



# Multimodal AI-based Assessment of Renal Allograft Biopsies

Sharifa Sahai, Richard J. Chen, Jana Lipova, Kuan Chen, Tiffany Chen, Judy J. Wang, Drew Williamson, Ming Y. Lu, Astrid Weins, Faisal Mahmood sharifasahai@g.harvard.edu - faisalmahmood@bwh.harvard.edu www.mahmoodlab.org







### Motivation

- >There are over 100,000 kidney transplants annually with over 24,000 transplants in the United States alone [1].
- >Manual assessment of renal biopsies is the standard for the assessment of pretransplant kidneys and post-transplant rejection, which suffers from inter- and intra-observer variability ( $\alpha$ =0.22) [2].
- >Observer variability can have dire consequences ranging from under and over treatment to partial or full transplant rejection or even death.
- >Renal allograft assessment is a complicated process involving multiple tissue stains and several modalities and requires the expertise of renal pathologists. Such expertise is often not available in low resource settings which can result in delays in diagnosis and treatment [3].

### MANTA

### (Multimodal AI for Renal Transplant Assessment)

>We propose MANTA (Multimodal AI for Renal Transplant Assessment) an objective and automated method for assessment of renal allograft biopsies for screening of renal allograft rejection.

>MANTA utilizes weakly supervised deep learning multimodal fusion using gigapixel whole slide images and patients' diagnosis as labels.

>MANTA does not require pixel, patch or ROI-level labels for training.

>MANTA fuses morphological features from H&E, PAS, Masson Trichrome and Jones Silver stains to get holistic predictive results for Active Cell Mediated Rejection (TCMR), Antibody Meditated Rejection (AMR) and Interstitial Fibrosis and Tubular Atrophy (IFTA).

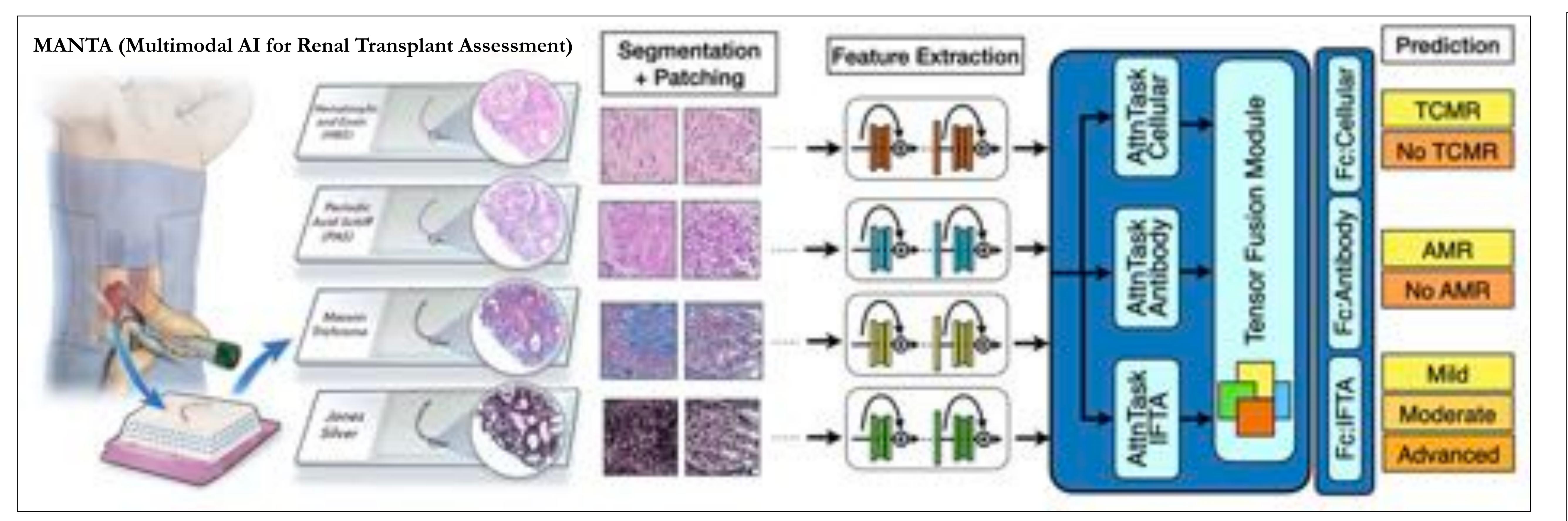
Antibody Mediated Rejection	75% Non AMR	25% AMR	
T Cell Mediated Rejection	71% Non TCMR	29% TCMR	
Interstitial Fibrosis and Tubular Atrophy	60% Mild IFTA	22% Moderate IFTA	18% Advanced IFTA

Distribution of renal allograft biopsy labels from 2013-2021 used in this study. (n = 11,974WSIs, N = 909 patients). Each patient has 6-18 slides. Assessment is based on the Banff Classification of Renal Allograft Pathology.

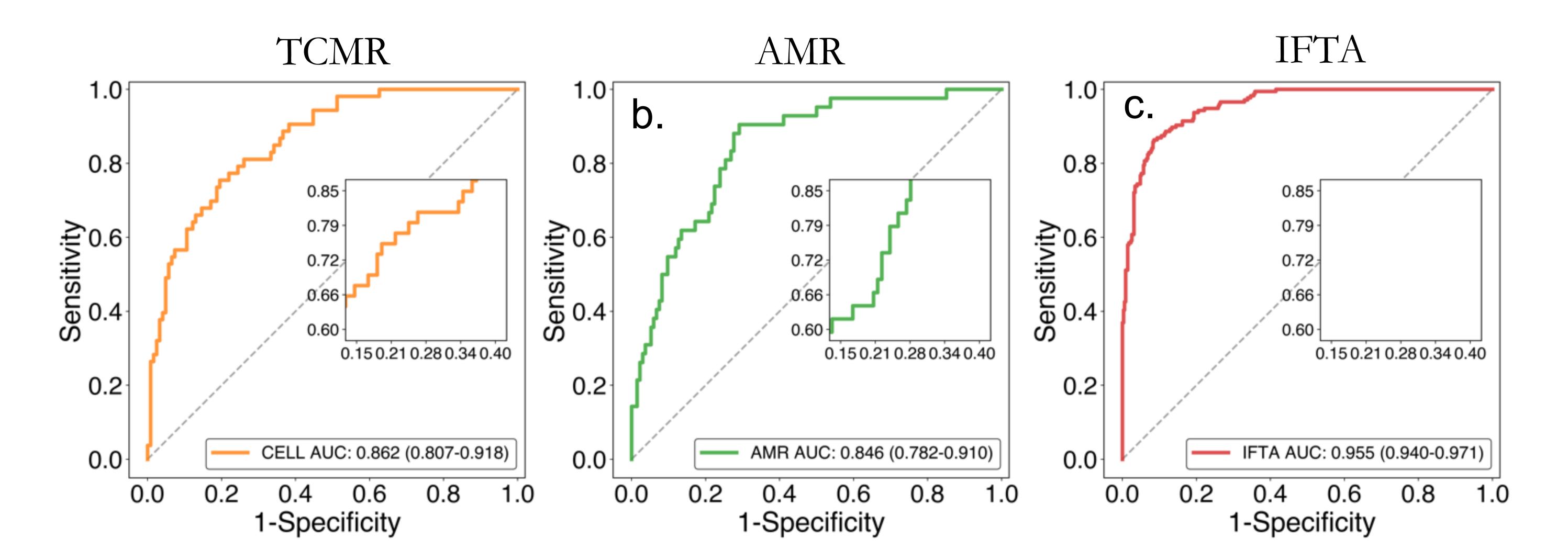
### References

[1]Global. (n.d.). Retrieved December 03, 2020, from <a href="http://www.transplant-observatory.org/global/">http://www.transplant-observatory.org/global/</a> [2] Furness, P.N., Taub, N. and Convergence of European Renal Transplant Pathology Assessment Procedures (CERTPAP) Project, 2001. International variation in the interpretation of renal transplant biopsies: report of the CERTPAP Project. Kidney international, 60(5), pp.1998-2012.

[3] Bijol, V., Farag, Y.M., Harris, D.C., Levin, A., Roy-Chaudhury, P., Singh, A.K. and Jhaveri, K.D., 2019. Renal pathology practice globally: identifying needs and meeting the challenge. Kidney international, 96(2), pp.258-261.



### Model Performance



- >We achieved a AUC-ROC of .86 for TCMR, .84 for AMR and .95 for IFTA.
- >The dataset was partitioned into 70/10/20% splits for training, validation (i.e. model tuning) and held out testing, respectively.
- >The model's training and evaluation was performed on the patient level using tensor fusion on multiple slide modalities.. 5-fold cross validation was used, and the best model was selected based on validation performance.

## Interpretability

>The multiple instance learning (MIL) module assigns attention scores to each tissue patch through ranking the importance of each patch corresponding to the label provided for the slide.

>The attention scores can be translated to WSI attention heatmaps reflecting relevance of each biopsy region towards the model predictions, which can be used for validation and interpretation by pathologists.

### Conclusion and Future Directions

Our results are promising with regards to all three tasks of Active Cell Mediated Rejection, Antibody Mediated Rejection and Interstitial Fibrosis and Tubular Atrophy classification and interpretability.

Heatmaps derived from the patch attention scores were shown to pathologists who confirmed that the high attention regions corresponded to clinically relevant regions.

>We will continue to consult with renal pathologists to rate the accuracy of our model and conduct a reader study.

>We can also extend the model architecture to the Banff subcategories and create other predictive labels. Additionally, causes which are not related to chronic or acute rejection, could also be incorporated.

>We believe that once deployed this model could be a great asset to both clinical and technical methods driven research.

# Attention Based Heatmap Visualization

Attention based heatmaps on H&E, PAS, Masson Trichrome and Jones Silver stains for a Mild IFTA case based on the Banff Classification for Renal Allograft. All slides are from the same patient biopsy. The greatest attention is on areas where there are chronic inflammatory cells, especially where there's damage to the renal tubules and infiltration of the glomeruli.