Patient and Encounter-level Analyses for Predicting Social Worker Referral Need: A Comparison of Alternative Machine Learning Prediction Approaches

Joshua R. Vest, PhD, MPH\textsuperscript{1,2}, Ofir Ben-Assuli, PhD\textsuperscript{3}, Shaun Grannis MD, MS\textsuperscript{1,2}, Suranga N. Kasthurirathne, PhD\textsuperscript{1,2}

\textsuperscript{1}Indiana University, Indianapolis, IN, USA; \textsuperscript{2}Regenstrief Institute, Indianapolis, IN, USA; \textsuperscript{3}Ono Academic College, Israel

Introduction. Machine learning approaches are well suited to leverage large volumes of varied patient datasets for modeling the complex relationships underlying disease status, care delivery, patient factors, and health outcomes. ML approaches have been used to successfully predict various health care outcomes and patient events\textsuperscript{1,3}. Problematically, health and health care information is inherently longitudinal and repetitive. Further, patient data may be clustered around a single event such as an encounter or lab test. A common approach in ML modeling research is to use the patient as the unit of analysis and aggregate, or summarize, historical diagnoses and experiences, e.g. total number of admissions during a recent time period or only the most recent observations per patient. This minimizes the statistical impact of non-independence among observations by removing multiple observations per individual. However, this does not exploit the longitudinal and repeated measure nature of the information. An alternative to patient-level analyses would be an encounter-level analysis that includes multiple observations per encounter, and retains the potential advantages of longitudinal and repeated measurement. In research design terms, the patient-level is analogous to cross-sectional information, while the encounter-level is reflective of longitudinal or panel information. This analysis explores the differential performance of patient-level and encounter-level analytic approaches for predicting need of referral to a social worker.

Methods. We compared the performance of the Random Forest (RF) classification algorithm applied to patient-level information with cluster adjusted Random Forest (RF++) algorithm\textsuperscript{4}, a novel variation of Random Forest to encounter-level information. The patient sample consisted of adults seeking care from an urban, Federally Qualified Health Center (FQHC) over a two-year period. We obtained a wide range of clinical, behavioral, and demographic data consisting of patient demographics, chronic conditions, addictions and narcotics use, medication history, encounter and visit history including hospitalizations, emergency visits and past referrals for machine learning. These features were extracted from the FQHC’s Electronic Health Record (EHR) and the Indiana Network for Patient Care (INPC), the statewide Health Information Exchange (HIE) of Indiana. The primary outcome of interest was a referral to a social worker\textsuperscript{7}, which was identified using provider orders. We created patient-level (i.e. one patient-per-row) and encounter-level (i.e. one encounter-per-row) data vectors consisting of binary or count variables for each feature. Encounter-level data vectors preserved the sequence of occurrence. For each dataset, we only included patient data that had been recorded prior to an occurrence of the outcome (referral to a social worker). For each machine learning approach, we split patient and encounter-level data into training and holdout test datasets using a 5:1 ratio. We applied the RF classification algorithm to the patient-level data and RF++ to the encounter-level data. The RF and RF++ algorithms leveraged their internal feature selection methods to identify optimal features. We calculated performance measures and 95\% confidence intervals for each model using thresholds that optimized F1-Score, the harmonic mean between Precision and Recall (see Figure 1).

\textbf{Figure 1.} Workflow representing our study approach
**Results.** The training dataset consisted of 50,000 patients and 165,423 encounters. The holdout test dataset consisted of 10,000 patients and 38,848 encounters. Prevalence of need of social services were 5.1% (patient level) and 5.2% (encounter level) across the training dataset, and 5.2% (patient level) and 5.0% (encounter level) across the test datasets. Both decision models performed well on the holdout test dataset with predictive metrics between 65%-100%, with ROC scores > 90%. (see Figure 2). However, the encounter-level model significantly outperformed the patient-level model across F1-Score, Recall and Area under the ROC curve (AUC ROC) measures. Precision-Recall curves (figure 3) present model performance across various cutoff thresholds. Area under the Precision Recall curve scores (AUCPR) were 78.91% for the patient-level model, and 88.32% for the encounter-level model.

**Figure 2. Performance metrics**

**Figure 3. Precision-Recall curve**

**Discussion.** Results suggest that both patient and encounter-level models were capable of predicting need of social work referrals with high performance metrics despite low prevalence of the target outcome. However, encounter-level based analysis yields statistically superior performance to the patient-level model. The Precision-Recall curve indicates that the model yields high performance across various thresholds, allowing users to select cutoff thresholds that best reflect their ideal precision or recall scores for implementation. However, it is unclear if these results are generalizable across other non-FQHC populations. Further investigations will explore the roles of patient and population-level social determinants factors in improving predictive performance. These findings provide a basis for further comparison of alternative ML modeling strategies that leverage the longitudinal and clustered nature of health care data for non-clinical outcomes.

**References**