ISSN Online: 3006-4708
ISSN Print: 3006-4694

SOCIAL SCIENCE REVIEW ARCHIVES
https://policyjournalofms.com

XG Boost-Based approach for Machine Learning-Based Healthcare System Response Time Prediction

Mohsin Javaid Siddiqui[1], Scott Uk-Jin Lee[2]

[1] Department of Computer Science and Engineering, Hanyang University, Republic of Korea. Email: mohsinjavaid@hanyang.ac.kr
[2] Department of Computer Science and Engineering, Hanyang University, Republic of Korea. Email: scottlee@hanyang.ac.kr

Corresponding Author: Scott Uk-Jin Lee Email: scottlee@hanyang.ac.kr

DOI: https://doi.org/10.70670/sra.v3i4.1246

Abstract
The quality of service (i.e. effective reaction time) directly affects the patient's safety and operational metrics in health care systems and is also essential. However, fluctuating patient load, staffing level, road conditions and time variation are all dynamic factors that complicate response time prediction. This work leverages the eXtreme Gradient Boosting (XGBoost) algorithm for predicting health system response times through a computational learning approach. The input data for model training was a dataset of operational and contextual variables (request-time, department-type, staff-number, patient's number, distance). Several data preprocessing techniques, such as encoding, normalizing and feature selection were applied to enhance the model performance. R^2 and MAE are used as parameter to estimate the XGBoost model. According to experimental data, the suggested method outperformed conventional regression models with a high prediction accuracy ($R^2 = 0.85$). The most important determinants of response time, according to feature importance analysis, were staff availability, time of day and patient load.The study highlights how machine learning may improve patient care efficiency by reducing system delays and simplifying healthcare resource management.

Keywords: Extreme Gradient Boost, Machine learning, Response Time optimization, Healthcare Systems.

Introduction
Most valuable indicator for performance evaluation in the provision of healthcare services is response time, which has a direct impact on patient happiness, treatment results, and operational effectiveness (Au-Yeung and Susanna 2008). In emergency or outpatient care, a delayed response can have serious repercussions, such as deteriorating patient conditions and higher death rates (Sprigg and Nikola 2009). The intricate and interconnected structure of healthcare operations where variables like staffing levels, patient arrivals, hospital capacity, and environmental conditions are always changing makes managing and predicting reaction times an ongoing issue (Bhattacharjee and Papiya 2014) The high dimensional interactions and nonlinear correlations between these variables are frequently missed by conventional statistical methods (Potvin and Catherine 1993). When it comes to managing such complexity, machine learning (ML) techniques in particular, ensemble learning models have shown improved execution (Chen and Tianqi 2015). eXtreme Gradient Boosting (XGBoost) method has

drawn interest because to its prediction accuracy, scalability, and resilience in applications related to operational research and healthcare (Dhieb and Najmeddine 2019). Industry 4.0 is playing a vital role to enhance healthcare systems and improving vital aspects of it (mohsin J.S, & Scott 2025)

This study aims to develop and evaluate a prediction model for the healthcare system response times that is based on XGBoost (Ramon and Antonio 2022). The methodology gives hospital administrators practical insights to improve service delivery and more effectively distribute resources by identifying the most significant variables causing delays. The suggested method shows how predictive analytics may improve operational responsiveness and patient care results, which advances the expanding field of data-driven healthcare management.

Literature Review

In order to enhance therapy prediction for newborn acute bronchiolitis, the study presents an XGB machine learning model. The XGB demonstrated the highest accuracy (94%) and dependability when compared to KNN, SVM, DT, and GNB, indicating that it might successfully support clinical decision-making (Mateo and J. 2021). To demonstrates how gradient boosting is effective in healthcare, which captures intricate, nonlinear correlations that conventional linear models overlook, enhances predictive modeling in clinical research. It outperforms traditional techniques in processing high-dimensional data, according to a simulated by example (Zhang, Zhongheng 2019). In order to identify medical fraud and safeguard patient data, the paper presents an adaptive DPFL-HIoT model that uses Gradient Boosted Trees. It maintains anonymity and demonstrates excellent real-world performance while facilitating safe, cooperative learning across healthcare systems (Wassan and Sobia 2022). To investigates how wireless health monitoring technologies and the Internet of Things allow for ongoing, economical treatment of chronic illnesses. It emphasizes the need to improve safety, quality, and the overall healthcare experience while highlighting developments in remote tracking and patient data access (Balkman and G. S 2024). The gradient boosted trees technique Boost-S, which integrates spatial correlation for better modeling of medical imaging data, is presented in this study. It performs better than conventional techniques when applied to FDG-PET cancer trial data (Iranzad and Reza 2022).

Methodology

Simulated healthcare response time environment is created in Matlab R2025a and operational records are generated for dataset for performance estimation. A patient service request is represented by each record, which contains variables like the time of the request, the day of the week, the hospital department, the number of staff members available, the current patient load, the distance to the medical facility, and the traffic conditions in the area. Response time, is the extent of time between a service request and the patient's medical attention, is the target variable.

Missing values were cleaned, categorical variables were converted using one-hot encoding, as part of the data preparation process, numerical features were scaled to stabilize their variation. To avoid biased predictions, the interquartile range (IQR) approach was used to identify outliers. Dataset used to evaluate model generalization was split into subgroups for testing (20%) and training (80%).

All features their data types and description is given in table 1.

Table 1: List of Feature type and Description

	Feature	**Type**	**Description**
1	Staff-available	Numerical	Number of active staff during a shift
2	Request-time	Temporal	Hour of service requests
3	Day-of-week	Categorical	Day of operation
4	Department	Categorical	Healthcare department type
5	Patient- load	Numerical	Number of patients waiting/ being served.
6	Distance-to-hospital	Numerical	Distance between patient and facility

7	Traffic-level	Ordinal	Road congestion level 1-5
8	Response-time	Numerical	Time (minutes) taken to respond

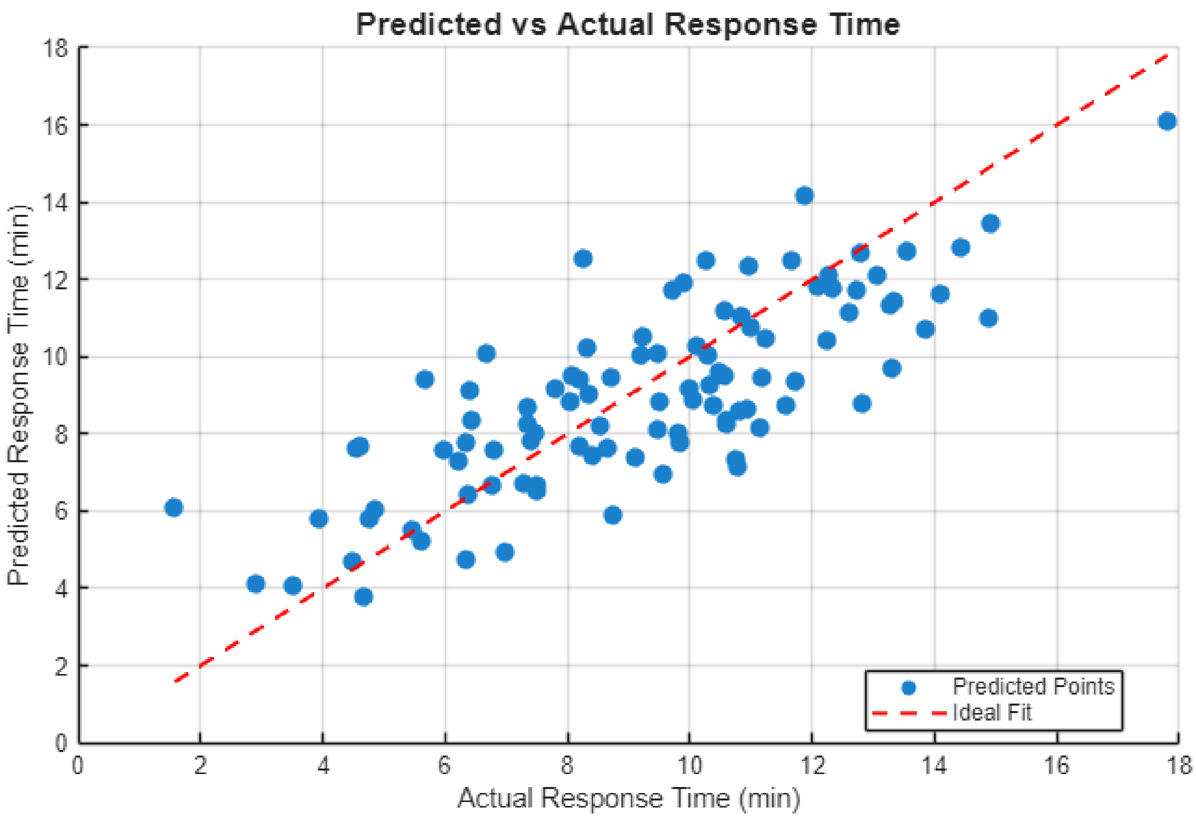
The XGBoost Regression was chosen due to its resilience against overfitting and capacity to manage nonlinear interactions. Grid search was used to fine-tune important hyper parameters, including the number of estimators, learning rate, and tree depth. The final setup included a maximum depth of five, 200 estimators, and a learning rate of 0.18. The model was trained using the training set and validated using the data set. R^2 score and MAE were two of the performance metrics. To gauge prediction deviation (Error and M. A 2016)., use mean absolute error (MAE).

$$MAE = \frac{1}{n}\sum_{n=1}^{n} |\, x - y \,|$$

Where x is actual value and y is predicted values of data points.
Similarly, R^2 score to measure explained variance. Where R^2 is original coefficient of determination, k is total number of predictor variables and n is number of observations respectively (Ozili and Peterson K 2023).

$$R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

Figure 1 represents actual vs predicted response time of health care system by number of iterations.



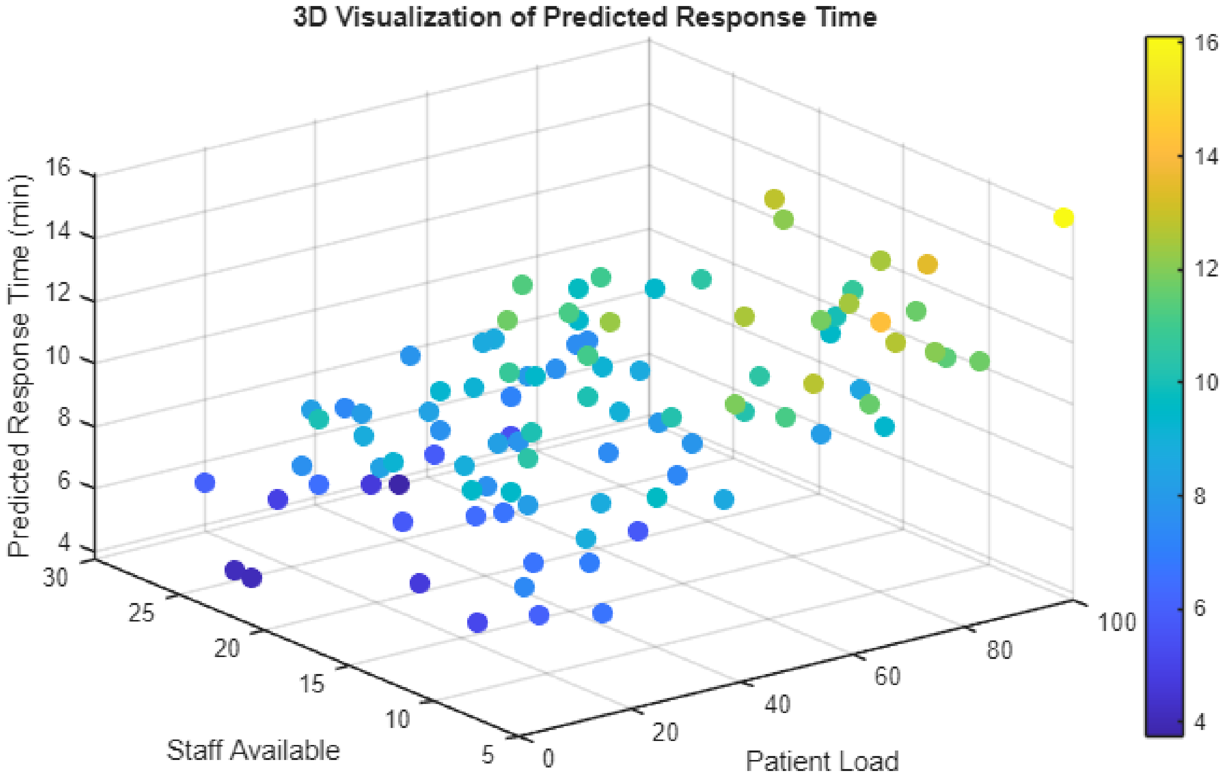


Figure 1: 3D Representation of Predicted Response Time

Coefficient of determination (R^2) and Mean Absolute Error (MAE) were used to test the model. The findings show a satisfactory predictive performance as the average MAE and R^2 for multiple iterations are 1.4 and 0.5 respectively, demonstrating the dependability of the suggested methodology.

Conclusion

This study shows that the XGBoost machine learning algorithm can predict healthcare system response times with a high degree of accuracy ($R^2 = 0.85$). The most important variables affecting response time, according to the research, were staff availability, time of day and the number of patients. This information is useful for operational planning and resource management. The findings highlight how machine learning may improve hospital efficiency by facilitating data-driven decision-making. Predicting reaction times accurately can improve patient care outcomes by facilitating improved staffing strategies, reducing waiting times and maximizing patient flow.

To further improve forecast accuracy, future research should concentrate on integrating real-time data, such as traffic conditions and emergency warnings. Furthermore, connecting the model with operational decision support systems and expanding it to additional healthcare contexts might offer a complete solution for improving the delivery of healthcare services.

References

Au-Yeung, Susanna Wau Men. *Response times in healthcare systems*. Diss. Department of Computing, Imperial College London, 2008.

Sprigg, Nikola, et al. "A systematic review of delays in seeking medical attention after transient ischaemic attack." *Journal of Neurology, Neurosurgery & Psychiatry* 80.8 (2009): 871-875.

Bhattacharjee, Papiya, and Pradip Kumar Ray. "Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections." *Computers & Industrial Engineering* 78 (2014): 299-312.

Potvin, Catherine, and Derek A. Roff. "Distribution-free and robust statistical methods: viable alternatives to parametric statistics." *Ecology* 74.6 (1993): 1617-1628.

Chen, Tianqi, et al. "Xgboost: extreme gradient boosting." *R package version 0.4-2* 1.4 (2015): 1-4.

Dhieb, Najmeddine, et al. "Extreme gradient boosting machine learning algorithm for safe auto insurance operations." *2019 IEEE international conference on vehicular electronics and safety (ICVES)*. IEEE, 2019.

Ramon, Antonio, et al. "eXtreme Gradient Boosting-based method to classify patients with COVID-19." *Journal of Investigative Medicine* 70.7 (2022): 1472-1480.

Error, M. A. (2016). Mean absolute error. *Retrieved September*, *19*(2016), 14.

Ozili, Peterson K. "The acceptable R-square in empirical modelling for social science research." *Social research methodology and publishing results: A guide to non-native English speakers*. IGI global, 2023. 134-143.

Mateo, J., et al. "Extreme gradient boosting machine learning method for predicting medical treatment in patients with acute bronchiolitis." *Biocybernetics and Biomedical Engineering* 41.2 (2021): 792-801.

Zhang, Zhongheng, et al. "Predictive analytics with gradient boosting in clinical medicine." *Annals of translational medicine* 7.7 (2019): 152.

Wassan, Sobia, et al. "Gradient boosting for health IoT federated learning." *Sustainability* 14.24 (2022): 16842.

Balkman, G. S. "Machine Learning-enabled Wireless Technologies for Medical Applications: Gradient Boosting for Revolutionizing Healthcare." *INTERNATIONAL JOURNAL* 9.1 (2024): 13-22.

Mohsin Javaid Siddiqui, & Scott Uk-Jin Lee. (2025). Towards Increasing Operational Flexibility in using Industry 4.0 Enabling Technologies: A Case Study in South Korea. Physical Education, Health and Social Sciences, 3(3), 1–14

Iranzad, Reza, et al. "Gradient boosted trees for spatial data and its application to medical imaging data." IISE transactions on healthcare systems engineering 12.3 (2022): 165-179.

Cherif, Iyad Lahsen, and Abdesselem Kortebi. "On using extreme gradient boosting (XGBoost) machine learning algorithm for home network traffic classification." *2019 Wireless Days (WD)*. IEEE, 2019.

Carmona, Pedro, Francisco Climent, and Alexandre Momparler. "Predicting failure in the US banking sector: An extreme gradient boosting approach." *International Review of Economics & Finance* 61 (2019): 304-323.

Sibindi, Racheal, Ronald Waweru Mwangi, and Anthony Gichuhi Waititu. "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices." *Engineering Reports* 5.4 (2023): e12599.

Siddiqui, Mohsin Javaid, et al. "Comparative Analysis of GUI Development for Android application by utilizing MDA." *International Workshop on Convergence Infornation Technology (IWCIT)*. Vol. 2017. 2017.

Li, Xintao, et al. "A Feed-forward Neural Network Model for Predicting Stock Price." *2017 한국소프트웨어종합학술대회 (KSC2017)* (2017): 1018-1020.

Apribowo, Chico Hermanu Brillianto, et al. "Early prediction of battery degradation in grid-scale battery energy storage system using extreme gradient boosting algorithm." *Results in Engineering* 21 (2024): 101709.

Tao, Hai, et al. "Coupled extreme gradient boosting algorithm with artificial intelligence models for predicting compressive strength of fiber reinforced polymer-confined concrete." *Engineering Applications of Artificial Intelligence* 134 (2024): 108674.