

ORIGINAL RESEARCH

Open access

Graph Convolutional Network with Attention for Predicting Chronic Kidney Disease Progression Using Longitudinal Laboratory Values, Medications, and Comorbidity Networks

Carlos Ramirez^{1*}, Elena Torres², Pablo Ortega¹, Sofia Mendes³

Abstract

Chronic kidney disease (CKD) affects 10–15% of adults worldwide and often progresses silently toward kidney failure requiring dialysis or transplantation. Monitoring longitudinal markers such as estimated glomerular filtration rate, creatinine, and albuminuria is essential for early intervention and delaying disease progression. However, current predictive models typically rely on static or isolated clinical features, limiting their ability to capture dynamic interactions between laboratory trends, medications, and comorbidities, which leads to incomplete risk assessment. To address this limitation, a conceptual framework based on a graph convolutional network with attention mechanisms is proposed to integrate longitudinal laboratory data, medication networks, and comorbidity structures for CKD progression prediction. Patient records from 2017–2023 are represented as a heterogeneous graph, where nodes include laboratory values, drugs, and diagnoses, and edges encode clinical and pharmacological relationships. Graph convolutional layers capture relational patterns, while attention mechanisms highlight the most clinically relevant interactions, enabling more informative patient-level representations for risk prediction across CKD stages. This approach improves interpretability by revealing which laboratory trends, medications, and comorbidities most influence predicted outcomes, aligning model behavior with clinical nephrology knowledge. Overall, the framework provides a unified and scalable strategy for more accurate and interpretable CKD progression risk prediction by leveraging relational and temporal data structures that traditional models fail to exploit.

Keywords Attention mechanisms, Graph convolutional networks, Chronic kidney disease progression, Longitudinal laboratory values, Medication networks, Comorbidity networks

*Correspondence:

Carlos Ramirez
carlos.ramirez@gmail.com

¹ Department of Intelligent Healthcare Systems, Faculty of Medicine, University of Barcelona, Barcelona, Spain

² Department of AI Healthcare Analytics, University of Lisbon, Lisbon, Portugal

³ Department of Clinical Informatics and AI, University of Porto, Porto, Portugal

Introduction

Chronic kidney disease is defined by abnormalities of kidney structure or function that persist for more than three months, typically identified through an estimated glomerular filtration rate below 60 mL/min/1.73 m² or albuminuria exceeding 30 mg/g of creatinine. Staging follows KDIGO guidelines that classify the disease into five levels based on

eGFR thresholds and albuminuria categories, providing a standardized foundation for clinical assessment and risk communication. Progression from earlier to advanced stages imposes substantial burdens on patients through increased morbidity, mortality, and healthcare costs worldwide. Longitudinal monitoring remains essential for

timely intervention under these established diagnostic criteria [1, 2].

Major predictors of chronic kidney disease progression encompass the slope of eGFR decline over time, persistent albuminuria, uncontrolled blood pressure, and exposure to renoprotective medications including ACE inhibitors, ARBs, and SGLT2 inhibitors. Comorbid conditions such as diabetes mellitus, hypertension, and cardiovascular disease further accelerate trajectories by amplifying renal stress through interconnected pathophysiological pathways. These elements interact dynamically, where medication efficacy often varies according to the presence and severity of specific comorbidities. Comprehensive modeling must therefore account for such interdependencies to generate accurate progression forecasts [3-5].

A key limitation of current machine learning models lies in their treatment of clinical features as independent variables, which prevents adequate representation of relational structures among laboratory measurements, medication histories, and comorbidity profiles. For instance, the impact of a diuretic on potassium levels may depend on concurrent hypertension or heart failure status, yet most approaches ignore these contextual edges. This independence assumption leads to suboptimal predictive performance and reduced clinical utility in heterogeneous patient populations. Graph-based alternatives are therefore required to overcome these structural shortcomings [6, 7].

This conceptual framework proposes a graph convolutional network enhanced with attention mechanisms that operates on a heterogeneous graph constructed from longitudinal laboratory values, medication networks, and comorbidity networks. The architecture systematically encodes patient data as nodes and relational edges to predict progression risks across CKD stages. Subsequent sections detail the background, graph construction process, network layers, attention integration, and evaluation strategy. This roadmap establishes a principled foundation for advancing AI-driven nephrology decision support [8-10].

The conceptual distinctions between conventional CKD prediction approaches and the proposed graph-based framework are systematically outlined in **Table 1**.

Table 1. Structural Comparison of Conventional CKD Prediction Models and the Proposed Attention-Based Graph Framework

Dimension	Conventional Models	Deep Learning (Non-Graph)	Proposed Graph Attention Framework
Data Representation	Independent variables	Sequential tensors	Heterogeneous graph (multiple node types)
Temporal Modeling	Limited or static	Sequence-based (RNN/LSTM)	Temporal graph with dynamic updates
Feature Interaction	Implicit or ignored	Learned but unstructured	Explicit relational edges (laboratory-medication-comorbidity)
Medication Effects	Isolated variables	Embedded but context-limited	Context-aware via pharmacological edges
Comorbidity Modeling	Binary flags	Feature vectors	Structural network relationships
Interpretability	Moderate (coefficients)	Low (black-box)	High (attention weights)
Scalability	High	Moderate	High with modular graph design
Clinical Alignment	Partial	Variable	Strong (KD alignment, relation modeling)
Handling Longitudinal Data	Aggregated summaries	Sequential modeling	Time-evolving graph structure
Prediction Capability	Static risk scores	Dynamic but feature-limited	Relational temporal prediction

Background

CKD progression prediction

Conventional risk tools for chronic kidney disease progression, such as the kidney failure risk equation and its variants, rely on a limited set of baseline variables to

estimate future outcomes. More recent machine learning efforts have incorporated electronic health record data to improve discrimination, yet they frequently process features in isolation without modeling temporal or relational dependencies. Limitations arise when longitudinal laboratory trajectories or medication changes are simplified into static summaries, reducing sensitivity to subtle progression signals. These shortcomings highlight the need for frameworks that preserve the full complexity of patient histories [1, 6, 11].

Advanced predictive schemes now explore deep learning on longitudinal laboratory sequences to forecast eGFR trajectories and renal failure endpoints. Despite promising conceptual directions, existing models struggle to integrate medication interaction networks or comorbidity clusters that exert multiplicative effects on disease course. The absence of graph representations means that clinically meaningful associations remain underexploited. A unified graph-centric perspective can therefore address these gaps in progression modeling [4, 12, 13].

Graph neural networks in healthcare

Graph convolutional networks enable efficient propagation of information across structured data by updating node representations through neighborhood aggregation, as originally formalized for semi-supervised learning tasks. In healthcare applications, GCN variants have been adapted to model disease prediction from imaging or molecular graphs, demonstrating their capacity to capture higher-order relationships. Extensions such as GraphSAGE and graph attention networks further support inductive learning on heterogeneous medical datasets. These architectures provide a natural fit for encoding patient-level interactions in chronic conditions [8, 14, 15].

Applications of graph neural networks in healthcare extend to drug interaction modeling and comorbidity network construction, where nodes represent entities like medications or diagnoses and edges reflect documented associations. Longitudinal extensions allow graphs to evolve with new laboratory results or treatment adjustments, supporting dynamic disease progression analysis. Such models have shown conceptual promise in integrating multi-modal data sources for outcome prediction. The healthcare domain thus benefits from graph methods that respect the relational nature of clinical information [5, 16, 17].

Attention mechanisms

Attention mechanisms in graph neural networks assign learnable coefficients to neighboring nodes, allowing the model to focus on the most relevant connections during message passing. This selective weighting improves representation quality by emphasizing clinically salient relationships while down-weighting noise in large graphs. Graph attention networks and their variants have been applied to disease prediction tasks where interpretability of node importance is required. The mechanism thereby enhances both accuracy and transparency in healthcare modeling [9, 10].

Task-level attention further refines predictions by weighting contributions from different node types such as laboratory, medication, or comorbidity nodes within the overall graph. This hierarchical focus aligns model decisions with domain knowledge, enabling clinicians to inspect which factors drive risk scores. Self-attention across time steps can also capture evolving importance in longitudinal sequences. Overall, attention integration promotes interpretable graph-based reasoning in complex medical prediction scenarios [9, 10].

Framework Overview

High-level architecture

The high-level architecture begins with longitudinal electronic health record data that are transformed into a heterogeneous graph structure encompassing laboratory nodes, medication nodes, and comorbidity nodes. Edges connect these entities according to pharmacological effects, temporal proximity, and clinical associations derived from established medical knowledge. Graph convolutional layers then propagate features across the graph while an attention mechanism dynamically prioritizes informative neighbors at each propagation step. The resulting embeddings feed into a prediction head that outputs progression probabilities or time-to-event estimates [8, 14].

Downstream, the framework aggregates node-level representations into patient-level embeddings through attention-weighted readout functions that preserve temporal context. This end-to-end flow ensures that longitudinal laboratory trends, medication changes, and comorbidity evolutions are jointly considered during inference. The architecture remains modular to accommodate varying data

availability across healthcare systems. Such a design conceptually supports scalable deployment for chronic kidney disease risk stratification [9, 10].

At the system level, the proposed framework follows a strictly hierarchical data-to-prediction pipeline integrating heterogeneous clinical inputs through graph-based learning, as illustrated in **Figure 1**.

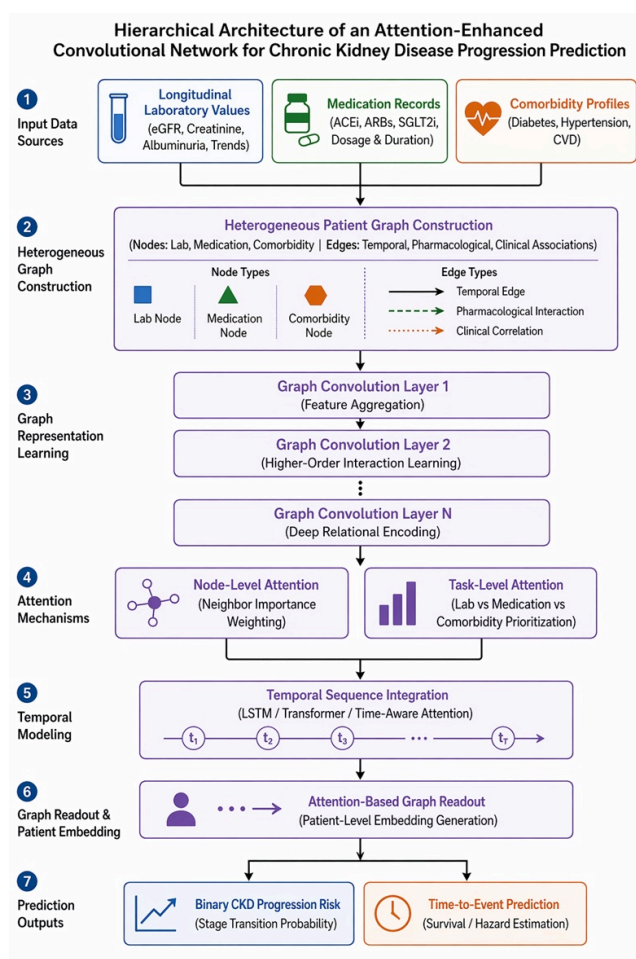


Figure 1. Hierarchical architecture of an attention-enhanced graph convolutional network integrating longitudinal laboratory values, medication networks, and comorbidity structures for chronic kidney disease progression prediction.

Core assumptions

The framework assumes availability of structured electronic health record data that include time-stamped laboratory measurements, medication prescriptions with start and stop dates, and coded diagnoses for comorbidities. These data elements must be sufficiently longitudinal to construct meaningful temporal graphs spanning multiple clinical

encounters. Standardization of laboratory units and medication coding systems is presumed to enable consistent node feature extraction across patients. Under these conditions, the graph construction process can reliably encode patient histories [2, 4, 16].

Additional assumptions involve the existence of baseline comorbidity and medication profiles that remain relatively stable or evolve predictably within the observation window. Missing values in longitudinal sequences are handled through imputation or masking strategies that preserve graph integrity. The framework does not require imaging or genomic data, focusing exclusively on routine laboratory, pharmacy, and diagnosis records. These assumptions align the architecture with widely available clinical datasets [3, 5].

Design principles

Heterogeneous graph representations form a central design principle, allowing distinct node types for laboratories, medications, and comorbidities while permitting typed edges that reflect domain-specific semantics. Temporal dynamics are incorporated by updating the graph at each time step to reflect new laboratory results or treatment modifications. Attention mechanisms are embedded at both node and task levels to ensure interpretability without sacrificing predictive power. This combination yields a clinically grounded architecture [8, 10].

Another principle emphasizes modularity so that individual components such as graph construction or attention layers can be refined independently as new medical knowledge emerges. Scalability to large patient cohorts is maintained through efficient message-passing formulations that avoid prohibitive computational costs. The framework prioritizes alignment with KDIGO staging and established risk factors to maintain translational relevance. Collectively, these principles guide the development of a robust conceptual model for chronic kidney disease progression [9, 15].

The functional contributions of each architectural component within the proposed framework are detailed in **Table 2**.

Table 2. Functional Roles of Core Components in the Attention-Based Graph CKD Prediction Framework

Component	Input Data	Core Operation	Out Represent
-----------	------------	----------------	------------------

Graph Construction	Lab, medication, comorbidity data	Node/edge formation with domain knowledge	Heterog gra
Node Feature Initialization	Raw EHR variables	Feature engineering (trend, duration, variance)	No embed
Graph Convolution Layers	Node embeddings + adjacency	Neighborhood aggregation	Update repre
Node-Level Attention	Neighboring nodes	Weighted message passing	Refine embed
Task-Level Attention	Node-type embeddings	Cross-domain weighting	Prior feat contrib
Temporal Module	Time-stamped graphs	Sequence learning (LSTM/Transformer)	Temp embed
Graph Readout	Node embeddings	Attention pooling	Patient vec
Prediction Head	Patient embedding	Classification / survival modeling	Risk s haz estim
Dynamic Graph Update	Incoming EHR data	Incremental graph modification	Update sta

ensure comparability across patients and time. Laboratory nodes thus serve as the foundational quantitative backbone of the patient graph [2-4].

Medication nodes represent prescribed agents including ACE inhibitors, ARBs, SGLT2 inhibitors, diuretics, and statins, encoded either as binary presence or dosage-adjusted embeddings. Each medication node carries attributes reflecting start dates, duration, and potential interactions derived from pharmacological databases. Comorbidity nodes indicate the presence and severity of conditions such as diabetes, hypertension, cardiovascular disease, and anemia through one-hot or embedding representations. The tripartite node structure enables comprehensive modeling of patient state [5, 16, 17].

Edge definitions

Edges between laboratory nodes capture statistical correlations or temporal dependencies such as the inverse relationship between creatinine and eGFR over sequential measurements. Medication-to-laboratory edges encode known pharmacological effects, for example linking SGLT2 inhibitors to expected reductions in albuminuria or potassium fluctuations associated with diuretics. Comorbidity-to-laboratory edges reflect established clinical associations, such as diabetes accelerating eGFR decline. These typed edges embed domain knowledge directly into the graph topology [4, 16].

Temporal edges connect the same node type across consecutive time steps to model disease evolution, while cross-type edges link current medications to contemporaneous comorbidities and laboratory values. Edge weights can be initialized from literature-derived strengths or learned during training to reflect patient-specific interaction intensities. This rich edge definition scheme ensures that the graph structure mirrors real-world clinical relationships. The resulting connectivity pattern facilitates effective message passing for progression prediction [3, 5].

Graph Construction

Node types

Laboratory nodes encode quantitative measurements such as estimated glomerular filtration rate, serum creatinine, albuminuria, potassium, and hemoglobin levels at each time point. These nodes incorporate derived features including recent trends, variance, and rate of change to capture dynamic renal function. Normalization techniques

Dynamic graph updates

The graph evolves dynamically as new laboratory results arrive, medication regimens change, or additional comorbidities are diagnosed within the longitudinal record. At each time window, nodes and edges are added or updated while preserving historical context through recurrent connections. This incremental construction

maintains temporal consistency without reconstructing the entire graph from scratch. Dynamic updates are essential for capturing the progressive nature of chronic kidney disease [2, 13].

Variants of temporal graph convolutional networks can process these evolving structures by incorporating time-aware propagation rules that discount older information or emphasize recent changes. Masking strategies handle irregular sampling intervals common in outpatient laboratory monitoring. The framework therefore supports realistic clinical timelines where data accrual is asynchronous. Such dynamic graph handling conceptually strengthens long-term progression forecasting [14, 15].

Graph Convolutional Network

Node feature initialization

Node features for laboratory measurements are initialized by concatenating raw values with engineered temporal summaries such as slope, intercept, and variance computed over recent windows. Medication nodes receive binary or continuous embeddings that incorporate dosage and duration attributes derived from prescription records. Comorbidity nodes are represented through indicator vectors augmented with duration-since-diagnosis features to reflect chronicity. Normalization and embedding layers ensure that heterogeneous feature scales are harmonized before graph processing [2-4].

Longitudinal trends are further encoded by appending time-difference features or recurrent summaries from prior time steps to each node. This initialization strategy preserves the sequential context that is critical for progression modeling. Random or pre-trained embeddings can initialize medication and comorbidity nodes when domain-specific knowledge graphs are available. The resulting feature vectors provide a rich starting point for subsequent convolution operations [5, 16].

Graph convolution layers

Graph convolution layers implement message passing where each node aggregates transformed features from its neighbors according to the defined edges and attention coefficients. The update rule follows the standard GCN formulation that normalizes adjacency information to prevent exploding or vanishing signals across multiple layers. Stacking several layers enables the model to

capture higher-order interactions, such as indirect effects of a comorbidity on laboratory values mediated through medication nodes. Layer-wise propagation thus builds increasingly abstract representations of patient state [8, 14, 15].

Residual connections and normalization techniques stabilize training on large heterogeneous graphs typical of longitudinal patient records. The convolution operation is applied independently at each time step before temporal fusion, ensuring that dynamic updates do not disrupt learned filters. Multiple layers also allow the network to integrate information across laboratory, medication, and comorbidity subgraphs simultaneously. This layered architecture forms the computational core of the conceptual framework [9, 10].

Graph-level readout

Graph-level readout aggregates all node embeddings into a fixed-size patient representation through attention-based pooling that weights nodes according to their relevance to the prediction task. This mechanism outperforms simple mean or max pooling by focusing on the most discriminative laboratory trends, medication patterns, or comorbidity clusters. The pooled embedding incorporates both local node information and global graph structure learned through prior convolutions. Patient-level representations are therefore contextually informed [9, 10].

Final readout can further incorporate a global attention vector that summarizes the entire graph at each time step before concatenation with temporal sequence models. The resulting vector serves as input to downstream prediction heads for binary progression or time-to-event tasks. Attention-guided pooling ensures that the readout remains interpretable by highlighting which nodes contributed most to the aggregated embedding. This step completes the transformation from raw graph data to clinically actionable predictions [14, 15].

Attention Mechanism

Node-level attention

Node-level attention computes learnable coefficients that determine the relative importance of each neighbor during message passing within the graph convolutional layers. For every node representing a laboratory value, medication, or comorbidity, the mechanism evaluates edge-specific

features and node attributes to assign higher weights to the most clinically relevant connections, such as the influence of an SGLT2 inhibitor on eGFR trajectories in the presence of diabetes. This selective aggregation prevents dilution of signals from less informative neighbors and produces refined embeddings that reflect personalized interaction strengths. The resulting attention scores offer a direct pathway to interpret which relationships drive individual node updates in the progression model [6, 10, 15].

Interpretability emerges naturally as clinicians can inspect the attention coefficients to trace how specific laboratory trends or medication adjustments propagate through the comorbidity network. By learning these weights end-to-end, the framework aligns model focus with established pharmacological and pathophysiological knowledge without manual feature engineering. Multiple attention heads can be employed to capture diverse relational aspects simultaneously, further enriching the representational capacity of the graph. This node-level mechanism thus forms a cornerstone for making graph convolutions both powerful and transparent in chronic kidney disease contexts [8, 14].

Task-level attention

Task-level attention operates at a higher granularity by weighting the contributions of entire node types—laboratory measurements, medication profiles, or comorbidity clusters—toward the final progression prediction. For example, it can emphasize how diabetes comorbidity nodes interact with longitudinal albuminuria trends more strongly than isolated hypertension flags when forecasting advancement to later CKD stages. This hierarchical weighting ensures that the prediction head receives a balanced yet clinically prioritized summary of heterogeneous inputs. The approach enhances the framework's ability to mirror nephrology decision-making processes where certain data domains dominate risk assessment [5, 16, 17].

Clinical interpretability is amplified because task-level attention weights can be visualized as importance rankings of node categories across patients or time windows. Such transparency allows nephrologists to validate model reasoning against KDIGO guidelines and known risk modifiers without requiring post-hoc explanation techniques. By integrating these weights into the readout step, the framework maintains end-to-end differentiability while providing actionable insights for treatment planning. Task-level attention therefore bridges the gap between

complex graph representations and practical healthcare utility [9, 10].

Temporal Dynamics

Handling longitudinal data

Longitudinal data are incorporated by constructing a sequence of time-stamped graphs where each snapshot reflects the most recent laboratory values, active medication regimens, and evolving comorbidity profiles. Time-stamped laboratory nodes such as serial eGFR and creatinine measurements are linked via temporal edges that encode the direction and magnitude of change between consecutive visits. Medication start and stop dates generate dynamic edges that appear or disappear as treatment regimens adjust, while comorbidity onset timestamps anchor new nodes to specific time points in the patient timeline. This temporal graph construction preserves the chronological order essential for modeling disease progression [2, 4, 13].

Irregular sampling intervals common in outpatient CKD monitoring are addressed through masking or interpolation strategies that maintain graph connectivity without introducing artificial data points. Historical context is retained by propagating embeddings across time steps via recurrent connections, ensuring that earlier laboratory trends continue to influence later predictions. The framework therefore captures the cumulative impact of sustained medication exposure or progressive comorbidity burden on renal trajectories. Such handling of longitudinal structure conceptually equips the model to forecast realistic clinical courses observed in electronic health records [3, 16].

Sequence modeling integration

Sequence modeling is integrated by combining graph convolutional outputs at each time step with recurrent architectures such as LSTM units or temporal transformers that process the ordered patient embeddings. Temporal attention layers further refine this fusion by assigning importance scores across the entire sequence of graph snapshots, highlighting critical inflection points like abrupt medication changes or accelerating eGFR decline. The combined module learns both intra-graph relational patterns and inter-time dependencies within a unified forward pass. This hybrid design leverages the strengths of

graph convolutions for structural data and sequence models for temporal ordering [14, 15].

Cross-time attention also enables the framework to weigh the relevance of past comorbidity networks when predicting future laboratory responses to ongoing therapies. By maintaining a memory of prior states, the model can detect subtle progression signals that static or single-snapshot approaches would miss. The integration remains modular, allowing substitution of different sequence backbones as new temporal neural network variants emerge. Overall, sequence modeling integration ensures that the conceptual framework fully exploits the time-varying nature of CKD data sources [8, 9].

Prediction Tasks

Binary progression prediction

Binary progression prediction focuses on classifying whether a patient will advance from CKD stage 3 to stage 4 or 5 within a clinically relevant horizon of two to five years. The final patient-level embedding produced by the graph convolutional network with attention is passed through a lightweight classification head that outputs a probability score for this binary outcome. Thresholds can be calibrated to balance sensitivity and specificity according to healthcare system priorities for early intervention. This task formulation aligns directly with routine nephrology decision points where timely referral or therapy intensification is required [1, 6, 12].

The binary setup simplifies deployment in electronic health record systems by providing a clear risk flag that can trigger alerts or decision-support recommendations. Longitudinal graph inputs ensure that the prediction reflects cumulative effects of laboratory trends, medication adherence, and comorbidity evolution rather than isolated measurements. By framing the problem as binary classification on heterogeneous graphs, the framework conceptually supports scalable screening of large CKD populations. Such an approach maintains clinical relevance while leveraging the full expressive power of the attention-augmented architecture [18-21].

Time-to-event prediction

Time-to-event prediction employs survival analysis techniques to estimate the expected duration until kidney failure or initiation of renal replacement therapy. The graph-

derived patient embeddings are fed into a survival head that parameterizes a hazard function while properly accounting for right-censoring in patients who remain event-free at the end of follow-up. Temporal dynamics encoded in the graph sequence allow the model to update hazard estimates as new laboratory, medication, or comorbidity information becomes available. This formulation provides richer prognostic information than binary labels alone [11, 22, 23].

Censoring handling is achieved through established likelihood formulations that incorporate both observed event times and censored observations without biasing the learned graph representations. The framework can output personalized survival curves that illustrate how attention-weighted factors influence projected timelines. By modeling time continuously, the approach supports dynamic risk reassessment at each clinical encounter. Time-to-event prediction thus extends the conceptual utility of the architecture to long-term planning and resource allocation in nephrology care [4, 13].

Evaluation Strategy

Prediction metrics

Prediction metrics for the binary progression task include area under the receiver operating characteristic curve and area under the precision-recall curve to assess discrimination across varying risk thresholds. For time-to-event prediction, the concordance index quantifies how well the model ranks patients according to their observed or censored survival times. Calibration plots and Brier scores further verify that predicted probabilities align with empirically observed event rates in held-out cohorts. These complementary metrics ensure comprehensive assessment of both ranking quality and absolute risk accuracy [1, 2, 12].

Additional metrics such as time-dependent AUROC at specific horizons provide insight into the framework's ability to forecast near-term versus long-term progression. The evaluation protocol emphasizes clinical utility by prioritizing metrics that reflect decision-making impact rather than abstract statistical performance. By reporting these measures on heterogeneous graph outputs, the conceptual framework can be benchmarked against existing CKD risk tools in a standardized manner. The chosen metrics therefore support transparent comparison and iterative refinement of the attention-based architecture [6, 18].

Validation protocols

Validation protocols rely on temporal splits that train on earlier records and test on subsequent time periods to simulate prospective deployment in evolving clinical environments. External validation on independent cohorts drawn from diverse geographic or demographic settings further tests the generalizability of the graph construction and attention mechanisms. Ablation studies systematically remove attention layers, medication nodes, or comorbidity subgraphs to quantify their individual contributions to predictive capability. Such protocols establish the robustness of each framework component [3-5].

Cross-validation within large longitudinal datasets can incorporate patient-level clustering to avoid leakage from repeated measures on the same individual. Sensitivity analyses explore the impact of varying graph density or temporal window lengths on overall performance. These rigorous validation strategies ensure that the conceptual design remains applicable across different electronic health record systems and CKD populations. The protocols thereby provide a blueprint for future empirical studies while maintaining the framework's focus on relational and temporal modeling [16, 17].

Conclusion

The proposed conceptual framework synthesizes graph convolutional networks with attention mechanisms operating on heterogeneous graphs built from longitudinal laboratory values, medication networks, and comorbidity structures to predict chronic kidney disease progression. By modeling patients as evolving graphs with typed nodes and clinically informed edges, the architecture captures the multifaceted interactions that drive renal decline in a manner unattainable by conventional independent-feature approaches. The integration of temporal dynamics and attention layers ensures that predictions remain both accurate and interpretable across varying time horizons. This unified design establishes a principled foundation for next-generation AI tools in nephrology.

Key advantages include the explicit representation of medication-laboratory dependencies and comorbidity-driven pathways, coupled with attention-based interpretability that aligns model outputs with established medical knowledge. Temporal graph updates further allow the framework to adapt to real-world data accrual patterns observed in routine care. These strengths collectively

position the architecture as a scalable and transparent solution for personalized risk assessment that respects the relational nature of electronic health records. The conceptual emphasis on modularity also facilitates future extensions as new data modalities or pharmacological insights emerge.

Limitations of the framework stem from the complexity of constructing and maintaining large heterogeneous graphs, which may impose computational demands on resource-constrained healthcare settings. Reliance on structured electronic health record data assumes consistent coding practices for medications and diagnoses, potentially limiting applicability in fragmented record systems. Additionally, the absence of direct experimental validation at this conceptual stage leaves open questions regarding real-world calibration across diverse populations. These considerations highlight areas for refinement in subsequent implementation phases.

Future work should prioritize implementation and prospective evaluation of the framework on established CKD cohorts such as CKiD, CRIC, and UK Biobank to translate the conceptual design into practical clinical decision support. Collaborative efforts between AI researchers and nephrologists can further refine node and edge definitions based on emerging KDIGO updates and novel therapeutic agents. By advancing graph-based methods for disease progression, the framework contributes to the broader goal of precision medicine in chronic kidney disease management. Ultimately, this graph convolutional network with attention paves the way for more proactive, interpretable, and patient-centered approaches to slowing CKD progression.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 01 Nov 2022 Revised: 09 Dec 2022 Accepted: 25 Jan 2023

Published online: 20 July 2023

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ferguson T, Ravani P, Sood MM, Clarke A, Komenda P, Rigatto C, et al. Development and external validation of a machine learning model for progression of CKD. *Kidney Int Rep.* 2022;7(8):1772-81. <https://doi.org/10.1016/j.ekir.2022.05.018>.
- Zhong J, Yang HC, Fogo AB. A perspective on chronic kidney disease progression. *Am J Physiol Renal Physiol.* 2017;312(3):F375-84. <https://doi.org/10.1152/ajprenal.00266.2016>.
- Zee J, Mansfield S, Mariani LH, Gillespie BW. Using all longitudinal data to define time to specified percentages of estimated GFR decline: a simulation study. *Am J Kidney Dis.* 2019;73(1):82-9. <https://doi.org/10.1053/j.ajkd.2018.07.019>.
- Tsai MH, Zhou MJ, Liu TC, Fang YW, Lu CJ. An integrated machine learning predictive scheme for longitudinal laboratory data to evaluate the factors determining renal function changes in patients with different chronic kidney disease stages. *Front Med (Lausanne).* 2023;10:1155426. <https://doi.org/10.3389/fmed.2023.1155426>.
- Khan A, Uddin S, Srinivasan U. Comorbidity network for chronic disease: a novel approach to understand type 2 diabetes progression. *Int J Med Inform.* 2018;115:1-9. <https://doi.org/10.1016/j.ijmedinf.2018.04.001>.
- Su CT, Chang YP, Ku YT, Lin CM. Machine learning models for the prediction of renal failure in chronic kidney disease: a retrospective cohort study. *Diagnostics (Basel).* 2022;12(10):2454. <https://doi.org/10.3390/diagnostics12102454>.
- Aoki J, Kaya C, Khalid O, Kothari T, Silberman MA, Skordis C, et al. CKD progression prediction in a diverse US population: a machine-learning model. *Kidney Med.* 2023;5(9):100692. <https://doi.org/10.1016/j.xkme.2023.100692>.
- Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Netw.* 2019;6(1):11. <https://doi.org/10.1186/s40649-019-0069-y>.
- Kazi A, Farghadani S, Aganj I, Navab N. IA-GCN: interpretable attention based graph convolutional network for disease prediction. In: *International Workshop on Machine Learning in Medical Imaging*; 2023 Oct 8; Vancouver, Canada. Cham: Springer Nature Switzerland; 2023. p. 382-92. https://doi.org/10.1007/978-3-031-44521-7_38.
- Xing X, Yang F, Li H, Zhang J, Zhao Y, Gao M, et al. Multi-level attention graph neural network based on co-expression gene modules for disease diagnosis and prognosis. *Bioinformatics.* 2022;38(8):2178-86.
- Inaguma D, Hayashi H, Yanagiya R, Koseki A, Iwamori T, Kudo M, et al. Development of a machine learning-based prediction model for extremely rapid decline in estimated glomerular filtration rate in patients with chronic kidney disease: a retrospective cohort study using a large data set from a hospital in Japan. *BMJ Open.* 2022;12(6):e058833. <https://doi.org/10.1136/bmjopen-2021-058833>.
- Gregorich M, Kammer M, Heinzel A, Böger CA, Eckardt KU, Heerspink HJL, et al. Development and validation of a prediction model for future estimated glomerular filtration rate in people with type 2 diabetes and chronic kidney disease. *JAMA Netw Open.* 2023;6(4):e231870. <https://doi.org/10.1001/jamanetworkopen.2023.1870>.
- Zhu Y, Bi D, Saunders M, Ji Y. Prediction of chronic kidney disease progression using recurrent neural network and electronic health records. *Sci Rep.* 2023;13(1):22091. <https://doi.org/10.1038/s41598-023-49252-1>.

Lee J, Warner E, Shaikhouni S, Bitzer M, Kretzler M, Gipson D, et al. Clustering-based spatial analysis framework through graph neural network for chronic kidney disease prediction using histopathology images. *Sci Rep.* 2023;13(1):12701. <https://doi.org/10.1038/s41598-023-39854-0>.

Ghorbani M, Kazi A, Baghshah MS, Rabiee HR, Navab N. RA-GCN: graph convolutional network for disease prediction problems with imbalanced data. *Med Image Anal.* 2022;75:102272. <https://doi.org/10.1016/j.media.2021.102272>.

Zhu W, Barreto EF, Li J, Lee HK, Kashani K. Drug-drug interaction and acute kidney injury development: a correlation-based network analysis. *PLoS One.* 2023;18(1):e0279928. <https://doi.org/10.1371/journal.pone.0279928>.

Shin EK, Choi HY, Hayes N. The anatomy of COVID-19 comorbidity networks among hospitalized Korean patients. *Epidemiol Health.* 2021;43:e2021035. <https://doi.org/10.4178/epih.e2021035>.

Dritsas E, Trigka M. Machine learning techniques for chronic kidney disease risk prediction. *Big Data Cogn Comput.* 2022;6(3):98. <https://doi.org/10.3390/bdcc6030098>.

Kanda E, Epureanu BI, Adachi T, Kashihara N. Machine-learning-based web system for the prediction of chronic kidney disease progression and mortality. *PLOS Digit Health.* 2023;2(1):e0000188. <https://doi.org/10.1371/journal.pdig.0000188>.

Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V. Chronic kidney disease. *Lancet.* 2021;398(10302):786-802. [https://doi.org/10.1016/S0140-6736\(21\)00519-5](https://doi.org/10.1016/S0140-6736(21)00519-5).

Anders HJ, Huber TB, Isermann B, Schiffer M. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat Rev Nephrol.* 2018;14(6):361-77. <https://doi.org/10.1038/s41581-018-0001-y>.

Yuan Y, Bar-Joseph Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol.* 2020;21(1):300. <https://doi.org/10.1186/s13059-020-02214-3>.

Iftikhar H, Khan M, Khan Z, Khan F, Alshanbari HM, Ahmad Z. A comparative analysis of machine learning models: a case study in predicting chronic kidney disease. *Sustainability.* 2023;15(3):2754. <https://doi.org/10.3390/su15032754>.