Development of a multi-institutional kidney biopsy report registry via a natural language processing pipeline

Rodrigo Azuero-Dajud[1], Vitaly Lorman[1], Amy Goodwin Davies[1], Mohan Kashyap Pargi[1], Laura S. Finn[1,2], Rebecca Scobell[1], Adya Maddox[1], Qiwei Shen[1], Aliyah Jones[1], Grace Park[1], Pete Camacho[1], Hanieh Razzaghi[1], L. Charles Bailey[1,2], Michelle Denburg[1,2]
[1]Children's Hospital of Philadelphia, Philadelphia, PA;
[2]Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA

Background

PEDSnet is a national, multi-institutional network of 10 children's hospitals, created to improve the conduct of clinical research [1]. PEDSnet maintains an electronic health record (EHR) database which has been utilized for multiple nephrology studies (e.g., [2]-[8]). The database primarily consists of structured data such as diagnosis codes, procedure codes, and lab results. Histopathological features are typically contained within free text reports and are more difficult to extract from structured data. The traditional process for extracting these features from biopsy reports involves medical chart reviews performed by subject matter experts. This is an expensive process that can take months to complete. In the case of kidney biopsy reports, we aim to improve this process by developing a natural language processing (NLP) pipeline to automate the extraction of histopathological concepts from free text data. This includes concepts such as kidney-related disease presence or antibody/cellular mediated rejection status. The extracted features will be represented as structured data (e.g. the OMOP NOTE_NLP table) in a multi-institutional kidney biopsy registry. This registry can be queried for use in future nephrology EHR research applications. We report on ongoing pilot work at CHOP and future directions as we expand across institutions.

Methods

The pipeline is broken down into 4 components detailed in Figure 1

1. *Kidney biopsy report extraction:* Extract kidney biopsy reports from general biopsy reports by applying specific filtering and exclusion criteria.

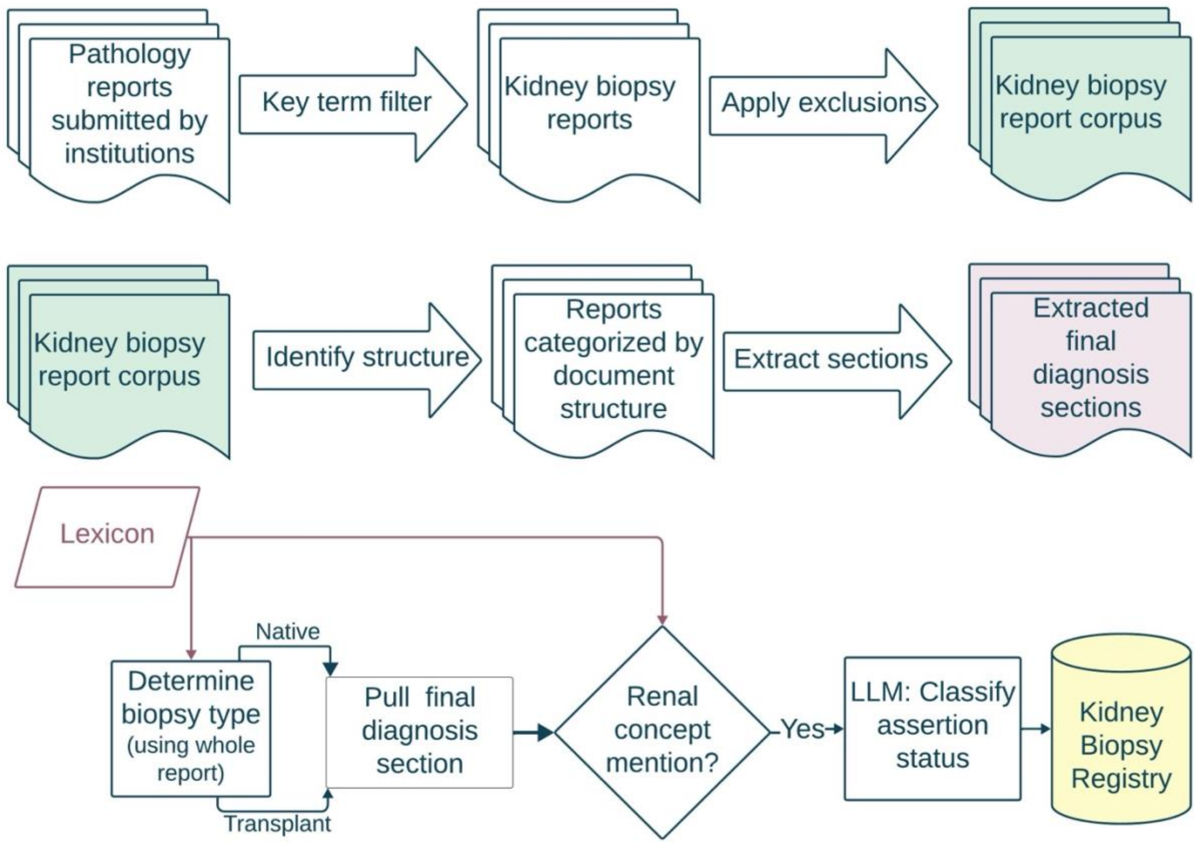
2. *Biopsy type classification:* Reports are categorized into transplant or native biopsies using a combination of pattern matching and rulesets

3. *Section identification in kidney biopsy reports*: Employing regex-based pattern matching techniques, different sections like the final diagnosis and comments are identified and extracted from classified native, transplant kidney biopsy reports.

4. *Assertion status extraction using a Large Language Model (LLM):* A lexicon of renal concepts is used with regex-based methods to identify relevant terms within the final diagnosis section. Subsequently, LLM queries are triggered to determine each concept's assertion status (positive/negative for native, positive/negative/borderline for transplant).

The reports themselves are stored in a PEDSnet database that follow the OMOP note table format and are stored on local servers.

Figure 1: Current NLP Pipeline Architecture



Results

An initial round of validation was completed for both native and transplant reports at 1 institution. 260 reports in total were reviewed by nephrology and pathology clinical subject matter experts. Satisfactory performance was observed for all renal concepts across both native and transplant condition concepts. The first version of the pipeline utilized rulesets based on report patterns to perform final diagnosis extraction and to perform biopsy type classification. Once this classification was performed, the report text ws then fed into a large language model to assess the assertion status of concepts identified in the final diagnosis section. More recent versions have incorporated using a large language model to perform biopsy type classification, final diagnosis section extraction, as well as improved prompt language. In comparison to the first version of the pipeline, the most recent iteration yields an increase across all performance metric values (Table 1).

Common issues found were inconsistencies from the LLM in handling ambiguous language such as ‘consistent with’, ‘suggestive of’, and similar vague language.

Additionally, there are limitations to performance due to variations in report structures across sites or report types. Further validation is planned to evaluate pipeline performance on reports from other sites, and future modifications to LLM prompt language will be made to account for current performance deficits.

Table 1: Pipeline Metric Comparison Outcome from Pipeline Versions 1.1 to 1.7

Sensitivity	**Specificity**	**PPV**	**NPV**	**F1**
Performance across all renal concepts for both report types: V1.1				
81.5%	98.2%	85.3%	97.7%	0.830
Performance across all renal concepts for both report types: V1.7				
84.5%	98.3%	87.6%	97.9%	0.860

Conclusion

Once a pipeline with satisfactory performance is developed and validated across multiple institutions, it can be run each data refresh to maintain an up-to-date multi-institutional registry across participating PEDSnet institutions. It is anticipated that the registry will contain renal concepts for up to 10,000 biopsy reports. Additionally, the standardized format of the OMOP Note NLP table would be an excellent resource to utilize for pipeline output of the biopsy reports. The registry should increase research efficiency and reduce costs for future studies which require histopathologic information from kidney biopsy reports by avoiding expensive chart review and automating the extraction process with a comparable level of reliability. Additionally, the registry would be able to link histopathological features from the unstructured data to the structured EHR data, which may make studies feasible that weren't previously. The registry would be a resource for nephrology research studies to enrich analysis of the structured data, minimize the necessity of costly chart review, as well as provide researchers access to histopathological features not captured in the structured data.

References

1. PEDSnet: A Pediatric Learning Health System. Accessed January 29, 2025. https://PEDSnet.org/

2. Denburg MR, Razzaghi H, Bailey LC, et al. Using Electronic Health Record Data to Rapidly Identify Children with Glomerular Disease for Clinical Research. J Am Soc Nephrol. 2019;30(12):2427-2435. doi:10.1681/ASN.2019040365

3. Goodwin Davies AJ, Xiao R, Razzaghi H, et al. Skeletal Outcomes in Children and Young Adults with Glomerular Disease. J Am Soc Nephrol. 2022;33(12):2233-2246. doi:10.1681/ASN.2021101372

4. Wenderfer SE, Chang JC, Goodwin Davies A, et al. Using a Multi-Institutional Pediatric Learning Health System to Identify Systemic Lupus Erythematosus and Lupus Nephritis: Development and Validation of Computable Phenotypes. Clin J Am Soc Nephrol. 2022;17(1):65-74. doi:10.2215/CJN.07810621

5. Gluck CA, Forrest CB, Davies AG, et al. Evaluating Kidney Function Decline in Children with Chronic Kidney Disease Using a Multi-Institutional Electronic Health Record Database. Clin J Am Soc Nephrol. 2023;18(2):173-182. doi:10.2215/CJN.0000000000000051

6. Mitsnefes MM, Maltenfort M, Denburg MR, et al. Derivation of paediatric blood pressure percentiles from electronic health records. EBioMedicine. 2023;98:104885. doi:10.1016/j.ebiom.2023.104885

7. Stone HK, Mitsnefes M, Dickinson K, et al. Clinical course and management of children with IgA vasculitis with nephritis. Pediatr Nephrol. 2023;38(11):3721-3733. doi:10.1007/s00467-023-06023-8

8. Dharnidharka VR, Scobell RR, Kallash M, et al. Clinical characteristics and favorable treatment responses of recurrent focal segmental glomerulosclerosis or steroid-resistant nephrotic syndrome in children after kidney transplantation. Pediatr Nephrol. 2024;39(11):3317-3331. doi:10.1007/s00467-024-06452-z