Journal of Global Information Management
Volume 31 • Issue 1

Predicting Patient Length of Stay Using Artificial Intelligence to Assist Healthcare Professionals in Resource Planning and Scheduling Decisions

Yazan Alnsour, University of Wisconsin-Oshkosh, USA*
https://orcid.org/0000-0001-7198-7197

Marina Johnson, Montclair State University, USA

Abdullah Albizri, Montclair State University, USA

Antoine Harfouch, Paris Nanterre University, France
https://orcid.org/0000-0002-0407-9217

ABSTRACT

Artificial intelligence (AI) significantly revolutionizes and transforms the global healthcare industry by improving outcomes, increasing efficiency, and enhancing resource utilization. The applications of AI impact every aspect of healthcare operation, particularly resource allocation and capacity planning. This study proposes a multi-step AI-based framework and applies it to a real dataset to predict the length of stay (LOS) for hospitalized patients. The results show that the proposed framework can predict the LOS categories with an AUC of 0.85 and their actual LOS with a mean absolute error of 0.85 days. This framework can support decision-makers in healthcare facilities providing inpatient care to make better front-end operational decisions, such as resource capacity planning and scheduling decisions. Predicting LOS is pivotal in today's healthcare supply chain (HSC) systems where resources are scarce, and demand is abundant due to various global crises and pandemics. Thus, this research's findings have practical and theoretical implications in AI and HSC management.

KEYWORDS:

Artificial Intelligence, Predictive Analytics, Length of Stay, Healthcare Supply Chain, Clinical Decision Support

1. INTRODUCTION

The increasing demand for value-based healthcare due to the recent COVID-19 pandemic put pressure on healthcare providers to provide fast service while maintaining quality. Healthcare providers reported that Electronic Healthcare Records (EHR) systems help enhance communication, eliminate redundancy in the delivery of care, and allow quick access to critical information (Silow-Carroll et al., 2012; Zhang et al., 2019). As a result, EHR systems enable providers to become more efficient in delivering care and reduce unnecessary hospital stays (Thompson et al., 2006; Von Wedel & Hagist,

DOI: 10.4018/JGIM.323059
*Corresponding Author

2020). Some healthcare practitioners are concerned that reducing hospital length of stay (LOS) may lead to higher rates of readmissions and mortality; however, recent studies show that there is no significant correlation between LOS and readmission and mortality rates and that a longer LOS does not necessarily mean better quality of care (Kaboli et al., 2012). Also, Thomas et al. (1997) argue that LOS is primarily used to measure efficiency and costs. Researchers discuss that hospitals with longer LOS are considered inefficient in using resources and to have a higher risk of exposing patients to hospital-acquired conditions (HACs), such as falls and trauma (Orsi et al., 2002; Services, 2014).

The recent advancements in AI have introduced new analytical tools capable of assisting Healthcare Professionals (HCP) in their day-to-day processes. There is an extensive body of research in the Digital Health literature that investigated the impact of using Artificial Intelligence (AI) and machine learning in different workflows (Fernandez-Luque & Imran, 2018; Salah et al., 2019). In the context of Clinical Decision Support (CDS) systems and AI applications in Healthcare Supply Chain (HSC) Management, this study investigates the complementary effects of AI, namely machine learning and healthcare. We posit that using machine learning in care delivery enables providers to predict patients' hospital LOS and invoke some needed interventions to reduce it. Furthermore, we argue that the artifact developed in this manuscript can be used to better manage patient LOS for enhanced resource capacity planning and creating efficient discharge plans. Additionally, this study's result can be integrated into optimization-based capacity allocation models aiming to improve resource distribution.

The rest of the manuscript progress as follows: the next section discusses the research framework, kernel theory, and related literature, focusing on AI applications in HSC management. The third section presents the proposed framework used to predict LOS. The fourth section deals with the findings, while the last section presents the conclusions, limitations, further research, and implications to the theory and practice of Information Systems.

2. RESEARCH FRAMEWORK

In order to attain our research goal, we employ the design science research paradigm to guide the development of the IT artifact, as it is an overarching framework for constructing IT artifacts (Abbasi et al., 2012; Hevner et al., 2004). The IT artifact is "a thing that has, or can be transformed into, material existence as an artificially made object (e.g., model, instantiation) or process (e.g., method, software)" (Gregor & Hevner, 2013). According to the design science research paradigm, this paper presents an innovative artifact – an AI-based framework utilizing state-of-the-art machine learning algorithms – to predict LOS and help hospitals better manage patient LOS for improved resource capacity planning and attaining efficient discharge plans (see Figure 1).

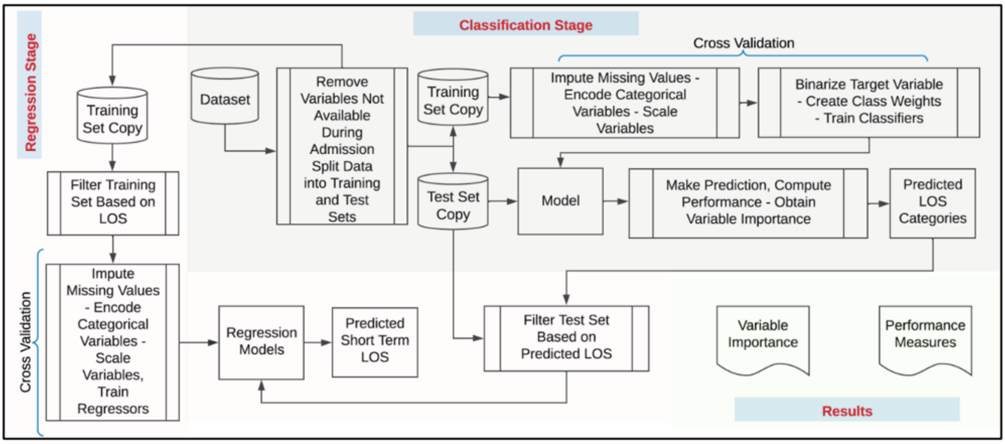
3. LITERATURE REVIEW: PROBLEM MOTIVATION & SOLUTION DEFINITION

3.1. EHR

According to HealthIT.gov the electronic healthcare record (EHR) is a digitalized patient's chart. EHR systems are built to provide HCPs with real time data. EHR systems go beyond standard clinical information and help HCPs by providing and inclusive view of the patient's care. Many EHR systems provide HCPs, staff, or other individuals with clinical decision support (CDS). HealthIT.gov also explains how CDS systems can provide knowledge and information intelligently presented at appropriate times, to enhance health and health care. In addition, CDS systems can deliver timely reminders and information to physicians, and may recommend screening tests, flag drug-drug interactions and drug-allergy information or discourage the provider from repeating a test by highlighting a previous result (Agha, 2014).

In most cases adopting the EHR revolves around cost and the implementation of the EHR system. Most health providers state that factors such as high startup cost and lack of capital are the financial

Figure 1. AI-based Artifact: Proposed framework



barriers behind not acquiring an EHR system (Fleming et al., 2011). A report by the RAND Corporation estimated that the adoption of EHR by hospitals could reduce annual health spending by as much as $81 billion and simultaneously improve healthcare quality (Hillestad et al., 2005). In response to the report by the RAND Corporation, an article published in the Journal Health Affairs state that the adoption of an EHR system would lead to high costs, lower productivity, and increased health spending (Sidorov 2006.) The effectiveness and cost of EHR systems stay debatable as more reports and studies are conducted.

There have been numerous case studies examining the cost and benefits of EHR adoption. According to Wang et al. (2003), the use of EHR systems generate a net benefit of $86,400 per provider within five years. Most researchers find that EHR's financial benefits may not occur until after years of its implementation. Chaudhry (2006) discusses three major benefits of EHR on quality: increased commitment to guideline-based care, enhanced surveillance and monitoring, and decreased medication errors. According to a survey conducted by the Centers for Disease Control and Prevention (CDC) in 2015, 86.9% of office-based physicians are using an EHR system in the US. In 2017, a survey showed that small to large hospitals had a certified EHR system in place ranging from 93-99%.

3.2. Reduction in Resources Cost and Improvement in Quality

Despite a study arguing that EHR systems should be viewed as a tool to reinforce patient-centered care rather than a cost-saving tool Sidorov (2006), various studies show that hospitals' operational costs decrease after several years of EHR system implementation. For example, Hillestad et al. (2005) and Fleming et al. (2011) can help hospitals save millions of dollars despite their high implementation costs (e.g., physical infrastructure, employee training, and a range of other expenses). A study demonstrates that the hospitals in the IT-intensive markets had a 3.4 percent decrease in costs within three years of basic EHR adoption (Dranove et al. 2014; (Khalil & Jones, 2007). EHR also improves the legibility of clinical notes, provides a decision support tool for drug ordering (e.g., including allergy warnings and drug incompatibilities), raises alerts for abnormal lab results, and helps providers manage chronic diseases, such as diabetes, hypertension, and heart failure (Fraser et al. 2005). Other benefits of EHR include gathering accurate clinical information, coordinating care, minimizing communication error, improving patient safety (Tubaishat and Omar 2017).

Numerous case studies and evidence reviews have been conducted to examine various aspects of EHRs. In a systematic review of HIT, Chaudry et al. (2006) determined that a broadly designed and implemented HIT system could provide benefits in terms of delivery of care, particularly in the domain of preventive health. A case study by Wang et al. (2003) determined that the small provider network used in their study saw modest to immediate cost benefits after implementing an EHR system. The researchers also acknowledged that any significant savings were projected to take years. Hillestad et al. (2005) raise concerns about "uneven adoption of EHR" and note the difficulty of adequately assessing short- and long-term costs and fairly apportioning cost burdens among stakeholders.

Although various factors may impact each hospital's LOS, research has shown that EHR & CDS systems can significantly affect the average patient's hospital LOS. EHR & CDS systems can influence various factors that can, as a result, reduce unnecessary hospital extended stays (Thompson et al., 2014). Using computerized physician order entry (CPOE) and clinical decision support (CDS) help healthcare professionals quickly identify medication allergies, recognize medication-to-medication interactions, easily access medical guidelines and thus reduce medication errors (Thompson et al., 2014). Patient movement between facilities, providers, and healthcare professionals limits the ability to access accurate information quickly. It may also yield duplicate efforts and delay the treatment process and, thus, the hospital discharge (Thompson et al., 2014). Integration of EHR & CDS systems and LOS prediction tools between different facilities enables richer and more accurate patient information to travel among the healthcare professionals in a timely manner and reduces overhead time. More integration yields a faster transfer of patients' data and will improve the efficiency of healthcare processes and care delivery and lower idle time (Angst et al., 2011). In addition, physicians who have access to such tools in the EHR are more efficient in their care delivery and will be associated with superior results (Dobrzykowski & Tarafdar, 2015).

To this extent, we conduct a literature review and examine recently published articles (2013-2019) regarding LOS prediction using EHR records. Table 1 summarizes the list of identified journal articles in our literature review. According to our literature review, and to the best of our knowledge, none of the prior LOS prediction studies attempt to examine the different factors that help planning available resources for optimal capacity management. Therefore, this study contributes the extant body of the literature by developing an artifact — AI models that predict the LOS. In addition to the development of the AI model, the importance of each variable from each model is extracted. The obtained importance values are then discussed in the context of the components of healthcare supply chain management.

4. DESIGN AND DEVELOPMENT OF THE ARTIFACT

4.1 Data Description

The dataset used in this study is obtained from Open New York (NY), which is an initiative of policies and programs that aims to provide public access to digital data for collaboration and analysis. This dataset – called the Statewide Planning and Research Cooperative System (SPARCS) Inpatient De-identified data – is comprised of patient characteristics, diagnoses, treatments, and services. Because this dataset is de-identified, it does not contain variables or identifiable elements and ensures the confidentiality, integrity of protected health information (PHI) under HIPAA. This study only considers the variables that are available to providers during the patient admission stage and removes the variables that can only be known to providers during the discharge stage (e.g., total charges). Then, the dataset is partitioned into training and test sets. The training and test sets are comprised of 70% and 30% of the instances, respectively. The 70:30 split guarantees that the training data has enough observations to reliably fit AI models. Additionally, the 70:30 split ensures that there are enough instances in the test set to measure the AI models' performance and generalizability on the unseen dataset. The variables, their definitions, and summary statistics are given in Table 2.

4.2. The Modeling Logic (Regressor & Classifiers)

Due to their superior performance obtained through the preliminary analysis, we employ two AI algorithms: Extreme Gradient Boosting (XGBoost) and Artificial Neural Networks to predict patient LOS. Since the LOS variable in this dataset is highly skewed and can take values between one day and 120 days, we first discretize it into six categories as follows:

Table 1. Summary of studies aiming to predict LOS

Authors	Methods used	Primary Contribution & Results
Barnes et al (2016)	Logistic regression, Random Forest	Introduced automated tool to predict LOS for discharge predictions. Compared model results to clinician predictions (results show higher sensitivity and lower specificity).
Cai et al. (2016)	Bayesian Network	Develop a Bayesian network model for EHR systems to estimate daily probabilities of LOS, mortality, and readmission.
Huang et al. (2013)	Case-based Reasoning methodology	Use prior information of treatment processes that have temporal similarity. Evaluated with 284 pulmonary infection CTPs patient.
Livieris et al. (2018)	Naive Bayes, ANN, SVM, kNN, Random Forest	Create a decision support software to predict LOS using a two-level classification algorithm (accuracy ~ 78.53%).
Osnabrugge et al. (2014)	Linear regression models	Develop an LOS prediction model using coronary artery bypass grafting patients (R^2 =0.51).
Rajkomar et al. (2018)	Deep learning neural network	Use deep learning techniques on raw EHR records to predict several medical events across different centers without site-specific data harmonization
Sotoodeh & Ho (2019)	Hidden Markov model-based framework	Use first 48 hours admission patients' physiological measurements to predict LOS of intensive care unit patients.
Turgeman et al. (2017)	Neural network CART tree CHAID tree Support vector machine Regression tree (Cubist)	Cubist model gave best LOS prediction results and is also interpretable. Prediction error is higher for patients with more recent admissions and longer stays

- Category 1 (i.e., less than 5 days)
- Category 2 (i.e., between 6 and 10 days)
- Category 3 (i.e., between 11 and 20 days)
- Category 4 (between 21 and 30 days)
- Category 5 (between 31 and 50 days)
- Category 6 (more than 50 days)

We then develop two classification models – called XGBoost and ANN classifiers – through Cross Validation (CV) using the training dataset to predict the patient LOS category. Afterward, we develop two regression models – called XGBoost and ANN regressors – through CV using the training set to forecast the patients' LOS in Categories 1 and 2. In other words, we utilize the classifiers to identify the patient LOS category. Then, we filter the patients in Categories 1 and 2 and feed them through the regressors to compute their LOS in days. There are several reasons to first predict LOS categories and then LOS in days using classifiers and regressors.

1. Regressors do not produce reliable results for patients expected to stay longer (i.e., Categories 3 through 6) due to only using admission level variables. For example, complications and procedures that occurred within the first couple weeks of these patients' stays may drastically impact their LOS.
2. Hospitals are interested in predicting the LOS of patients expected to stay short-term (i.e., Categories 1 and 2) in the admission stage so they can promptly plan their bed capacity and resources.
3. Hospitals can use the classification models to predict the patients in Categories 3 through 6. Even though the classification model predictions are not as granular as the regression model predictions, they can still help hospitals plan their long-term resources.

Table 2. Summary of the dataset used to predict LOS

Variable Name	Variable Description	Summary Statistics			
LOS	Total number of patient days	Standard Deviation	8	Average 6	
Age Group	Age grouped into categories	70 or Older	29%	0 to 17	14%
		50 to 69	28%	18 to 29	11%
		30 to 49	18%		
Gender	Patient gender	Female	56%		
		Male	44%		
Race	Patient race	White	69%	Multi-racial	1
		African American	14%	Other Race	16%
Ethnicity	Patient ethnicity	Not Hispanic	86%	Unknown	6%
		Hispanic	8%		
Type of Admission	Manner in which the patient was admitted to health care facility	Emergency	62%	Newborn	9%
		Elective	21%	Urgent	8%
CCS Diagnosis Code	AHRQ Clinical Classification Software Diagnosis Category Code	Liveborn	9%	Mood disorders	3%
		Osteoarthritis	3%	Heart failure	3%
		Others combined	67%	Septicemia	5%
CCS Procedure Code	AHRQ Clinical Classification Software ICD-9 Procedure Category Code	No Procedure	35%	Ventilator	3%
		Assisting delivery	3%	Drug detox	3%
		Others combined	55%	Therapeutic	4%
APR DRG Code	APR-DRG Classification Code	Neonatal birth	8%	Heart failure	3%
		Delivery	9%	Infections	5%
		Others combined	75%		
APR MDC Code	Patient Refined Major Diagnostic Category Code.	Circulatory	11%	Respiratory	10%
		Musculoskeletal	10%	Digestive	9%
		Others combined	50%	Childbirth	10%
APR Severity of Illness Code	APR-DRG Severity of Illness Code	Moderate	39%	Major	23%
		Minor	32%	Extreme	6%
APR Risk of Mortality	Patient Refined Risk of Mortality	Moderate	57%	Major	15%
		Minor	22%	Extreme	5%
APR Medical Surgical Description	Patient Refined Severity of Illness	Medical	77%	Surgical 23%	
Emergency Department Indicator	Code showing if patients come from emergency department	Yes	58%	No 42%	
Payment Typology 1 through 3	Descriptions of payment type	Private Health Insur.	11%	Blue Cross	13%
		Medicare	41%	Other combined	7%
		Medicaid	28%		

4.3. AI Algorithms

XGBoost is an ensemble AI technique that utilizes the improved version of the Gradient Boosting Algorithm. Boosting algorithms aim to create a robust model (i.e., a model with high predictive power) by iteratively building weak models (i.e., models with low predictive power) that reduce the previous model's error rates. XGBoost first constructs an initial model using the training data and computes the error between the actual and predicted target variables. It starts training the weak learners by giving more importance to the observations with high error rates. In each iteration of the XGBoost algorithm, a weak learner is trained on a newly sampled distribution: the remaining errors of the previous weak learner, which is called pseudo-residuals. Finally, the contribution of the weak learner to the final model is computed using the gradient descent optimization process so that the overall error rate of the final model is minimized. In order to ensure that the XGBoost model generates good performance, we tune one of its most essential hyperparameters: the number of estimators. The number of estimators hyperparameter in XGBoost determines the number of weak learners in the final model. In general, the XGBoost performance increases with respect to the number of estimators; however, this performance increase flattens out at a certain point. Additionally, high values for the number of estimators may cause overfitting and drastically increase the computational time required to train the model. Thus, we vary the number of estimators between 10 and 250 through cross-validation as explained below to identify the XGBoost model giving the best performance.

ANN is a multilayer perceptron-based AI algorithm that can accurately model complex datasets. Typically, an ANN model is comprised of input, hidden, and output layers with multiple neurons. The neurons in the input and output layers represent the input and target variables, respectively. The neurons in the hidden layer represent the latent features derived as linear combinations of input variables (James et al., 2013). The neurons in different layers are via weights, and the ANN algorithm generally uses backpropagation to find the neurons' optimal weights. In this study, we adopt an ANN structure with a single hidden layer. Increasing the number of neurons in the hidden layer allows the ANN model to incorporate more latent features and increasing model performance. However, this can cause the ANN structure to become overly complex and result in overfitting. Therefore, the number of neurons in the hidden layer is varied between 2 to 150 via CV to identify its optimal value.

4.4. Cross-Validation

zIn 5-fold CV, the training dataset is randomly split into 5-subsets of nearly equal size. Then, an ANN classifier with a particular number of neurons and an XGBoost classifier with a specific number of estimators are instantiated. Afterward, the observations in the $5-1$ folds are preprocessed. The preprocessing stage within CV includes imputing missing values, encoding the categorical input variables, scaling the dataset, and balancing it with Synthetic Minority Oversampling Technique (SMOTE). The XGBoost and ANN classifiers are then trained using the discretized LOS variable as the target variable, and their performance measures are computed on the last remaining fold. It is critical to note that the last remaining fold is also preprocessed using the pipeline obtained from the training stage. The preprocessing steps for the last fold include imputing missing values, encoding the categorical input variables, and scaling the dataset. These steps are repeated for every fold in 5-folds for once, allowing each fold to become a test case for once. The performance of XGBoost and ANN classifiers with their respective parameters can be obtained by averaging their performance measures throughout the CV process. The XGBoost and ANN classifiers with the best performance metrics are selected as the final models (Simsek et al. 2020). Typically, accuracy, recall, precision, specificity, F-score, and the AUC are used to assess the performance of AI models used for classification. Because AUC incorporates both recall and specificity, we choose it as the primary evaluation metric for the XGBoost and ANN classifiers.

For the XGBoost and ANN regressors, we use the same logic as described above. However, we employ the original LOS variable as the target variable and only include the observations in Categories 1 and 2 to train the regressors. It should also be noted regressors do not require balancing the dataset. We utilize the mean absolute error (MAE), and the root mean square error (RMSE) metrics as the XGBoost and ANN regressors' primary performance measures. The formulas to compute the ANN and XGBoost algorithms' performance measures through CV are given in formulas 1 and 2, where h is the number of hidden layers in the ANN algorithm and e is the number of estimators in the XGBoost algorithm. After finding the performance measures for each AI model with a specific hyperparameter, the model with the best performance metric is selected, as provided in Equations 3 and 4.

$$P^{ANN_h} = \frac{1}{5}\sum_{k=1}^{5} AUC_k, \mathrm{h} \in \left\{2, 3, 4, ...150\right\} \quad (1)$$

$$P^{XGBoost_e} = \frac{1}{k}\sum_{i=1}^{k} AUC_i, \mathrm{e} \in \left\{2, 3, 4, ...150\right\} \quad (2)$$

$$Optimal\, h = \underset{h}{\operatorname{argmax}}\, P^{ANN_h}, \mathrm{h} \in \left\{2, 3, 4, ...150\right\} \quad (3)$$

$$Optimal\, e = \underset{e}{\operatorname{argmax}}\, P^{XGBoost_e}, \mathrm{e} \in \left\{2, 3, 4, ...150\right\} \quad (4)$$

4.5. Feature Importance

Because ANN has a Blackbox nature, it does not provide insights regarding its internal structure and how it obtains a particular prediction. XGBoost provides importance values for input variables. However, the feature importance values directly obtained from the XGBoost model may be inflated for the categorical variables with many levels (Parr et al. 2018). Hence, we choose to use the permutation importance algorithm to interpret the models. In order to obtain the feature importance of a variable using the permutation importance algorithm, an AI model is trained using the training set, and its accuracy value for the test set is calculated. This accuracy is called original accuracy. Then, the values of the variable whose permutation importance is computed are randomly shuffled in the test set. Using this new test set, a new accuracy, called shuffled accuracy, is computed. Afterward, the difference between the original accuracy and the shuffled accuracy is obtained to determine the mean decrease in accuracy (MDA). The permutation importance of a feature is proportional to its MDA. A variable is considered important if shuffling its values results in a high MDA value because the model heavily relies on the feature to generate accurate predictions.

5. DEMONSTRATION OF THE SOLUTION

In this section, we report the results of the developed artifact and discuss the results of different algorithms. As shown in Figure 2, we plot the different number of parameters versus the AUC values for the XGBoost algorithm. We notice that the AUC will reach above 0.84 after 50 estimators and keep improving with the increase of the number of estimators until 150, which will be the optimal AUC value for the average and the different categories. Regarding the model's accuracy, the optimal value can be found closer to the 200 estimators. For the ANN model, we plot the number of neurons versus the AUC. The optimal AUC for the average across various categories is 60 neurons and optimal accuracy around 115 hidden neurons (see Figure 2).

In terms of evaluating the performance of the different models, we report the Area Under the Curve (AUC) for each model. The receiver operating characteristic (ROC) curve is plotted using the true positive (TP) against the false positive (FP) rates for each model. In Figures 4, we plot the XGBoost ROC curve and the ANN ROC curve; the curves describe the discriminative power of each classifier independently of the class distribution. The earlier classifier shows a superior AUC of 0.85 comparing to the latter with an AUC equivalent to 0.82 (see Table 3). In general, the AUC is considered a more desirable metric for comparing different models than accuracy. Besides, we noticed that the XGBoost model is superior in predicting patients in category 6 with an AUC of 0.93 and still better than the ANN classifier for the same group (and all other categories, as shown in Figure 3).

We also use the XGBoost and ANN regressors on categories 1 (LOST less than 6 days) and category 2 (LOS between 6 and 11 days). A widely used metric to examine regression-based models used in predictions and forecasting is the root mean square error (RMSE). The metric represents the standard deviation of the differences between the forecasted or predicted value and the observed value, referred to as residuals. The residual, in general, is how far or spread out an actual value from the model prediction

Figure 2. Cross-validation and hyperparameter tuning results for XGBoost classifier

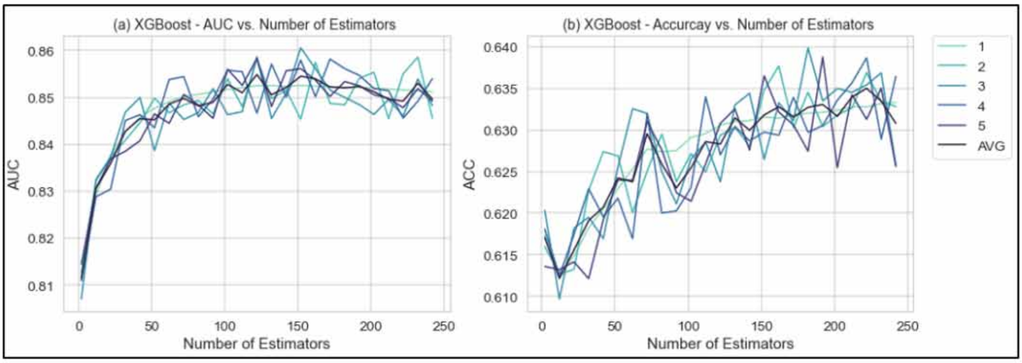


Figure 3. Cross-validation and hyperparameter tuning results for ANN classifier

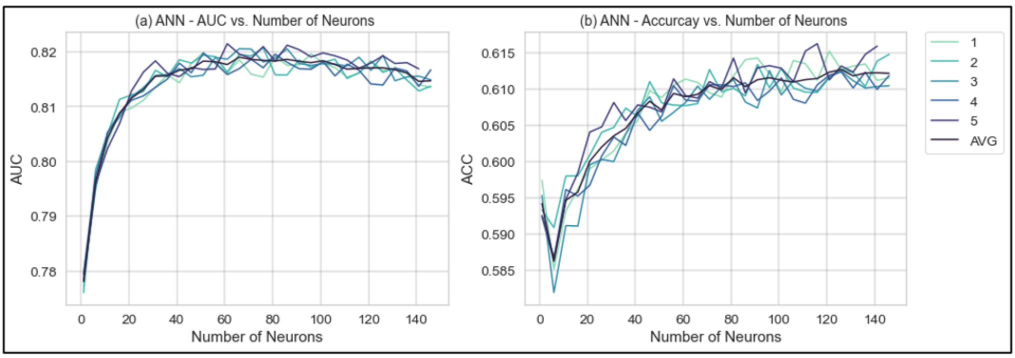


Figure 4. AUC of the best XGBoost and ANN classification models

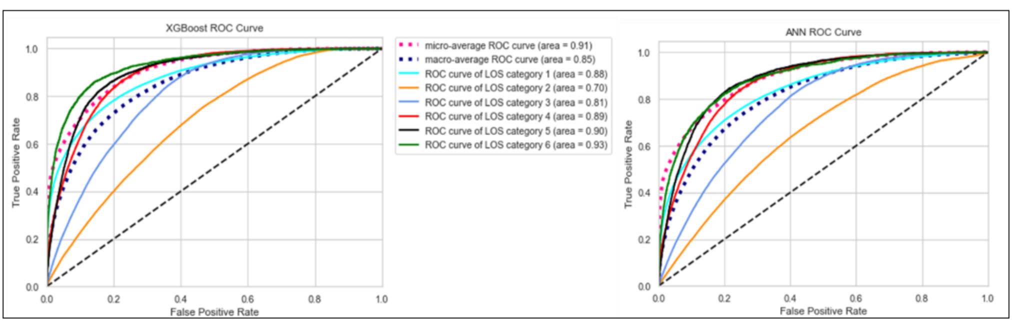


Table 3. Performance results of the best classification models

Model	AUC	Accuracy	Recall	Precision	F- Score
XGBOOST Classifier	0.85	0.64	0.64	0.75	0.68
ANN Classifier	0.82	0.62	0.62	0.73	0.66

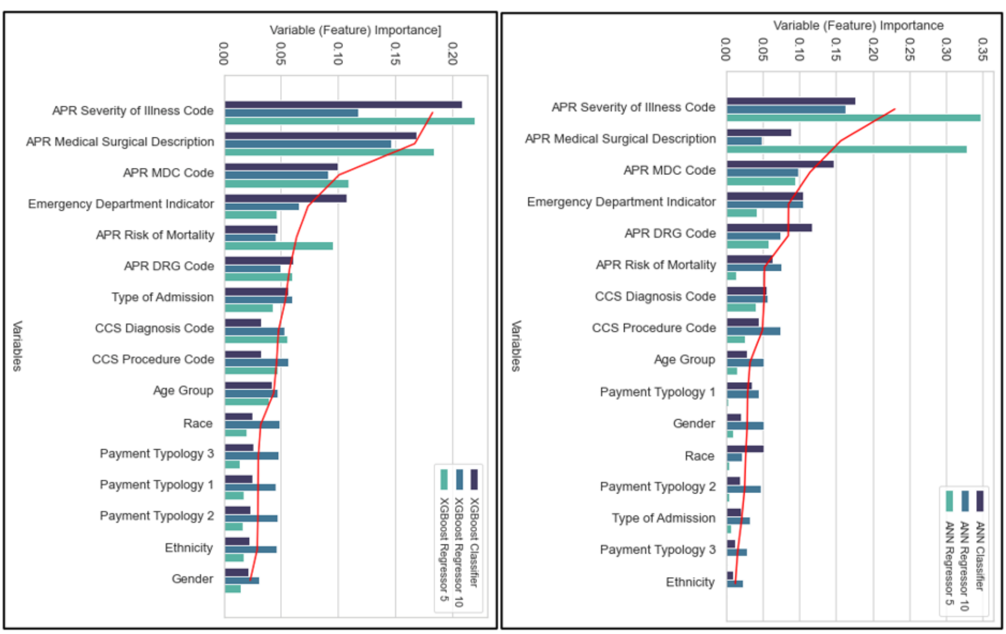
line. The XGBoost regressor with an RMSE of 1.07 days outperforms the ANN regressor (see Table 4, column 3). We also include another metric to compare between the candidate models. The metric known as the mean absolute error (MAE) represents an arithmetic average of the absolute differences between the predicted values and the observed ones. The XGBoost remains consistently showing better results for LOS of less than 6 days with an MAE of 0.85 days and for LOS of more than 6 days and less than 11 days with an MAE equal to 1.12 days. The ANN showed inferior results of MAE equivalent to 0.90 of a day for Category 1 and MAE of 1.12 days for Category 2 (see Table 4, column 4).

It is a common practice in data mining projects, specifically in predictive analytics, to analyze the different features used in a model to predict a certain target or outcome to determine the importance of features in terms of their effect on the model predictability. As seen below in Figures 5.a and 5.b, the APR severity of illness comes in the first place in terms of importance for both the XGBoost and the ANN classifiers. The APR medical surgical description falls in second place for the XGBoost classifier, but the APR MDC code is the second in importance for the ANN classifier. In terms of regressors, we found that the APR severity of illness is the highest in terms of feature importance for the XGBoost regressor 5 and ANN regressor 5. We also notice that it is still the case for the ANN regressor 10 but not for the XGBoost regressor 10 where APR medical surgical description is the most important feature for that model.

Table 4. Performance results of the best regression models

LOS Category	Model	RMSE	MAE
Category 1 (LOS Less than 6)	XGBoost Regressor	1.07 days	0.85 days
	ANN Regressor	1.11 days	0.90 days
Category 2 (LOS between 6 and 11)	XGBoost Regressor	1.34 days	1.12 days
	ANN Regressor	1.58 days	1.13 days

Figure 5. (a) Variable importance obtained from XGBoost Models (b) Variable importance obtained from ANN Models



(a) (b)

6. DISCUSSION AND EVALUATION OF THE SOLUTION

The patient length of stay is considered an essential metric for healthcare supply chain management. Healthcare professionals will need to consider different factors when planning available resources for optimal capacity management. The continuous demand and pressure on healthcare institutions and facilities force managers and decision-makers to allocate and distribute resources efficiently. Decision support systems equipped with predictive models can help Healthcare Professionals, managers, and decision-makers apply timely interventions and make informed decisions. In this study, we used advanced Machine Learning techniques, namely XGBoost and ANN, to predict patients' LOS. Our results showed that XGBoost can be used on 16 features (variables) to predict LOS with an average AUC of 0.85. As shown in Figure 5.a developed solution uses a list of needs as an input in the developed model ranked from the most important to the least. The extreme gradient boosting model has been dominating the ML world with so many solutions and applications trying to leverage the power of the model. This paper has employed design science research to develop an AI model to predict LOS in terms of research contributions. This study contributes to the body of literature by creating a model with high prediction accuracy and identifying the variables that contribute most to predicting LOS.

7. CONCLUSION, IMPLICATIONS, AND LIMITATIONS

This study aimed to design and develop an artifact that will help healthcare professionals, managers, and decision-makers employ needful interventions and better plan and allocate resources in the supply chain context. To that extent, we utilize the design science research paradigm to develop an AI model using the well-known extreme gradient boost algorithm. We fine-tune the algorithm to establish the model that delivers an AUC of 0.85 using 16 widely available medical features that are favorable compared to the existing literature. The developed solution and the findings of this paper provide implications for using Artificial Intelligence, particularly Machine Learning, to enhance hospitalization metrics like LOS. The paper also provides practical implications that can improve resource capacity planning in healthcare organizations to increase operational efficiency. Finally, the study has some limitation which future research can address. The data used in this study was secondary data. Future projects can address this limitation by collecting primary data that includes other or additional features. Future research should demonstrate the accuracy of the framework using multiple datasets. This will prove the generalizability of the results.

REFERENCES

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A meta-learning framework for detecting financial fraud. *Management Information Systems Quarterly*, *36*(4), 1293–1327. doi:10.2307/41703508

Angst, C. M., Devaraj, S., Queenan, C. C., & Greenwood, B. (2011). Performance effects related to the sequence of integration of healthcare technologies. *Production and Operations Management*, *20*(3), 319–333. doi:10.1111/j.1937-5956.2011.01218.x

Barnes, S., Hamrock, E., Toerper, M., Siddiqui, S., & Levin, S. (2016). Real-time prediction of inpatient length of stay for discharge prioritization. *Journal of the American Medical Informatics Association : JAMIA*, *23*(e1), e2–e10. doi:10.1093/jamia/ocv106 PMID:26253131

Cai, X., Perez-Concha, O., Coiera, E., Martin-Sanchez, F., Day, R., Roffe, D., & Gallego, B. (2016). Real-time prediction of mortality, readmission, and length of stay using electronic health record data. *Journal of the American Medical Informatics Association : JAMIA*, *23*(3), 553–561. doi:10.1093/jamia/ocv110 PMID:26374704

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., & Shekelle, P. G. (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine*, *144*(10), 742–752. doi:10.7326/0003-4819-144-10-200605160-00125 PMID:16702590

Dobrzykowski, D. D., & Tarafdar, M. (2015). Understanding information exchange in healthcare operations: Evidence from hospitals and patients. *Journal of Operations Management*, *36*(1), 201–214. doi:10.1016/j.jom.2014.12.003

Fernandez-Luque, L., & Imran, M. (2018). Humanitarian health computing using artificial intelligence and social media: A narrative literature review. *International Journal of Medical Informatics*, *114*, 136–142. doi:10.1016/j.ijmedinf.2018.01.015 PMID:29395987

Fleming, N. S., Culler, S. D., McCorkle, R., Becker, E. R., & Ballard, D. J. (2011). The financial and nonfinancial costs of implementing electronic health records in primary care practices. *Health Affairs*, *30*(3), 481–489. doi:10.1377/hlthaff.2010.0768 PMID:21383367

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *Management Information Systems Quarterly*, *37*(2), 337–355. doi:10.25300/MISQ/2013/37.2.01

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly*, *28*(1), 75–105. doi:10.2307/25148625

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., & Taylor, R. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs*, *24*(5), 1103–1117. doi:10.1377/hlthaff.24.5.1103 PMID:16162551

Huang, Z., Juarez, J. M., Duan, H., & Li, H. (2013). Length of stay prediction for clinical treatment process using temporal similarity. *Expert Systems with Applications*, *40*(16), 6330–6339. doi:10.1016/j.eswa.2013.05.066

Kaboli, P. J., Go, J. T., Hockenberry, J., Glasgow, J. M., Johnson, S. R., Rosenthal, G. E., Jones, M. P., & Vaughan-Sarrazin, M. (2012). Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 Veterans Affairs hospitals. *Annals of Internal Medicine*, *157*(12), 837–845. doi:10.7326/0003-4819-157-12-201212180-00003 PMID:23247937

Khalil, M. M., & Jones, R. (2007). Electronic Health Services An Introduction to Theory and Application. *The Libyan Journal of Medicine*, *2*(4), 202–210. doi:10.3402/ljm.v2i4.4732 PMID:21503245

Livieris, I. E., Kotsilieris, T., Dimopoulos, I., & Pintelas, P. (2018). Decision support software for forecasting patient's length of stay. *Algorithms*, *11*(12), 199. doi:10.3390/a11120199

Orsi, G. B., Di Stefano, L., & Noah, N. (2002). Hospital-acquired, laboratory-confirmed bloodstream infection: Increased hospital stay and direct costs. *Infection Control and Hospital Epidemiology*, *23*(04), 190–197. doi:10.1086/502034 PMID:12002233

Osnabrugge, R. L., Speir, A. M., Head, S. J., Jones, P. G., Ailawadi, G., Fonner, C. E., Fonner, E. Jr, Kappetein, A. P., & Rich, J. B. (2014). Prediction of costs and length of stay in coronary artery bypass grafting. *The Annals of Thoracic Surgery*, *98*(4), 1286–1293. doi:10.1016/j.athoracsur.2014.05.073 PMID:25110338

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, *1*(1), 1–10. doi:10.1038/s41746-018-0029-1 PMID:31304302

Salah, H. T., Muhsen, I. N., Salama, M. E., Owaidah, T., & Hashmi, S. K. (2019). Machine learning applications in the diagnosis of leukemia: Current trends and future directions. *International Journal of Laboratory Hematology*, *41*(6), 717–725. doi:10.1111/ijlh.13089 PMID:31498973

Sidorov, J. (2006). It ain't necessarily so: The electronic health record and the unlikely prospect of reducing health care costs. *Health Affairs*, *25*(4), 1079–1085. doi:10.1377/hlthaff.25.4.1079 PMID:16835189

Silow-Carroll, S., Edwards, J. N., & Rodin, D. (2012). Using electronic health records to improve quality and efficiency: The experiences of leading hospitals. *Issue Brief (Commonwealth Fund)*, *17*(1), 40. PMID:22826903

Sotoodeh, M., & Ho, J. C. (2019). Improving length of stay prediction using a hidden Markov model. *AMIA Joint Summits on Translational Science Proceedings AMIA Summit on Translational Science*, *2019*, 425. PMID:31258996

Thomas, J. W., Guire, K. E., & Horvat, G. G. (1997). Is patient length of stay related to quality of care? *Journal of Healthcare Management*, *42*(4), 489. PMID:10174462

Thompson, D., Mohan, S., & Goddard, K. (2014). *Reducing Hospital Length of Stay with EMR and Other Information Technologies*. NIH.

Thompson, D. I., Classen, D. C., & Haug, P. J. (2006). EMRs in the fourth stage: The future of electronic medical records based on the experience at intermountain health care. *Journal of healthcare information management. Journal of Healthcare Information Management*, *21*(3), 49–60. PMID:19195294

Turgeman, L., May, J. H., & Sciulli, R. (2017). Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission. *Expert Systems with Applications*, *78*, 376–385. doi:10.1016/j.eswa.2017.02.023

Von Wedel, P., & Hagist, C. (2020). Economic Value of Data and Analytics for Health Care Providers: Hermeneutic Systematic Literature Review. *Journal of Medical Internet Research*, *22*(11), e23315. doi:10.2196/23315 PMID:33206056

Wang, S. J., Middleton, B., Prosser, L. A., Bardon, C. G., Spurr, C. D., Carchidi, P. J., Kittler, A. F., Goldszer, R. C., Fairchild, D. G., Sussman, A. J., Kuperman, G. J., & Bates, D. W. (2003). A cost-benefit analysis of electronic medical records in primary care. *The American Journal of Medicine*, *114*(5), 397–403. doi:10.1016/S0002-9343(03)00057-3 PMID:12714130

Zhang, R., Burgess, E. R., Reddy, M. C., Rothrock, N. E., Bhatt, S., Rasmussen, L. V., Butt, Z., & Starren, J. B. (2019). Provider perspectives on the integration of patient-reported outcomes in an electronic health record. *JAMIA Open*, *2*(1), 73–80. doi:10.1093/jamiaopen/ooz001 PMID:30976756

Yazan Alnsour received his Ph.D. in Information Systems from the University of Colorado Denver, Business School. He has a B.S. in Computer Engineering, Master of Business Administration (MBA), and industry experience as a database programmer, Business Intelligence (BI) developer, and Business Analyst. Yazan's research interests are in Data & Text Analytics, Information Systems Security & Privacy, Health Information Systems, and Business Value of Information Systems. He also has valuable teaching experience delivering face-to-face and online graduate and undergraduate courses in the areas of Data Analytics, Big Data, Information Management, Information Security, Business Process Management, IT Governance, and System Analysis and Design.

Marina Johnson is an assistant professor at Montclair state University

Abdullah Albizri is an Assistant Professor of Information Management & Business Analytics at Feliciano School of Business, Montclair State University. He holds a PhD degree in Management Information Systems from Sheldon B. Lubar School of Business, University of Wisconsin - Milwaukee. He has authored and co-authored several papers that appeared in Communications of the Association for Information Systems, Journal of Information Systems, Annals of Operations Research, Journal of Business Analytics, Journal of Enterprise Information Management, and the proceedings of major IS conferences such as ICIS and AMCIS. His research interests focus on AI solutions, blockchain applications, and the role of Information Systems in sustainability and human well-being. Prior to joining academia, Dr. Albizri worked as an IT business consultant in the banking industry.

Antoine Harfouche is an Associate Professor of Information Systems (IS) and Artificial Intelligence (AI) at University Paris Nanterre where he teaches undergraduate and graduate courses in various areas, including IS, AI, Big Data, and Quantitative Methods. Dr. Harfouche has contributed to the IS & AI community through his excellent teaching, cutting-edge research, and outstanding service. Dr. Harfouche completed his M.Sc. and Ph.D. in Management Information Systems at Paris Dauphine University FR) / PhD done in collaboration with Georgia State University. Dr. Harfouche's research primarily examines how IS and AI impact individuals, organizations, countries, and societies in general. His publications appeared in peer-reviewed journals (e.g., the Annals of Operation Research, Trends in Biotechnology, Information Technology and People, Lecture Notes in Information Systems and Organization) and renowned conference proceedings (e.g., ICIS, PACIS, MCIS). Throughout his career, Dr. Harfouche has obtained a considerable amount of funding - above 2 million dollars - from the French Research Council (ANR) and the European ERANSMUS+. In the ANR SCHOPPER project (ANR-DS0701/2016), he designed a new AI framework called Informed AI, in which human input is considered as an inextricable part of AI applications. In the ERASMUS+ project VRAILEXIA, Dr. Harfouche has co-designed a strategic partnership project called "Partnership outside of the box: integrating artificial intelligence tools with Virtual reality to support higher education students with dyslexia." This innovative project was given a score of 98 out of 100 by the ERASMUS+ agency, thus achieved first place in 2020.