Multimodal AI-based Assessment of Renal Allograft Biopsies


sharifasahai@g.harvard.edu - faisalmahmood@bwh.harvard.edu

www.mahmoodlab.org

Abstract

According to the global observatory of transplantation and donation there are over 100,000 kidney transplants annually, with over 24,000 transplants in the United States alone [2]. The standard of care for the assessment of pre-transplant kidneys and post-transplant rejection is the manual assessment of renal biopsies. However, such manual assessment suffers from large inter- and intra-observer variability (κ=0.22) [2]. Such variability can have dire consequences ranging from under and over treatment to partial or full transplant rejection or even death. Moreover, renal allograft assessment is a complex process involving multiple tissue states 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]. Here, we propose MANTA (Multimodal AI for Renal Transplant Assessment) an objective and automated method for assessment of renal allograft biopsies using screening of post-allograft rejection. MANTA utilizes weakly supervised deep learning multimodal fusion using gigapixel whole slide images and patient’s diagnosis as labels. MANTA does not require path or ROI-level labels for training. MANTA uses morphological features from H&E, PAS, Masson Trichrome and Jones Silver stains to get holistic predictive results for acute cell mediated rejection, antibody mediated rejection and interstitial fibrosis and tubular atrophy (IFTA). We achieved AUC ROC of 0.86 for TCMR, 0.84 for AMR and 0.84 for IFTA. Heatmaps derived from the patch attention scores were shown to pathologists who confirmed that the high attention regions corresponded to clinically relevant regions. We believe that once deployed this model could be a great asset to both clinical and technical methods driven research.

Materials and Methods

We propose a multimodal AI based assessment of renal allograft biopsies for screening of post-allograft kidneys for renal allograft rejection. We utilize multimodal fusion to account for information from Hematoxylin and eosin (H&E), Periodic acid Schiff (PAS), Masson trichrome and Jones silver stains data images (WSI) for overall transplant assessment. 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 parameters used to determine how much attention is shared to the classification tasks which determine evidence of rejection by antibody mediated, cell mediated and IFTA. The attention score 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 to pathologists.

We used a dataset of 109 cases (11,374 WSI’s) from 2013-2021. Each patient had between 6-18 slides. Assessment is based on the Banff Classification of Renal Allograft Pathology. the model architecture is set up for evidence of antibody mediated rejection (AMR), T cell mediated rejection (TCMR) and interstitial Fibrosis and Tubular Atrophy (IFTA) as labels. Of the 109 cases, 28% had evidence of TCMR, 28% had evidence of AMR and 7% had evidence of IFTA. For IFTA assessment, 40% had medium to high evidence of IFTA and 18% had absent IFTA. The dataset was partitioned into 70/20/10 split for training, validation and testing, respectively. The model’s training was performed on each slide level with each slide as an independent data point and it was evaluated at the patient level. 5-fold cross-validation was used, and the best model was selected based on validation performance.

Conclusion

Our results are promising with regards to all three tasks of active cell mediated reaction, antibody mediated rejection and interstitial fibrosis and tubular atrophy classification and interpretability. Heatmaps derived from the patch attention scores were shared to the classification tasks which determine evidence of rejection by antibody mediated, cell mediated and IFTA. The attention score 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 to pathologists. Additionally, Banff category 6, 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.

Weakly-Supervised, Multi-Task, Deep-Learning Model Workflow

Model Performance

Our results are promising with regards to all three tasks of acute cell mediated rejection, antibody mediated rejection and interstitial fibrosis and tubular atrophy classification and interpretability. Heatmaps derived from the patch attention scores were shared to the classification tasks which determine evidence of rejection by antibody mediated, cell mediated and IFTA. The attention score 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 to pathologists.

References


http://www.transplant-observatory.org/global/

