patches from 190 annotated WSIs, representing various pathologies and healthy kidney tissues. Once trained, the model is available in the ComPRePS model zoo for prediction

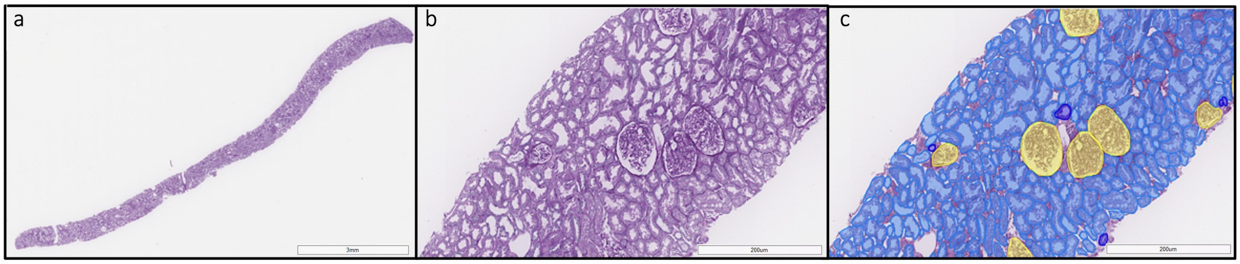


Figure 1. Representative T0 kidney biopsy images demonstrating the analysis pipeline. (a) Low-magnification view of the entire biopsy section, highlighting the overall structure. (b) Higher-magnification view of a region of interest, showing glomeruli and tubulointerstitial compartments. (c) Computational segmentation overlay, with glomeruli (yellow), tubules (blue), and arteries/arterioles (dark blue).

on new WSIs, where it segments large WSIs by processing smaller patches and then stitches the results into a complete segmentation mask, saved as JSON for visualization.

ComPRePS is accessible through a user-friendly interface that allows customizable training and prediction workflows, mask editing, and hyperparameter tuning. This tool, offered open source, democratizes access to AI-driven pathology analysis, allowing researchers and clinicians to use, refine, and apply segmentation models without additional software or third-party tools.[11]

## 2.3 Feature Extraction

To analyze glomerular health and predict kidney transplant outcomes, we leveraged HIPT (Hierarchical Image Pyramidal Transformer)[15] trained on TCGA dataset and ViT (Vision Transformer) trained on Image-Net dataset for deep feature extraction and representation. The HIPT model enabled efficient hierarchical processing of WSIs, allowing us to extract relevant features at multiple scales, which is crucial to capture both local and global patterns in complex tissue structures.

For feature extraction, we first used the ViT model[16](specifically, the pretrained "vit-base-patch16-224-in21k") image to obtain embeddings from segmented glomeruli. This Vision Transformer operates by dividing the glomerular image into fixed-size patches, transforming each patch into a linear embedding space, and employing self-attention mechanisms to capture intricate contextual relationships within and across patches. The resulting embeddings capture essential structural and contextual features of the glomeruli, including patterns indicative of disease progression and transplant health.

## 2.4 Classification Model

These embeddings were then utilized to train an Artificial Neural Network (ANN) model for patient outcome prediction. The dataset was split based on unique patient IDs, ensuring no overlap of patient data in training and testing sets, which mitigates the risk of data leakage and preserves model generalizability. The ANN model architecture included an input layer, two hidden layers with ReLU activations, and a sigmoid-activated output layer for binary classification. This network was trained on standardized feature data to classify patients based on the percentage change in eGFR from 3 months to 12 months post-transplant, an important metric for assessing kidney function over time.

$$Percent\ Decline\ in\ eGFR = \frac{\mathrm{eGFR}_{12\ \mathrm{months}} - \mathrm{eGFR}_{3\ \mathrm{months}}}{\mathrm{eGFR}_{12\ \mathrm{months}}} \tag{1}$$

We divided our percent decline in eGFR into 2 classes. Class 0 if there is a negative decline in eGFR and Class 1 if eGFR remains stable or increases. To derive patient-level outcomes from glomerulus-level predictions, we used a majority voting strategy. For each patient, the final prediction was determined by the most frequent glomerulus-level prediction, enhancing robustness by averaging out individual prediction variances at the glomerulus-level. This method as depicted in ***Fig*** 2, improves reliability by minimizing the impact of isolated prediction errors,

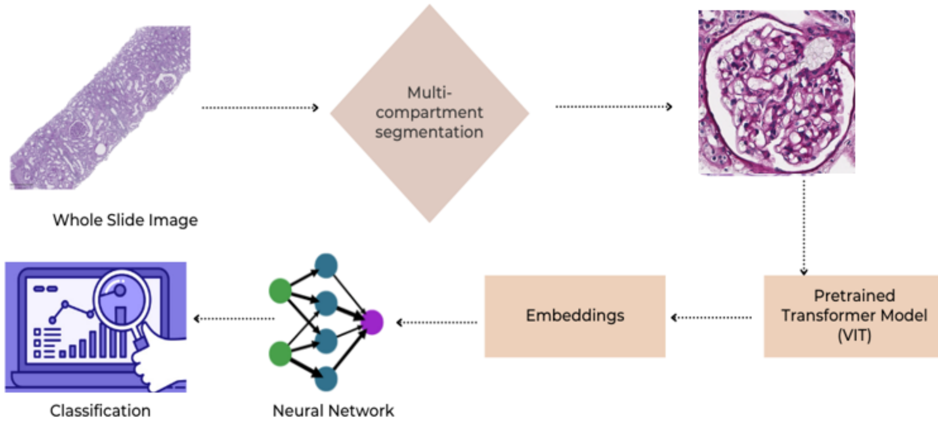


Figure 2. **Workflow for analyzing Whole Slide Images (WSIs).** The pipeline begins with multi-compartment segmentation of WSIs to isolate Glomerulus regions, followed by feature extraction using a pretrained Vision Transformer (ViT) model. The extracted embeddings are then processed by a neural network for classification tasks.

resulting in a more stable patient-level classification. The model's performance was evaluated using key metrics: accuracy, precision, recall, and F1 score, offering a comprehensive assessment of its predictive effectiveness for patient outcomes post-transplant. This pipeline, combining HIPT and ViT embeddings with ANN classification, demonstrates a powerful approach for leveraging advanced deep learning techniques in digital pathology to support clinical decision-making in transplant medicine.

## 3. RESULTS

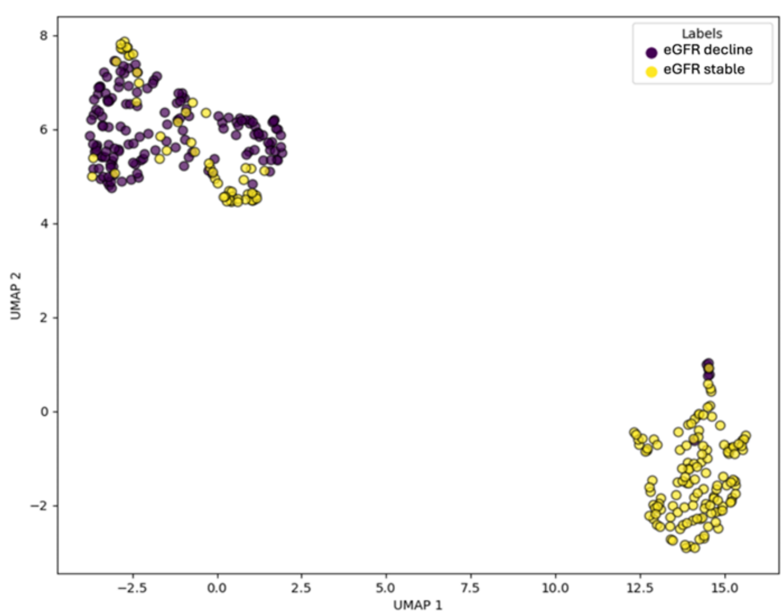


Figure 3. UMAP visualization of embeddings from a Vision Transformer (ViT) of each glomerulus.

We achieved an accuracy of 0.73 in predicting eGFR changes from time-zero biopsies, which is significant given the challenges of early graft assessment. This model demonstrates the potential to provide early insights into kidney function, which can guide immediate clinical decisions and personalized treatments. ***Fig*** 3 depicts

UMAP visualization which reveals a clear separation between eGFR decline and stability, indicating that the embeddings capture meaningful predictive features. Future work should focus on expanding the dataset, refining the model with additional features, and conducting longitudinal studies to enhance accuracy and generalizability, thus improving early intervention and patient management.

| | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| ViT embedding | 0.73 | 0.80 | 0.67 | 0.73 |

Table 1. This table summarizes the performance of the trained model, presenting key evaluation metrics including Accuracy, Precision, and Recall on test set.

## 4. CONCLUSION

Predicting kidney function from time-zero biopsies offers significant benefits for patient care. Early detection of potential graft issues allows for timely, personalized interventions, potentially improving long-term transplant outcomes and reducing the need for additional invasive follow-up biopsies. This proactive approach not only minimizes patient discomfort but also supports better clinical decision-making and optimizes treatment strategies.

To enhance this work, we will consider expanding the dataset to include a broader and more diverse patient population, which will improve model generalizability. Incorporate additional clinical and genetic data to enrich the assessment of graft health. Refining the model by exploring advanced neural network architectures and optimizing hyperparameters could further enhance predictive accuracy. Conducting longitudinal validation will ensure the model's robustness over time. Lastly, integrating the model into clinical workflows and seeking clinician feedback will help refine its practical application and effectiveness in patient care.

## 5. ACKNOWLEDGMENT

This work was supported by the National Institutes of Health [NIH/NIDDK R21 DK11229541].

## REFERENCES

[1] Kher, V., “End-stage renal disease in developing countries,” *Kidney international* **62**(1), 350–362 (2002).

[2] Administration, B.-H., “Biden-harris administration acts to improve access to kidney transplants,” (May 2024).

[3] Rao, P. S., Schaubel, D. E., Guidinger, M. K., Andreoni, K. A., Wolfe, R. A., Merion, R. M., Port, F. K., and Sung, R. S., “A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index,” *Transplantation* **88**(2), 231–236 (2009).

[4] Williams, W. W., Taheri, D., Tolkoff-Rubin, N., and Colvin, R. B., “Clinical role of the renal transplant biopsy,” *Nature Reviews Nephrology* **8**(2), 110–121 (2012).

[5] Racusen, L. C., Solez, K., Colvin, R. B., Bonsib, S. M., Castro, M. C., Cavallo, T., Croker, B. P., Demetris, A. J., Drachenberg, C. B., Fogo, A. B., et al., “The banff 97 working classification of renal allograft pathology,” *Kidney international* **55**(2), 713–723 (1999).

[6] Solez, K., Colvin, R., Racusen, L., Haas, M., Sis, B., Mengel, M., Halloran, P., Baldwin, W., Banfi, G., Collins, A., et al., “Banff 07 classification of renal allograft pathology: updates and future directions,” *American journal of transplantation* **8**(4), 753–760 (2008).

[7] Loupy, A., Haas, M., Solez, K., Racusen, L., Glotz, D., Seron, D., Nankivell, B., Colvin, R., Afrouzian, M., Akalin, E., et al., “The banff 2015 kidney meeting report: current challenges in rejection classification and prospects for adopting molecular pathology,” *American journal of transplantation* **17**(1), 28–41 (2017).

[8] Roufosse, C., Simmonds, N., Clahsen-van Groningen, M., Haas, M., Henriksen, K. J., Horsfield, C., Loupy, A., Mengel, M., Perkowska-Ptasińska, A., Rabant, M., et al., “A 2018 reference guide to the banff classification of renal allograft pathology,” *Transplantation* **102**(11), 1795–1814 (2018).

[9] Munivenkatappa, R. B., Schweitzer, E., Papadimitriou, J., Drachenberg, C., Thom, K., Perencevich, E., Haririan, A., Rasetto, F., Cooper, M., Campos, L., et al., “The maryland aggregate pathology index: a deceased donor kidney biopsy scoring system for predicting graft failure,” *American Journal of Transplantation* **8**(11), 2316–2324 (2008).

[10] Remuzzi, G., Cravedi, P., Perna, A., Dimitrov, B. D., Turturro, M., Locatelli, G., Rigotti, P., Baldan, N., Beatini, M., Valente, U., et al., “Long-term outcome of renal transplantation from older donors,” *New England Journal of Medicine* **354**(4), 343–352 (2006).

[11] Mimar, S., Paul, A. S., Lucarelli, N., Border, S., Santo, B. A., Naglah, A., Barisoni, L., Hodgin, J., Rosenberg, A. Z., Clapp, W., et al., “Compreps: An automated cloud-based image analysis tool to democratize ai in digital pathology,” *bioRxiv* (2024).

[12] Dosovitskiy, A., “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929* (2020).

[13] Zupan, J., “Introduction to artificial neural network (ann) methods: what they are and how to use them,” *Acta Chimica Slovenica* **41**(3), 327 (1994).

[14] McInnes, L., Healy, J., and Melville, J., “Umap: Uniform manifold approximation and projection for dimension reduction,” *arXiv preprint arXiv:1802.03426* (2018).

[15] Chen, R. J., Chen, C., Li, Y., Chen, T. Y., Trister, A. D., Krishnan, R. G., and Mahmood, F., “Scaling vision transformers to gigapixel images via hierarchical self-supervised learning,” in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 16144–16155 (2022).

[16] Alexey, D., “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv: 2010.11929* (2020).