

Predicting inpatient pharmacy order interventions using provider action data

Corresponding Author:

Martina Balestra

NYU Center for Urban Science and Progress

370 Jay St.

Brooklyn NY 11201

mbalestra@nyu.edu

415-260-9758

Authors:

Martina Balestra¹

Ji Chen²

Eduardo Iturrate³

Yindalon Aphinyanaphongs^{2,5}

Oded Nov^{1,4}

¹ NYU Center for Urban Science and Progress, Brooklyn, New York, U.S.A.

² Department of Population Health, NYU Grossman School of Medicine, New York, NY, U.S.A.

³ Department of Medicine, NYU Langone Health, New York, NY, U.S.A.

⁴ Department of Technology Management and Innovation, Tandon School of Engineering, New York University, Brooklyn, NY, U.S.A.

⁵ Center for Healthcare Innovation and Delivery Science, NYU Langone Health, New York, NY, U.S.A.

LAY SUMMARY

The widespread deployment of electronic health records (EHR) has introduced new sources of error to the process of ordering medications in the hospital setting. Prior research identifies order errors by comparing orders against a patient's medical records, but this approach may miss certain sources of error and may put patient data privacy and security at risk. In this work we develop a machine learning model for identifying medication orders that require pharmacy intervention using only provider behavior and other contextual features that may reflect these new sources of inefficiencies. We collected this data from the EHR system of a major metropolitan hospital system over a 2-week period and trained a gradient boosted tree model. The resultant model had an area under the ROC curve of 0.91 and an area under the precision-recall curve of 0.44. We tune the model to the context in which it would be deployed, and evaluate global and local feature importance. We find that providers' actions can serve as useful predictors in identifying suboptimal medication orders, and careful tuning to the clinical context in which the model is deployed can help to create improved health outcomes.

ABSTRACT

Objective: The widespread deployment of electronic health records (EHR) has introduced new sources of error and inefficiencies to the process of ordering medications in the hospital setting. Existing work identifies orders that require pharmacy intervention by comparing them to a patient’s medical records. In this work we develop a machine learning model for identifying medication orders requiring intervention using only provider behavior and other contextual features that may reflect these new sources of inefficiencies.

Methods: Data on providers’ actions in the EHR system and pharmacy orders were collected over a 2-week period in a major metropolitan hospital system. A classification model was then built to identify orders requiring pharmacist intervention. We tune the model to the context in which it would be deployed, and evaluate global and local feature importance.

Results: The resultant model had an area under the ROC curve of 0.91 and an area under the precision-recall curve of 0.44.

Conclusions: Providers’ actions can serve as useful predictors in identifying medication orders that require pharmacy intervention. Careful model tuning for the clinical context in which the model is deployed can help to create an effective tool for improving health outcomes without using sensitive patient data.

INTRODUCTION

Significant amounts of heterogeneous data from electronic health record (EHR) platforms have led to a surge in interest in using machine learning in clinical decision support (CDS) tools [1]. This has resulted in the development of tools that provide clinicians with information which, when embedded at the appropriate point in their workflows, can improve healthcare [2, 3, 4, 5]. CDS tools encompass a variety of systems that can assist in the interpretation, diagnosis, and treatment of patients through the use of various tools, including: alerts and reminders, clinical guidelines, recommendations, order sets, patient data reports and dashboards, documentation templates, diagnostic support, and other clinical workflow tools [3, 4].

One interesting application area lies in detecting medication order errors [6]. Medical errors (including invasive procedures and hospital-acquired infections as well as those involving drugs and medical devices) are a significant public health problem and a leading cause of death [7]. For medication order errors, manual review of incoming pharmacy orders is the ‘gold standard’ [8] for improving the use of medications and minimizing prescribing errors [9, 8], though a series of recent studies have shown that pharmacies in both inpatient and outpatient settings are often understaffed [10, 11, 12, 13]. As a result, pharmacists experience a high degree of burnout [10] and are at higher risk of making errors or not detecting problems with incoming orders. Supporting clinical pharmacists with CDS tools could therefore improve health outcomes for patients and effectively help pharmacists in under-resourced institutions to manage the load of detecting and correcting orders requiring intervention.

The widespread deployment of computerized physician order entry systems is believed to have resulted in significant declines in traditional sources of medication order errors [14, 15, 5]. Recent research, however, has suggested ways in which these tools may have introduced or contributed to other sources of error (e.g. alert fatigue, orders in the wrong medical records, etc.) [16, 14, 17]. These types of errors tend to fall into two categories [18]: (1) errors in the process of entering and retrieving information (e.g. interfaces that are not suitable for a highly interruptive use context, that produce cognitive overload by requiring structured information entry, that fragment information onto different screens, and interfaces that overemphasize information about a patient that is not useful), or (2) errors that come from a mismatch between the structured communication and coordination processes embedded in digital systems relative to the highly flexible and fluid ways in which clinical work happens in reality.

Prior work has centered on comparing the order to the patient’s medical records, and omits features representing provider behavior. For example, Corny et al. [8] evaluate orders in the medical context of the individual patient’s laboratory reports, demographics, medical and allergy history, and physiological data. Similarly, Segal et al. [19] screen patients’ medical records and corresponding orders to identify time-dependent irregularities, clinical outliers, dosage outliers, and drug overlaps that might be indicative of medication order errors. Nguyen et al. [20] use an alternative approach in which patient information is used to identify those who are at high-risk of receiving an inappropriate therapeutic and to prioritize review of orders made on their behalf.

These approaches may be missing critical signals from the provider’s suboptimal interactions with the EHR leading up to the submission of an order, and therefore entire ‘phenotypes’ of errors. In this paper we develop a predictive model for flagging orders requiring intervention using only information about the ordering provider’s interaction with the EHR. The model is then tuned and evaluated within the clinical context in which it would be deployed.

METHODS

Setting

This work was conducted within a large urban academic hospital system comprised of 3 hospitals (a quaternary care, a tertiary care, and an orthopedic sub-specialty hospital) with over 1,600 beds (combined) and numerous satellite locations.

Pharmacy orders were submitted via the electronic medications management system, EPIC. Auto-verification was implemented only for dietary supplements. Interventions were made by pharmacists in a main dispensing pharmacy or unit-based satellites, as well as by clinical pharmacists that rounded directly with the medical teams in certain specialty areas. The frequency of each type of clinical intervention may vary in different functional areas. For example, a pharmacist rounding with a medical team is more likely to see patient-specific details and optimize a medication therapy plan to improve patient tolerance. Alternatively, a pharmacist dispensing medications from the pharmacy may not have access to as many patient-specific details, but may intervene on general medication issues (e.g. reviewing medication dose, checking for appropriateness given patient renal function/age/weight, advising of allergy risk). There may also be variability in documentation of interventions between pharmacists and at times of increased workload, however supervision of clinical pharmacists and review of interventions by supervisors finds a high rate of adherence to the expected workflow.

Analysis

Data

This study relies on inpatient data from July 10th through July 24th 2017. The dataset included a total of 181,407 individual orders submitted by 2,708 providers. On average, providers submit 4.96 orders at once. We therefore consolidated orders made simultaneously into 36,585 order “batches”. Of these, 2,054 (5.61%) contained at least one order that required intervention. The sample is described in Table 1.

Type	Inpatient
Dates	July 10-24, 2017
# orders	181,407
# order batches	38,966
# order batches requiring intervention	2,054 (5.61%)
# providers	2,708
# departments	183
# therapeutic classes	45
# patients	16,714

Table 1: Description of sample used.

Feature Construction

Model features were determined using descriptive analysis and clinical expertise. Conversations with clinician informants revealed that they experience high administrative workload and a high degree of disruption and fragmentation in their workflows. This can tax working memory, which acts as a temporary storage for task-relevant information (often in the face of distractions) [21, 22, 23]. This is consistent with prior research [14, 17] and informed our selection of 17 features (continuous features described in Table 2).

1. *Measures of clinician engagement with patients and the EHR in the hour preceding the order.* These features capture behaviors that may require the provider to store more information in working memory [22, 23] (e.g. seeing many patients in a short period of time), or periods in which a provider is multitasking or their workflow is disrupted (e.g. needing to communicate information about a patient to a colleague while engaged with a different patient) [24]. Interruptions and multitasking increase the demands on working memory by requiring them to process information unrelated to their primary task, increasing the potential for error [21, 25, 21].
 - Number patient encounters
 - Number workstations accessed by a provider
 - Number general administrative actions the provider engaged in in the EHR (e.g. reviewing a patient list, checking messages, using the chat tool).
 - Number administrative actions in a patient's EHR (e.g. reviewing a prior patient's history) outside of the current patient encounter
 - Number of administrative actions in the EHR related to a specific patient during an in-person visit (e.g. reviewing their chart)
2. *Details of the orders contained in the order batch.* Creating orders is one of the more complex parts of the order-prescribing task [21] and taxes a provider's working memory (e.g. when choosing clinical elements for an order), particularly when they are being made for multiple patients. This is especially true when the provider is disrupted or distracted as they are placing the order, increasing the probability of making an error or sub-optimal order.
 - Number of orders in batch (submitted at once)
 - Number of orders in a batch using a predefined order template, called ordersets
 - Number of medications in a batch which are re-orders of a prior medication order, called reorders
 - Number of orders in a batch intended to keep a patient's medications up-to-date, commonly from old or outdated prescriptions, when the patient's care is transferred from one context to another (e.g. when transferred from the ambulatory context to the inpatient context or from one team in the hospital to another), called reconciliations
 - Number of time sensitive ("STAT") orders in a batch
 - Number of patients for whom orders within a batch are being made
3. Contextual data related to the clinician and the order. The context in which orders are made may influence providers' behavior; for example, clinicians in certain specialties (e.g. Emergency Department) may experience higher levels of cognitive load. We include the following contextual features:
 - Ordering clinician type (e.g. Nurse Practitioner, Resident)

- Ordering clinician specialty (e.g. Emergency Medicine, Oncology)
- Day of the week
- Time of day (e.g. early morning, late morning)
- Hospital
- Order therapeutic class (e.g. Antivirals, Antibiotics)

Feature	Mean	Median	Std. Dev.
# administrative actions in hour preceding order	15.42	2	29.62
# administrative actions in patient files in hour preceding order	30.26	7	46.30
# orders in batch	3.09	2	3.22
# ordersets in batch	0.05	0	0.44
# actions related to patient encounters in hour preceding order	49.65	8	79.81
# patients in batch	1.07	1	0.31
# reconciliations in batch	1.23	0	3.96
# STAT orders in batch	0.68	0	1.80
# unique patient encounters in hour preceding order	3.63	1	5.70
# unique workstations	1.62	1	1.06

Table 2: Descriptive statistics of continuous features included in the model

Machine Learning Model

The model was designed as a binary classification task where the target labels corresponded to whether an order batch required intervention.

The data were split randomly into 70%, 15%, and 15% sets for model training, validation, and testing, respectively. Logistic regression with L1 and L2 regularization were used as baselines, and gradient-boosted trees (XGBoost) were employed as our focal ML algorithm. Cross-validation was used to identify the λ penalty in the regressions with regularization where the value selected gives the simplest model but also lies within one standard error of the optimal value of lambda ($\lambda = 0.013$ and $\lambda = 0.28$, respectively).

The XGBoost model was implemented with nested 5-fold cross-validation with early stopping (maximum of 50 rounds). Grid search was used to tune model hyperparameters, resulting in a maximum tree depth = 7, minimum child weight = 1, subsampling = 0.84, evaluation metric = auc, eta = 0.17, gamma = 0.32. The validation set was used to monitor the model training through early stopping to evaluate the model performance and the effectiveness of the decision boundary; the test set was then used to evaluate the generalizability of the trained model. Due to a significant class imbalance in the training data (5.53% of batches required intervention), synthetic minority oversampling technique (SMOTE) was used to generate a pseudo-synthetic training set with 5,672 order batches in which 2,836 (50%) required intervention.

RESULTS

The XGBoost algorithm outperformed both of the logistic regressions as well as the random forest algorithm by a significant margin in both area under the receiver-operator (AUROC) and precision-recall (AUPR) curves (Table 3). This indicates complex and non-linear relationships between predictors that are not captured by linear classifiers or simple decision trees.

Model	AUROC	AUPR
Logistic regression with L1 (Lasso) regularization	0.528	0.276
Logistic regression with L2 (Ridge) regularization	0.530	0.278
Random Forest with pruning	0.579	0.180
Extreme gradient boosted trees (XGBoost)	0.908	0.439

Table 3: Model performance metrics for baseline (Lasso, Ridge, and Random Forest regression) and focal models (XGBoost).

The trained XGBoost model had an AUROC of 0.908 (Figure 1); however, the high AUROC performance is likely a result of the imbalance data set. In this case, the Precision-Recall curve can offer us a more accurate representation of model performance (visualized in Figure 2). The AUPR curve is 0.439. The resultant Lift curve shows that flagging a small fraction of orders for review results in the detection of a large number of the orders that required intervention (e.g. selecting the top 20% of orders according to the model results in approximately 4 times the total number of orders requiring intervention being identified). This is visualized in Figure 3.

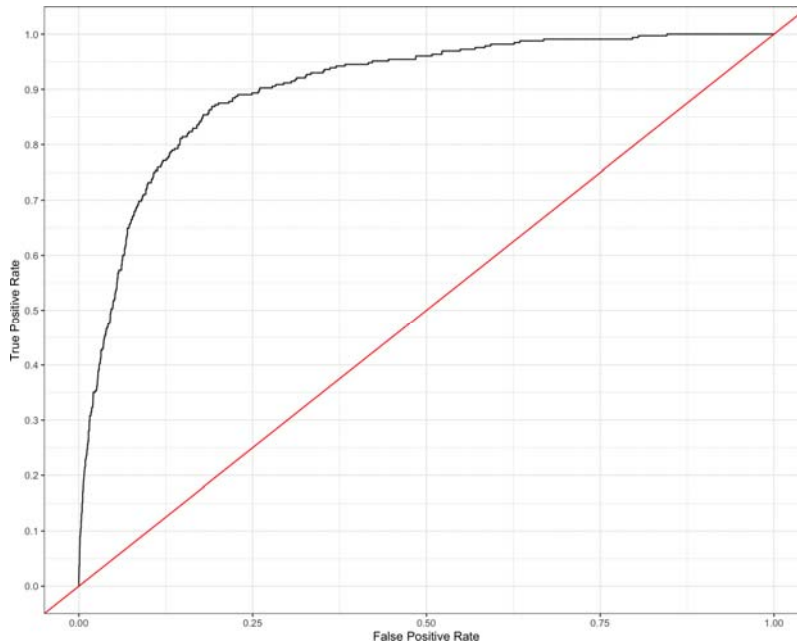


Figure 1: Average Receiver-Operator Characteristic (ROC) curve. AUROC is 0.908.

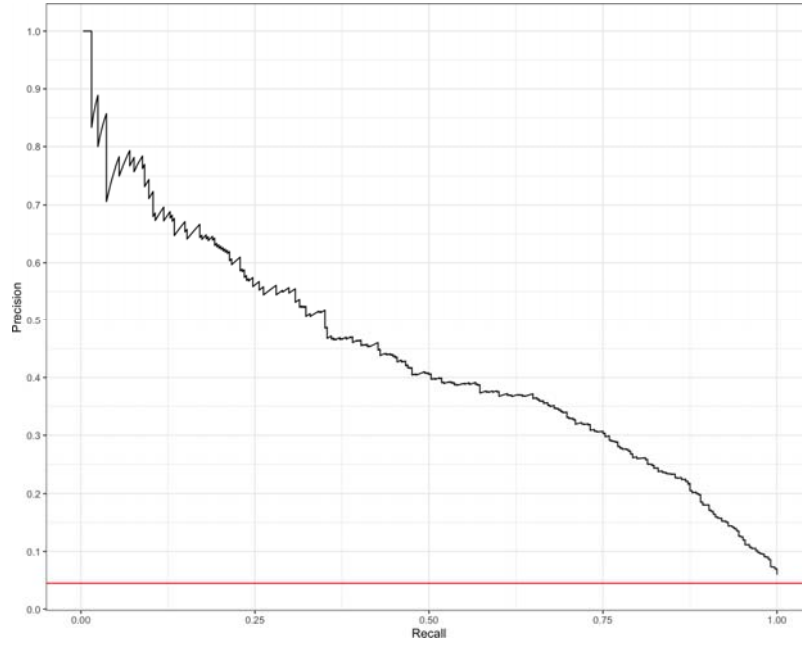


Figure 2: Precision-Recall curve. AUPR is 0.439.

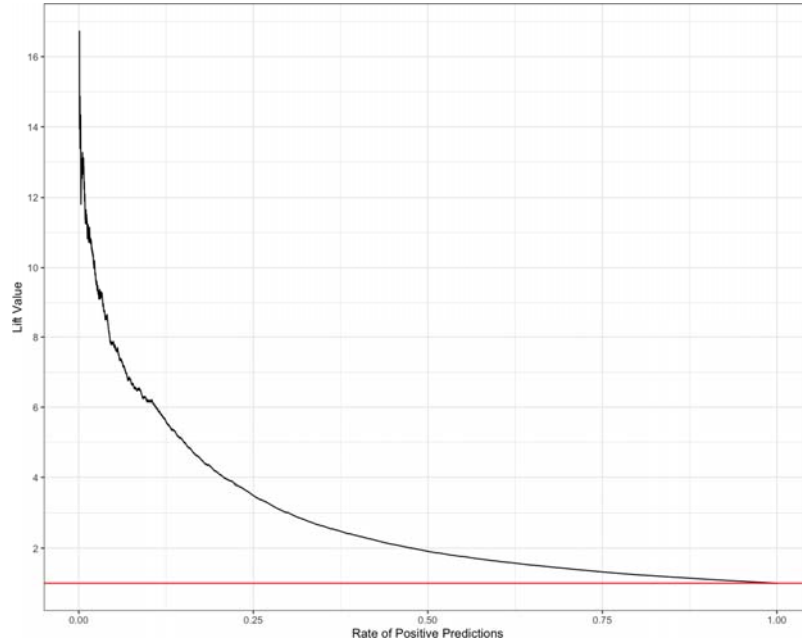


Figure 3: Lift curve

The choice of decision boundary will influence the performance of the model in the context of the clinical workflow. We therefore want to consider constraints on model error that exist at different stages in the workflow, then evaluate the model performance within the context of each of these stages. The hospital system on which this analysis is centered follows a well-established, digitally-mediated process for prescribing and dispensing medications in the inpatient setting that involves the verification of all orders that are placed by clinicians to the pharmacy. Based on this workflow, we evaluate the model at two potential points:

1. *Intervene at the point of order submission by the clinician*; for example, alerting the clinician at the time of order entry that it may contain an error or require optimization. In this scenario, a high number of incorrectly flagged orders (false positives) could become burdensome and lead the provider to distrust or disregard the system [26, 27, 28]. Within these constraints the precision rate should be relatively high, but the precision-recall curve for our model (Figure 1) suggests that this model may not adequately minimize both false positives and false negatives (AUPR= 0.439); we therefore consider the alternative:
2. *Intervene at the point of order receipt by the pharmacist*; for example, by removing the requirement for pharmacist verification on incoming orders that the model has identified as having a low risk of requiring an intervention. Clinical pharmacists currently review incoming orders assuming equal probability of requiring an intervention. Their order verification queue could be reduced by de-prioritizing these orders. An intervention targeting pharmacists therefore has a requirement for a low false negative rate, but pharmacists can likely tolerate a higher rate of false positives (relative to prescribing clinicians) because they already review a large number of orders that do not require intervention in the current workflow so any reduction in this workload is an improvement.

In light of this, we compute the classification threshold for the model that optimizes for model recall performance using the F10 score (0.076). Table 4(a) displays the corresponding confusion matrix and other model performance metrics.

Though the model accuracy is only 0.41, it has a sensitivity of 0.99 and specificity of 0.37 on the test set. The associated confusion matrix shows a very low number of false negatives (2) when de-prioritizing 35% (1,933 + 2) of orders identified by the model as not requiring intervention.

(a)		<i>Actual</i>		<i>Model Performance Metrics</i>			
<i>Decision boundary = 0.076</i>		Intervention	No Intervention	Accuracy	Recall	Specificity	Precision
Prediction	Intervention	312	3,241	0.41	0.99	0.37	0.09
	No Intervention	2	1,933				
(b)		<i>Actual</i>		<i>Model Performance Metrics</i>			
<i>Decision boundary = 0.5</i>		Intervention	No Intervention	Accuracy	Recall	Specificity	Precision
Prediction	Intervention	242	606	0.88	0.77	0.88	0.29
	No Intervention	72	4,568				
(c)		<i>Actual</i>		<i>Model Performance Metrics</i>			
<i>Decision boundary = 0.83</i>		Intervention	No Intervention	Accuracy	Recall	Specificity	Precision
Prediction	Intervention	130	120	0.94	0.41	0.98	0.52
	No Intervention	184	5,054				

Table 4: Confusion matrix associated with decision boundaries displayed on the left side of the double lines in the table, and the corresponding model performance metrics are displayed on the right. (a) Displays model performance associated with the selected decision boundary (0.076). (b) displays the model performance with a decision boundary of 0.5; and (c) displays the model performance with a decision boundary of 0.83.

To demonstrate how alternative thresholds would perform consider (a) 0.5 and (b) 0.83 (model performance metrics in Table 4 (b, c)). The model performance metrics associated with threshold (a) (Table 4(a)) show an accuracy and specificity rate of 0.88. This is a significant increase in specificity compared with the selected threshold (0.37) and corresponds to a substantial decrease in false positives (from 3,241 to 606 with threshold (b)). However, the number of false negatives increases from 2 to 72, corresponding to a decrease in sensitivity from 0.99 to 0.77 with threshold (b). These trends continue when the decision boundary is increased to 0.83 (Table 4 (c)). Though the percentage of orders that can be de-prioritized increases from 85% (with threshold (a)) to 95% (with threshold (b)), there is a simultaneous and substantial increase in false negatives (the number of errors that are incorrectly labeled as not requiring intervention) - well beyond what we have identified as an acceptable rate of missing errors (maximum 5%).

Finally, we examine the global and local feature importance of model features. In Figure 4 we display the 20 features with the highest gain. Whether the provider is a Resident has the greatest impact, followed closely by the number of reconciliations contained in an order batch. The number of individual orders in a batch and orders for antibiotics have moderate gain, whereas remaining features have relatively low influence over the model at the aggregate level.

To understand how these features impact individual predictions, we visualize local feature importance (Figures 5(a) to 6(b)) as the log-odds of an order requiring pharmacist intervention with the addition of each feature to the model. The baseline intercept corresponds to the naive probability that any single order batch requires intervention (5.61%). Figure 5(a) visualizes these effects for a randomly sampled order from the test set where the outcome of the model resulted in a true positive, Figure 5(b) visualizes a true negative, Figure 6(a) visualizes a false positive, and Figure 6(b) visualizes a false negative. The y-axis corresponds to the probabilities associated with the calculated log-odds, red bars correspond to a decrease in the log-odds of an order requiring intervention, and the blue bars signify an increase in the log-odds. The order represented in Figure 5(a) contains two cardiovascular medications, which is associated with an increase in the log-odds of this order requiring intervention of 0.65 over the baseline rate (5.61% in the sample). The plot also showed significant increases in the log-odds attributed to the number of reconciliations contained in the batch, orders for antibiotics, and providers who are residents (+0.55, +0.46, and +0.44, respectively). The number of patients the provider saw in the hour preceding the submission of the order and the day of the week (Monday) were among the features that reduced the log-odds of the order requiring intervention (by -0.45 and -0.41, respectively).



Figure 4: Top 20 features with highest global importance. Gain represents the improvement in accuracy brought by a feature to the branches it is on.

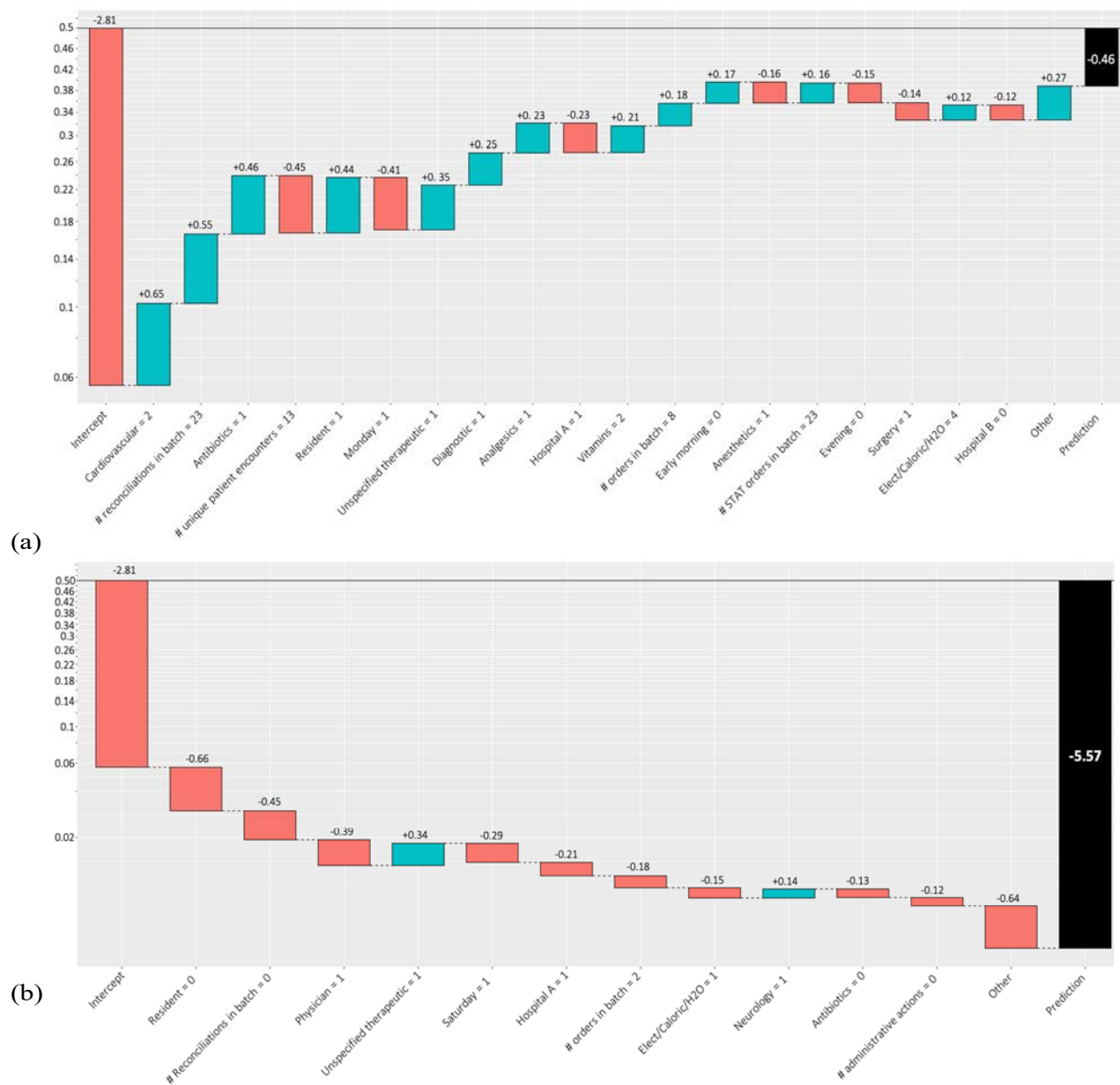
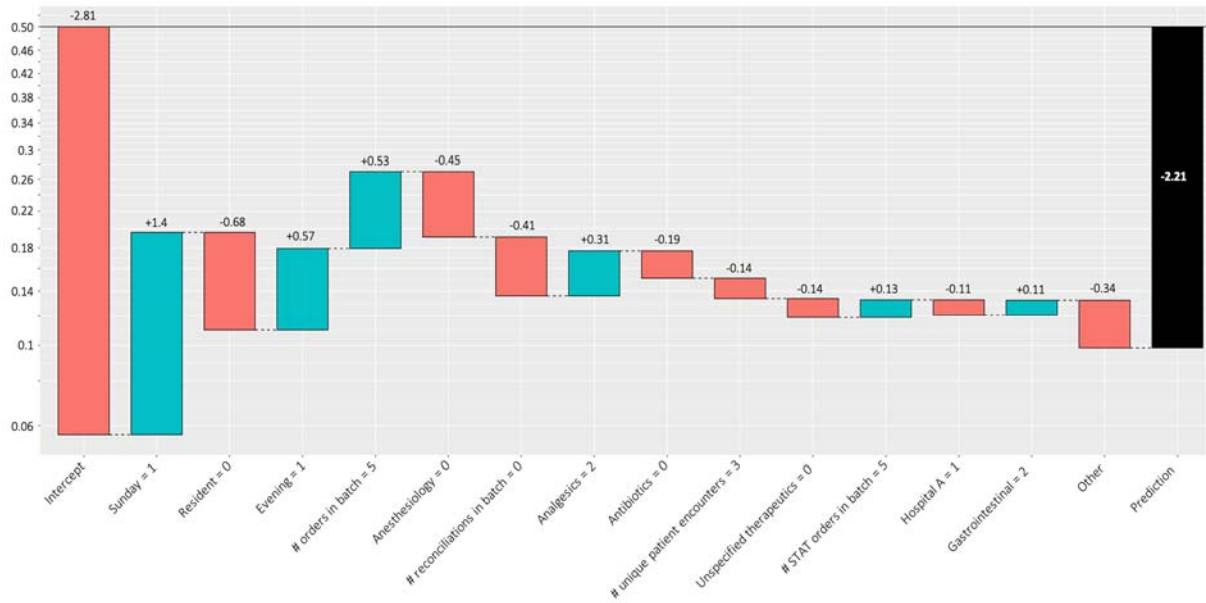


Figure 5: The y-axis represents the change in the probability of the order requiring intervention. Red bars represent a decrease in the log-odds, whereas blue bars represent an increase. (a) Contribution of features to the log-odds of a true positive order. The estimated probability of this order requiring intervention is 0.39, well above the 0.076 decision boundary and consistent with the observed outcome. (b) Contribution of features to the log-odds of a true negative order. The estimated probability of this order requiring intervention is 0.004, below the 0.076 decision boundary.

(a)



(b)

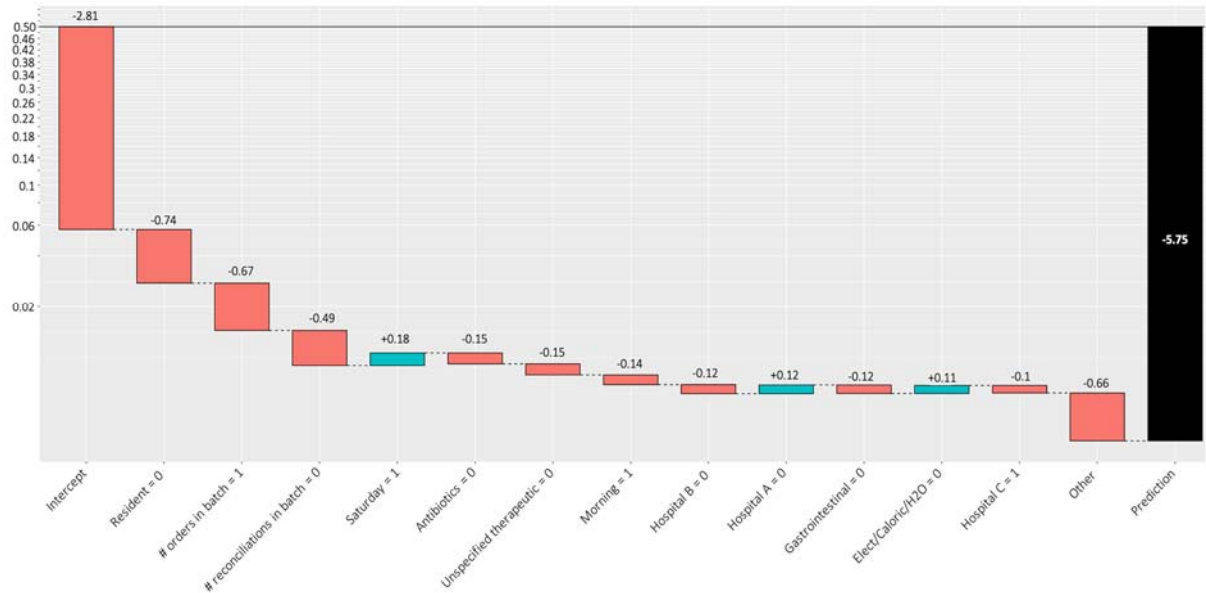


Figure 6: The y-axis represents the change in the probability of the order requiring intervention. Red bars represent a decrease in the log-odds, whereas blue bars represent an increase. (a) Contribution made by individual features to the log-odds of a false positive order. The estimated probability of this order requiring intervention is 0.10, which is above the 0.076 decision boundary, though this particular order was not observed to require intervention. (b) Contribution made by individual features to the log-odds of a false negative order. The estimated probability of this order requiring intervention is 0.003. Despite being well below the 0.076 decision boundary, this particular order did require intervention.

DISCUSSION

The widespread deployment of electronic health records and computerized order entry systems have largely reduced medication order errors and inefficiencies in the inpatient setting. Emerging research suggests, however, that they have also introduced new sources of error related to the interaction between the provider and the platform. Despite this, prior work has centered on the compatibility of the medication ordered with the patient's medical records. We instead developed a machine learning model using provider behavioral data and other contextual features related to the usage of these systems. We collected data on providers' actions in the EHR and pharmacy orders over a 2-week period within the hospital system. We then built a classification model to identify incoming orders requiring pharmacist intervention and

tuned the model considering the constraints of the pharmacist’s workflow. The resultant model had an AUROC curve of 0.91 and an AUPR curve of 0.44.

Though the precision-recall curve in Figure 2 suggests that false positive and negatives are not entirely separable, we show that by strategically tuning the model to the clinical context in which it is deployed we can still bring significant value to users. The model performance metrics (Table 4 (a)) show a very low number of false negatives (2) when de-prioritizing 35% ($1,933 + 2$) of orders deemed by the model as not requiring intervention. While this reduction may not be as meaningful to a well-resourced institution, many hospitals have understaffed pharmacies. In these settings, such an intervention may be critical for managing the workload on pharmacy staff. Future work may further improve the accuracy of the model by taking into consideration the severity of the intervention or stratifying by type of intervention. This may improve the model’s utility in a broader range of clinical contexts.

The features that have the highest gain in our model (Figures 5(a) to 6(b)) are consistent with clinical experience. For example, Residents are often the least experienced class of provider and can have high workloads. This could account for their orders having a higher likelihood of requiring intervention. We also see order reconciliations providing substantial gain in this model. The reconciliation of medications involves reviewing medications that patients were receiving at an earlier phase of care (either as an outpatient prior to being admitted or on a different unit prior to being transferred) and decide which ones to continue and discontinue depending on the change of clinical context. Clinicians may anchor on the patient’s previous prescriptions, biasing them towards continuing these medications and not thoughtfully assessing whether the change in context should correspond to an adjustment.

The total number of orders in a batch might correlate to a higher rate of intervention because each individual order has some baseline probability of containing an error and the probability of requiring intervention increases for batches containing more orders. It could also represent cognitive overburdening of the submitting clinician; orders often accumulate as the clinician is rounding and are submitted together at the end, providing opportunity for errors associated with switching among patients and their different needs. Finally, antibiotic ordering is understood to be complex and many institutions have structured guidelines and oversight groups that direct their approach to ordering antibiotics. It is thus not a surprise that antibiotic ordering is associated to increased gain in our model.

This modeling approach offers us a novel perspective on the factors influencing order entry by focusing on the behavior of the provider and errors that arise from the workflow around the EHR. Whereas previous models predicting errors ingest patients’ medical records, by focusing on the behavior of the clinician we also reduce the risk to the privacy and security of these patients’ data while still being useful to pharmacists.

LIMITATIONS & FUTURE WORK

In our analysis, we used data from two weeks in July, 2017. While we believe that this sample is sufficient to demonstrate the use of clinician behavior and other contextual features in predicting the occurrence of order errors, there may be seasonality and other effects that limit the generalizability of the model. For example, July is the start of the year for medical and surgical trainees. This introduces the possibility of increased order errors due to new interns being less knowledgeable about medications, the EHR system, and about processes in busy health systems [29]. Further analysis using longer periods of data collection is needed to account for new trainees’ evolving clinical experience and knowledge.

Future work may also consider comparing the results of these models across hospital systems to gain a better understanding of how different clinical contexts influence the occurrence of such errors; however, each hospital system is likely idiosyncratic in terms of the culture, processes, and procedures which may affect the occurrence and treatment of order errors. We therefore believe that models for detecting these errors will likely need to be tuned for the specific hospital in which they are deployed. Additional refinement of order error by type (e.g. improper dosage, incorrect medication, etc.) may bring further precision to these models, and improve their utility to the clinician in context.

The implementation and deployment of this model in the hospital setting remains an open question; specifically, what features are available at the time the batch of orders is submitted to run the model and provide decision support. Three types of features outlined in the analysis section include measures of HER clinical engagement, details of orders, and contextual data related to clinician and order. Both of the latter feature sets are readily available at order time. The most difficult feature is the former (EHR clinical engagement) that are not routinely available in real time from our EHR vendor. A custom build from the EHR would be required to access these data elements. Future work could analyze more sub-feature sets to see how to maintain performance given features more readily available.

CONCLUSION

Errors involving medications are a significant challenge which is well suited to interventions based on machine learning models. Conventional approaches have opted to compare medication orders to the contents of the patient's medical records and does not consider sources of error that are produced by EHRs and reflected in the provider's behavior.

In this work, we develop a machine learning model that relies only on features related to the provider's behavior and basic contextual information related to the order to demonstrate a well-performing model. Further, we show that with proper tuning, such models can significantly improve the workload on pharmacists without risking the privacy and security of sensitive patient data.

FUNDING STATEMENT

This work was supported by the National Science Foundation grant number 1928614

COMPETING INTEREST STATEMENT

The authors have no competing interests to declare.

CONTRIBUTORSHIP STATEMENT

All authors provided critical feedback and helped to shape the research, analysis, and manuscript. M.B., Y.A., and O.N. conceived and planned the analysis. J.C. and E.I. collected the data. M.B. carried out the development of the ML model with input from J.C., E.I., Y.A. and O.N. M.B., E.I., Y.A., and O.N. contributed to the interpretation of the results; E.I. contributed clinical insight. M.B. took the lead in writing the manuscript with input from all co-authors.

DATA AVAILABILITY

The data that support the findings of this study are not available due to the sensitive context of this study.

REFERENCES

- [1] Steven G Johnson et al. “Application of an ontology for characterizing data quality for a secondary use of EHR data”. In: *Applied clinical informatics* 7.1 (2016), p. 69.
- [2] Saif Khairat et al. “Reasons for physicians not adopting clinical decision support systems: critical analysis”. In: *JMIR medical informatics* 6.2 (2018), e24.
- [3] Mark A Musen, Blackford Middleton, and Robert A Greenes. “Clinical decision-support systems”. In: *Biomedical informatics*. Springer, 2014, pp. 643–674.
- [4] Jerome A Osheroff et al. “A roadmap for national action on clinical decision support”. In: *Journal of the American medical informatics association* 14.2 (2007), pp. 141–145.
- [5] Clare L Tolley et al. “Factors contributing to medication errors made when using computerized order entry in pediatrics: a systematic review”. In: *Journal of the American Medical Informatics Association* 25.5 (2018), pp. 575–584.
- [6] “Patient Safety Primer: Systems Approach”. In: *Patient Safety Network* (Sept. 2019).
url:<https://psnet.ahrq.gov/primer/systems-approach>.
- [7] Martin A Makary and Michael Daniel. “Medical error—the third leading cause of death in the US”. In: *Bmj* 353 (2016).
- [8] Jennifer Corny et al. “A machine learning–based clinical decision support system to identify prescriptions with a high risk of medication error”. In: *Journal of the American Medical Informatics Association* (2020).
- [9] David M Benjamin. “Reducing medication errors and increasing patient safety: case studies in clinical pharmacology”. In: *The Journal of Clinical Pharmacology* 43.7 (2003), pp. 768–783.
- [10] Paul R Muir and Diana A Bortoletto. “Burnout among Australian hospital pharmacists”. In: *Journal of Pharmacy Practice and Research* (2007), pp. 187–189.17
- [11] Tana Wuliji. “Current status of human resources and training in hospital pharmacy”. In: *American Journal of Health-System Pharmacy* 66.5 Supplement3 (2009), s56–s60.
- [12] Ellen Gabler. “At Walgreens, Complaints of Medication Errors Go Missing”. In: *New York Times* (Feb. 2020). url:<https://www.nytimes.com/2020/02/21/health/pharmacies-prescription-errors.html>.
- [13] Ellen Gabler. “How chaos at chain pharmacies is putting patients at risk”. In: *New York Times* (Jan.2020). url:<https://www.nytimes.com/2020/01/31/health/pharmacists-medication-errors.html>.
- [14] Ross Koppel et al. “Role of computerized physician order entry systems in facilitating medication errors”. In: *Jama* 293.10 (2005), pp. 1197–1203.
- [15] David W Bates et al. “Effect of computerized physician order entry and a team intervention on prevention of serious medication errors”. In: *Jama* 280.15 (1998), pp. 1311–1316.
- [16] GD Schiff et al. “Computerised physician order entry-related medication errors: analysis of reported errors and vulnerability testing of current systems”. In: *BMJ quality & safety* 24.4 (2015), pp. 264–271.
- [17] Anne Bobb et al. “The epidemiology of prescribing errors: the potential impact of computerized pre-scriber order entry”. In: *Archives of internal medicine* 164.7 (2004), pp. 785–792.
- [18] Joan S Ash, Marc Berg, and Enrico Coiera. “Some unintended consequences of information technology in health care: the nature of patient care information system-related errors”. In: *Journal of the American Medical Informatics Association* 11.2 (2004), pp. 104–112.

- [19] Gad Segal et al. “Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine learning based clinical decision support system in an inpatient setting”. In: *Journal of the American Medical Informatics Association* 26.12 (2019), pp. 1560–1565.
- [20] Tri-Long Nguyen et al. “Improving medication safety: Development and impact of a multivariate model-based strategy to target high-risk patients”. In: *PloS one* 12.2 (2017), e0171995.
- [21] Johanna I Westbrook et al. “Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study”. In: *BMJ quality & safety* 27.8 (2018), pp. 655–663.
- [22] Alan Baddeley. “Working memory”. In: *Science* 255.5044 (1992), pp. 556–559.
- [23] Alan Baddeley. *Working memory, thought, and action*. Vol. 45. OuP Oxford, 2007.
- [24] Patricia Potter et al. “Understanding the cognitive work of nursing in the acute care environment”. In: *JONA: the journal of nursing administration* 35.7 (2005), pp. 327–335.18
- [25] Lily Thomas, Patricia Donohue-Porter, and Joanna Stein Fishbein. “Impact of interruptions, distractions, and cognitive load on procedure failures and medication administration errors”. In: *Journal of Nursing Care Quality* 32.4 (2017), pp. 309–317.
- [26] C William Carspecken et al. “A clinical case of electronic health record drug alert fatigue: consequences for patient outcome”. In: *Pediatrics* 131.6 (2013), e1970–e1973.
- [27] Allison B McCoy et al. “Clinical decision support alert appropriateness: a review and proposal for improvement”. In: *Ochsner Journal* 14.2 (2014), pp. 195–202.
- [28] Sundas Khan et al. “Improving provider adoption with adaptive clinical decision support surveillance: an observational study”. In: *JMIR human factors* 6.1 (2019), e10245.
- [29] David P Phillips and Gwendolyn EC Barker. “A July spike in fatal medication errors: a possible effect of new medical residents”. In: *Journal of general internal medicine* 25.8 (2010), pp. 774–779.