Dec 11th, 12:00 AM

Analyzing Patients' EHR: Predicting and Explaining Admission Consequences for COPD and Liver Disease Patients

Ofir Ben-Assuli
Ono Academic College, ofir.benassuli@gmail.com

Gaya Geva
Ono Academic College, geva.gaya@gmail.com

Analyzing Patients' EHR: Predicting and Explaining Admission Consequences for COPD and Liver Disease Patients

Short Paper

Ofir Ben-Assuli
Ono Academic College
Tzahal 104, Kiriat Ono, Israel
ofir@ono.ac.il

Gaya Geva
Ono Academic College
Tzahal 104, Kiriat Ono, Israel
geva.gaya@gmail.com

Abstract

This study analyzed the admission outcomes in chronic patients (with COPD, and Liver disease) to demonstrate the feasibility of applying prediction methods on EHR records while incorporating an explainable AI technique. We predicted three target variables: 30-day readmission, Medium&Long Length of Stay and Single-day admission and analyzed the features using an explainable AI technique, the SHapley Additive exPlanations (SHAP). The results show that Readmission had higher prediction scores than all other dependent variables. Some features affected all target variables with either positive or negative influence including: Age, Charlson comorbidity index, Day-Shift, Gender, using EHR screens and Insurance cover level. These findings thus point to the value of using Machine-Learning combined with an explainable AI method to understand and assess the risks factors. The assessment of the potential factors leading to multiple complications can bolster prevention-oriented medical decisions to groups of patients but can also be tailored to the patient level.

Keywords: Data-Mining; Readmissions; Chronic Obstructive Pulmonary Disease; Liver disease

Introduction

Electronic Health Records (EHR) is a central hospital-based information system used also in health information system research (Bardhan et al. 2020). EHR data, when incorporated with Machine Learning (ML) models, makes crucial medical information available that can save lives in many ways (Lin et al. 2017). Previous studies have reported a significant association between EHR use and better clinical consequences (Ben-Assuli and Vest 2020; Vest et al. 2021) as well as patient care and clinical management (Nguyen et al. 2014; Yaraghi et al. 2014). These include treating chronic diseases such as Chronic Obstructive Pulmonary Disease (COPD) and liver disease which were used as the test cases here. Various methodologies and ML capabilities are highly recommended for management of chronic disease patients (Bardhan et al. 2020). Numerous IS works have implemented EHR data using advanced ML methods (Bardhan et al. 2015; Ben-Assuli and Padman 2020; Lin et al. 2017), as we suggest doing here by exploiting the recent development of explainable artificial intelligence (AI). Explainable AI is a growing area of research that aims to create more transparent and interpretable AI systems. This study applied predictive analytics methods to predict admission outcomes (30-day readmission, Medium&Long Length of Stay (LOS) and Single-day admissions (avoidable admissions)). It innovates by incorporating a quite new explainable AI technique that determines which features are the most powerful in predicting each target variable in a transparent and coherent way. The EHR data were drawn from a real nationwide EHR which thus provides a real opportunity for investigation in hospital settings. We collected data on 15,016 admissions of patients with chronic conditions (COPD and Liver disease). Of these, 4,262 had been readmitted within 30 days, 3,409 had Medium&Long LOS and 2,868 were Single-day admissions. After compiling and preparing the data, we generated prediction models (using XGBoost, Logistic regression and Neural Networks) by incorporating all the predictor information into multiple factors causing readmission and other related outcome risks to generate actionable insights. Finally, Feature importance was used for XGBoost and

Logistic regression to analyze the key features for each observation, followed by an explainable AI method, the SHapley Additive exPlanations (SHAP) analysis which is used in the XGBoost prediction for both groups of patients but also tailored to the patient level.

Literature Survey

Studies predicting COPD complications: Min et al. (2019) indicated that the combination of knowledge and data driven variables yielded the best prediction of readmission outcomes from Logistic regression (LR) and Support Vector Machine (SVM) methods, and achieved an area under the ROC curve (AUC) of 0.65. Goto et al. (2020) obtained an AUC of 0.57 and 0.63 in two works predicting readmissions using LR. Liu et al. (2021) tracked the risk of Single-day admissions and its factors in COPD patients but did not generate predictions. One work highlighted the need to study and predict prolonged LOS (between 3-16 days) to determine the factors causing long LOS (e.g. low albumin level and other chronic conditions as such as heart failure) but the authors did not generate predictions (Wang et al. 2014).
Studies predicting liver disease complications: According to Singal et al. (2013), the prediction of 30-day readmission using the Indiana readmission model (Berman et al. 2011) generated C-statistic scores of 0.57 to 0.67 with various models. Overall, previous works have mostly been based on narrow data (mostly from one visit per patient), rather than patient-level big data (Mitchell et al. 2016) and additionally, more papers are needed using complex statistical models to predict readmission or mortality likelihood such as Neural network (Kalgotra and Sharda 2021).
Advanced models can help physicians and hospital management develop early interventions to identify and treat patients at high risk of admission complications, especially with visualization and explainable tools. To respond to these needs, the current study applied an approach for tracking future admission complications by assessing the potential factors leading to these complications. It accessed a comprehensive dataset of two leading chronic diseases and included all hospital stays. An analysis of several reviews of the literature on EHR prediction (Goldstein et al. 2017) and integration of explainable AI tools in medicine (Loh et al. 2022) only revealed one paper that has incorporated a very recent explainable AI tool for predicting COPD on a small dataset (Kor et al. 2022) but did not use EHR data. We included explainable AI in this work to enhance its contribution to medicine by using the SHAP method, a powerful tool for understanding the predictions made by ML models (Lundberg and Lee 2017). Explainable AI is relevant to the field of IS as a whole as reflected in several papers (Meske et al. 2022) which emphasize its ability to improve predictions, for instance, by identifying the contribution of explainable AI to enhancing customer service quality (Guo et al. 2023).

Objectives

This study applied predictive analytics methods (e.g., data preparation, data aggregation across visits of the same patients, parameter optimization and classifications) to analyze admission outcomes (30-day readmission, Medium&Long LOS and Single-day admissions) in COPD and Liver disease patients to demonstrate the feasibility of using these prediction classification methods on Big-Data. This research in progress (RIP) incorporates an explainable AI technique that pinpoints the most powerful features in predicting admission outcomes. By using explainable AI and SHAP in particular, the analysis can be expanded beyond the importance of the features to the target variables, as well as custom preventative treatment for individual patients (Davazdahemami et al. 2022; Kor et al. 2022). It exploited a comprehensive database on two prevalent chronic diseases and took all hospital visits into account in a chain of sophisticated models. Specifically, it aimed to: 1) Demonstrate the feasibility of ML methods, 2) Compare these methods in terms of their predictive power, 3) Depict the factors significantly affecting key outcome measures, such as likelihood of readmission. The three dependent variables predicted for these chronic diseases were as follows.
1. 30-day readmission – One of the most telling managerial and medical outcomes, in particular in the case of chronic diseases, is patient readmission to hospitals (Chamberlain RS et al. 2018; Goto et al. 2020; Shang et al. 2021; Soh et al. 2020). The readmission rate is an important indicator of quality of care. To reduce readmissions and elevate quality of care, it is vital to identify the factors leading to hospital readmissions.
2. Medium&Long LOS – Hilton et al. (2020) used a LOS of five days or more as a key variable for prediction which was considered to represent the threshold for medium or long LOS. This threshold was adopted here.

3. Single-day admissions – Quantifies whether a patient, as a result of the decision to admit, was admitted for a single day or for a longer period of time. It is likely that some single-day admissions are redundant and could easily be avoided if physicians had access to proper medical history. Previous works have shown that such short-term admissions are important for measurement and prediction purposes (Liu et al. 2021).

Materials and Methods

Procedures

The procedure was as follows: 1. Assemble the data. Thousands of admissions for COPD, and Liver disease as the primary diagnosis were extracted from an existing EHR with several independent variables including demographics, main EHR data components ((I) Medical test results; (II) Medications taken; (III) Prior visits/hospitalizations and (IV) Surgical history), the Charlson comorbidity index score, Number of ED visits in the last half year and Admitting department (e.g., Internal Medicine, Surgical, Orthopedics, etc.). 2. Data preparation. The data were cleaned, computed or treated if missing (based on Predictive Mean Matching imputation technique (Landerman et al. 1997) with the mice R package (Van Buuren and Groothuis-Oudshoorn 2011) and scaled when needed. 3. Generate the independent and dependent variables. The data were aggregated, as needed, into predictors and the dependent variables were created. 4. Flagging previous differential diagnoses. The focal differential diagnoses (COPD and Liver disease) were added. 5. Predictive analytics followed be explainable AI analytics.

Methods

After assembling the data and completing the data preparation processes, we utilized all the predictors to build prediction models using XGBoost (Chen and Guestrin 2016), Logistic regression and Neural Networks. We combined our prediction models with SHAP. The SHAP method is a powerful tool for understanding the predictions made by ML models (Lundberg and Lee 2017). This approach offers a transparent and coherent explanation of the contribution of each feature to the overall prediction of the model. The SHAP values are based on the principles of Shapley values from cooperative game theory, which determines a fair distribution of the value generated by a cooperative system among its individual members. In a similar manner, the SHAP values allocate the prediction of a machine learning model among its input features. In our analysis, we utilized the SHAP method to evaluate our target variable as done in recent papers (Lopez et al. 2023).

Results

Indicator and Description	COPD (N = 12112)	Liver Disease (N = 2904)
Patient age ± SD	45.377 ± 29.211	61.801 ± 16.461
Gender (Male) (%)	5523 (45.599 %)	1859 (64.015 %)
Insurance Cover Level[1] (%)	10081 (83.232 %)	2204 (75.895 %)
Previous Labs Retrieved (%)	1900 (15.687 %)	746 (25.689 %)
Permanent Medications Retrieved (%)	109 (0.899 %)	22 (0.758 %)
Count of total main EHR screens	1298 (26.511 %)	319 (49.277 %)
Charlson comorbidity index score[2] ±SD	0.460 ± 0.533	0.922 ± 1.717
Number of ED Visits in the last half year[3]	1.725 ± 1.538	2.326 ± 1.572
Day Shift	5687 (46.953 %)	1665 (57.335 %)
Internal Medicine Physician Involved	8069 (66.620 %)	2552 (87.879 %)
30-day readmission	3040 (25.099 %)	1237 (42.596 %)
Medium&Long LOS	2640 (21.797 %)	778 (26.791 %)
Single-day admission	2274 (18.775 %)	596 (20.523 %)

Table 1. Data Description

[1] This variable is categorized with two levels. The first for insurance from main HMO and the second is from all other existing HMO.
[2] This variable represents the prediction for medical complications. Higher score indicated a higher chance for risks.
[3] This variable represents the number of times that a patient presented at the ED in the six months prior to this admission.

Descriptive Statistics

14,988 admissions of patients with COPD and Liver disease as the primary diagnosis were extracted as well as their associated records. Table 1 lists the general descriptive statistics and distributions of the dependent variables. It shows patients who were readmitted within 30 days, with a LOS below 5 days and those who were admitted for longer LOS (five days or more) and whether the visit lasted one day or more.

Prediction Results

The tables below lists the prediction outcomes for all three dependent variables. Five-fold cross-validation was used to obtain the prediction scores using Python for each method (Pedregosa et al. 2011). We used the XGBoost, Scikit-learn and SHAP packages for our analysis. Hyper-parameter validation was implemented to tune the parameters (such as in XGBoost: the number of trees in the forest, the maximum depth of the tree and the number of features to consider when looking for the best split). Tables 2-3 show the individual prediction results for each condition.

The results (Tables 2-3) for all predictions indicated that predicting 30-day readmission had the best AUC scores for each diagnosis (0.678 -0.738). The highest score was for Liver disease. Specifically, Table 3 shows that predicting 30-day readmission had the highest AUC results for each model (0.711-0.716), where the best score was achieved by XGBoost. The next best in predicting 30-day readmission was Single-day admission with the highest AUC score of 0.702 with Logistic regression but less than a 1% difference between models. The Medium&Long LOS prediction obtained similar AUC results in the 0.678-0.681 range. Table 2 lists the prediction of COPD admission complications, where the highest AUC score was obtained with XGBOOST with a score of 0.716. The Logistic regression and NN scores were lower by less than 1% in predicting 30-day readmission. The second highest target variable predicted was single-day admission with an AUC score of 0.701 achieved by NN**.**

Model	**XGBoost**		**Logistic regression**		**Neural Network (NN)**	
(N = 12,112)	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
30-day readmission	0.749	**0.716**	0.758	0.715	0.753	0.712
Medium&Long LOS	0.782	0.680	0.780	0.682	0.780	0.679
Single-day admission	0.812	0.696	0.811	0.702	0.812	0.701

Table 2. COPD: Prediction of admission outcomes across all models

Table 3 shows the prediction results for Liver disease admission complications. It indicates that the highest AUC score was a 0.738 with XGBoost on 30-day readmission prediction. The differences between each model within the same dependent variable were roughly 2%-13%, whereas the differences across dependent variables ranged from 4% to 35%. Single-day admission prediction achieved the best 0.7 AUC score using Logistic regression. Medium&Long LOS had an AUC score of 0.599 using Logistic regression as well.

Model	**XGBoost**		**Logistic regression**		**Neural Network (NN)**	
(N = 2,904)	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**
30-day readmission	0.691	**0.738**	0.679	0.731	0.655	0.726
Medium&Long LOS	0.732	0.586	0.731	0.599	0.731	0.532
Single-day admission	0.796	0.685	0.799	0.701	0.794	0.684

Table 3. Liver disease: Prediction of admission outcomes across all models

Overall, the results here (Tables 2-3) for 30-day readmission predictions (albeit not receiving high AUC) were comparable to previous studies. Single-day admissions and Medium&Long LOS predictions for COPD (which have not been reported to date to the best of our knowledge) had an accuracy of 81% and 78% and an AUC of 69% and 67% respectively.

Explainable AI using SHAP

Figures (1-3)A present the 10 highest ranked features according to their importance, which was measured using the mean absolute SHAP values based on XGBoost including both diagnoses. Figures (1-3)B present the significance of each feature in our model by displaying the global SHAP values, which reflect the positive or negative impact of each feature to the target variable. The SHAP value indicates the contribution of each feature, with a positive value indicating a positive contribution and a negative value indicating a negative contribution. Each patient is represented as a dot in the plot. The horizontal position of the dot reflects the corresponding SHAP value, while the color of the dot (red indicates high risk and blue indicates low risk) represents the relative value of the feature and its average in the dataset. This representation provides a

clear and concise understanding of the contribution of each feature to the prediction model. As indicated by Figures 1A and 1B, the variables with the high values that are leading to positive impact on the target variable of readmission are: Number ED visits in the last half year (with some negative values indicated for some patients), Age (also with several negative values for some patients), Internal Medicine Physician Involved, Gender and Insurance cover level. Conversely, high values variables with a negative effect on the target variable of readmission are: Charlson comorbidity index, admission to Pediatrics department, Number of cases per physician, and Blood test ordered.

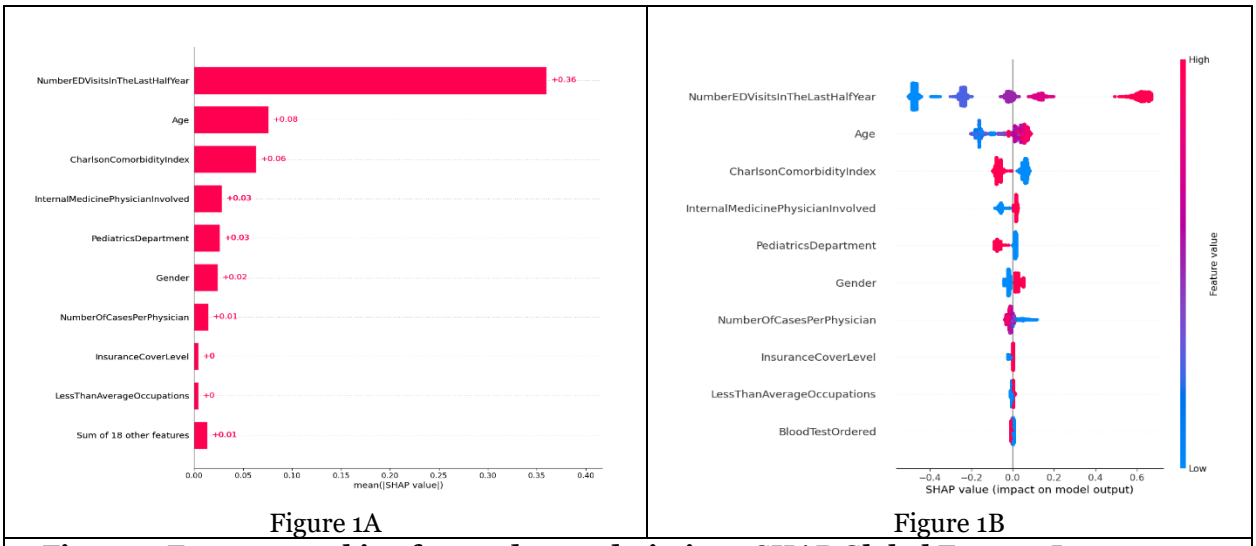

Figure 1A Figure 1B

Figure 1. Features ranking for 30-day readmission - SHAP Global Feature Importance

According to Figures 2A and 2B, the high value in the following variables have positive effect on Medium&Long LOS variable: Age, Day-Shift, Internal medicine physician involved, Number ED visits in the half year and Charlson comorbidity index (both with few with few negative values for some patients), Count of total main EHR screens and also specifically Laboratory test screens and admission to Surgical departments. In difference, for high values of the variables: Number of cases per physician and admission to Pediatrics department, there is negative contribution on Medium&Long LOS variable.

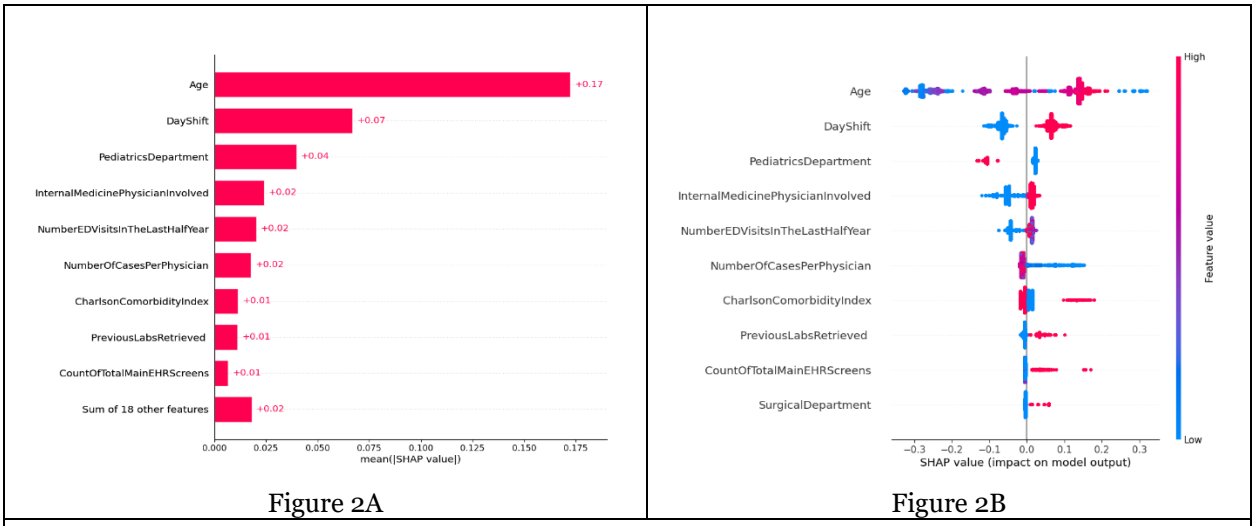

Figure 2A Figure 2B

Figure 2. Features ranking for Medium&Long LOS - SHAP Global Feature Importance

As shown in Figure 3A and Figure 3B, admission to Pediatrics department, Number of cases per physician (with some negative values for some patients) and Number of ED visits in the last half year, have a high positive impact on single day hospitalization when they have high values. The variables with a large negative

influence on single day hospitalization variable when they have high values are: Age (with some positive values for some patients), Blood test ordered (with a share of positive values), Internal medicine physician involved, Insurance cover level, Count of total main EHR screens and specifically Laboratory test screens.

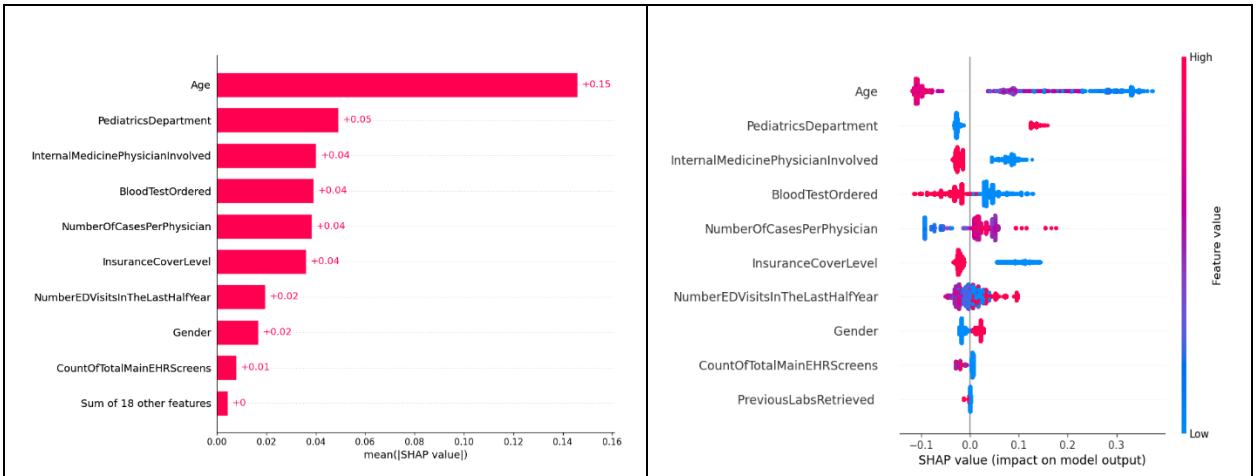

Figure 3A Figure 3B

Figure 3. Features ranking for Single day admission - SHAP Global Feature Importance

Figure 4 depicts the tailored explanation results (for one randomly selected patient) for the readmission target variable. The green and red bars are shaded in terms of the impact of the variable on the likelihood of the target variable. Red indicates the positive values (that enhance the predicted probability) and green indicates the negative values (that decrease the predicted probability). Note that for individual patients, the influence of the features (e.g. the Number of previous ED visits) is different from the global feature importance (Figure 1), which shows the importance of using explainable AI at in individual level as well.

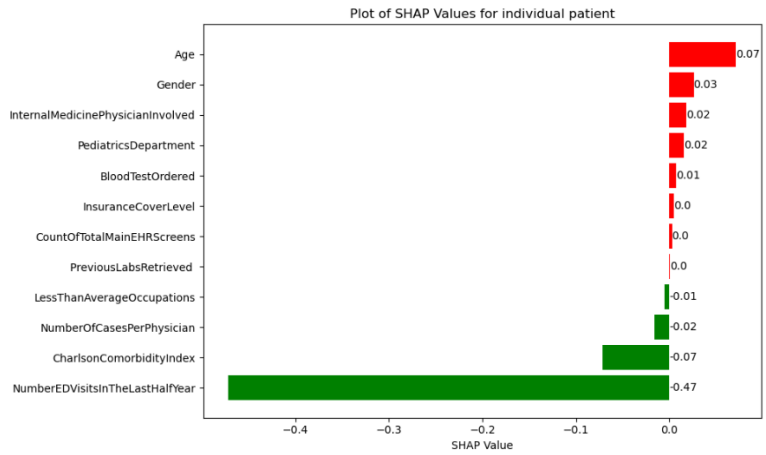

Figure 4. Feature ranks for an individual patient by XGBoost for 30-day readmission

Preliminary Discussion

Using new ML methodologies is crucial in the field of IS as a whole and specifically in the healthcare IT as has recently been reported for fall detection and prevention (Yu et al. 2021; Yu et al. 2023).

The prediction results for the standard ML prediction models were assessed in terms of 30-day readmission, Medium&Long LOS and Single-day admission. The results showed that 30-day readmission variable received higher prediction scores for both diagnoses. The variables most impacting 30-day readmission, Medium&Long LOS and Single-day admission for chronic patients were admission to

Pediatrics unit, patient age, Number of ED visits in the last half year, the Charlson comorbidity index score, previous medical procedures, and treatment during the day versus at night. The results for 30-day readmission predictions in terms of AUC were comparable to previous works.

Using an explainable AI new technique, namely the SHAP, indicates some powerful features in predicting each one of the target variables. In common for 30-day readmission and LOS, Age has a positive influence on 30-day readmission incorporated with less negative values, whereas in Single day admission it is mostly negative influence on one day admissions with small positive values. It is reasonable since older patients are more likely to be readmitted and also admitted for longer time than one day. Another variable with considerable great influence is the number of ED visits in the last half year, with positive influence (with some negative values) for all target variables. Charlson comorbidity index showed negative influence on readmission (surprisingly leading to less readmissions), and positive impact on longer LOS, leading to longer LOS (the effects were somewhat inconclusive, with some values from the opposite effect). Additional impactful variables were Day-Shift, Gender, using EHR operations screens (previous medical procedures), Insurance cover level and etc.

One example of a limitation of this RIP that will be dealt with in future research is that we only examined two diagnoses. The next steps in this RIP will involve studying other chronic conditions and developing more complex models for better quality prediction (Meyer et al. 2014). Future research could also ask physicians to evaluate the key features generated by the model that could influence admission outcomes and allow human-AI interaction towards the development of a probable AI-based tool. The other dependent variables besides readmissions have received less attention in previous prediction studies. Here they enhanced the contribution of this paper by showing how additional admission complication variables can constitute new targets for prediction.

Policy makers should support integrating models as proposed here to be implemented to improve current scores and to enable them also for practice actual use. Using prediction models complemented with explainable AI methods is likely to predict better medical outcomes and suggest preventable medical decisions tailored to the patient level and enhance the availability of alternative nonhospital services can reduce the admission complications.

Acknowledgement: This work was funded by a regular research grant from the Chief Scientist Office of the Ministry of Health, Israel (grant number: 3-0000-18435).

References

Bardhan, I., Chen, H., and Karahanna, E. 2020. "Connecting Systems, Data, and People: A Multidisciplinary Research Roadmap for Chronic Disease Managment," *Management Information Systems Quarterly* (44:1), pp. 185-200.

Bardhan, I., Oh, J.-h., Zheng, Z., and Kirksey, K. 2015. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure," *Information Systems Research* (26:1), pp. 19-39.

Ben-Assuli, O., and Padman, R. 2020. "Trajectories of Repeated Readmissions of Chronic Disease Patients: Risk Stratification, Profiling, and Prediction," *MIS Quarterly* (44:1), pp. 201-226.

Ben-Assuli, O., and Vest, J. R. 2020. "Data Mining Techniques Utilizing Latent Class Models to Evaluate Emergency Department Revisits," *Journal of biomedical informatics*), p. 103341.

Berman, K., Tandra, S., Forssell, K., Vuppalanchi, R., Burton Jr, J. R., Nguyen, J., Mullis, D., Kwo, P., and Chalasani, N. 2011. "Incidence and Predictors of 30-Day Readmission among Patients Hospitalized for Advanced Liver Disease," *Clinical Gastroenterology and Hepatology* (9:3), pp. 254-259.

Chamberlain RS, Sond J, Mahendraraj K, Lau CS, and BL, S. 2018. "Determining 30-Day Readmission Risk for Heart Failure Patients: The Readmission after Heart Failure Scale," *International Journal of General Medicine* (11), pp. 127–141.

Chen, T., and Guestrin, C. 2016. "Xgboost: A Scalable Tree Boosting System," *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794.

Davazdahemami, B., Zolbanin, H. M., and Delen, D. 2022. "An Explanatory Machine Learning Framework for Studying Pandemics: The Case of Covid-19 Emergency Department Readmissions," *Decision Support Systems* (161), p. 113730.

Goldstein, B. A., Navar, A. M., Pencina, M. J., and Ioannidis, J. 2017. "Opportunities and Challenges in Developing Risk Prediction Models with Electronic Health Records Data: A Systematic Review," *Journal of the American Medical Informatics Association* (24:1), pp. 198-208.

Goto, T., Yoshida, K., Faridi, M. K., Camargo, C. A., and Hasegawa, K. 2020. "Contribution of Social Factors to Readmissions within 30 Days after Hospitalization for Copd Exacerbation," *BMC Pulmonary Medicine* (20), pp. 1-10.

Guo, Y., Li, Y., Liu, D., and Xu, S. X. 2023. "Measuring Service Quality Based on Customer Emotion: An Explainable Ai Approach," *Decision Support Systems*), p. 114051.

Hilton, C. B., Milinovich, A., Felix, C., Vakharia, N., Crone, T., Donovan, C., Proctor, A., and Nazha, A. 2020. "Personalized Predictions of Patient Outcomes During and after Hospitalization Using Artificial Intelligence," *NPJ digital medicine* (3:1), pp. 1-8.

Kalgotra, P., and Sharda, R. 2021. "When Will I Get out of the Hospital? Modeling Length of Stay Using Comorbidity Networks," *Journal of Management Information Systems* (38:4), pp. 1150-1184.

Kor, C.-T., Li, Y.-R., Lin, P.-R., Lin, S.-H., Wang, B.-Y., and Lin, C.-H. 2022. "Explainable Machine Learning Model for Predicting First-Time Acute Exacerbation in Patients with Chronic Obstructive Pulmonary Disease," *Journal of personalized medicine* (12:2), p. 228.

Landerman, L. R., Land, K. C., and Pieper, C. F. 1997. "An Empirical Evaluation of the Predictive Mean Matching Method for Imputing Missing Values," *Sociological methods & research* (26:1), pp. 3-33.

Lin, Y.-K., Chen, H., Brown, R. A., Li, S.-H., and Yang, H.-J. 2017. "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach," *MIS Quarterly* (41:2), pp. 473-495.

Liu, Y., Chen, Y., Kong, D., Liu, X., Fu, J., Zhang, Y., Zhao, Y., Zhao, X., Xu, K., and Jiang, C. 2021. "Short-Term Effects of Cold Spells on Hospitalisations for Acute Exacerbation of Chronic Obstructive Pulmonary Disease: A Time-Series Study in Beijing, China," *BMJ open* (11:1), p. e039745.

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., and Acharya, U. R. 2022. "Application of Explainable Artificial Intelligence for Healthcare: A Systematic Review of the Last Decade (2011–2022)," *Computer Methods and Programs in Biomedicine*), p. 107161.

Lopez, K., Li, H., Paek, H., Williams, B., Nath, B., Melnick, E. R., and Loza, A. J. 2023. "Predicting Physician Departure with Machine Learning on Ehr Use Patterns: A Longitudinal Cohort from a Large Multi-Specialty Ambulatory Practice," *Plos one* (18:2), p. e0280251.

Lundberg, S. M., and Lee, S.-I. 2017. "A Unified Approach to Interpreting Model Predictions," *Advances in neural information processing systems* (30).

Meske, C., Bunde, E., Schneider, J., and Gersch, M. 2022. "Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities," *Information Systems Management* (39:1), pp. 53-63.

Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., and O'Connor, P. J. 2014. "A Machine Learning Approach to Improving Dynamic Decision Making," *Information Systems Research* (25:2), pp. 239-263.

Min, X., Yu, B., and Wang, F. 2019. "Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on Copd," *Scientific reports* (9:1), pp. 1-10.

Mitchell, S. E., Martin, J., Holmes, S., van Deusen Lukas, C., Cancino, R., Paasche-Orlow, M., Brach, C., and Jack, B. 2016. "How Hospitals Reengineer Their Discharge Processes to Reduce Readmissions," *Journal for healthcare quality: official publication of the National Association for Healthcare Quality* (38:2), p. 116.

Nguyen, L., Bellucci, E., and Nguyen, L. T. 2014. "Electronic Health Records Implementation: An Evaluation of Information System Impact and Contingency Factors," *International journal of medical informatics* (83:11), pp. 779-796.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. 2011. "Scikit-Learn: Machine Learning in Python," *the Journal of machine Learning research* (12), pp. 2825-2830.

Shang, Y., Jiang, K., Wang, L., Zhang, Z., Zhou, S., Liu, Y., Dong, J., and Wu, H. 2021. "The 30-Days Hospital Readmission Risk in Diabetic Patients: Predictive Modeling with Machine Learning Classifiers," *BMC medical informatics and decision making* (21:2), pp. 1-11.

Singal, A. G., Rahimi, R. S., Clark, C., Ma, Y., Cuthbert, J. A., Rockey, D. C., and Amarasingham, R. 2013. "An Automated Model Using Electronic Medical Record Data Identifies Patients with Cirrhosis at High Risk for Readmission," *Clinical Gastroenterology and Hepatology* (11:10), pp. 1335-1341. e1331.

Soh, J. G. S., Wong, W. P., Mukhopadhyay, A., Quek, S. C., and Tai, B. C. 2020. "Predictors of 30-Day Unplanned Hospital Readmission among Adult Patients with Diabetes Mellitus: A Systematic Review with Meta-Analysis," *BMJ Open Diabetes Research and Care* (8:1), p. e001227.

Van Buuren, S., and Groothuis-Oudshoorn, K. 2011. "Mice: Multivariate Imputation by Chained Equations in R," *Journal of statistical software* (45), pp. 1-67.

Vest, J., Kasthurirathne, S., Ge, W., Gutta, J., Ben-Assuli, O., and Halverson, P. 2021. "Choice of Measurement Approach for Area-Level Social Determinants of Health and Risk Prediction Model Performance," *Informatics for Health and Social Care*), pp. 1-12.

Wang, Y., Stavem, K., Dahl, F. A., Humerfelt, S., and Haugen, T. 2014. "Factors Associated with a Prolonged Length of Stay after Acute Exacerbation of Chronic Obstructive Pulmonary Disease (Aecopd)," *International journal of chronic obstructive pulmonary disease* (9), p. 99.

Yaraghi, N., Du, A. Y., Sharman, R., Gopal, R. D., and Ramesh, R. 2014. "Health Information Exchange as a Multisided Platform: Adoption, Usage, and Practice Involvement in Service Co-Production," *Information Systems Research* (26:1), pp. 1-18.

Yu, S., Chai, Y., Chen, H., Brown, R. A., Sherman, S. J., and Nunamaker Jr, J. F. 2021. "Fall Detection with Wearable Sensors: A Hierarchical Attention-Based Convolutional Neural Network Approach," *Journal of Management Information Systems* (38:4), pp. 1095-1121.

Yu, S., Chai, Y., Samtani, S., Liu, H., and Chen, H. 2023. "Motion Sensor–Based Fall Prevention for Senior Care: A Hidden Markov Model with Generative Adversarial Network Approach," *Information Systems Research*).