

---

# Towards Multimodal Representation Learning in Paediatric Kidney Disease

---

**Ana Durica**

Institute of Health Informatics  
University College London  
London, United Kingdom  
ana.durica.21@ucl.ac.uk

**John Booth**

Data Research, Innovation and Virtual Environments Unit  
Great Ormond Street Hospital  
London, United Kingdom  
john.booth@gosh.nhs.uk

**Ivana Drobniak**

Department of Computer Science  
University College London  
London, United Kingdom  
i.drobniak@ucl.ac.uk

## Abstract

Paediatric kidney disease varies widely in its presentation and progression, which calls for continuous monitoring of renal function. Using electronic health records collected between 2019 and 2025 at Great Ormond Street Hospital, a leading UK paediatric hospital, we explored a temporal modelling approach that integrates longitudinal laboratory sequences with demographic information. A recurrent neural model trained on these data was used to predict whether a child would record an abnormal serum creatinine value within the following thirty days. Framed as a pilot study, this work provides an initial demonstration that simple temporal representations can capture useful patterns in routine paediatric data and lays the groundwork for future multimodal extensions using additional clinical signals and more detailed renal outcomes.

## 1 Introduction

Paediatric kidney disease varies widely in its clinical presentation and rate of progression [2], which requires continuous monitoring of renal function to detect early deterioration. Serum creatinine, the standard marker of kidney function, reflects the glomerular filtration rate and underpins diagnosis and staging [5, 6]. Anticipating short-term changes in creatinine could support several paediatric nephrology decisions, including early identification of kidney injury, assessment of treatment response, and monitoring of disease progression.

Great Ormond Street Hospital (GOSH) is a leading paediatric centre with an advanced electronic health record (EHR) system and a trusted digital research environment [3]. The dataset used in this study contains multi-year clinical and biochemical information, including demographics, laboratory results, hospital episodes and procedures, all linked at the patient level. Representation learning provides a way to combine such heterogeneous inputs into patient embeddings that capture temporal and cross-modal structure. Recurrent neural networks (RNNs) have shown promising performance in adult EHR tasks, including chronic kidney disease progression and heart-failure prediction [4, 7], but similar work in paediatric nephrology is limited.

In this study, we develop a simple pipeline to organise EHR data into a format suitable for RNNs. As an initial prediction task, we focus on whether a patient will record an abnormal creatinine

value within the following 30 days, using only two modalities: routine laboratory tests and basic demographic information. This work serves as an initial step toward multimodal representation learning in paediatric kidney disease, with the longer-term aim of incorporating additional data sources and more clinically meaningful renal outcomes.

## 2 Methodology

This study developed a pipeline for combining longitudinal laboratory data and demographic features to predict whether a patient would record an abnormal creatinine value within the following 30 days. The workflow included: (1) cohort selection and outcome definition, (2) temporal structuring of longitudinal records, and (3) sequence modelling with an RNN and evaluation of the learned representations.

### 2.1 Cohort selection and outcome definition

Data were extracted from the EHR for all patients with at least one nephrology episode between 2019 and 2025 ( $n = 1,032$ ), of whom 38 (3.5%) died during follow-up. Follow-up was defined from the first to the last recorded serum creatinine, or to the date of death if applicable. To form a 30-day prediction window ending at the last follow-up time ( $t_{\text{end}}$ ), patients were included if they (1) had at least three creatinine measurements on separate days before the window and (2) if deceased, had at least one measurement within the final 30 days of life. These criteria yielded a final cohort of  $n = 826$ , including 17 patients (2.1%) who died but had a valid measurement within the window. The prediction target was the occurrence of any abnormal serum creatinine value within the 30-day window. Abnormality status was taken directly from EHR laboratory flags without additional processing. Patients with no abnormal values were labelled as 0 ( $n = 370$ , 44.8%), and those with one or more abnormal values as 1 ( $n = 456$ , 55.2%).

### 2.2 Data extraction and preprocessing

For the selected cohort, a longitudinal dataset was created by identifying all valid creatinine measurements prior to the 30-day prediction window and using their dates as event time points. At each time point, all laboratory tests performed on the same calendar day were extracted, and their presence and abnormality status were recorded. Fifteen clinically relevant laboratory markers were included, each represented by two binary indicators (presence and abnormality), giving 30 features, together with age and sex as static variables. Absent tests were encoded as presence = 0 and abnormal = 0, while performed tests had presence = 1 with the abnormality flag reflecting test status. This produced temporally ordered sequences of laboratory information for each patient, forming the input to the time-series model.

### 2.3 Sequence modelling

Given a sequence of patient records  $X = (x_1, x_2, \dots, x_n)$ , where  $x_t$  is the multi-hot laboratory feature vector at time  $t$ , the model aimed to learn a latent representation  $c$ . Sequences were limited to 100 time points, with shorter sequences left-padded with zeros to preserve temporal order. A gated recurrent unit (GRU) encoder [1] processed  $X$  to capture temporal dependencies, and the resulting embedding  $c$  was concatenated with demographic features and passed to a linear classifier to predict 30-day outcomes. The model was trained using a stratified 70/10/20 train-validation-test split to maintain outcome balance.

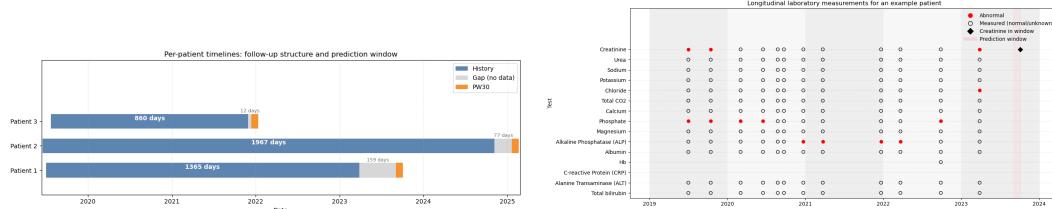
### 2.4 Evaluation

Model performance was evaluated on the held-out 20% test set. Discrimination was assessed using the area under the ROC curve (AUC), with 95% confidence intervals estimated from 2,000 bootstrap resamples. Using a 0.5 decision threshold, we computed and displayed the confusion matrix to illustrate classification behaviour. Predicted probabilities were obtained from sigmoid-transformed logits, and test-set embeddings were visualised using t-SNE.

### 3 Results

We present the cohort’s temporal and laboratory structure, the model’s performance in predicting abnormal creatinine within the following 30 days, and an exploratory visualisation of the learned embeddings.

#### 3.1 Cohort and temporal characteristics



(a) Follow-up timelines for example patients. Blue bars show the available pre-window history, grey indicates periods without recorded laboratory data, and orange marks the fixed 30-day prediction window. Timelines are shown in calendar time, corresponding to each patient’s recorded laboratory events.

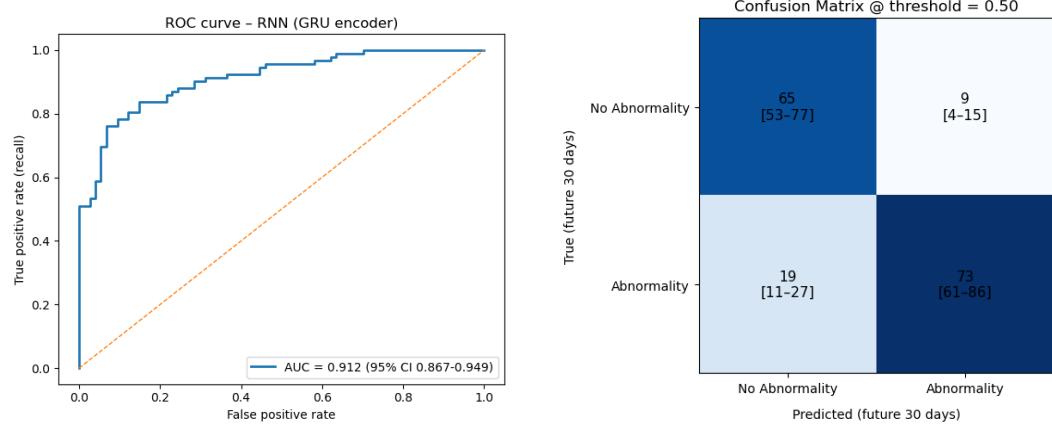
(b) Example patient’s laboratory measurements over time. Red circles indicate abnormal results, open circles denote measured tests with non-abnormal values, and missing markers indicate that a test was not performed at that event time point. Alternating grey bands show calendar years, and the pink region marks the 30-day prediction window.

Figure 1: Overview of longitudinal data used for temporal modelling.

Figure 1 illustrates the cohort’s longitudinal patterns: follow-up durations varied substantially across patients, and laboratory measurements were recorded at irregular intervals with occasional data gaps. Consequently, the 30-day prediction window frequently followed periods without recent laboratory observations, mirroring real clinical practice.

#### 3.2 Predictive performance

The GRU-based model was trained to predict abnormal creatinine within the following 30-day window. Figure 2 shows that the model achieved good discrimination, providing preliminary support for this short-term prediction task.



(a) ROC curve for the GRU model on the test set.

(b) Confusion matrix at a 0.5 decision threshold.

Figure 2: Predictive performance of the GRU model. Bootstrapped 95% confidence intervals are shown for both the ROC curve and the confusion-matrix cell counts.

### 3.3 Embedding visualisation

The GRU embeddings from the test set were projected into two dimensions using t-SNE (Figure 3).

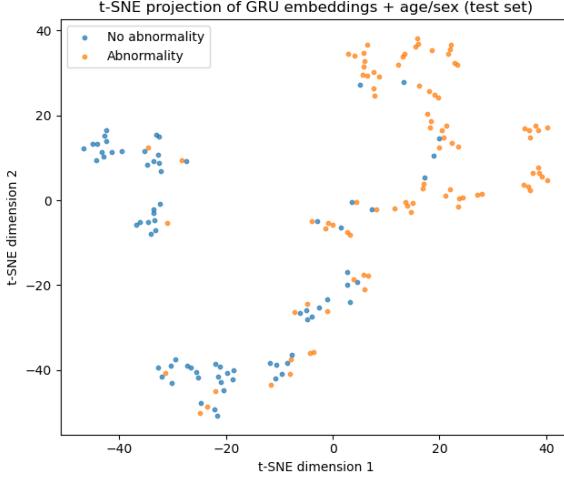


Figure 3: t-SNE projection of GRU embeddings for the test set. Each point represents a patient embedding coloured by 30-day outcome.

This exploratory visualisation shows partially separated regions for patients with and without future abnormal creatinine values, suggesting that the model captured preliminary predictive signal.

## 4 Discussion

This work provides a preliminary demonstration that temporal modelling of longitudinal EHR data can support the near-term prediction of creatinine abnormalities in paediatric patients. Using a GRU encoder applied to laboratory event sequences, together with static demographic features, we achieved good discrimination in predicting whether a patient would have an abnormal creatinine test within 30 days. These findings suggest that recurrent models can extract clinically relevant temporal signals from laboratory histories.

Several simplifications shaped this pilot. Sequences were treated as step-indexed events without explicit timing, despite irregular sampling. Laboratory features were anchored to creatinine measurement dates, blending physiological information with patterns of clinical monitoring. Tests were encoded as binary indicators of presence and abnormality, without quantitative values. These design choices enabled a tractable first implementation but reduce the clinical detail captured in the representations.

Although the current model uses only two input modalities (laboratory data and basic demographics), the underlying GOSH cohort contains a far broader set of signals. Multi-year trajectories include laboratory, medication, diagnosis, vital-sign and procedural data, each with distinct temporal structure. Future work will extend this framework to multimodal and time-aware models, and to more clinically meaningful renal endpoints such as sustained creatinine elevation, acute kidney injury, treatment response or post-transplant outcomes. These directions aim to make fuller use of the depth of paediatric EHR data and support more proactive renal care.

## Acknowledgments and Disclosure of Funding

This research was supported by UKRI (EP/S021612/1) through the UCL Centre for Doctoral Training in AI-enabled Healthcare Systems and by GOSH Children’s Charity Award VS0618. Research at GOSH NHS Foundation Trust and the UCL Great Ormond Street Institute of Child Health is supported by the NIHR GOSH Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

## References

- [1] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014. URL <https://arxiv.org/abs/1412.3555>.
- [2] Luigi Cirillo, Letizia De Chiara, Samantha Innocenti, Carmela Errichiello, Paola Romagnani, and Francesca Becherucci. Chronic kidney disease in children: an update. *Clinical Kidney Journal*, 16(10):1600–1611, 2023. doi: 10.1093/ckj/sfad097.
- [3] Great Ormond Street Hospital for Children NHS Foundation Trust, Data Research, Innovation and Virtual Environments (DRIVE) Unit. 5 years of digital innovation: Impact report, 2023. URL [https://www.goshdrive.com/\\_files/ugd/074fdb\\_46f2b3d1edef4b4ea630e5d07d06f017.pdf](https://www.goshdrive.com/_files/ugd/074fdb_46f2b3d1edef4b4ea630e5d07d06f017.pdf). [Accessed 4 November 2025].
- [4] Liqi Lei, Yangming Zhou, Jie Zhai, Le Zhang, Zhijia Fang, Ping He, and Ju Gao. An effective patient representation learning for time-series prediction tasks based on ehrs. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 885–892, Madrid, Spain, 2018. IEEE. ISBN 978-1-5386-5488-0. doi: 10.1109/BIBM.2018.8621284. URL <https://doi.org/10.1109/BIBM.2018.8621284>.
- [5] Adeera Levin, Sofia B. Ahmed, Juan Jesus Carrero, Bethany Foster, Anna Francis, Rasheeda K. Hall, Will G. Herrington, Guy Hill, Lesley A. Inker, Rümeysa Kazancioğlu, Edmund Lamb, Peter Lin, Magdalena Madero, Natasha McIntyre, Kelly Morrow, Glenda Roberts, Dharshana Sabanayagam, Elke Schaeffner, Michael Shlipak, Rukshana Shroff, Navdeep Tangri, Teerawat Thanachayanont, Ifeoma Ulasi, Germaine Wong, Chih-Wei Yang, Luxia Zhang, Karen A. Robinson, Lisa Wilson, Renee F. Wilson, Bertram L. Kasiske, Michael Cheung, Amy Earley, and Paul E. Stevens. Executive summary of the kdigo 2024 clinical practice guideline for the evaluation and management of chronic kidney disease: known knowns and known unknowns. *Kidney International*, 105(4S):684–701, 2024. doi: 10.1016/j.kint.2023.10.016. URL <https://doi.org/10.1016/j.kint.2023.10.016>. Published by Kidney Disease: Improving Global Outcomes (KDIGO).
- [6] George J. Schwartz, Alvaro Muñoz, Michael F. Schneider, Robert H. Mak, Frederick Kaskel, Bradley A. Warady, and Susan L. Furth. New equations to estimate gfr in children with ckd. *Journal of the American Society of Nephrology*, 20(3):629–637, 2009. ISSN 1046-6673. doi: 10.1681/ASN.2008030287. URL <https://doi.org/10.1681/ASN.2008030287>.
- [7] Yitan Zhu, Dehua Bi, Milda Saunders, and Yuan Ji. Prediction of chronic kidney disease progression using recurrent neural network and electronic health records. *Scientific Reports*, 13(1):22091, 2023. doi: 10.1038/s41598-023-49271-2. URL <https://doi.org/10.1038/s41598-023-49271-2>. Open Access under Creative Commons Attribution 4.0 International License.