Explainable Tree-Based Predictions for Unplanned 30-Day Readmission of Patients With Cancer Using Clinical Embeddings

Chi Wah Wong, PhD[1]; Chen Chen, MS[1]; Lorenzo A. Rossi, PhD[1]; Monga Abila, MPH, BSN, RN[2]; Janet Munu, MBA, BSN, RN[2]; Ryotaro Nakamura, MD[3]; and Zahra Eftekhari, MS[1]

PURPOSE Thirty-day unplanned readmission is one of the key components in measuring quality in patient care. Risk of readmission in oncology patients may be associated with a wide variety of specific factors including laboratory results and diagnoses, and it is hard to include all such features using traditional approaches such as one-hot encoding in predictive models.

METHODS We used clinical embeddings to represent complex medical concepts in lower dimensional spaces. For predictive modeling, we used gradient-boosted trees and adopted the shapley additive explanation framework to offer consistent individualized predictions. We used retrospective inpatient data between 2013 and 2018 with temporal split for training and testing.

RESULTS Our best performing model predicting readmission at discharge using clinical embeddings showed a testing area under receiver operating characteristic curve of 0.78 (95% CI, 0.77 to 0.80). Use of clinical embeddings led to up to 23.1% gain in area under precision-recall curve and 6% in area under receiver operating characteristic curve. Hematology models had more performance gain over surgery and medical oncology. Our study was the first to develop (1) explainable predictive models for the hematology population and (2) dynamic models to keep track of readmission risk throughout the duration of patient visit.

CONCLUSION To our knowledge, our study was the first to develop (1) explainable predictive models for the hematology population and (2) dynamic models to keep track of readmission risk throughout the duration of patient visit.

INTRODUCTION

Thirty-day unplanned readmission rate is often used as a measure for quality of care.[1] Patients with cancer are particularly at high risk for hospital readmission,[2] associated with higher cost and safety concerns.[3,4] A study by Jencks et al[5] reported that the 30-day readmission rate for Medicare beneficiaries was 19.6%. The Medicare Payment Advisory Commission estimated that readmissions result in $15 billion US dollars in additional annual Medicare expenditures, which accounts for approximately 10% of overall Medicare spending on hospital inpatient care.[6] Furthermore, readmission can be linked with negative outcomes, such as higher mortality rates.

Artificial intelligence enables drawing insights from a vast amount of data without the explicit need to generate a hypothesis or to make previous assumptions on data distribution as in traditional research. Machine learning and deep learning models have been increasingly used in precision oncology.[7] Tree-based machine learning models are among the most commonly used nonparametric and nonlinear models.[8-10] Efficient approaches for post hoc explanations of individual predictions from tree-based models are gaining popularity and can provide critical insights into applications in clinical decision support.[11-14]

Embedding techniques, eg, Word2Vec,[15] provide low-dimensional representations of words from their context to preserve semantic similarity properties. Those techniques can also be applied to clinical codes in the electronic health record (EHR), such as Logical Observation Identifiers Names and Codes (LOINC) and International Classification of Diseases (ICD)-9 codes,[16] to extract low-dimensional features for clinical prediction tasks.[17] Liu et al[18] showed an improvement in the performance of predicting 30-day readmission by using embeddings of diagnosis codes and artificial neural networks. It is unclear whether using clinical embeddings will improve predictive performance in the cancer population.

In this study, we made three major contributions toward predictive modeling of 30-day unplanned readmission

CONTEXT

Key Objective

Thirty-day unplanned readmission is one of the key quality measures in cancer care. In this study, we developed a dynamic, explainable machine learning model to predict the risk of readmission at various time points throughout a patient's visit. We also investigated the benefits of using clinical embeddings in our predictive models.

Knowledge Generated

Predictive performance at admission was modestly lower than at discharge. Addition of International Classification of Diseases-9 codes embeddings improved performance in most cases. Within services, incorporating clinical embeddings led to more performance gains in hematology than surgery and medical oncology.

Relevance

A dynamic model potentially allows a patient's readmission risk to be identified early, hence making intervention possible. Risk assessment and prediction explanations could potentially be incorporated in patient education and discharge planning.

for cancer population. First, we developed multiple explainable tree-based predictive models associated with different time points throughout patient visit. Explainable predictions were made available shortly after individual patient's admission, and the predictions were refined throughout the course of hospital stay and toward discharge when additional information was available. Second, we investigated the benefits of using clinical embeddings derived from ICD-9 (diagnoses) and LOINC (laboratory tests) codes. Third, we compared model performance among different cancer specialties (hematology, medical oncology, and surgery). To the best of our knowledge, our study is the first to develop a predictive model for readmission in hematology population.

METHODS

Study Design

Using EHR data, we conducted a single institutional study on a cohort of all adult patients (age ≥ 18 years) admitted to the hospital between 2013 and 2017 (n = 18,811). We performed a temporal split on the data such that the visit data with discharge date between 2013 and 2016 were used for model training (n = 15,524), and the data with discharge date in 2017 were used for model testing (n = 3,287). In the testing data set, we only considered new patients admitted to the hospital after January 1, 2017, to avoid potential data leakage with the training data set.

The outcome is a two-class variable denoting whether 30-day unplanned readmission occurred. The overall rate of 30-day unplanned readmission for the data set based on our inclusion criteria is 16.4%, with hematology having the highest readmission rate (23.8% among 6,916 visits), followed by medical oncology (18.7% among 3,089 visits) and surgery (9.7% among 8,693 visits). Patient characteristics are summarized in Table 1.

Admission and Discharge Factors

We assigned input features into three groups: baseline factors, discharge factors, and clinical embeddings. The features were identified based on the literature[13,19-21] and potential clinical relevance. In the first group, we identified multiple baseline factors that are typically available shortly (within 24 hours) after admission: age at admission, gender, number of admissions to the study institution in the past 6 months, presence of emergency treatment center visit to the study institution in the past 30 days, service line (admission department: hematology, medical oncology, surgery, and others), primary insurance (MEDICARE, MEDI-CAL, and others), and bone marrow transplant type (allogeneic or autologous for hematology only). In the second group, we identified three discharge factors that were typically obtained during discharge planning of the inpatient visit: discharge disposition (home, other hospitals, skilled nursing facilities, rehabilitation, and others), discharge reconciliation (medication review), and length of stay. Summary statistics of the baseline and discharge factors are listed in Table 1.

Features From Embeddings of Clinical Codes

For ICD-9 codes, we use the 300-dimensional Word2Vec embedding model trained by Choi et al[16] on the insurance claims of 4 millions of patients. For each inpatient visit, we considered up to the most recent 50 diagnoses within the 6 months preceding admission and a time point (TP) during the hospital visit. The choice of TP depends on the prediction time and will be discussed in the next section. Between 2013 and 2017, our medical center used diagnoses associated with 2,945 unique ICD-9 codes, where 0.5% of the ICD-9 codes were out of vocabulary for the insurance claims model. We produced a 300-dimensional feature vector for each visit by averaging the embeddings of the ICD-9s.

For LOINC codes, we considered up to the most recent 50 laboratory tests with abnormal results ordered between admission and TP. The laboratory tests corresponded to 425 unique LOINC codes, where 44% of the LOINC codes were out of vocabulary for the insurance claims model. As a result, we decided to exclude LOINC code embeddings

TABLE 1. Patient Characteristics and Outcome

Variables	Training Data Set 2013-2016 (n = 15,524)	Testing Data Set 2017[a] (n = 3,287)	Total (N = 18,811)
Outcome			
Unplanned readmission in days			
< 7 days	825 (5.3%)	158 (4.8%)	983 (5.2%)
≥ 7 days and < 14 days	716 (4.6%)	162 (4.9%)	878 (4.7%)
≥ 14 days and < 21 days	508 (3.3%)	115 (3.5%)	623 (3.3%)
≥ 21 days and ≤ 30 days	465 (3.0%)	128 (3.9%)	593 (3.2%)
Rate of unplanned readmission			
Overall	16.2%	17.1%	16.4%
Hematology	24.0%	23.1%	23.8%
Medical oncology	17.3%	25.5%	18.7%
Surgery	9.8%	9.4%	9.7%
Baseline factors			
Sex			
Male	7,965 (51.3%)	1,665 (50.7%)	9,630 (51.2%)
Female	7,559 (48.7%)	1,622 (49.3%)	9,181 (48.8%)
Age at index admission			
Median (range)	60 (18-97)	61 (18-99)	60 (18-99)
Interquartile range	50-68	50-69	50-68
Number of admissions in the previous 6 months			
= 0	11,007 (70.9%)	2,441 (74.3%)	13,448 (71.5%)
= 1	2,463 (15.9%)	494 (15.0%)	2,957 (15.7%)
= 2	949 (6.1%)	167 (5.1%)	1,116 (5.9%)
≥ 3	1,105 (7.1%)	185 (5.6%)	1,290 (6.9%)
ETC visit in the previous 30 days			
True	792 (5.1%)	470 (14.3%)	1,262 (6.7%)
Service line			
Hematology	5,682 (36.6%)	1,234 (37.5%)	6,916 (36.8%)
Medical oncology	2,555 (16.5%)	534 (16.2%)	3,089 (16.4%)
Surgery	7,187 (46.3%)	1,506 (45.8%)	8,693 (46.2%)
Others	100 (0.6%)	13 (0.4%)	113 (0.6%)
Primary insurance			
MEDICARE	5,521 (35.6%)	1,084 (33.0%)	6,605 (35.1%)
MEDI-CAL	1,564 (10.1%)	255 (7.8%)	1819 (9.7%)
Others	8,439 (54.3%)	1948 (59.3%)	10,387 (55.2%)
Bone marrow transplant			
Autologous	1,058 (6.8%)	245 (7.5%)	1,303 (6.9%)
Allogeneic	768 (4.9%)	192 (5.8%)	960 (5.1%)
Discharge factors			
Length of stay at index admission			
Median (range)	5 (0-232)	5 (0-406)	5 (0-406)
Interquartile range	2-10	3-11	2-10
Discharge to location			
Home	11,673 (75.2%)	2,385 (72.6%)	14,058 (74.7%)

(Continued on following page)

TABLE 1. Patient Characteristics and Outcome (Continued)

Variables	Training Data Set 2013-2016 (n = 15,524)	Testing Data Set 2017[a] (n = 3,287)	Total (N = 18,811)
Other hospitals	3,200 (20.6%)	763 (23.2%)	3,963 (21.1%)
Skilled nursing facilities	354 (2.3%)	97 (3.0%)	451 (2.4%)
Rehabilitation	265 (1.7%)	38 (1.2%)	303 (1.6%)
Others	32 (0.2%)	4 (0.1%)	36 (0.2%)
Discharge reconciliation			
True	8,441 (54.4%)	2,635 (80.2%)	11,076 (58.9%)
Medical codes at discharge			
Number of ICD-9 codes[b]			
Median	2	5	3
Interquartile range	0-6	2-10	1-6
Number of LOINC codes[c]			
Median	50	50	50
Interquartile range	20-50	22-50	20-50

Abbreviations: ETC, emergency treatment center; ICD, International Classification of Diseases; LOINC, Logical Observation Identifiers Names and Codes.
[a]For the data set in 2017, only new patients as of January 1, 2017, were considered.
[b]Total count of ICD-9 codes between 6 months before admission and discharge.
[c]Total count of LOINC codes between admission and discharge.

from the insurance claims model for most of our analyses and developed a custom LOINC embeddings model based on our center's laboratory data. We trained 100-dimensional Wor2Vec embeddings from the LOINCs of laboratory orders associated with encounters in our EHR. In the context of word embeddings, an LOINC was equivalent to a word and an encounter was equivalent to a sentence. We used the LOINCs associated with 703,839 encounters recorded from July 2009 to December 2016. We used the skip-gram approach with a context window equal to 5. LOINCs occurring five times or less were excluded. This led to embedding representations for 1,005 distinct LOINCs. We produced a 100-dimensional feature vector for each visit by averaging the embeddings of the LOINCs.

We concatenated the ICD-9 and LOINC embedding feature vectors to the baseline and discharge features. Summary of count of ICD-9 and LOINC codes is listed in Table 1.

Explainable Tree-Based Predictive Modeling

Various combinations of the input features were fed into machine learning models to predict whether 30-day unplanned readmission will occur for an inpatient visit. Specifically, we investigated six scenarios, and clinical embeddings were used in scenarios 3-6:

FIG 1. Comparison of testing data set prediction performance of scenarios 1 and 2. Orange curves represent the performance of scenario 1, whereas blue curves represent the performance of scenario 2. (A) ROC curve plot. (B) Precision-recall curve plot. AUC, area under the curve; ROC, receiver operating characteristic.
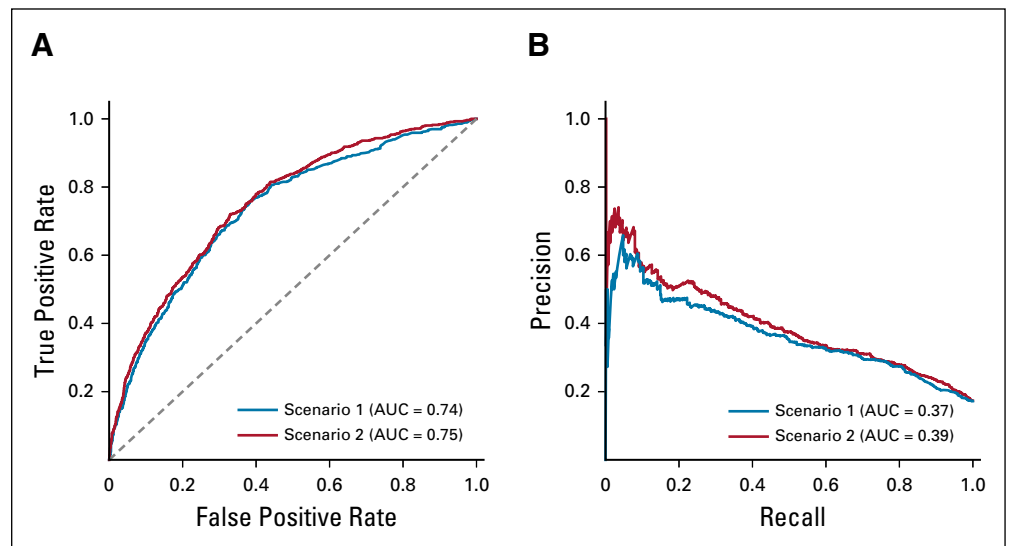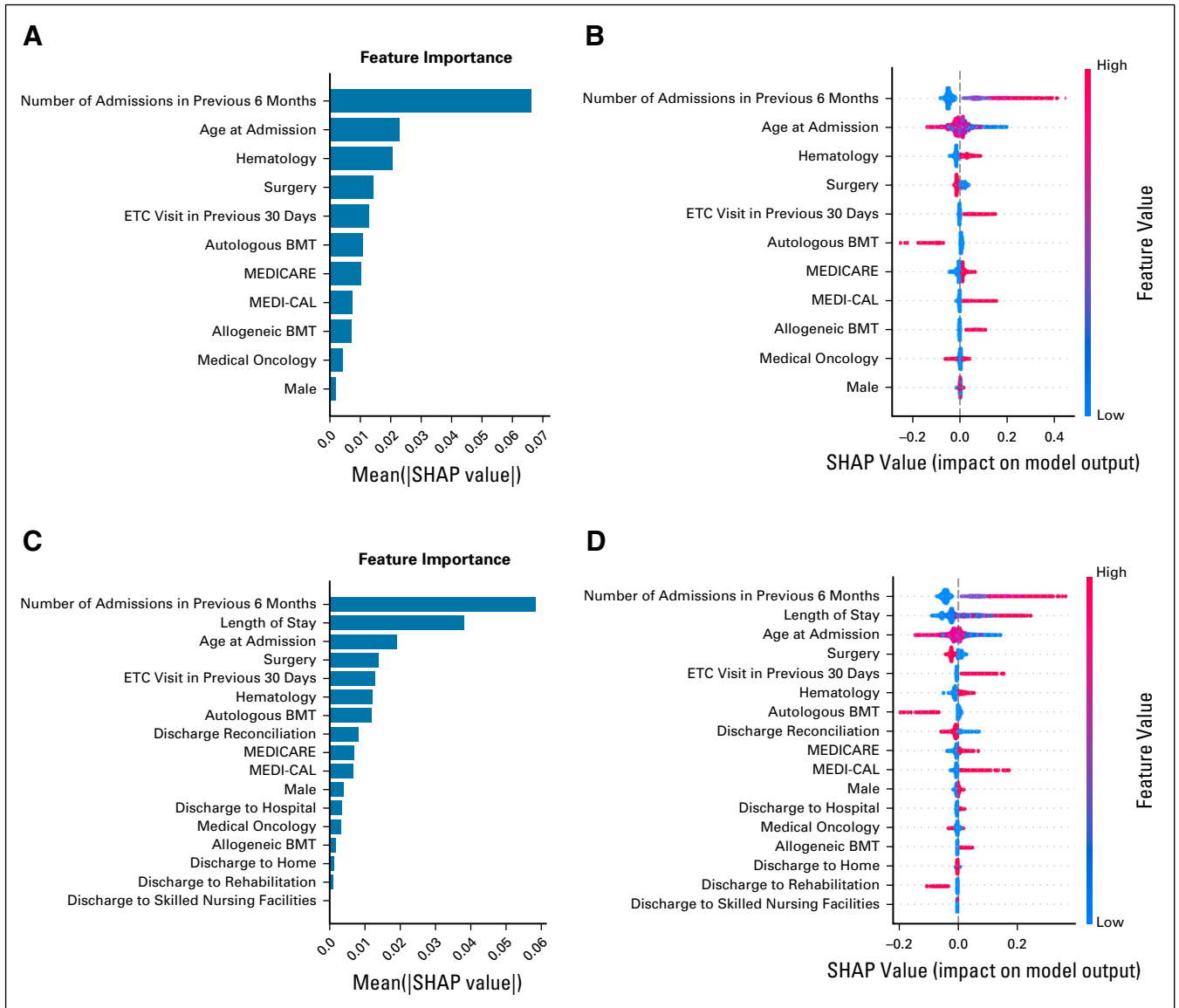
FIG 2. Summary plots combining local explanation values from individual predictions (scenarios 1 and 2). (A, C) Bar chart showing mean absolute of all shap values of the model: (A) scenario 1 and (C) scenario 2. (B and D) Beeswarm plots where in each row (corresponds to one feature), one dot corresponds to one patient in the testing data set. The color of each dot corresponds to a normalized feature value (qualitative: blue for low values and red for high values), whereas its position along the x-axis depicts the shap value describing the impact on the model prediction: (B) scenario 1 and (D) scenario 2. BMT, bone marrow transplant; ETC, emergency treatment center; SHAP, shapley additive explanation.

1. Prediction performed using baseline factors only.
2. Prediction performed at discharge using baseline and discharge factors.
3. Prediction performed 24 hours after admission using baseline factors and clinical embeddings based on ICD-9 codes available until 24 hours after admission.
4. Prediction performed dynamically. In this case, we randomly selected a time point (TP, median = 2.4 days) during each visit. We considered ICD-9 codes up till TP to generate clinical embeddings. We also used baseline factors in this case.
5. Prediction at discharge using baseline factors, discharge factors, and clinical embeddings based on ICD-9 codes until discharge.

6. We appended the LOINC embedding feature vector into the feature vector of scenario 5 to assess potential improvement in predictive performance.

We chose nonparametric gradient-boosted tree models[10] for classification, along with tree-explainer for local explanations of the model.[11,12] For the training data set, we used 10-fold cross-validation to identify optimal model hyperparameters. Appendix Table A1 shows the training data set model performance for various combinations of clinical and embeddings factors. The optimal model was then applied on the testing data set. Receiver operating characteristic (ROC) and precision-recall curves were used for model assessment. The associated 95% CIs were calculated by using 1,000 times of
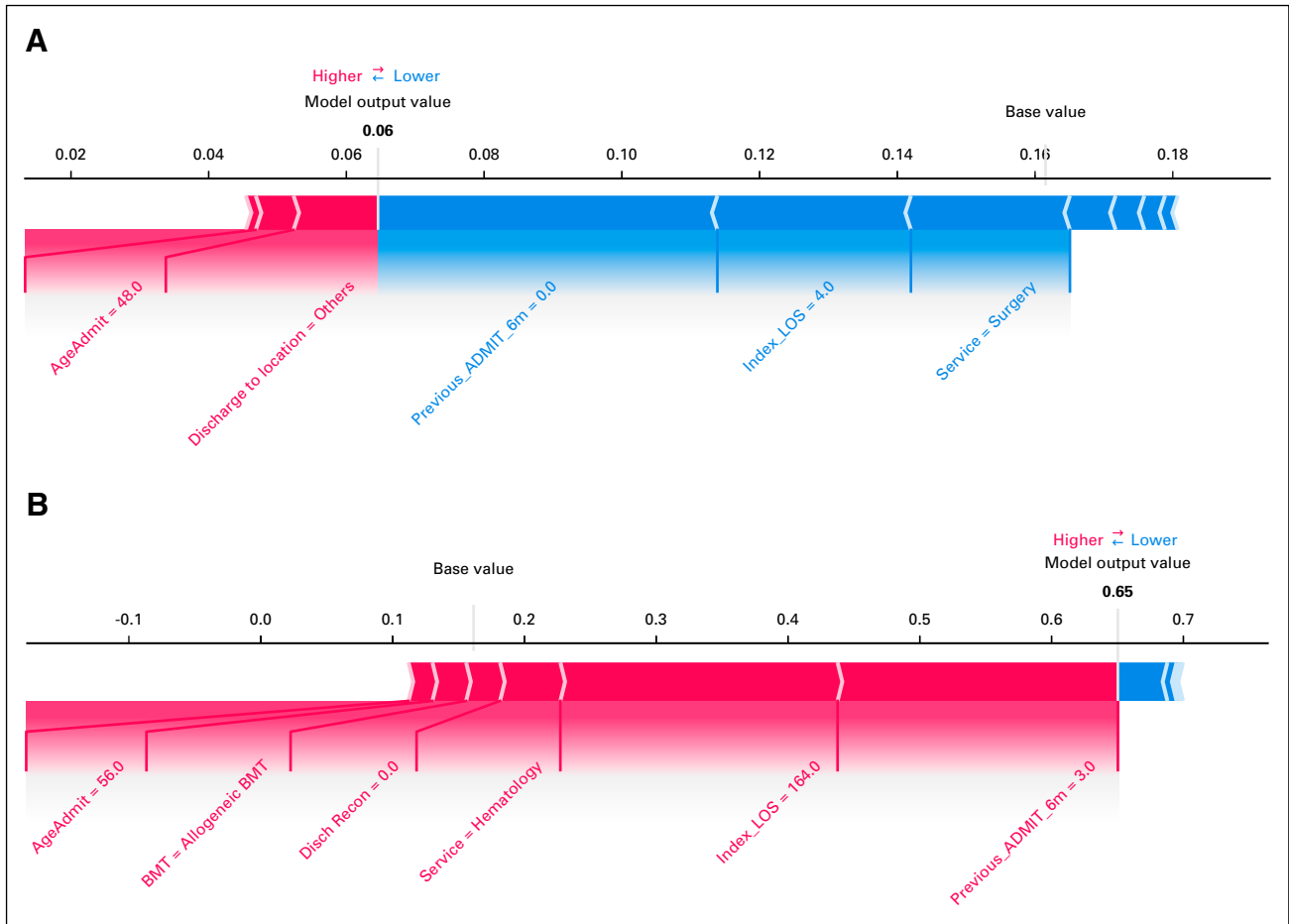
FIG 3. Composition of individual predictions from two sample patients using the model in scenario 2. Explained prediction for each patient. Blue arrows depict the factors contributed to decreased risk of readmission, whereas red arrows depict the factors contributed to increased risk of readmission. (A) The prediction of readmission risk was 0.06 with explanations that 30-day unplanned readmission was false for this patient. The patient of age 48 years was admitted to inpatient service in the surgery department for 4 days. This patient had not been admitted to inpatient visit in the previous 6 months. The patient was discharged to other hospitals, and discharge reconciliation is complete. (B) The prediction of readmission risk was 0.65 with explanations. Thirty-day unplanned readmission was true for this patient. The patient was admitted to Hematology for Allogeneic Bone Marrow Transplant. The patient stayed in the inpatient service for 164 days. The patient had been admitted to inpatient visit once in the past 6 months. Discharge reconciliation was not complete. BMT, bone marrow transplant; LOS, length of stay.

bootstrapped resampling. We then applied tree-explainer upon the optimal model to compute explanations for each individual prediction (probabilistic shapley additive explanation [SHAP] value) based on the associated exact shapley values generated for each feature from individual patient's visit data in the testing set. The summation of all SHAP values associated with all features for an individual patient's visit is equal to the model outcome prediction. Features with the highest SHAP values could be displayed to clinicians along with each prediction to show the factors contributing the most to risk of readmission. This could potentially be useful to tailor intervention. Because individual embedding features do not offer much insight into model explanation, for each patient's visit data, we summed up all SHAP values associated with embeddings for ICD-9 and LOINC codes, respectively, to create one SHAP value for ICD-9 and one SHAP value for LOINC. The overall feature importance was obtained by

calculating the mean absolute SHAP values of individual features. Python 3.7.6, scikit-learn 0.22.1, lightGBM 2.3.1, and shap 0.35.0 were used for machine learning classification and explanation.

RESULTS

Explainable Predictions from Admission to Discharge

Figure 1A shows that the area under receiver operating characteristic curve (AUROC) of the model using both baseline and discharge factors is 0.75 (95% CI, 0.73 to 0.77, scenario 2), just 1.4% more than that of using baseline (0.74, 95% CI, 0.72 to 0.75, scenario 1). On the other hand, Figure 1B shows that the area under precision-recall curve (AUPRC) of the model using both baseline and discharge factors is 0.39 (95% CI, 0.36 to 0.43, scenario 2), with 5.4% improvement over using baseline factors only
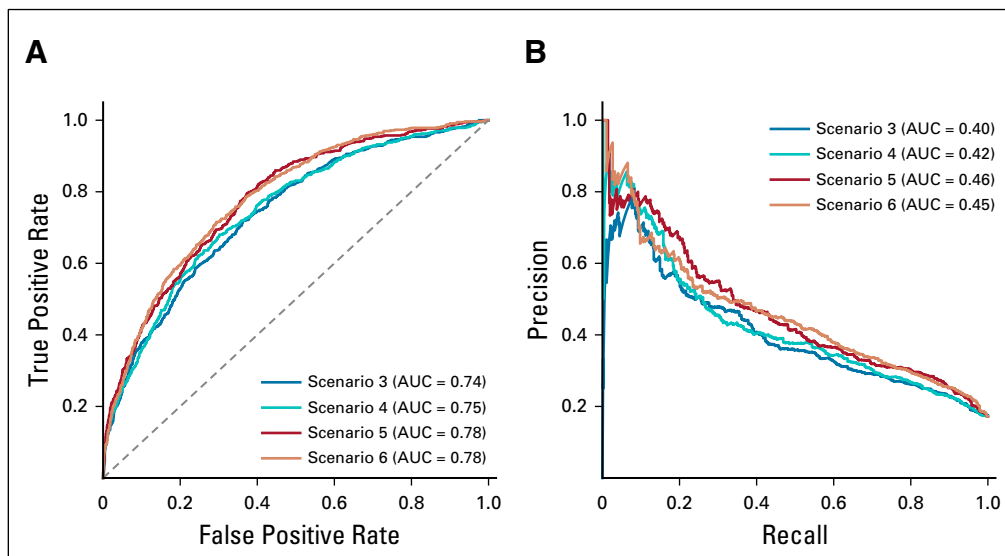
FIG 4. Comparison of testing data set prediction performance of using embeddings on top of various combinations of baseline and discharge factors (scenarios 3-6). (A) Receiver operating characteristic curve plot. (B) Precision-recall curve plot. AUC, area under the curve; ROC, receiver operating characteristic.

(0.37, 95% CI, 0.33 to 0.40, scenario 1). It should be noted that for an imbalanced data set, and real-world applications such as readmission, PRC is a more valuable metric for model assessment than ROC, as ROC tends to generate overly promising interpretation of predictive performance.[22]

Figures 2A and 2B display feature importance bar plots of mean absolute SHAP values, grouped by individual variables associated with models using baseline factors and baseline along with discharge factors, respectively. In both cases, the number of admissions in the past 6 months before admission and patient's age at admission are among the most important features. For the model with discharge factors, length of stay is the second most important feature. Figures 2C and 2D display beeswarm plots[11] where in each row (corresponds to one feature), one dot corresponds to one patient in the testing data set. The color of each dot corresponds to a normalized feature value (qualitative: blue for low values and red for high values), whereas its position along the x-axis depicts the SHAP value describing the impact on model performance. Using the beeswarm plot, we can observe the directionality of the relation between individual factors and outcome (readmission in our case), along with the magnitude of the relation. For instance, we observe patients with a larger number of admissions in the 6 months leading to the visit have increased probability of readmission. Additionally, patients with lower age have increased risk of readmission in our cohort. In hematology population, patients with allogeneic bone marrow transplant have increased risk, whereas patients with autologous bone marrow transplant have reduced risk of readmission. For the model using discharge factors (Fig 2B), a larger length of stay corresponds to an increased risk of readmission, whereas patients with discharge reconciliation have slightly reduced the risk of readmission.

Figure 3 visualizes the composition of individual explained predictions from 2 sample patients using the model with baseline and discharge factors. In each plot, blue arrows depict the factors contributing to decreased risk of readmission whereas red arrows depict the factors contributing to increased risk of readmission.

Performance Improvement Using Clinical Embeddings

Figure 4A displays the AUROC and AUPRC values of the model using various combinations of factors with embeddings. Using dynamic embeddings (scenario 4, green curves: AUROC = 0.75 [95% CI, 0.73 to 0.77], AUPRC = 0.42 [95% CI, 0.38 to 0.45]) shows only a 1.3% improvement in AUROC compared with using embeddings at baseline (scenario 3, orange curves: AUROC = 0.74 [95% CI, 0.72 to 0.76], AUPRC = 0.40 [95% CI, 0.36 to 0.43]). Incorporating embeddings and factors at discharge shows 5.4% and 15% improvements at AUROC and AUPRC, respectively (scenario 5, blue curves: AUROC = 0.78 [95% CI, 0.76 to 0.79], AUPRC = 0.46 [95% CI, 0.42 to 0.49]), compared with using baseline factors and embeddings (scenario 3).

If we compare the performance between models with and without embeddings (Figs 2 and 4), we observe 8.1% AUPRC improvements for baseline (scenario 1 v 3). The embedding-related performance gain between scenarios 2 and 5 was 4% AUROC and 18% AUPRC, suggesting that using embeddings offers notable improvements in overall precision. No notable performance gain was observed compared with scenario 5 when adding LOINC embeddings to the predictive models (scenario 6, red curves: AUROC = 0.78 [95% CI, 0.77 to 0.80], AUPRC = 0.45 [95% CI, 0.42 to 0.49]).
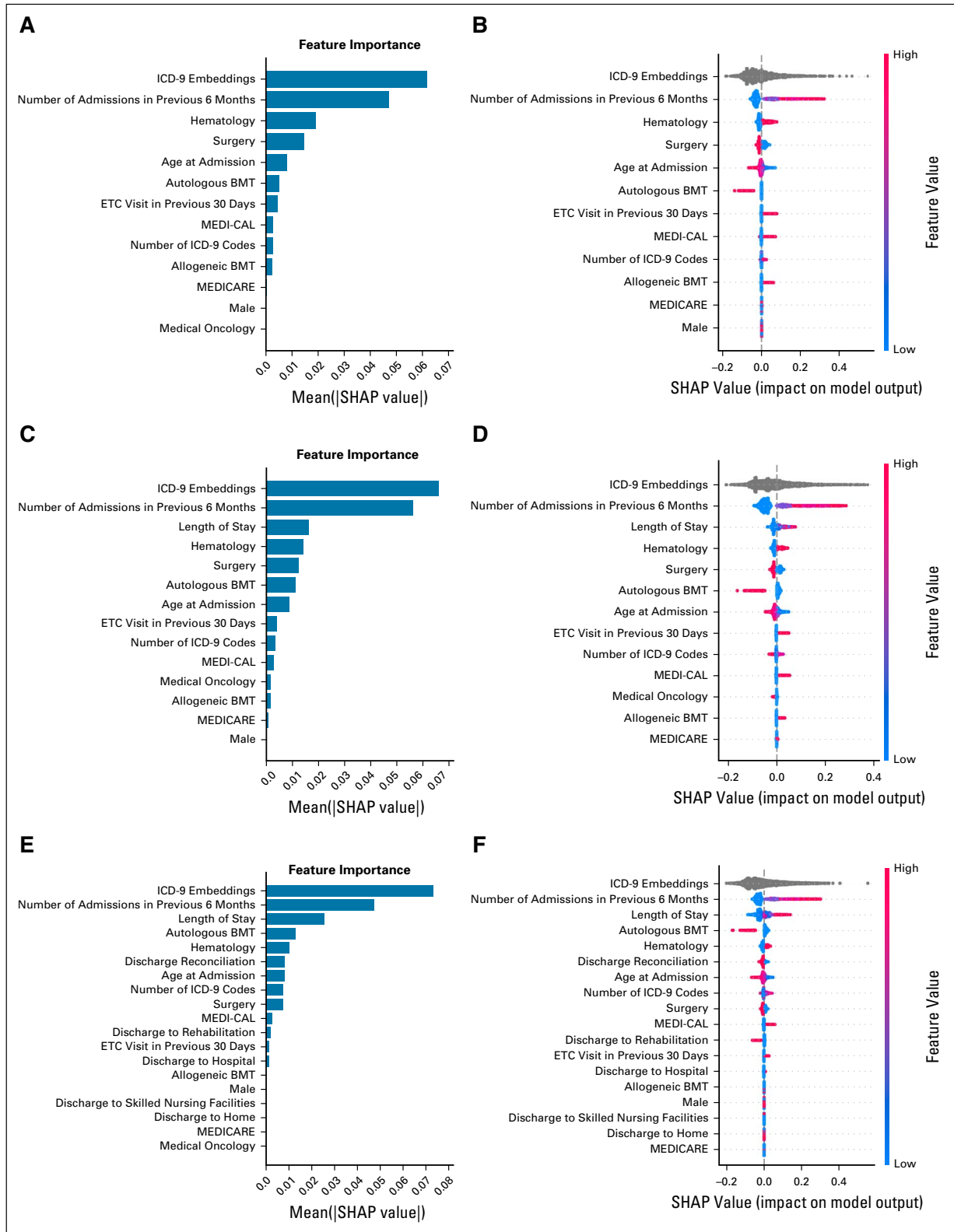
FIG 5. Summary plots combining local explanation values from many individual predictions (scenarios 3-6). (A, C, E, and G) Bar chart showing mean absolute of all shap values of the model: (A) scenario 3, (C) scenario 4, (E) scenario 5, and (G) scenario 6. (B, D, F, and H) Beeswarm plots where in each row (corresponds to one feature), one dot corresponds to one patient in the testing data set. The color of each dot corresponds to a normalized feature value (qualitative: blue for low values and red for high values), whereas its position along the x-axis depicts the shap value describing the impact on the model prediction: (B) scenario 3, (D) scenario 4, (F) scenario 5, and (H) scenario 6. BMT, bone marrow transplant; ETC, emergency treatment center; SHAP, shapley additive explanation.
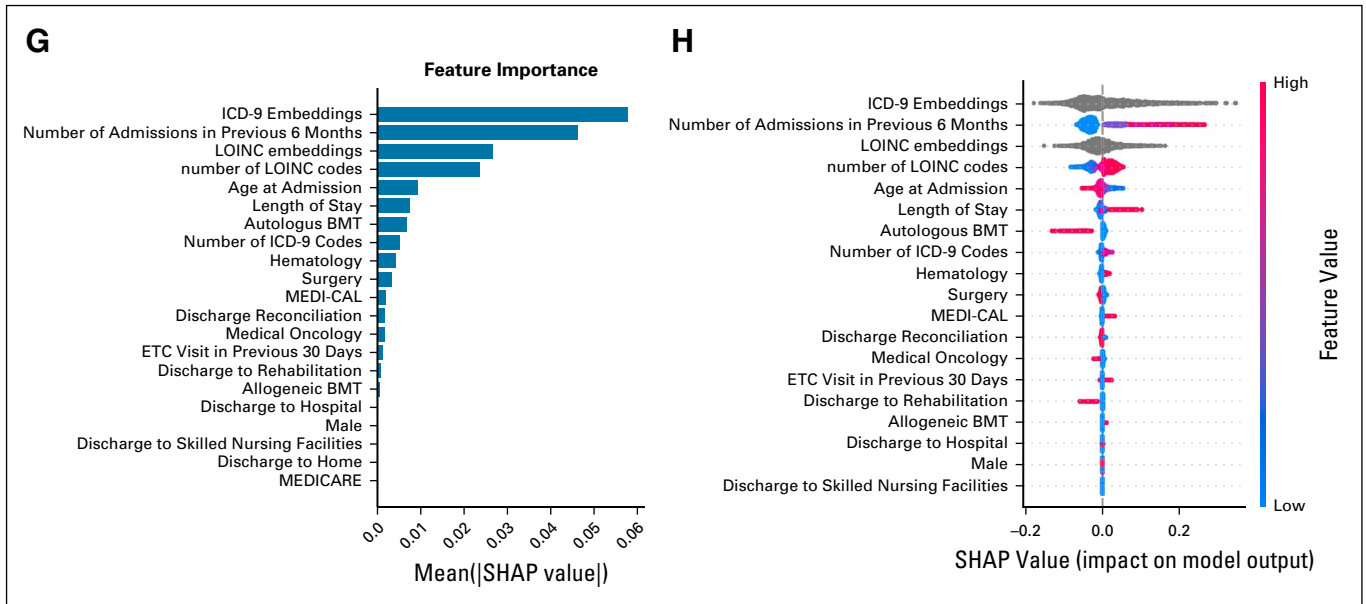
FIG 5. (Continued).

The associated feature importance bar plots and local explanation beeswarm plots are displayed in Figure 5. For the purpose of visualization, the feature values associated with ICD-9 and LOINC embeddings were set to zero.

Model Assessment for Various Cancer Services

Using the same predictive model, Table 2 summarizes testing data set model performance for hematology, surgery, and medical oncology services. We observe more notable difference in AUPRC among the services than the associated AUROC. For instance, in the case where only baseline and discharge factors were used (scenario 2), we observe 5.7% difference in AUROC among services (0.70-0.74), whereas there is 96% difference in AUPRC among services (0.28-0.55). The interservice differences in AUPRC could be partly related to the interservice differences in the rate of unplanned readmission. Surgery has a lower rate of readmission (9.4%) as well as a lower AUPRC (0.28) comparing with hematology (readmission rate = 23.1%, AUPRC = 0.39) and medical oncology

(readmission rate = 25.5%, AUPRC = 0.55). For improvements using clinical embeddings (scenario 2 v 5), we found 5.7% gain for AUROC and 23.1% gain AUPRC for hematology. For surgery, the associated gain was 2.7% for AUROC and 7.1% for AUPRC. For medical oncology, the gain was up to 4.2% for AUROC and only up to 1.8% gain for AUPRC.

DISCUSSION

In this study, we have developed multiple models predicting risk of readmission at different time points throughout the duration of an inpatient visit. Individual prediction explanation is available shortly after admission (scenarios 1 and 3). Dynamic predictions are then provided throughout the course of the visit (scenarios 4), and a final prediction is available around time of discharge (scenarios 2, 5, and 6). Predictive performance at admission is modestly lower than at discharge. However, performing prediction at admission has the advantage of estimating individual's readmission risk early in the visit and makes

TABLE 2. Testing Data Set Model Assessment for Various Combinations of Input Factors

Scenario	1	2	3	4	5	6
Area under receiver operating characteristic curve						
Hematology	0.67	0.70	0.71	0.71	0.74	0.75
Medical oncology	0.70	0.71	0.69	0.72	0.74	0.73
Surgery	0.67	0.74	0.69	0.69	0.76	0.78
Area under precision-recall curve						
Hematology	0.35	0.39	0.42	0.44	0.48	0.47
Medical oncology	0.51	0.55	0.50	0.54	0.56	0.55
Surgery	0.25	0.28	0.23	0.23	0.30	0.30

potential intervention possible. Readmission risk assessment with model explanations may also be incorporated into discharge planning.

We also found that the inclusion of clinical embeddings improves the predictive performance in most cases. Specifically, higher positive predictive value (precision) would increase clinical utility of our predictive model. By diving into individual departments of hematology, surgery, and medical oncology, we found notably different readmission rate and predictive performance. Performance gains by including clinical embeddings are different for different services, with hematological patients benefiting more than patients in surgery and medical oncology.

Although our study was retrospective in nature, we have performed temporal split on our data set for training and testing. In addition, we used only new patients in the testing data, making sure that no patient in the testing data was present in the training data. This approach minimized the risk of data leakage and ensured that the testing was closer to a prospective setting.

A few studies have been conducted to identify the likelihood of unplanned readmission and the associated risk factors in general medicine and a variety of specialties.[13,20,21,23-34] With the advancement of artificial intelligence and computational technologies, more complex algorithms can be used, and the predictions become more individualized. For instance, using deep learning on a rich set of EHRs, Rajkomar et al.[32] achieved approximately 10% AUROC improvement in predicting 30-day unplanned readmission at discharge than traditional approaches. A more recent work using explainable gradient-boosted trees approach also achieved similar performance, with an AUROC of 0.76 and the ability to generate personalized predictions.[13] For oncological models, Schmidt et al.[21] were the first to develop a statistical model using logistic regression with validation c-statistics of 0.70. However, the prospective analysis was descriptive, and no prospective predictive performance was reported. In addition, the cohort focused on patients in surgery and medical oncology services, and it

was unclear whether the model works on hematology patients. A few studies were found on identifying risk factors related to inpatient hematological unplanned readmission,[23,28,33,34] and, to our knowledge, our study was probably the first to explicitly predict 30-day unplanned readmission for hematology population. At the same time, the highest AUROC achieved for our study using the explainable tree-based model is 0.78.

The majority of studies on machine learning and deep learning modeling used the ROC and the associated area under the ROC as the key metric for assessment. However, the measure could be misleading in cases where the data are imbalanced, leading to overly promising interpretation of predictive performance.[22,35] Precision-based metrics, relying on the proportion of true positives among all positive predictions, is a better alternative for imbalanced data sets. Average precision and precision-recall curve are common alternatives for ROC and should be more widely adopted.

Recent studies using features from embeddings of clinical codes or concepts tend to adopt deep learning approaches.[18,36-38] Our results show that even nondeep learning algorithms, such as gradient-boosted trees, can yield significant performance gains from the inclusion of features from clinical embeddings. With a limited number of patients in a single institution, it seems we were not able to train embeddings that could significantly improve performance for readmission prediction. A multi-institutional study should be conducted across cancer centers, using a larger number of patient visit transactions to develop more complex embedding models that can be used for the oncology population. In addition, external validation for the readmission models should be conducted. Another limitation for clinical utility of this approach is that it is a fixed model trained on historical data. We are in the process of deploying this model in our real-time data infrastructure, which allows for retraining the model regularly with more recent data, monitoring data consistency, data or concept drifts, etc, and presenting predictions and explanations in the EHR.

AFFILIATIONS
[1]Department of Applied AI and Data Science, City of Hope National Medical Center, Duarte, CA
[2]Department of Clinical Informatics, City of Hope National Medical Center, Duarte, CA
[3]Department of Hematology and HCT, City of Hope National Medical Center, Duarte, CA

CORRESPONDING AUTHOR
Chi Wah Wong, PhD, City of Hope National Medical Center, 1500 E Duarte Rd, Duarte, CA 91010; e-mail: alecwong@coh.org.

DATA SHARING STATEMENT
The data associated with the current study is not publicly available as the data are not legally certified as being de-identified. Summary data may be available by contacting corresponding author.

AUTHOR CONTRIBUTIONS
Conception and design: Chi Wah Wong, Zahra Eftekhari
Collection and assembly of data: Chi Wah Wong, Chen Chen
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors
Other: Lorenzo A. Rossi [Clinical embeddings pipeline], Chi Wah Wong [Predictive modeling]

REFERENCES

1. Group, U. C. S. W. CDC. Atlanta, GA, USDHHS, National Cancer Institute, 2013

2. Donzé JD, Lipsitz S, Schnipper JL: Risk factors and patterns of potentially avoidable readmission in patients with cancer. J Oncol Pract 13:e68-e76, 2017

3. Friedman B, Basu J: The rate and cost of hospital readmissions for preventable conditions. Med Care Res Rev 61:225-240, 2004

4. Rocque GB, Barnett AE, Illig LC, et al: Inpatient hospitalization of oncology patients: Are we missing an opportunity for end-of-life care? J Oncol Pract 9, 51-54, 2013

5. Jencks SF, Williams MV, Coleman EA: Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 360:1418-1428, 2009

6. Medicare Payment Advisory Commission: Payment policy for inpatient readmissions. In report to the Congress: promoting greater efficiency in medicare. Washington, DC, MedPAC: 102-120, 2007

7. Azuaje F: Artificial intelligence for precision oncology: Beyond patient stratification. NPJ Precis Oncol 3:6, 2019

8. Prokhorenkova L, Gusev G, Vorobev A, et al: CatBoost: Unbiased boosting with categorical features. Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, Canada, 6639-6649, 2018

9. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794, 2016.

10. Ke G, Meng Q, Finley T, et al: LightGBM: A highly efficient gradient boosting decision tree. Proceedings of the 31st International Conference on Neural Information Processing Systems, 3149-3157, 2017

11. Lundberg SM, Erion G, Chen H, et al: From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2:56-67, 2020

12. Lundberg SM, Nair B, Vavilala MS, et al: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2:749-760, 2018

13. Hilton CB, Milinovich A, Felix C, et al: Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. NPJ Digit Med 3:51, 2020

14. Li R, Shinde A, Liu A, et al: Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival. JCO Clin Cancer Inform 4:637-646, 2020

15. Mikolov T, Sutskever I, Chen K, et al: Distributed representations of words and phrases and their compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems—Volume 2, Lake Tahoe, CA, 3111-3119, 2013

16. Choi Y, Chiu CY, Sontag D: Learning low-dimensional representations of medical concepts. AMIA Jt Summits Transl Sci Proc 2016:41-50, 2016

17. Rossi LA, Shawber C, Munu J, et al: Evaluation of embeddings of laboratory test codes for patients at a cancer center. KDD Workshop on Applied Data Science for Healthcare (DSHealth), 2019

18. Liu W, Stansbury C, Singh K, et al: Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. PLoS One 15:e0221606, 2020

19. Burhenn P, Suna CL, Scherb KS, et al: Predictors of hospital readmission among older adults with cancer. J Geriatr Oncol 11:1108-1114, 2020

20. Ashfaq A, Sant'Anna A, Lingman M, et al: Readmission prediction using deep learning on electronic health records. J Biomed Inform 97:103256, 2019

21. Schmidt CR, Hefner J, McAlearney AS, et al: Development and prospective validation of a model estimating risk of readmission in cancer patients. J Surg Oncol 117:1113-1118, 2018

22. Saito T, Rehmsmeier M: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10:e0118432, 2015

23. Dhakal B, Giri S, Levin A, et al: Factors associated with unplanned 30-day readmissions after hematopoietic cell transplantation among US hospitals. JAMA Netw Open 2:e196476, 2019

24. Jiang W, Siddiqui S, Barnes S, et al: Readmission risk trajectories for patients with heart failure using a dynamic prediction approach: Retrospective study. JMIR Med Inform 7:e14756, 2019

25. Rojas JC, Carey KA, Edelson DP, et al: Predicting intensive care unit readmission with machine learning using electronic health record data. Ann Am Thorac Soc 15:846-853, 2018

26. Socwell CP, Bucci L, Patchell S, et al: Utility of Mayo Clinic's early screen for discharge planning tool for predicting patient length of stay, discharge destination, and readmission risk in an inpatient oncology cohort. Support Care Cancer 26:3843-3849, 2018

27. Fisher AV, Fernandes-Taylor S, Campbell-Flohr SA, et al: 30-day readmission after pancreatic resection: A systematic review of the literature and meta-analysis. Ann Surg 266:242-250, 2017

28. Kunapareddy G, Ahmed H, Patel SS, et al: Predictors for recurrent 30-day unplanned readmissions in patients with hematologic malignancies. Blood 130:532, 2017

29. McCoy A, Das R: Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 6:e000158, 2017

30. Leppin AL, Gionfriddo MR, Kessler M, et al: Preventing 30-day hospital readmissions: A systematic review and meta-analysis of randomized trials. JAMA Intern Med 174:1095-1107, 2014

31. van Walraven C, Wong J, Forster AJ: LACE+ index: Extension of a validated index to predict early death or urgent readmission after hospital discharge using administrative data. Open Med 6:e80-e90, 2012

32. Rajkomar A, Oren E, Chen K, et al: Scalable and accurate deep learning with electronic health records. npj Digital Med 1:18, 2018

33. Jaglowski SM, Ruppert AS, Hofmeister CC, et al: The hematopoietic stem cell transplant comorbidity index (HCT-CI) can predict for readmission following autologous stem cell transplant for lymphoma and multiple myeloma. Blood 120:4286, 2012

34. Mohan SR, Rybicki L, Smith SD, et al: Analysis of readmission after autologous HCT: Predictive factors and clinical consequences. Blood 116:932, 2010

35. Davis J, Goadrich M: The relationship between precision-recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, 233-240, 2006

36. Miotto R, Wang F, Wang S, et al: Deep learning for healthcare: Review, opportunities and challenges. Brief Bioinform 19:1236-1246, 2018

37. Shickel B, Tighe PJ, Bihorac A, et al: Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 22:1589-1604, 2017

38. Xiao C, Choi E, Sun J: Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. J Am Med Inform Assoc 25:1419-1428, 2018

APPENDIX

TABLE A1. Cross-Validation Data Set Model Assessment for Various Combinations of Input Factors

Scenario	1	2	3	4	5	6
Area under receiver operating characteristic curve	0.74	0.76	0.76	0.78	0.80	0.80