Developing a Patient Similarity Network for Predicting Post-stroke Urinary Tract Infection Risk in Hospitalized Immobile Patients

Zidu Xu, MMed, BS, RN[1], Yaowen Gu, MMed, BS[2], Chen Zhu, MSN, RN[3], Maxim Topaz, PhD, RN[1,4,5]
[1] School of Nursing, Columbia University, New York, NY; [2]Department of Chemistry, New York University, New York, NY; [3]Department of Nursing, Peking Union Medical College Hospital, Beijing, China; [4]Center for Home Care Policy & Research, Visiting Nurse Service of New York, New York, NY; [5]Data Science Institute, Columbia University, New York, NY

Background. Identifying high-risk groups for urinary tract infections (UTIs) is crucial for managing complications in immobile stroke patients to avoid poor prognoses. In our previous work, an ensemble learning model following the SHapley Additive exPlanations (SHAP) approach was adopted for UTI risk prediction and clinical interpretation of important features across the population, yet the nuances of individual cases were uncaptured[1]. Thus, we introduce a case-based reasoning method - patient Similarity Network (PSN) - that is inherently interpretable to provide clinically meaningful insights aside from individualized clinical risk modelling[2]. PSN calculates patient similarity scores based on their shared similar patterns in their electronic health records (EHR) data and constructs a network where each node represents a patient and the edges represent the degree of similarity between patients to identify similar cohorts and support clinical decision-making. Meanwhile, graph-based models such as Graph Attention Networks (GATs) have been applied to improve PSN representation learning on high-dimensional EHR data with multiple relationships, aiding accurate risk prediction and clinical interpretation[3]. So far, no studies have specifically used PSN or its extensions for risk prediction of post-stroke comorbidities.

Aim. To develop a PSN that combines GAT to create an individualized UTI risk prediction model for hospitalized immobile stroke patients.

Methods. We utilized EHR data from our previous multicenter investigation in China from 11/01/2015 to 06/30/2016, which enrolled 23,985 patients (⩾18 years old) immobile for over 24 consecutive hours after admission. With a primary diagnosis of stroke (ICD-10 I60~I64.x), 3982 immobile stroke patients from stage I were included to build the derivation cohort. Our outcome, UTI, was defined according to the European Association of Urology Urologic Infections Guidelines. We included 18 features based on research and literature evidence, including demographics, hospitalization information, medication status, and disease-related information. The proposed PSN regards each patient as a node, and simultaneously learns the node embeddings and pairwise similarities between nodes based on given features, using a defined similarity metric. To rank the pairwise similarities between patient nodes, a GAT is then employed to calculate attention coefficients as importance weights, following a Multi-layer Perceptron to output UTI risk probability based on learned pairwise similarities. We examined the strengths of PSN by making comparisons with 6 traditional machine learning (ML) models. We evaluate all prediction models using Sensitivity, Specificity, Accuracy, Precision, F1 score, AUC, and AUPRC. Visualization analysis examined PAN's capacity in representation learning and similarity measuring. The released source code is accessible at https://github.com/Somewhat120/Stroke_PSN.

Results. In the derivation cohort, 103 patients had UTIs during hospitalization (incidence, 2.59%). Table 1 shows the results of averaged five-fold cross-validation. PSN outperformed in AUC and sensitivity metrics on this patient pairwise similarity classification task. Particularly, the desirable sensitivity performance suggested that PSN can be clinically insightful since missing or delayed diagnosis and treatment of UTI can be severe and expensive. The representative case depicted in Fig 1 showed that PSN effectively learned similarities between candidate and target nodes to facilitate high-risk patient identification. In addition, ensembling GAT corrected predictions by recalculating attention coefficients to adjust similarity rankings for target nodes (patients), despite PSN causing incorrect connections between neighbor nodes (patients) with different labels. A feature-level analysis further reported the top-5 risk factors (female gender, duration of catheterization, pneumonia, and urinary incontinence) shared between the target node(patient) and most similar neighbor nodes (patients), which were well aligned with important risk factors identified by SHAP approach[1], while providing more details describing individual clinical characteristics.

Table 1. Prediction Model Performances.

Prediction Models	AUC	F1-score	Accuracy	Recall	Specificity
Logistic Regression	0.782±0.008	0.112±0.003	0.714±0.003	0.695±0.025	0.715±0.004
Random Forest	0.800±0.005	0.122±0.003	0.746±0.004	0.683±0.018	0.683±0.018
XGBoost	0.786±0.012	0.191±0.030	0.955±0.004	0.203±0.027	0.203±0.027
LightGBM	0.788±0.012	0.111±0.005	0.708±0.004	0.709±0.036	0.708±0.005
CatBoost	0.782±0.013	0.149±0.020	0.970±0.001	0.102±0.015	0.993±0.001
PSN	0.792±0.007	0.108±0.003	0.686±0.009	0.737±0.015	0.685±0.009

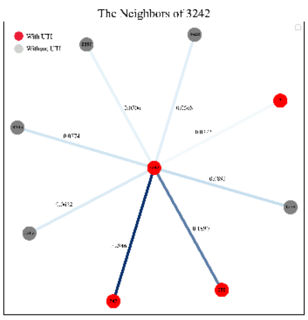

Figure 1. Visualization of pairwise similarities between target episode and neighbor episodes.

Conclusion. The experimental results demonstrated that PSN has comparable performance in post-stroke UTI risk prediction for immobile hospitalized patients compared to other traditional classification algorithms. However, its superiority of inherent interpretability allows for more precise identification of clinical patterns shared among similar individual patients, which is critical for further customization of UTI preventative interventions.

References

1. Zhu C, Xu Z, Gu Y, Zheng S, Sun X, Cao J, Song B, Jin J, Liu Y, Wen X, Cheng S, Li J, Wu X. Prediction of post-stroke urinary tract infection risk in immobile patients using machine learning: an observational cohort study. Journal of Hospital Infection. 2022;122:96-107. doi: https://doi.org/10.1016/j.jhin.2022.01.002.
2. Pai S, Hui S, Isserlin R, Shah MA, Kaka H, Bader GD. netDx: interpretable patient classification using integrated patient similarity networks. Molecular Systems Biology. 2019;15(3):e8497. doi: https://doi.org/10.15252/msb.20188497.
3. Brody S, Alon U, Yahav E. How attentive are graph attention networks? arXiv preprint arXiv:210514491. 2021.