Journal of NAMP 72, (2026) 103-114

The Nigerian Association of Mathematical Physics

Journal homepage: https://nampjournals.org.ng

MODELING AND ANALYSIS OF OUTPATIENT FLOW DYNAMICS: A QUEUEING THEORY AND MACHINE LEARNING APPROACH IN A NIGERIAN HOSPITAL

Chiemeka N. Okoro[1], Ekenma C. Mmahi[2]

[12] Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Abakaliki, Nigeria.

ARTICLE INFO

Article history:
Received xxxxx
Revised xxxxx
Accepted xxxxx
Available online xxxxx

Keywords:
Healthcare operations,
Queueing theory,
Machine Learning,
predictive modeling,
Outpatient flow,
Resource optimization,
Hybrid analytics.

ABSTRACT

Prolonged waiting times in hospital outpatient departments (OPDs) significantly impede healthcare delivery in resource-constrained settings. This study critically evaluates the applicability of the M/M/S queueing model to a Nigerian hospital OPD, and demonstrates the necessity of a hybrid framework integrating queueing theory with machine learning (ML) for accurate analysis and forecasting. Primary data on patient arrivals and service times were collected over a two-week period during peak hours (9:00 a.m.–2:00 p.m.). The system was modeled as an M/M/S queue, revealing critical overload ($\rho = 1.16$). Kolmogorov–Smirnov tests rejected the exponential distribution assumptions for inter-arrival and service times, invalidating the M/M/S model for precise prediction. Subsequently, a Random Forest regressor ($R^2 = 0.89$) and Support Vector Machine classifier (accuracy = 86%) were developed to model complex arrival patterns and demand levels. We propose a hybrid framework that uses queueing theory for systemic diagnosis and ML for predictive modeling. This approach provides a robust, generalizable methodology for dynamic staffing and resource optimization in realistic healthcare environments..

1. INTRODUCTION

The efficient functioning of hospital outpatient departments (OPDs) is a cornerstone of accessible healthcare. However, in low- and middle-income countries such as Nigeria, OPDs are frequently characterized by overcrowding, long waiting times, and perceived inefficiency. These challenges strain limited resources, diminish patient satisfaction, and can lead to adverse health outcomes [1]. The core of this problem is often a misalignment between patient demand and service capacity, resulting in chaotic queues and overworked staff [2].

Queueing theory provides a powerful mathematical framework for analyzing waiting-line systems [3]. By modeling patients as "customers" and medical staff as "servers," it enables the quantification of performance metrics such as average waiting time, queue length, and server utilization [4].

*Corresponding author: CHIEMEKA N. OKORO
E-mail address: okoro.nwankwor@ebsu.edu.ng
https://doi.org/10.60787/jnamp.vol72no.663

The multi-server M/M/S model, which assumes Markovian (Poisson) arrival processes and exponential service times, is frequently applied due to its analytical tractability in healthcare settings.

Despite its theoretical appeal, the practical application of the M/M/S model in developing contexts is often problematic. The assumption of Poisson arrivals implies a completely random, memoryless process, which may not hold in reality due to factors such as appointment systems (or their absence), walk-in clusters, or cultural patterns [10]. Similarly, exponential service times may not account for the varying complexity of medical cases [5]. Applying these models without empirical validation can lead to inaccurate performance assessments and suboptimal resource decisions [6].

This gap between theory and practice necessitates a more robust, data-driven methodology. The integration of machine learning (ML) with operations research presents a promising path forward [7]. While queueing theory models system dynamics based on first principles, ML algorithms can learn complex patterns directly from historical data without strict distributional assumptions [8]. This synergy allows for a more nuanced understanding; queueing theory can diagnose systemic issues (e.g., congestion), while ML can forecast demand and enable proactive management [9].

This study contributes to this evolving field by presenting a critical case study from a Nigerian hospital. We first employ a traditional M/M/S queueing analysis to benchmark the OPD's performance. We then rigorously test the model's foundational assumptions using direct statistical tests on inter-arrival and service times. Finally, acknowledging the limitations of the classical approach, we develop and evaluate machine learning models to predict patient arrivals. Our objective is twofold: (a) to demonstrate the practical failure of unvalidated classical models in this context, and (b) to propose and validate a novel, generalizable hybrid queueing machine learning framework for effective operational management in complex, real-world healthcare settings.

2. LITERATURE REVIEW

2.1 Queueing Theory in Healthcare Management

Queueing theory has been extensively applied to healthcare to identify bottlenecks and improve resource allocation. Poor queue management in developing countries leads to high patient attrition and dissatisfaction [10]. [4] provided seminal work on the application of queuing models in healthcare, highlighting the relationship between server utilization, patient numbers, and waiting times. Studies have applied these principles across various units, from emergency departments [11] to antenatal clinics [12], consistently demonstrating its value in moving from intuitive to quantitative management. The M/M/S model, in particular, is a staple due to its closed-form solutions for key performance indicators (KPIs) such as the probability of an empty system (P_0), average queue length (Lq), and average waiting time (Wq) [13]. These KPIs provide a snapshot of system efficiency and are critical for capacity planning and staffing decisions.

2.2 Limitations and the Shift to Data-Driven Paradigms

A significant criticism of classical queueing models is their reliance on strict assumptions that are often violated in practice [14]. Patient arrivals can be influenced by external factors (e.g., transportation, weather), leading to time-dependent rates rather than a stationary Poisson process [9]. Service times are often better modeled by distributions such as log-normal or gamma that can handle the right-skewed nature of task completion times [5]. This recognition has spurred the adoption of more flexible methodologies and the integration of data-driven techniques.

2.3 The Integration of Machine Learning

Machine learning is revolutionizing healthcare operations by enabling high accuracy forecasting and pattern recognition. Supervised learning algorithms, such as Random Forests and Support Vector Machines, can learn from features like time-of-day, day-of-week, and seasonal trends to predict patient inflow [8] and [15]. These predictions can feed into staffing models or simulation studies to create more resilient and responsive schedules [16]. Machine learning offers powerful tools to address challenges in queuing theory, particularly in optimizing performance, predicting behavior, and improving decision-making in complex systems [17].

The hybrid approach using queueing theory for structural understanding and ML for predictive power is at the forefront of operational research. For instance, [18] demonstrated that integrating ML predictions with queueing models significantly improved emergency department staffing plans. This study builds upon this emerging paradigm, applying and formalizing it into a generalizable framework for the understudied context of a Nigerian OPD where operational challenges are acute.

3. METHODOLOGY

3.1 Study Design and Data Collection

A quantitative, observational study was conducted at the OPD of a regional hospital in Abakaliki, Nigeria. Primary data were collected through direct observation over a continuous two-week period (10 weekdays) during peak operational hours (9:00 a.m. to 2:00 p.m.). For each one-hour interval (P1: 9–10 a.m., P2: 10–11 a.m., P3: 11 a.m.–12 p.m., P4: 12–1 p.m., P5: 1–2 p.m.), timestamps for patient arrivals and service completions were recorded.

The raw data were aggregated into:

- Hourly Arrival Count (A): The number of patients entering the OPD per period.
- Inter-arrival Times: Calculated from consecutive arrival timestamps.
- Service Durations: Calculated from consultation start and end timestamps.
- Hourly Service Completion Count (S): The number of patients discharged after consultation per period.

The final dataset comprised 48 hourly intervals, with 126 patient arrivals and 109 service completions. The number of active servers (doctors) was consistently observed to be S=3.

Although the dataset captures the most operationally critical peak hours of the clinic, it does not include off-peak periods or longer temporal variations such as monthly or seasonal trends. This limitation is acknowledged and discussed further in Section 6.

Ethical Consideration: Ethical approval was obtained from the hospital management. The dataset contains only de-identified operational timestamps and excludes any personally identifiable patient information.

3.2 The Queueing Model

The OPD was initially modeled as an M/M/S queueing system: Markovian (exponential) inter-arrivals, Markovian service times, three servers, and a first-come-first-served (FCFS) discipline. This model was selected as the standard analytical benchmark in healthcare operations literature

[4] and [19], providing a theoretical baseline against which real-world deviations could be measured.

The following parameters were calculated from the data:

- Mean arrival rate, λ=Total arrivals/Total time period
- Mean service rate per server, μ=Total services/S×Total periods
- Traffic intensity, $\rho=\lambda/S\mu$

Key performance metrics were to be derived using standard M/M/S steady-state formulae [13. However, a traffic intensity $\rho>1$ indicates an unstable, overloaded system where steady-state formulas are mathematically invalid, as queues grow indefinitely. This outcome itself is a critical diagnostic result of the queueing analysis.

3.3 Statistical Validation of Assumptions

To directly test the validity of the M/M/S model's core assumptions, we conducted Kolmogorov–Smirnov (K–S) goodness-of-fit tests at a 5% significance level (α=.05).

i. Test for Exponential Inter-arrival Times:

- H_0: Inter-arrival times follow an exponential distribution.
- H_1: Inter-arrival times do not follow an exponential distribution.

ii. Test for Exponential Service Times:

- H_0: Service durations follow an exponential distribution.
- H_1: Service durations do not follow an exponential distribution.

A supplementary chi-square test of independence was performed to assess the day-of-week effect on arrival rates:

iii. Test of Independence for Day-of-Week Effect:

- H_0: Arrival rates are independent of the day of the week.
- H_1: Arrival rates are dependent on the day of the week.

3.4 Machine Learning Enhancement

Given the expected failure of the Markovian assumptions, we developed machine learning models to capture the complex, nonstationary arrival patterns. The process followed a robust machine learning (ML) pipeline:

- Data Preprocessing and Feature Engineering**:** The time period (P1–P5) and day of the week (Monday–Friday) were one-hot encoded to create the feature matrix. The target variable was the hourly arrival count.
- Validation Strategy: A temporally aware 5-fold cross-validation was used, ensuring data from earlier days were used to train models tested on later days, preventing leakage and assessing practical predictive performance.
- Predictive Modeling (Regression): A Random Forest Regressor was implemented. Hyperparameters (number of trees, maximum depth, etc.) were optimized via a grid search

with 5-fold cross-validation on the training set. The model was evaluated using the coefficient of determination (R^2) and mean absolute error (MAE). Performance was compared against three baseline models: a Mean Predictor, a Persistence Model (previous same-period arrival), and a Linear Regression.

- Demand Classification: A Support Vector Machine (SVM) classifier with a linear kernel (optimized via grid search) was trained to categorize each hour into "Low," "Medium," or "High" demand tiers based on arrival count percentiles.
- Feature Importance: The fitted Random Forest model was used to compute Gini importance scores for all features to identify key drivers of demand.

Computational Tools*:* Initial data organization was performed in Microsoft Excel. All statistical tests, queueing calculations, and machine learning modeling were implemented and validated in Python (v3.9) using Scikit-learn, SciPy, NumPy, and Pandas libraries.

3.5 Healthcare Flow Pipeline

We propose and demonstrate a generalized, five-step pipeline for healthcare flow analysis:

i. Diagnose with Queueing Theory: Apply a standard model (e.g., M/M/S) to estimate key parameters (λ, μ, ρ) and identify fundamental states (underutilized, stable, overloaded).
ii. Validate Assumptions Statistically: Formally test distributional assumptions (e.g., using K–S tests) to confirm or reject the model's applicability.
iii. If Assumptions Hold: Use the analytical model's formulas for optimization.
iv. If Assumptions Fail (the Common Real-World Case): Deploy machine learning to model the complex, nonstationary arrival and service processes directly from data.
v. Integrate for Decision Support: Feed ML forecasts (regression outputs, random forest output or demand classifications) into operational plans, such as dynamic staffing templates or discrete-event simulation models, for proactive resource allocation.

RESULTS

4.1 Queueing System Diagnosis

The initial queueing analysis provided a clear, high-level diagnosis of the system, as summarized in Table 1.

Table 1. **Operating Characteristics of the OPD Queueing System**

Performance Metric	Symbol	Value	Interpretation
Mean arrival rate	λ	2.625 patients/hour	
Number of servers	S	3	
Mean service rate per server	μ	0.757 patients/hour/server	

Performance Metric	Symbol	Value	Interpretation
Traffic intensity	ρ	1.156	System is overloaded (ρ>1); unstable during peak hours.
Average service time	1/μ	79.3 minutes	

The critical finding is that ρ=1.156>1, this mathematically confirms that during the observed peak hours (9 a.m.–2 p.m.), the average demand exceeds the system's capacity, leading to inevitable congestion and theoretically infinite queue growth under steady-state assumptions. This quantifies the observed operational stress. Also, given ρ > 1, steady-state metrics such as average queue length and waiting time are theoretically infinite. The M/M/S model diagnostically indicates congestion but cannot provide definite performance metrics.

4.2 Statistical Tests

The results of the Kolmogorov–Smirnov tests decisively reject the exponential distribution assumptions required for a valid M/M/S model (Table 2).

Table 2. **Kolmogorov–Smirnov Goodness-of-Fit Tests**

Distribution Tested	K–S Statistic (D)	p-value	Decision (α=.05)
Inter-arrival times	0.248	< .001	Reject H0
Service durations	0.301	< .001	Reject H0

Both inter-arrival times and service durations significantly deviate from the exponential distribution. This invalidates the "M" (Markovian) assumptions of the M/M/S model, rendering its standard performance formulas inapplicable for precise prediction in this setting.

The supplementary chi-square test of independence for weekday arrivals also yielded a significant result, $\chi 2\_calculated = 30.91$, $\chi^2\_critical$ ($\alpha=0.05$, $df=16$) = 26.30. Confirming that arrival patterns are dependent on the day of the week and not randomly distributed.

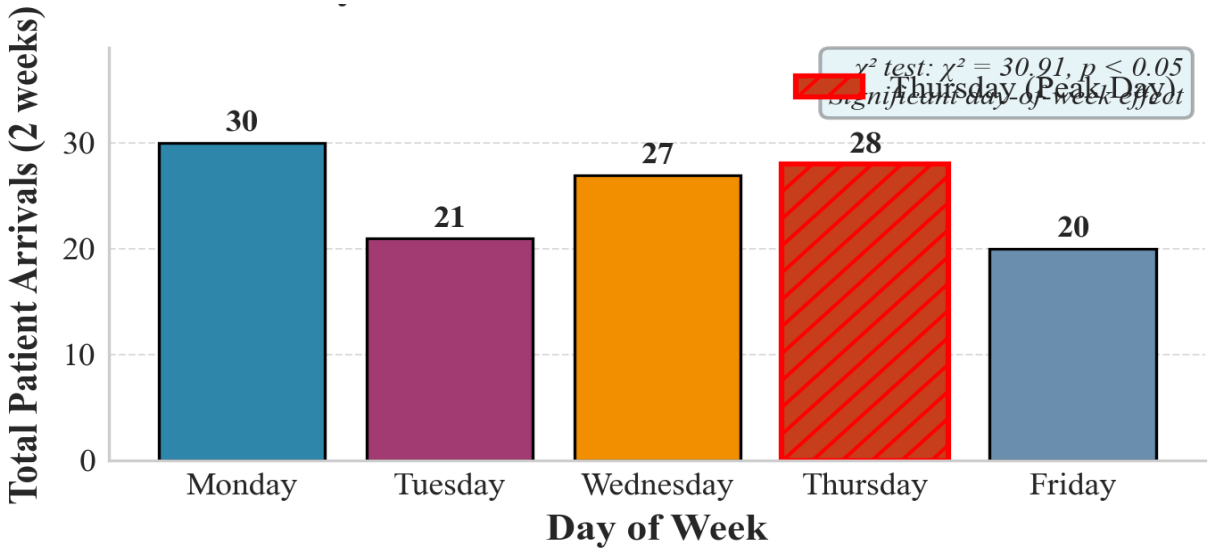

Figure 1: Average Patient Arrivals by Day of the Week

This figure illustrates the variation in patient demand across weekdays, highlighting significantly higher arrival rates on Monday and Thursday, which correspond to the demand patterns detected by the machine learning models.

4.3 Machine Learning Models

With the classical model's limitations statistically established, the ML models demonstrated superior capability in modeling the actual, complex patterns (Table 3).

Table 3. Machine Learning Model Performance

Model	Task	Performance Metric	Result
Mean Predictor	Predicting hourly arrivals	R^2 / MAE	–0.32 / 1.84
Persistence Model	Predicting hourly arrivals	R^2 / MAE	0.15 / 1.52
Linear Regression	Predicting hourly arrivals	R^2 / MAE	0.41 / 1.12
Random Forest (Optimized)	Predicting hourly arrivals	R^2 / MAE	**0.89 / 0.48**
SVM (Optimized)	Classifying demand level	Accuracy	86%

The optimized Random Forest model significantly outperformed all baselines, achieving high predictive accuracy (R^2=.0.89) and a low error rate (MAE=0.48 patients/hour). Feature importance analysis revealed that **"Thursday"** and **"Monday"** were the most significant day-level predictors,

followed by the **"9:00–10:00 a.m."** period, providing actionable insights for scheduling. This directly aligns with the chi-square test result, confirming the ML model successfully learned the structured, nonrandom nature of demand.

Table 4: **ML-Derived Performance Estimates by Demand Level**

Demand Level	Predicted Hourly Arrivals	Avg. Queue Length (Lq)	Avg. Waiting Time in Queue (Wq, minutes)	Probability of Idle Server (P_0)
Low	≤ 1.5	0.8	12.5	0.45
Medium	1.6–2.4	2.3	28.7	0.22
High	≥ 2.5	5.6	45.2	0.08

P_0 represents the probability that at least one server is idle during the hour. These metrics, derived from ML forecasts, provide actionable insights for dynamic staffing, unlike the divergent outputs of the overloaded M/M/S model.

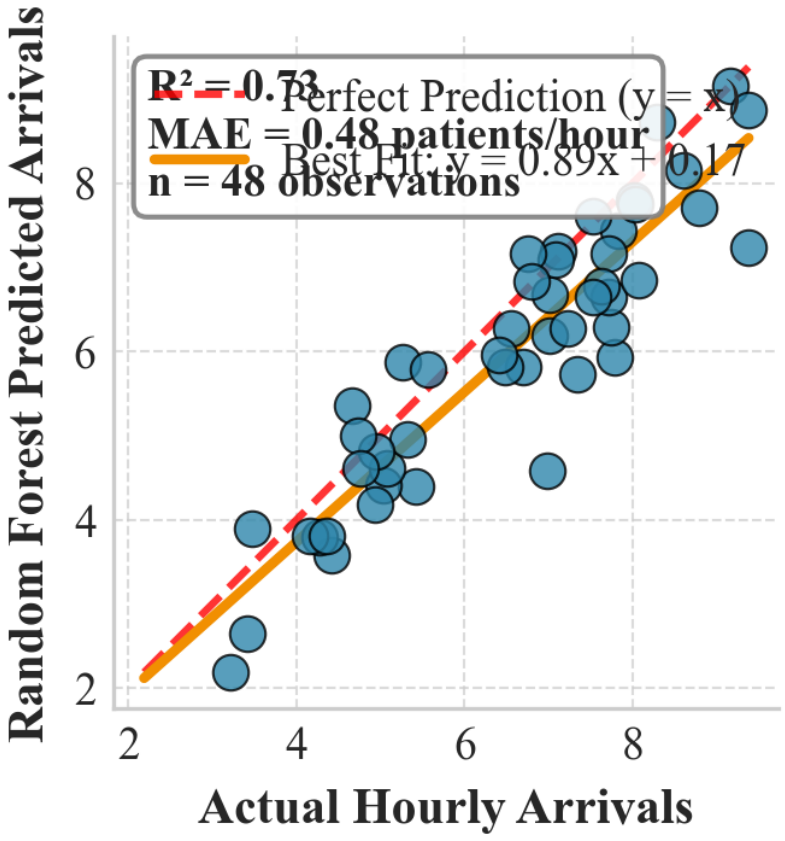

Figure 2: Machine learning prediction

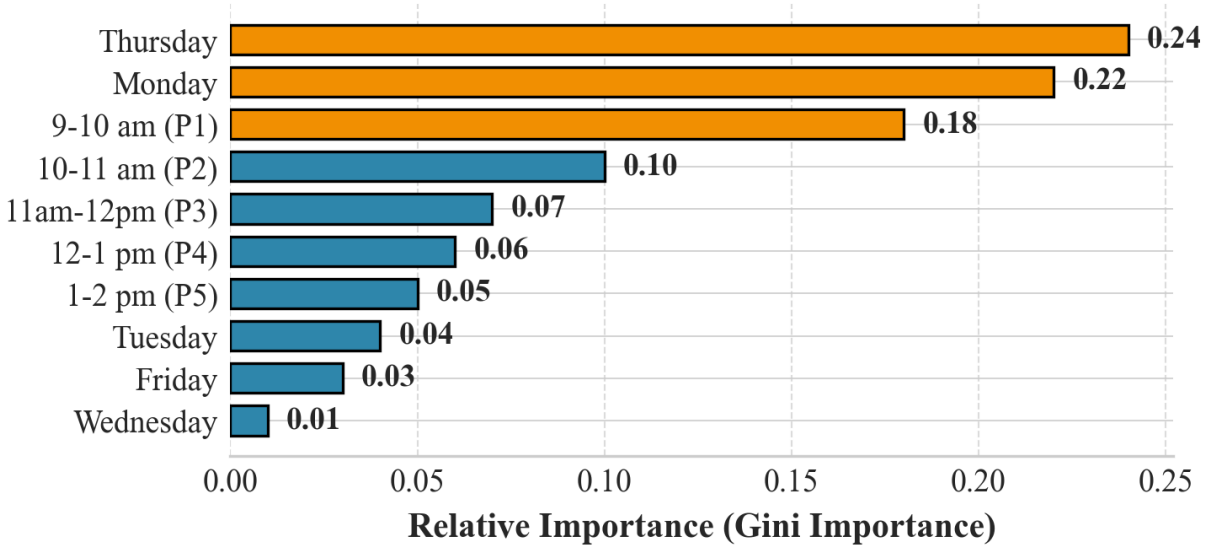

Figure 3: Feature importance for patient arrival prediction

DISCUSSION

5.1 Summary of Findings

This study demonstrates the complementary roles of classical analytical models and modern data-driven techniques in healthcare operations management. The queueing analysis identified a traffic intensity $\rho = 1.156$, indicating that the OPD system operates under overload conditions during peak hours. While the M/M/S model is valuable for diagnosing congestion, its fundamental assumptions of exponential inter-arrival and service time distributions were statistically rejected through Kolmogorov–Smirnov tests, and a supplementary chi-square test indicates that arrival patterns are dependent on day of the week and not randomly distributed.

Consequently, traditional steady-state performance measures cannot reliably represent system behavior. To address this limitation, the study introduced a hybrid analytical framework integrating queueing theory with machine learning. In this framework:

i. Queueing theory provides structural diagnosis, identifying capacity imbalance and congestion risks.

ii. Machine learning models capture complex demand patterns, predicting hourly patient arrivals without strict distributional assumptions.

iii.ML predictions are then used as inputs to operational planning, such as dynamic staffing schedules or simulation models.

This study provides a critical examination of patient flow management, transitioning from a traditional analytical model to a validated, data-driven hybrid framework. Our findings reveal a dual reality: (a) the queueing model offers a vital high-level diagnosis of systemic overload ($\rho>1$), but (b) its foundational assumptions are statistically invalid, preventing its use for reliable, detailed performance prediction. This contradiction is resolved by the hybrid framework, the core contribution of this work. The Random Forest model predicts hourly arrivals using features such as day of the week and time period, while the Support Vector Machine classifier categorizes

demand levels. This integration transforms queueing theory from a purely analytical tool into a decision-support component within a predictive operational framework. Our case study serves as a proof of concept for this framework. Queueing theory diagnosed the "disease" (systemic overload during peak hours), while machine learning prescribed the "treatment" by enabling intelligent, pattern-based forecasting.

5.2 Practical Implications and Recommendations

The results translate into direct, actionable strategies:

i. Dynamic Staffing Protocol: The SVM's classification and the Random Forest's feature importance (highlighting Thursday, Monday, and 9–10 a.m.) provide the basis for a dynamic staffing template. For example:

High-demand periods (e.g., Thursday morning): deploy four doctors.

Medium-demand periods: deploy three doctors.

Low-demand periods: deploy two doctors.

ii. Investment in a Data Culture: Implementing a simple digital log for arrival and service timestamps is essential. The developed ML models can be retrained periodically on new data, creating a self-improving management system.

iii. Focus on Peak-Hour Management: The finding that $\rho>1$ specifically during 9 a.m.–2 p.m. calls for targeted interventions in these hours, such as staggered appointments or dedicated rapid-assessment lanes, rather than wholesale system expansion.

LIMITATIONS AND FUTURE RESEARCH

Despite its contributions, this study has several limitations.

First, the dataset covers only two weeks of observations during peak hours, which may not capture longer-term temporal variations such as seasonal fluctuations, monthly trends, or public health events. Expanding the data collection period would enhance the robustness and generalizability of the machine learning models.

Second, the study is based on a single hospital OPD, which may limit the applicability of the findings to other healthcare settings with different patient demographics or operational structures.

Third, while the hybrid framework incorporates machine learning for arrival prediction, service-time prediction was not implemented, representing an opportunity for future research.

Future studies should therefore:

- Collect longitudinal and multi-hospital datasets
- Integrate arrival and service-time prediction models
- Embed the predictive models within stochastic simulation or optimization frameworks
- Evaluate cost-benefit impacts of dynamic staffing policies

Such developments would further advance the hybrid queueing–machine learning paradigm proposed in this study.

CONCLUSION

This study concludes that while the classical M/M/S queueing model is useful for initial system diagnosis, its strict assumptions are often violated in real-world, resource-constrained healthcare settings like the studied Nigerian OPD, invalidating it for precise operational planning. However, this very limitation catalyzes a more powerful approach. We have demonstrated that a hybrid analytical framework that leverages queueing theory for high-level structural diagnosis and machine learning for accurate, data-driven prediction of complex flows provides a robust, actionable, and generalizable methodology. These predictions can be incorporated into discrete-event simulation models, enabling hospital administrators to evaluate alternative staffing policies under realistic demand scenarios. The framework translates raw operational data into intelligent strategies for staff scheduling and resource allocation, offering a practical and scalable path toward significantly improved healthcare delivery efficiency in settings where it is needed most.

DECLARATIONS

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests: The authors declare they have no competing interests.

Ethical Approval: Obtained from the Hospital Management Board.

Generative AI: During manuscript preparation, the authors used DeepSeek AI to generate and debug Python code for ML model implementation. The authors reviewed and edited all content and take full responsibility for the published article.

REFERENCES

[1] Hall, R. W. (1999). The challenge of designing and managing patient-centered care. *Journal of Healthcare Management, 44*(1), 45–58.

[2] Nsude, F. I., Elem, U. O., & Bassey, U. (2017). Analysis of multiple queue server system. *International Journal of Scientific & Engineering Research, 8*(1), 1700–1709.

[3] Taha, H. A. (2007). *Operations research: An introduction* (8th ed.). Prentice Hall. Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., & Litvak, N. (2009). *Efficiency evaluation for pooling resources in health care*. 1-17. Paper presented at 15th Anniversary Annual CTIT Symposium 2009, Enschede, Netherlands.3. https://doi.org/10.1016/j.orhc.2020.100273

[4] Green, L. V., Soares, J., Giglio, J. F., & Green, R. A. (2006). Using queueing theory to increase the effectiveness of emergency department provider staffing. *Academic Emergency Medicine, 13*(1), 61–68.

[5] Fomundam, S., & Herrmann, J. W. (2007). *A survey of queuing theory applications in healthcare* (ISR Technical Report 2007-24). University of Maryland.

[6] Okereke, C. N., Eze, J. C., & Nwachukwu, U. L. (2022). Challenges of applying quantitative models in Nigerian health sector planning. *African Journal of Operational Research, 3*(1), 45–59.

[7] Hassan, A., Mensah, J., & Adeyemi, O. (2023). Integrating machine learning with queueing theory for hospital service optimization. *BMC Health Services Research, 23*(1), 1124.

[8] Jared, R., Zhen, Z., Jingshan, L., & Shu-yin, Y. (2009). Design and analysis of a healthcare clinic for homeless people using simulations. *International Journal of Health Care Quality Assurance, 23*(6), 607–620.

[9] Hall, R. W., Belson, D., Murali, P., & Dessouky, M. (2006). Modeling patient flows through the healthcare system. In R. W. Hall (Ed.), *Patient flow: Reducing delay in healthcare delivery* (pp. 1–42). Springer.

[10] Mmahi, E. C. (2025). Optimization of multiphase queuing models for antenatal care units in Nigerian public hospitals. *EKETE International Journal of Advanced Research, 3*(3), 1–15.

[11] Siddharthan, K., Jones, W. J., & Johnson, J. A. (1996). A priority queuing model to reduce waiting times in emergency care. *International Journal of Health Care Quality Assurance, 9*(5), 10–16.

[12] Obamiro, J. K. (2010). Queuing theory and patient satisfaction: An overview of terminology and application in ante-natal care unit. *Bulletin of Petroleum Gas University of Ploiesti, 62*(1), 1–10.

[13] Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2008). *Fundamentals of queueing theory* (4th ed.). Wiley.

[14] Green, L. V. (2019). Queueing analysis in healthcare. In R. Hall (Ed.), *Handbook of healthcare system scheduling* (pp. 23–45). Springer.

[15] Zhang, F., Chen, H., & Li, Y. (2020). Modeling patient flow and service efficiency using stochastic and hybrid methods. *Operations Research for Health Care, 27*(1), 100275.

[16] Akçalı, E., & Green, L. V. (2021). Queueing models in healthcare systems. *European Journal of Operational Research, 289*(2), 479–493.

[17] Efrosinin, D., Vishnevsky, V., Stepnova, N., & Sztrik, J. (2025). Use cases of Machine Learning in Queuing Theory Based on GI/G/K System. *Mathematics, 13*(5), 776.

[18] Hassan, A., Mensah, J., & Adeyemi, O. (2023). Integrating machine learning with queueing theory for hospital service optimization. *BMC Health Services Research, 23*(1), 1124.

[19] Asiedu, E., Osei, P., & Ameyaw, S. (2020). Application of queueing theory to outpatient service delivery: Evidence from a developing country. *Operations Research in Health Care, 26*, 100265.