Longitudinal Two-Center EHR Resource for Antibiotic Stewardship in Patients with Bacteremia

Seung Hwan Bae[1], Ki Tae Kwon[2]*, Inuk Jung[1]*

[1] School of Computer Science and Engineering, Kyungpook National University, Daegu, 41566, Republic of Korea

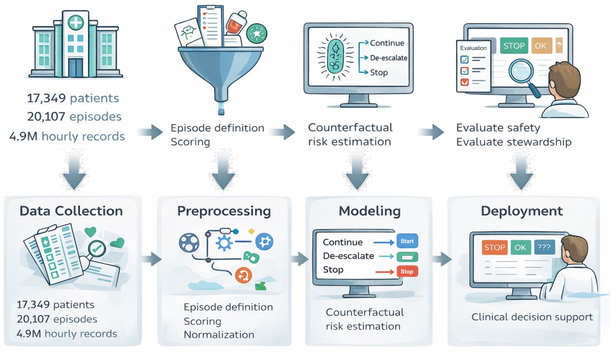
[2] Division of Infectious Diseases, Department of Internal Medicine, Kyungpook National University Chilgok Hospital, School of Medicine, Kyungpook National University, Daegu, 41404, Republic of Korea

Motivation

Bacteremia demands prompt, effective antibiotics, but prolonged or unnecessarily broad-spectrum therapy increases toxicity and drives antimicrobial resistance. Longitudinal data that couple physiologic response, labs, microbiology, and detailed regimens are needed to develop and validate antimicrobial stewardship intervention support tools.

Dataset

The resource comprises 17,349 patients and 20,107 admission episodes, including 19,749 culture-confirmed bacteremia episodes, with 4.9 million time-stamped records. The three most frequent pathogens were Escherichia coli, Klebsiella pneumoniae, and Staphylococcus aureus. A total of 31 features were collected, including static features, such as, age, sex, height and weight. A total of 6 vital sign features were used, including SBD, DBP, body temperature, SpO2. For lab features, while there were many missing ones, we first collected a comprehensive set of 21 features, including, WBC, neutrophils, platelets, bilirubin and creatinine.



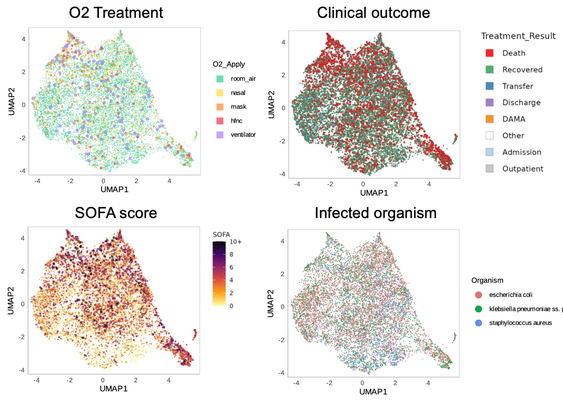
Methods

We curated a harmonized, hourly, two-center EHR cohort from Kyungpook National University Hospital (KNUH) and Kyungpook National University Chilgok Hospital (KNUHCH), Daegu, Korea (2014-2024). Eligible admissions required ≥1 blood-culture-confirmed bacteremia episode. Each record integrates vital signs and oxygen support, laboratory biomarkers, severity scores (NEWS2, qSOFA, SOFA, WHO clinical progression, Pitt Bacteremia Scores) with completeness indicators, species-level microbiology with resistance profiles, and structured antibiotic exposure for multi-drug regimens. Especially, we focused on tracking the severity trajectory after the antibiotic treatment started to later model a stewardship policy.

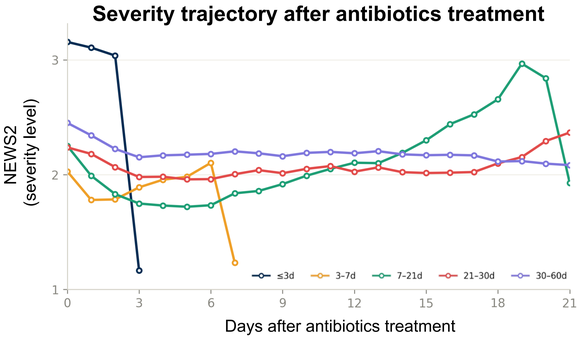
Results

Exploratory Uniform Manifold Approximation and Projection embeddings in the figure below of hourly clinical states showed separable clusters by severity and outcomes; severity/outcome distributions also differed across organisms, motivating organism-aware modeling. Also, the clinical outcome of recovery, deterioration and death was clearly observed in the collected longitudinal data in respect to the time point where antibiotics treatment has started. Especially, a clear recovery pattern was observed at day 3 and 7 post antibiotics treatment, which can be a strong indicator to be learned by a deep learning model to recommend a de-escalating the level of antibiotics treatment.

The UMAP representation of the dataset

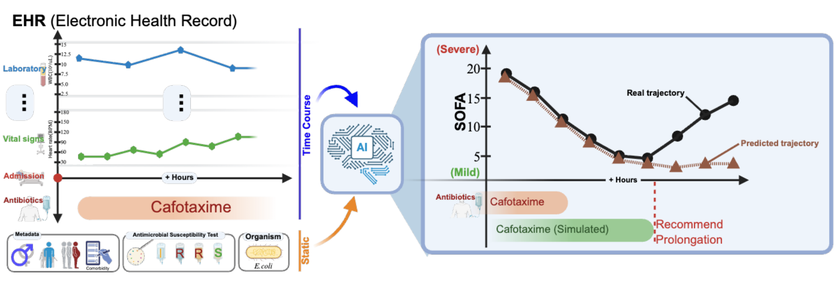


Severity trajectory after antibiotics treatment



Conclusion

This high-resolution dataset supports target-trial emulation and counterfactual, confidence-aware AI models to evaluate stewardship interventions (e.g., continue, de-escalate, or discontinue antibiotics) and to derive evidence-based prescribing guidelines for patients with bacteremia. A conceptual depiction is shown in Fig. 2.



Funding

This research was supported by a grant of the project for 'Research and Development for Enhancing Infectious Disease Response Capacity in Medical & Healthcare settings', funded by the Korea Disease Control and Prevention Agency, the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2025-02307351).