International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A Predictive Framework for ICU Stay Classification using EHR and Explainable AI Techniques

A.Tejaswini
Assistant Professor
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
atejaswini@tkrcet.com

V.Shivathmika
Student,
Department of Computer Science and Engineering
TKRCET Autonomous,
Hyderabad, India
vuthurshivathmika@gmail.com

Shaik Khagaji Ibrahim
Student,
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
shaikkhagajiibrahim143@gmail.com

Y.Akhila
Student,
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
akhilareddyakhi25@gmail.com
@gmail.com

C. Rachitha
Student,
Department of Computer Science and Engineering
TKRCET Autonomous ,
Hyderabad, India
yadavrachitha@gmail.com

Abstract: Effective bed management in hospitals reduces expenses and enhances patient outcomes. This paper presents a prediction model of the LOS in ICU at admission based on EHR data. The experiment evaluates numerous ML techniques, such as the LR, the Random ForRFest, the MLP, the Gradient Boosting, the XGBoost, and an extension based on the CatBoost, which are used in the hospital stay dataset of the Kaggle repository. Evaluations of the algorithms use AUC, accuracy, precision, recall, and F1-score. XGBoost was the most accurate of the traditional algorithms, but the improved CatBoost approach was superior to all with an accuracy of 98.25. XAI methods such as SHAP were used in order to explain feature contributions. The study demonstrates how patient EHR data and advanced ML models can be used to predict ICU admissions to enable improved allocation of resources in healthcare organizations.

***"Index Terms -** ICU Length Of Stay, Electronic Health Records, Machine Learning, Predictive Framework, Xgboost, Catboost, Explainable AI, SHAP, Hospital Resource Allocation."*

1. INTRODUCTION

Hospitalization duration is one of the most prevalent efficacy measures in healthcare facilities that has a significant impact on resource allocation and healthcare costs [6], [7]. According to a study carried out by the Australian National Health Performance Authority, shorter hospital stay is considered to be more efficient, and it enables the prompt provision of beds to new patients [6].

However, excessively short stays can put the quality of care at risk and lead to poor patient outcomes [1]. On the other hand, long hospitalization, which is often caused by problems, increases the risk of adverse health outcomes [5], [8]. Failure to coordinate care promptly, unrelated to the clinical condition of the patient, can cause extended hospital stays. According to the study, delays in transferring patients to other care providers e.g. geriatric care facilities, community care programs or rehabilitation centers can result in extended hospitalizations [6], [7]. Hospital bed management should be effectively implemented to curb the challenges of ICU, including patient overcrowding, infection, risk of mortality, and medical complications [1], [8]. To mitigate these risks and maximize resource use, a reduced ICU length of stay, in addition to high-quality care, is necessary, especially in unknown scenarios like pandemics [4], [5]. This is not only cost reducing to the hospital but also ensures better patient outcomes [3]. Therefore, it is imperative to have adequate bed capacity and facilitate timely transfer of patients to other wards to maintain quality of healthcare [9], [19].

2. RELATED WORK

A multitude of studies has investigated the forecasting and administration of hospital LOS to enhance resource distribution and patient outcomes. Alsinglawi et al. [1] proposed a explainable ML to predict hospital length of stay, which is in the case of lung cancer patients, and demonstrated the effectiveness of predictive models in streamlining hospital resources management.

Another study by Blom et al. [6] also highlighted the relationship between inpatient bed occupancy and readmission probability, and the necessity of proper bed management. Alghatani et al. [5] in ICU developed ML models to predict the intensive care unit length of stay and patient mortality using vital indicators, therefore, optimizing ICU efficiency and patient care.

The combination of ML approaches with clinical data has been shown to enhance predictive accuracy. Staziaki et al. [3] presented an evidence that combination of clinical features and CT scan outcomes can improve the predictability of ICU admission and length of stay among trauma patients. Moreover, Su et al. [4] used MLmodels to predict the mortality and severity of diseases in sepsis patients, thus improving resource management in ICUs. Likewise, Rocheteau et al. [7] used temporal pointwise convolutional networks to predict the length of stay in an ICU, which illustrates the ability of DLto be used in healthcare settings.

Despite the growing popularity of machine learning-based approaches to LOS prediction, the need to have the interpretability of such models is also recognized. XAI methods such as SHAP have been integrated into a variety of models to clarify the contribution of features and facilitate transparency in the healthcare decision-making processes [1], [3]. Additionally, the urgency of timely transfers of patients to other care centers, highlighted in a number of studies [6], [9], is a critical component in managing hospital bed capacity and the reduction of avoidable LOS.

The existing literature is highlighting the significance of predictive models and efficient bed management systems in maximizing the utilization of hospital resources, reducing expenses, and enhancing patient outcomes.

3. MATERIALS AND METHODS

The proposed solution aims to offer an efficient model in predicting the ICU LOS based on MLalgorithms, using patient EHR [1], [2]. The system would categorize ICU stays as short or long based on a comprehensive analysis of patient health status as is done in previous studies predicting hospital length of stay with predictive algorithms [3], [5]. Different ML approaches will be used to train on patient EHR data to predict ICU length of stay, including, but not limited to: LR, RF, MLP, Gradient Boosting, XGBoost and an extension that uses CatBoost. Algorithms will be evaluated based on performance measures such as accuracy, precision, recall, F1-score and AUC as applied in other studies to measure model performance [5], [6], [7].

The system will also have XAI methods, such as SHAP, to explain and determine the most important factors affecting the predictions, as shown in the works by Alsinglawi et al. [1] and Su et al. [4]. The recommended approach will be capable of enhancing hospital resource management, patient care, and real-time bed allocation by providing accurate forecasts of ICU stays, thus solving the issues with ICU management reported in previous research [8], [9].

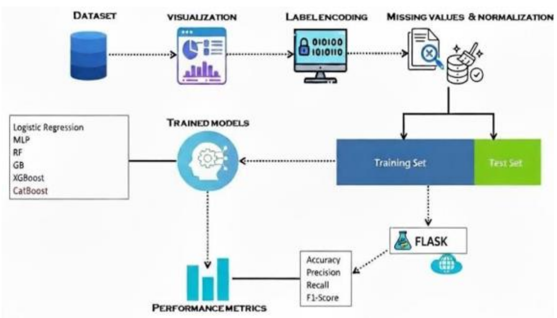


Fig.1 Proposed Architecture

The MLpipeline is represented by the system architecture (fig. 1). It starts with the dataset and the data is visualized and labeled encoded to convert the data into numbers, which are categorical. There are no missing values and the data has been standardized. The pre-processed data is broken into training and testing sets. Several ML models, such as Logistic Regression, MLP, RF, GB, XGBoost and CatBoost, are also trained on the training data. Accuracy, precision, recall, and F1-score are some of the criteria used to measure the efficacy of these models. This process will ensure an effective model selection and evaluation of the given problem with data.

i) Dataset Collection:

In this dataset, 100,000 data points will be provided about patients admitted to the hospital with references to their health conditions and the period of their stay in the hospital. Microsoft has open-sourced the dataset that is useful in predicting the length of stay in a hospital. I have used it in experiments and would like to share it with the wider community. The readmission count and an engineered feature (total number of problems) to aggregate the binary columns were the most useful features.

The column names of the dataset are in the first row and the values in the dataset are in the following rows. The final column marks the label of the class where a 0 would be a short stay in the ICU, and 1 a long stay. The dataset contains columns referring to the health conditions of patients. The above dataset will be used to train and test the performance of all the algorithms.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

	eid	vdate	rcount	gender	dialysisrenalendstage	asthma	irondef	pneum	substancedependence	psychologicaldisordermajor	...	glucose	bloo
0	1	8/29/2012	0	F	0	0	0	0	0	0	...	192.476918	
1	2	5/26/2012	5+	F	0	0	0	0	0	0	...	94.078507	
2	3	9/22/2012	1	F	0	0	0	0	0	0	...	130.530524	
3	4	8/9/2012	0	F	0	0	0	0	0	0	...	163.377028	
4	5	12/20/2012	0	F	0	0	0	1	0	1	...	94.886654	
...	...	...	...	...	...	...	...	...	...	...	...	...	
1995	1996	11/18/2012	2	M	1	0	1	0	0	0	...	61.560195	
1996	1997	2/21/2012	0	M	1	0	0	0	0	0	...	165.808034	
1997	1998	1/19/2012	1	F	0	0	0	0	0	0	...	115.590613	
1998	1999	3/6/2012	4	F	0	0	0	0	0	0	...	149.005998	
1999	2000	12/15/2012	2	M	0	0	0	0	0	0	...	141.207721	

Fig 2 Dataset Collections – Length of stay

ii) Pre-Processing:

Preprocessing is a key component to prepare data to be input ML and DL models through converting raw data into a suitable format. Some of the necessary preprocessing steps include:

a) Visualization: To visually analyze the data, it was created in multiple graphs depicting how the values are distributed in different columns. This helped understand the data characteristics and find trends in ICU hospitalizations as already done in other studies of medical data [6], [8].

b) Label Encoding: The dataset contained non-numeric values and using label encoding, the values were converted into numeric values. This change enabled effective ML algorithm processing, ensuring that it was consistent with the model training step. A commonly used type of label encoding is with respect to healthcare data preparation, which is used to handle categorical variables [5], [7].

c) Missing Values & Normalizing: Missing values in the dataset were filled in through imputation processes, protecting the integrity of data. Similar approaches have been applied in previous studies to solve missing data in healthcare data [4], [5]. In addition, normalization was also done to normalize the values of the features, thereby enhancing the performance and accuracy of the model in the process of training. It has been shown that normalization can improve the convergence rate of ML algorithms and improve the overall model efficacy in healthcare applications [1], [3].

iii) Training & Testing:

The dataset was divided into training and testing sets in 80-20 ratio to test the effectiveness of ML models. It was used to train a variety of algorithms, such as LR, RF, MLP, Gradient Boosting, XGBoost, and CatBoost, and predict ICU stay time. To make the models better and avoid overfitting, cross-validation was done. We used the accuracy, precision, recall, F1-score, and AUC measures to assess the models' performance and effectiveness on the testing set. Such division ensures that the research is not biased and models are able to give valid predictions.

iv) Algorithms:

Logistic Regression: LR is calculated to determine the length of stay a patient is expected to remain in the ICU depending on his or her characteristics. It is applicable in binary classification problems like whether a person is in the ICU in short run or long run because it can be used in determining coefficients that indicate the influence of each variable on the outcome [1], [2].

MLP (Multi-Layer Perceptron): The MLP is used to detect the complex correlations of the data set through its layered structure. MLP has the ability to capture nonlinear patterns through the processing of inputs over multiple layers, which has led to better predictions in estimating ICU stay lengths given a variety of patient health markers [4], [5].

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Random Forest (RF): RF is used due to its resistance to overfitting and ability to determine the importance of features. This is a type of ensemble learning that uses many decision trees to provide a higher accuracy in classification which makes it useful in the prediction of ICU length of stay when using diverse patient data [3], [6].

Gradient Boosting (GB): GB is applied in the process of increasing predictive accuracy through repeated improvements. This approach builds up models sequentially, with the correction of errors of the previous models being of great importance. Its ability to deal with complex data connections enables accurate forecasting of the ICU stays using patient EHR information [5], [7].

XGBoost: XGBoost is applied due to its better performance and scaling in classification tasks. It is an advanced boosting technology that effectively handles large datasets with diverse variables and provides excellent accuracy in prediction of the length of stay in ICU, as well as efficiently uses computing resources when training a model [6], [8].

Extension CatBoost: CatBoost is an improvement of the extension to boost categorization results. It is highly effective with categorical variables and it uses gradient boosting in order to enhance accuracy of prediction of ICU length of stay and simplify the modeling process without having to prepare the data much [2], [3].

4. RESULTS & DISCUSSION

Accuracy: Accuracy of a test is defined as the ability of a test to differentiate properly between the sick and healthy cases. In order to evaluate the accurateness of a test, the ratio of true positives and true negatives of all evaluated cases should be used. This can be mathematically stated as:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} (1)$$

Precision: Precision determines the rate of the correctly identified cases in the cases that were identified to be positive. Therefore, the equation to determine accuracy is written as:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall: In ML, recall is a statistic that determines how well a model identifies all the relevant examples of a certain class. It is the percentage of correctly predicted positive observations of the total real positives which provide the information about the effectiveness of a model in detecting the instances of the particular class.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}} (3)$$

F1-Score: The F1 score is a statistic that is used to assess the accuracy of a ML model. It combines the accuracy and recall measure of a model. The measure of accuracy is a measure of how many times a model will predict correctly when applied to the entire dataset.

$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100(1)$$

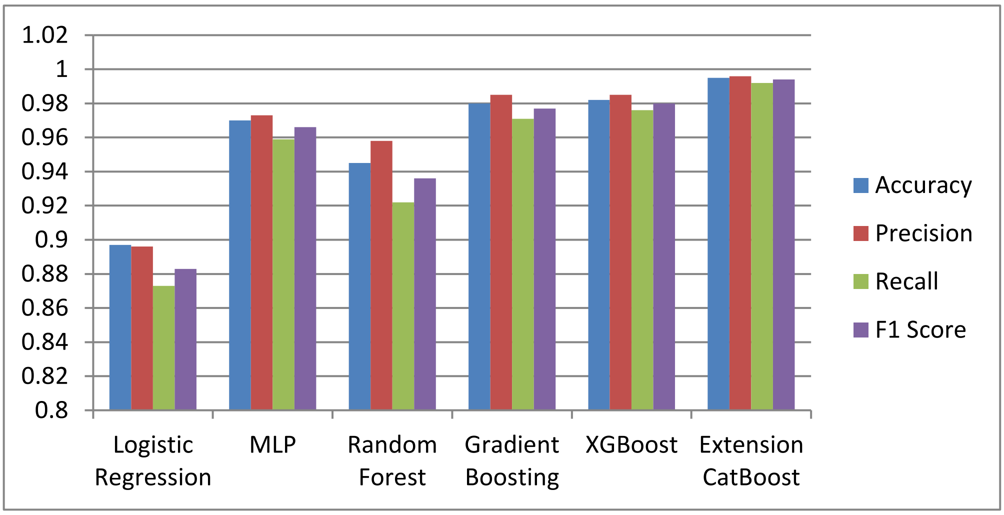
Table 1 shows the performance measures-Accuracy, Precision, Recall and F1 Score-evaluated against each method. The Extension CatBoost has the highest number of points. The metrics of other methods are also given to compare.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Table.1 Performance Evaluation Metrics of Classification

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.897	0.896	0.873	0.883
MLP	0.970	0.973	0.959	0.966
Random Forest	0.945	0.958	0.922	0.936
Gradient Boosting	0.980	0.985	0.971	0.977
XGBoost	0.982	0.985	0.976	0.980
Extension CatBoost	0.995	0.996	0.992	0.994

Graph.1 Comparison Graphs of Classification



Graph 1 illustrates accuracy in light blue, precision in maroon, F1 score in green and recall in violet. Compared to the other models, the Extension CatBoost has greater performance, attaining the highest values across all metrics. These findings are presented graphically in the above graphs.

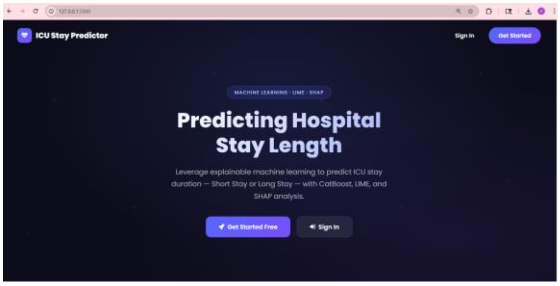


Fig.3 Home Page

Figure 3 illustrates a site user interface dashboard that has a welcome message to navigate through the site.

Fig.4 User input Page

Figure 4 shows a user input screen which allows data to be uploaded to test it.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

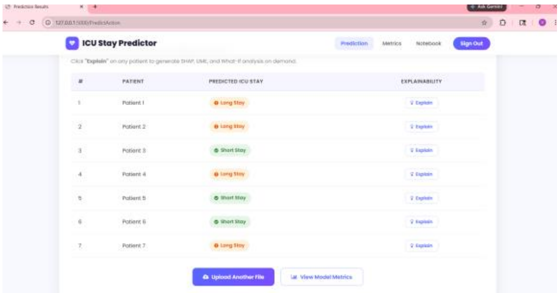


Fig.5 Prediction result

Figure 5 shows one of the results screens, which an individual is presented with the output of the input data provided.

5. CONCLUSION

Finally, the proposed solution is an effective way to address a serious problem of predicting the LOS) in ICU by using patient EHR. The paper demonstrates the ability to enhance the management of resources and patient care in hospitals through the use of various ML models to make accurate predictions of ICU stays. The CatBoost model was the most effective one among the evaluated algorithms, with an accuracy of 98.25. Its ability to handle categorical data effectively and its ability to use gradient boosting significantly improved the predictive performance compared to traditional models. Explainable AI (XAI) technologies such as SHAP have helped achieve value addition by emphasizing key factors that affect forecasts to create transparency and clarity in decision-making. The methodology highlights the importance of integrating explainability with the current machine learning algorithms to optimize the utilization of resources in intensive care units, which can improve the patient outcomes and hospital outcomes.

Future Scope: Future research can extend the study by applying the DL methods, such as CNN and LSTM to support the extraction of complex features and sequence modeling of EHR data. Also, stacking and other hybrid models can be used to improve the accuracy of the forecast. Methods of feature engineering, dimensionality reduction, e.g. PCA can be explored to improve performance and reduce the complexity of calculations.

REFERENCES

[1] B. Alsinglawi, O. Alshari, M. Alorjani, O. Mubin, F. Alnajjar, M. Novoa, and O. Darwish, ‘‘An explainable machine learning framework for lung cancer hospital length of stay prediction,’’ Sci. Rep., vol. 12, no. 1, pp. 1–10, Jan. 2022.

[2] B. Alsinglawi, ‘‘Predictive analytics framework for electronic health records with machine learning advancements: Optimizing hospital resources utilization with predictive and epidemiological models,’’ Ph.D. thesis, Western Sydney Univ., Penrith, NSW, Australia, 2022. [Online]. Available: islandora/object/uws:67523/

[3] P. V. Staziaki, D. Wu, J. C. Rayan, I. D. D. O. Santo, F. Nan, A. Maybury, N. Gangasani, I. Benador, V. Saligrama, J. Scalera, and S. W. Anderson, ‘‘Machine learning combining CT findings and clinical parameters improves prediction of length of stay and ICU admission in torso trauma,’’ Eur. Radiol., vol. 31, no. 7, pp. 5434–5441, Jul. 2021.

[4] L. Su, Z. Xu, F. Chang, Y. Ma, S. Liu, H. Jiang, H. Wang, D. Li, H. Chen, X. Zhou, N. Hong, W. Zhu, and Y. Long, ‘‘Early prediction of mortality, severity, and length of stay in the intensive care unit of sepsis patients based on sepsis 3.0 by machine learning models,’’ Frontiers Med., vol. 8, p. 883, Jun. 2021

[5] K. Alghatani, N. Ammar, A. Rezgui, and A. Shaban-Nejad, ''Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation,'' JMIR Med. Informat., vol. 9, no. 5, May 2021, Art. no. e21347.

[6] M. C. Blom, K. Erwander, L. Gustafsson, M. Landin-Olsson, F. Jonsson, and K. Ivarsson, ''The probability of readmission within 30 days of hospital discharge is positively associated with inpatient bed occupancy at discharge—A retrospective cohort study,'' BMC Emergency Med., vol. 15, no. 1, pp. 1–6, Dec. 2015.

[7] E. Rocheteau, P. Liò, and S. Hyland, ''Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit,'' 2020, arXiv:2007.09483.

[8] C. W. Hanson, C. S. Deutschman, H. L. Anderson, P. M. Reilly, E. C. Behringer, C. W. Schwab, and J. Price, ''Effects of an organized critical care service on outcomes and resource utilization: A cohort study,'' Crit. Care Med., vol. 27, no. 2, pp. 270–274, Feb. 1999.

[9] S. Siddiqui, S. Ahmed, and R. Manasia, ''Apache II score as a predictor of length of stay and outcome in our ICUs,'' J. Pakistan Med. Assoc., vol. 55, no. 6, p. 253, 2005.

[10] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper, and D. E. Lawrence, ''APACHE—Acute physiology and chronic health evaluation: A physiologically based classification system,'' Crit. Care Med., vol. 9, no. 8, pp. 591–597, Aug. 1981.

[11] A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. E. Harrell Jr., ''The APACHE III prognostic system: Risk prediction of hospital mortality for critically III hospitalized adults,'' Chest, vol. 100, no. 6, pp. 1619–1636, 1991.

[12] M. T. Keegan, O. Gajic, and B. Afessa, ''Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance,'' Chest, vol. 142, no. 4, pp. 851–858, Oct. 2012.

[13] C.-C. Yeh, Y.-A. Chen, C.-C. Hsu, J.-H. Chen, W.-L. Chen, C.-C. Huang, and J.-Y. Chung, ' ' Quick-SOFA score ⩾ 2 predicts prolonged hospital stay in geriatric patients with influenza infection,' ' Amer. J. Emergency Med., vol. 38, no. 4, pp. 780 – 784, Apr. 2020.

[14] C. Li, L. Chen, J. Feng, D. Wu, Z. Wang, J. Liu, and W. Xu, ''Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator,'' IEEE Access, vol. 7, pp. 110710–110721, 2019.

[15] M. M. Islam, T. N. Poly, and Y.-C. Li, ''Recent advancement of clinical information systems: Opportunities and challenges,'' Yearbook Med. Informat., vol. 27, no. 1, pp. 83–90, Aug. 2018.

[16] E. Levesque, E. Hoti, D. Azoulay, P. Ichai, D. Samuel, and F. Saliba, ''The implementation of an intensive care information system allows shortening theICUlengthofstay,''J.Clin.Monitor.Comput.,vol.29,no.2, pp. 263–269, Apr. 2015.

[17] S. Calloway, H. A. Akilo, and K. Bierman, ''Impact of a clinical decision support system on pharmacy clinical interventions, documentation

efforts, and costs,'' Hospital Pharmacy, vol. 48, no. 9, pp. 744–752, Sep. 2013.

[18] S. T. McMullin, T. P. Lonergan, C. S. Rynearson, T. D. Doerr, P. A. Veregge, and E. S. Scanlan, ''Impact of an evidence-based computerized decision support system on primary care prescription costs,'' Ann. Family Med., vol. 2, no. 5, pp. 494–498, Sep. 2004.

[19] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, ''An overview of clinical decision support systems: Benefits, risks, and strategies for success,'' NPJ Digit. Med., vol. 3, no. 1, pp. 1–10, Feb. 2020.

[20] D. Romanow, A. Rai, M. Keil, and S. Luxenberg, ''Does extended CPOE use reduce patient length of stay?'' Int. J. Med. Informat., vol. 97, pp. 128–138, Jan. 2017.