BRAIN. Broad Research in Artificial Intelligence and Neuroscience

e-ISSN: 2067-3957 | p-ISSN: 2068-0473

Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Center, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)

2024, Volume 15, Issue 3, pages: 180-187.

Submitted: August 10th, 2024| *Accepted for publication:* September 24th, 2024

A Simulation-based Analysis Using Machine Learning Models to Optimize Patient Flow and Treatment Costs

Kristijan Cincar [1]
Andrea Amalia Minda[2]
Marija Varga [3]

[1] West University of Timișoara, Faculty of Mathematics and Computer Science, Romania, SC Nikola Tesla, Vrsac, Serbia,
kristijan.cincar80@e-uvt.ro,
https://orcid.org/0000-0002-5740-6680
[2] Babeș-Bolyai University of Cluj Napoca, Engineering Faculty, Romania,
andrea.minda@ubbcluj.ro,
https://orcid.org/0000-0001-7219-827X
[3] University of Novi Sad, Faculty of Technical Sciences, Serbia,
marija.varga1234@uns.ac.rs,
https://orcid.org/0000-0002-6040-167X

Corresponding author:
andrea.minda@ubbcluj.ro

Abstract: *This study investigates the application of discrete event simulation in analyzing patient management and cost dynamics within a hospital system. A simulation that integrates machine learning models, specifically Decision Trees, Random Forests, Support Vector Machines, and Gradient Boosting methods, to predict treatment costs and appointment availability was developed. Conducted over 30 days, the simulation generates synthetic data for training the models. The results are assessed in terms of the total number of patients treated, cumulative costs incurred, and the cost-effectiveness of each predictive model. The findings reveal significant variations in the performance of different machine learning techniques, demonstrating that adopting advanced analytics can substantially improve hospital resource management. This research aims to develop more efficient patient care strategies, contributing to optimizing hospital operations and enhancing patient experiences.*

Keywords: *hospital scheduling; machine learning; patient management*

How to cite: Cincar, K., Minda, A. A., & Varga, M. (2024). A simulation-based analysis using machine learning models to optimize patient flow and treatment costs. *BRAIN: Broad Research in Artificial Intelligence and Neuroscience*, 15(3), 180-187.
https://doi.org/10.70594/brain/15.3/13

1. Introduction

In the current context, healthcare systems face increasingly complex challenges, particularly regarding resource management and patient flow. The rising demand for medical services, coupled with financial constraints and limited resources, necessitates adopting innovative solutions that enhance efficiency and quality of care. Discrete event simulation serves as an effective modeling approach for complex processes, allowing for the analysis of system dynamics and the identification of blockages in patient management.

In the specialized literature, a series of papers deal with the very important subject of scheduling patients, referring to their appointments in advance or to their appointments on the day they present themselves at the hospital (Yousefi et al., 2019).

Machine learning algorithms can handle extensive healthcare data, such as electronic health records, medical imaging, and patient demographics, revealing patterns and connections that traditional statistical approaches may overlook. By utilizing machine learning, healthcare providers can improve their predictions and offer personalized recommendations, ultimately leading to better patient outcomes. The implementation of machine learning technologies in healthcare has shown promising potential for improving decision-making and operational efficiency (Mehmood, 2023).

Accurate prediction of a patient's length of stay is essential for optimizing hospital resources, bed management, and discharge planning. In (Arabnia, 2024) recent advances in using machine learning models to predict a patient's length of stay based on patient attributes and clinical data are presented. The study presented by (Li et al., 2023) explored the prediction of outpatient waiting times at a pediatric hospital using four machine learning algorithms—linear regression (LR), random forest (RF), gradient boosting decision tree (GBDT), and K-nearest neighbor (KNN). Providing patients with advance notice of their waiting times allows them to make better-informed decisions and plan their visits more effectively.

Sometimes machine learning models are combined with mathematical programming to obtain a more complex model to minimize patient waiting time (Yousefi et al., 2019).

The study presented by (Yanamala et al., 2022) uses cost-sensitive deep learning to predict hospital readmission, addressing the critical issue of imbalanced prediction errors in healthcare analytics. By integrating advanced machine learning techniques with healthcare data, it can improve predictive accuracy and optimize resource allocation in clinical settings.

The application of several machine learning models, including logistic regression, random forests, gradient boosting machines, and artificial neural networks for predicting the individual risk of late arrival was presented in (Srinivas, 2020)

A hospital scheduling system based on multi-agent coordination improves decision-making in hospital planning by enhancing collaboration between doctors, hospital resources, care staff, and patients. The papers (Cincar, 2020), (Hsieh et al., 2014), (Khanna et al., 2012) and (Cincar et al., 2019) present a multi-agent system designed to optimize the coordination of these components in hospital operations.

This paper aims to explore the impact of integrating machine learning models into optimizing patient management and the costs associated with medical treatments. Predictive models can assist in anticipating patient needs, scheduling resources more effectively, and minimizing costs. In this study, we will analyze how these techniques can be integrated into a simulation system to evaluate the efficiency of hospital resource management and costs while improving patient experiences.

2. Models and results

Hospital scheduling is a complex optimization problem involving the allocation of staff, resources, and services to meet patient demand while respecting constraints such as labor regulations, staff preferences, and operational efficiency. Machine learning (ML) models are increasingly used to improve hospital scheduling by predicting patient flows, and resource requirements and optimizing staffing schedules.

The simulation presented in this paper was executed over 30 days, allowing for the treatment of 30 patients each day, resulting in a total of 900 patient interactions throughout the study. Each machine learning model was evaluated based on key performance metrics: the total number of patients treated, cumulative costs incurred, and average costs per patient.

The Python code developed simulates patient management in a hospital setting using four machine learning models: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB), to predict treatment costs and optimize scheduling.

A *patient function* was used to simulate the journey of each patient in the hospital through the following steps:

Arrival: the patient arrives at the hospital and the time is recorded,
Cost Prediction: the treatment cost is predicted using a machine-learning cost model,
Scheduling: the model determines if a time slot is available. If the slot is filled, the patient is rescheduled; otherwise, they are treated immediately,
Departure: after treatment, the patient leaves the hospital, and their statistics are updated.

The *patient_generator function* generates patients at regular intervals (every 15 minutes) and starts the simulation for each patient by calling the *patient function*.

Next, the *run_simulation function* performs the hospital simulation for multiple days: it initializes a *SimPy* environment and simulates a set number of patients each day, at the end of each day the environment is reset and the process continues for the desired number of days and the function also updates and prints the summary statistics at the end.

```
# Train the models
# Generate synthetic data for training
X_train_cost = np.array([i for i in range(390)]).reshape(-1, 1)
y_train_cost = np.random.uniform(100, 500, size=390)
X_train_schedule = np.array([[i] for i in range(26)])
y_train_schedule = np.random.randint(0, 2, size=26)

# Decision Tree
cost_model_dt = DecisionTreeRegressor().fit(X_train_cost, y_train_cost)
scheduling_model_dt = DecisionTreeClassifier().fit(X_train_schedule,
y_train_schedule)

# Random Forest
cost_model_rf = RandomForestRegressor().fit(X_train_cost, y_train_cost)
scheduling_model_rf = RandomForestClassifier().fit(X_train_schedule,
y_train_schedule)

# Support Vector Machine
cost_model_svm = SVR().fit(X_train_cost, y_train_cost)
scheduling_model_svm = SVC().fit(X_train_schedule, y_train_schedule)

# Gradient Boosting
cost_model_gb = GradientBoostingRegressor().fit(X_train_cost, y_train_cost)
scheduling_model_gb = GradientBoostingClassifier().fit(X_train_schedule,
y_train_schedule)
```

Figure 1. Machine learning models for cost prediction and scheduling

Before the simulation, synthetic data is generated for training the machine learning models for cost prediction and scheduling (see Figure 1). Four machine learning models for both tasks were used: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB). Each model is trained separately on the generated data.

Models are trained using synthetic cost data to predict the treatment costs and to predict whether a time slot is filled based on the current time.

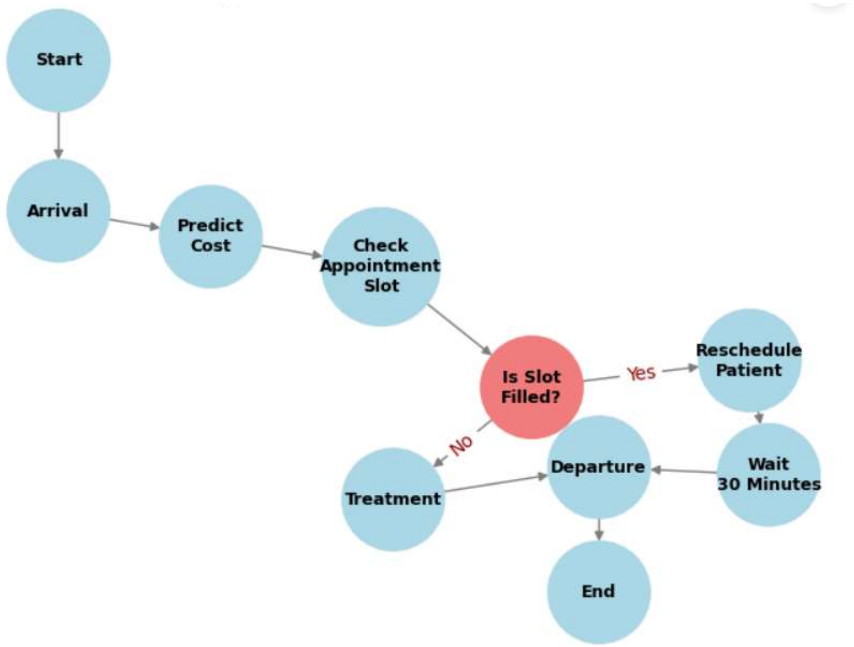

Figure 2. Patient flowchart

The *run_simulation function* is called for each model over 30 days, with 30 patients treated per day. The statistics for each model's performance are returned.

This code simulates hospital operations, using machine learning to predict treatment costs and manage scheduling. It trains different machine learning models and compares their performance based on patient throughput and total costs over 30 simulated days. The results are visualized to help assess which model performs best for managing hospital resources. Schematically the patient flow can be presented as in figure 2.

Finally, the following graphs will be plotted to visualize the simulation results: the average number of patients treated over time for each model (figure 3), the cumulative treatment costs over time (figure 4), treatment costs per patient over time (figure 5) and a bar chart with average total costs per day for every model (figure 6). Figures 3, 4, and 5 contain the graphical representations for the evolution of average patient flow, cumulative treatment costs per day, and cost per patient for different models in a day, more precisely in the working time from 7 a.m. to 2 p.m., i.e. during 420 minutes.

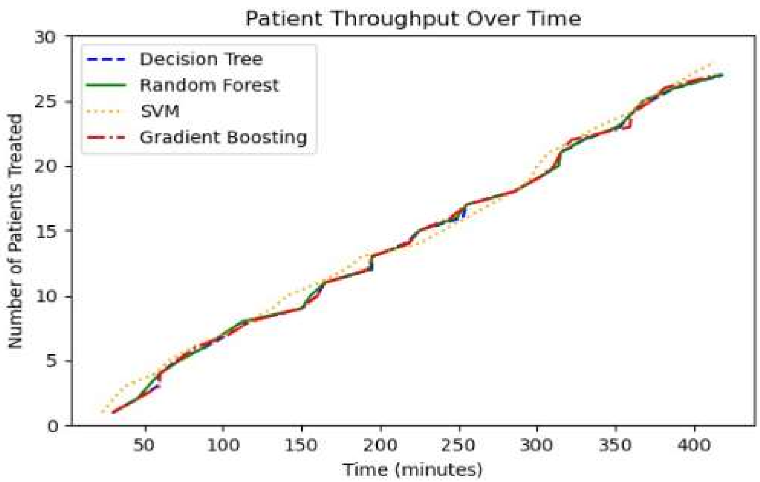

Figure 3. Average patient flow over time

The Random Forest (RF) model emerged as the most effective in terms of patient flow. It treated an average of 28 patients per day (in the working interval from 7 a.m. to 2 p.m., i.e. 420 minutes), resulting in a total of 840 patients treated over the simulation period. In contrast, the Decision Tree (DT) model treated significantly fewer patients, averaging around 20 patients per day, which totals approximately 600 patients. The Support Vector Machine (SVM) and Gradient Boosting (GB) models performed similarly, treating around 25 patients daily, with 750 patients treated. The evolution of the number of patients treated in one day, more precisely in the working interval from 7 a.m. to 2 p.m., i.e. 420 minutes, is shown in Figure 3. This disparity in patient flow illustrates the Random Forest's capability to handle complexities in patient scheduling and treatment more efficiently than the simpler Decision Tree model. Table 1 presents the key performance metrics for each model analyzed in this study, showing the average number of patients treated each day, the average cost per patient, and the cumulative costs.

Table 1. Performance metrics of each model

Models	The average number of patients treated per day	Average cost per patient (RON)	Cumulative costs (RON)
DT	20	250	150 000
RF	28	143	120120
SVM	25	165	123750
GB	25	180	135000

Cumulative treatment costs per day were also evaluated in this study and its evolution over time can be seen in Figure 4. The Random Forest model not only treated the most patients but also managed to maintain the lowest cumulative costs, totaling approximately 120000 RON over the simulation.

The SVM model followed, with cumulative costs nearing 125000 RON, while the GB model's costs reached about 135000 RON. The Decision Tree model had the highest cumulative costs at around 150000 RON, reflecting its inefficiencies in resource allocation and scheduling. This highlights how predictive accuracy in treatment cost estimation can lead to significant differences in overall expenses incurred by the hospital.

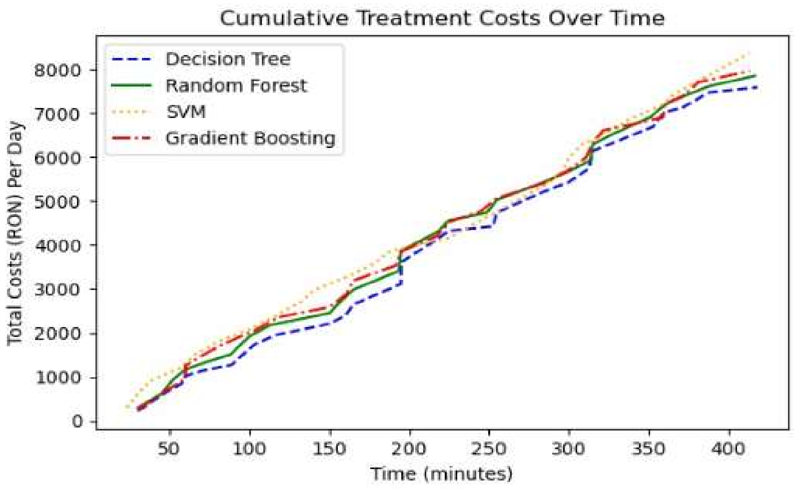

Figure 4. Cumulative treatment costs per day over time

When analyzing average costs per patient (see Figure 5 and Table 1), the RF model again outperformed the others, with an average cost of approximately 143 RON per patient. The SVM model produced an average of about 165 RON per patient, while the GB model showed average costs of around 180 RON.

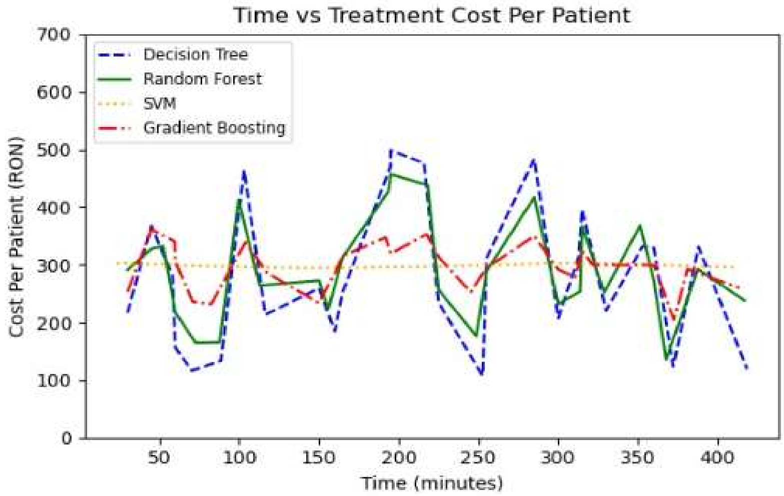

Figure 5. Cost per patient for different models used

The DT model had the highest average cost per patient, at around 250 RON. This substantial difference suggests that the RF model was more effective at accurately predicting both treatment needs and associated costs, thereby facilitating better resource management.

In addition to these metrics, the variability in costs and patient throughput across different models was examined. The RF model demonstrated lower variability in patient throughput, indicating a more consistent performance day-to-day. This consistency is crucial for hospital operations, as it allows for better resource planning and reduces uncertainty in patient management. Conversely, the DT model exhibited higher variability in both costs and patient treatment rates, which could complicate financial forecasting and resource allocation.

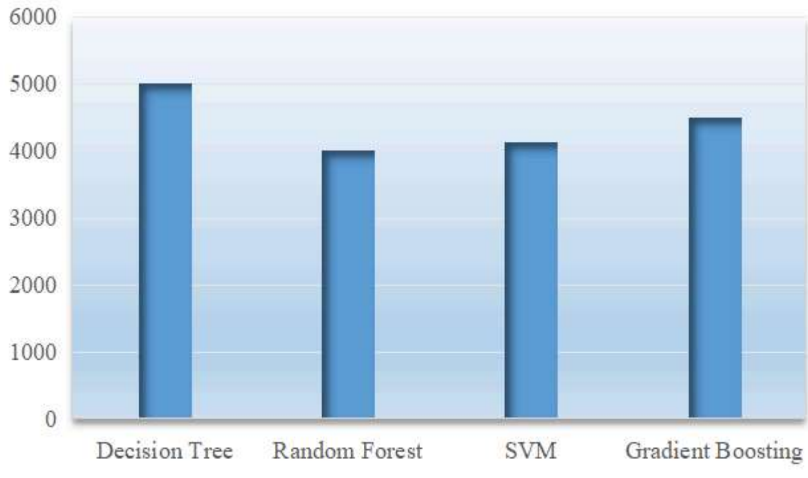

Figure 6. Average total costs per day

The graphical analysis of the simulation results provided clear insights into these differences. Plots illustrating patient flux over time revealed the RF's ability to maintain a steady increase in the number of patients treated, whereas the DT model's throughput fluctuated significantly. Similarly, cumulative cost graphs highlighted the cost-saving advantages of using more advanced machine learning techniques, with the RF consistently showing lower cumulative expenditures throughout the simulation period.

In summary, the results indicate that the RF model is superior in managing both patient flow and costs within a hospital simulation context. Its predictive capabilities lead to better treatment scheduling, lower cumulative costs, and reduced costs per patient, thereby reinforcing the importance of selecting the appropriate machine learning model for healthcare resource management. These findings not only highlight the effectiveness of advanced machine-learning techniques but also suggest practical implications for their application in real-world hospital settings.

The results of this study open several avenues for future research aimed at enhancing hospital management through advanced analytics and machine learning. Future studies could focus on integrating real-time data from hospital information systems to improve the accuracy of predictive models. By utilizing live patient data, such as admission rates, treatment times, and resource availability, models can be dynamically updated to provide more precise forecasts and enhance decision-making processes.

While this study utilized traditional machine learning models, there is significant potential for exploring deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models could analyze more complex patterns in patient data, potentially leading to better predictions of treatment outcomes and costs.

Incorporating multi-agent systems into the simulation could enhance the realism of patient interactions and resource management. By modeling individual patient behaviors and decisions, as well as staff responses, researchers could gain deeper insights into the complexities of hospital operations and the impact of various interventions.

3. Conclusions

This research demonstrates the advantages of implementing machine learning techniques in the management of hospital systems. The integration of predictive models within simulations not only allows for the evaluation of the impact of various management strategies but also enables rapid adaptation based on real-world requirements. The obtained results suggest that advanced methods, such as random forests and gradient boosting, offer superior predictive capabilities compared to simpler models like decision trees.

This study contributes significantly to understanding how advanced technologies can improve efficiency in healthcare systems and opens new avenues for future research. Subsequent studies could explore more complex applications, including the integration of real-time data and the use of deep learning techniques to refine predictions further. Implementing these solutions can help reduce costs, improve care quality, and create a more sustainable healthcare system capable of addressing current and future challenges.

References

Arabnia, H.R. (2024). Machine Learning for Predictive Modeling of Patient Length of Stay in Hospitals: Develops machine learning models to predict the length of hospital stays for patients, optimizing resource allocation and discharge planning in healthcare settings, Journal of AI in healthcare and medicine, Vol. 4, No. 1, pp.71-82.

Cincar, K. (2020). Hospital management and scheduling with multi agents approach. In *Proceedings of the 2020 International Conference on e-Health and Bioengineering (EHB)* (pp. 1-4). Iasi, Romania. https://doi.org/10.1109/EHB50910.2020.9280281

Cincar, K., & Ivascu, T. (2019). Agent-based hospital scheduling system. In *Proceedings of the 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)* (pp. 337–338). https://doi.org/10.1109/SYNASC49474.2019.00055

Hsieh, F.-S., & Lin, J.-B. (2014). Scheduling patients in hospitals based on multi-agent systems. In *Modern Advances in Applied Intelligence* (pp. 32–42). Springer International Publishing. https://doi.org/10.1007/978-3-319-07380-4_4

Khanna, S., Cleaver, T., Sattar, A., Hansen, D., & Stantic, B. (2012). Multiagent based scheduling of elective surgery. In *Principles and Practice of Multi-Agent Systems* (pp. 74–89). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30141-2_5

Li, X., Liu, W., Kong, W., Zhao, W., Wang, H., Tian, D., Jiao, J., Yu, Z., & Liu, S. (2023). Prediction of outpatient waiting time: Using machine learning in a tertiary children's hospital. *Translational Pediatrics*, *12*(11), 2030-2043. https://doi.org/10.21037/tp-23-58

Mehmood, A. M..(2023). International Journal of Computer Science and Information Technologies, Vol. 14 (1), pp. 3-7.

Srinivas, S. A (2020), Machine Learning-Based Approach for Predicting Patient Punctuality in Ambulatory Care Centers. Int J Environ Res Public Health., 17(10):3703 . https://doi.org/10.3390/ijerph17103703:contentReference[oaicite:0]{index=0}:contentReference[oaicite:1]{index=1

Yanamala, A.K.Y. & Suryadevara, S.(2022). Cost-Sensitive Deep Learning for Predicting Hospital Readmission: Enhancing Patient Care and Resource Allocation, International Journal of Advanced Engineering Technologies and Innovations Volume 01 Issue 03, pp.56-81.

Yousefi, N., Hasankhani, F., & Kiani, M. (2019). Appointment scheduling model in healthcare using clustering algorithms. https://doi.org/10.48550/arXiv.1905.03083