*MEDINFO 2023 — The Future Is Accessible*
*J. Bichel-Findlay et al. (Eds.)*

*doi:10.3233/SHTI231233*

# Prediction Models for Readmission Using Home Healthcare Notes and OMOP-CDM

Sujin Gan[a], Chungsoo Kim[a], Dong Yun Lee[b] and Rae Woong Park[a,b,1]
[a] *Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea*
[b] *Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea*
ORCiD ID: Sujin Gan https://orcid.org/0000-0003-3326-8428, Chungsoo Kim https://orcid.org/0000-0003-1802-1777, Dong Yun Lee https://orcid.org/0000-0002-3678-9862 Rae, Woong Park https://orcid.org/0000-0003-4989-3287

**Abstract**. This study developed readmission prediction models using Home Healthcare (HHC) documents via natural language processing (NLP). An electronic health record of Ajou University Hospital was used to develop prediction models (A reference model using only structured data, and an NLP-enriched model with structured and unstructured data). Among 573 patients, 63 were readmitted to the hospital. Five topics were extracted from HHC documents and improved the model performance (AUROC 0.740).

**Keywords.** Readmission, home healthcare, machine learning, prediction

## 1. Introduction

Readmission is an indicator of inpatient care quality and a major contributor to growing healthcare costs [1]. Therefore, identifying patients at high risk for readmission is crucial to reduce the likelihood of readmission. Home healthcare (HHC) is provided to discharged patients. Thus, HHC documents, which include post-discharge information from medical procedures to patient complaints, may contain hidden risk factors. Machine learning-based prediction models have been developed to assist in the identification of readmission risk factors, but they have not been applied to HHC documents.

## 2. Methods

We developed prediction models using an electronic health records database of Ajou University School of Medicine (AUSOM), which was converted to Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) version 5.3.1, constructed by the Observational Health Data Sciences and Informatics (OHDSI) [2].

This study included inpatients who had been admitted more than one day in the hospital and the outcome was defined as readmission within 90 days after discharge. Applying the patient-level-prediction framework of OHDSI, we developed a model only

[1] Corresponding Author: Rae Woong Park, email: veritas@ajou.ac.kr.

using structured variables (a reference model) and the other model with both structured and unstructured data from the HHC notes (an NLP-enriched model) [2]. Clinical data of demographics, condition, drug, procedure, and measurement were used as structured variables. The HHC notes were used to extract unstructured variables via Latent Dirichlet Allocation (LDA) topic modeling based on the degree of coherence scores [3]. Also, the model performance was evaluated according to the area under a receiver operating characteristic curve (AUROC). Descriptive analysis was conducted by comparing patient baseline characteristics between patients who revisited the hospital and those without readmissions, using frequencies for categorical variables. The two-sided $P < 0.05$ was considered statistically significant.

## 3. Results

Among 573 patients selected in this study, 63 (10.99%) patients revisited the hospital f or admission or ED. The mean age of them was 64.3 ± 13.3, and there were 22 (34.9%) males. The average age of the patients without readmission was 54.8 ± 15.6 and 145 (2 8.4%) of them were men. As structured variables, a feature of male sex, age group of 6 5-69, Immunocytochemical procedure, lesion of lung, and drugs for cardiovascular syst em were selected. Through LDA topic modeling on the HHC notes, five topics on Orth opedic operation, Chronic kidney diseases and diabetes, Diabetes mellitus foot, Gastroi ntestinal tract, and chemotherapy were determined according to the degree of coherenc e scores. The reference model had an AUROC of 0.654 (0.499-0.810). The NLP-enrich ed model, obtained by adding the variables from HHC documents to the structural varia bles, had an AUROC of 0.740 (0.630-0.852).

## 4. Conclusions

We developed the readmission prediction models using predictors extracted from unstructured HHC notes. Compared to the reference model created using only structured variables, prediction performance incorporating the five topics from HHC notes was higher. We could discover risk factors for readmission, which patients might have after their discharge, and observe how the topics are related to the readmission status.

## References

[1] Balla U, Malnick S, Schattner A. Early readmissions to the department of medicine as a screening tool for monitoring quality of care problems. Medicine. 2008 Sep;87(5):294-00, doi: 10.1097/MD.0b013e3181886f93.

[2] Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018 Aug;25(8):969-75, doi: 10.1093/jamia/ocy032.

[3] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research. 2003 Jan;3:993-22.