Decision Making in Manufacturing and Services
Vol. 18 • 2024 • pp. 39–64

Predictions and Application of Queueing Analysis: Case of Regional Hospital Limbe, Cameroon

Daphne T. Machangara*, Habiboulaye Amadou Boubacar**, Giovanni Andreatta***, Antony Ndolo****

Abstract. In this work, we applied queue analysis and the predictions of waiting times at Regional Hospital Limbe (RHL) in Cameroon. The main purpose of the work was to be able to make mathematical sense of a real-life scenario that concerned queues (waiting lines) and try to come up with models for performance measurements and improvements; this could be achieved by using queueing theory concepts that were composed of queueing models that provided some operational insights because of their analytical nature. The observations included studying patient arrival and waiting times, along with doctor service times; the results showed busy departments in the hospital, busy days, and busy times. Long waiting times were mainly found to exist in general practitioner (GP) and specialist consultations. The queueing concept was applied to only one service segment – GP consultation. Although strong scientific conclusions cannot be made on the queuing models that were obtained due to inefficient data, the value of this work lies mainly in the methodologies and proposals of different operating systems that could be adopted. Furthermore, some predictions were made using machine learning to see how long a patient could wait in a queue for service; the model predictions had an average of 10 minutes and 53 seconds of error.

Keywords: queueing theory, queueing model, queueing system, waiting time, prediction

Mathematics Subject Classification: 60K25

JEL Classification: C53

Submitted: July 15, 2024

Revised: October 21, 2024

* African Institute for Mathematical Sciences (AIMS), Limbe, Cameroon, e-mail: daphne.machangara@aims-cameroon.org
** Universite de Lille, Lille, France
*** Università degli Studi di Padova: Padova, Veneto, Italy
**** Karadeniz Technical University, Trabzon, Türkiye; Deep Learning Indaba, Africa, e-mail: ndolo@deeplearningindaba.com

DOI: https://doi.org/10.7494/dmms.2024.18.6442

1. INTRODUCTION

Waiting times are generally a problem, and many have tried to tackle this issue. Over the years, there has been growing evidence of long waiting times in healthcare organizations, thus leading to different waiting times for patients (Garcia-Corchero & Jimenez-Rubio, 2022); therefore, long queues and waiting times negatively impact healthcare service deliveries through patient outcomes and satisfaction in terms of increased risks of delays when attending to time-sensitive issues and the outcomes of patient frustrations. As a result, this work sought to make mathematical sense of queues, adopt models for performance measurements and improvements, and provide some waiting-time predictions for patients; this was achieved by understanding the arrival and waiting/service times as well as the busy times, busy departments, and busy days at our hospital. Moreover, at least 90% of patients should be seen within 30 minutes of their scheduled appointment times according to the recommendation by the Institute of Medicine (IOM) (O'Malley et al., 1983). Whether this is being implemented remains a challenge that is faced by many. We studied this problem in our thesis (Machangara, 2018), and this paper is a continuation of this thesis.

Queueing theory was pioneered in the 20th century by Agner Krarup Erlang – a Danish mathematician, statistician, and engineer (Brockmeyer et al., 1948; Cooper, 1981; Erlang, 1909; Lakatos et al., 2019; Saaty, 1957). It continues to be a vital interdisciplinary field of study that explores how queues (or waiting lines) form and examines how they can be efficiently managed. The fundamental principles of queueing theory revolve around reviewing waiting-line processes. The main principle of this theory by Erlang (1909) paved the way by developing models to analyze a Copenhagen telephone exchange. An improvement to the theory was done by Palm (1943), thus extending Erlang's work, introducing the concepts of random arrivals and intensity fluctuations in a queueing system. This was particularly pertinent in all of the service disciplines that covered various sectors, like telecommunications (network routing, load balancing, and packet-switching systems (Dshalalow, 1995) and manufacturing/production (determining the optimal service rates for machinery or labor) (Boukas et al., 1995; Liberopoulos et al., 2006; Yao, 1994). In the energy sector, queueing theory is in smart-grid management and power-plant scheduling (as was commented on in (Nair et al., 2021; Zavanella et al., 2015). At the same time, this theory is also used in financial service, retail, and service industries to optimize the operations of trading systems, ATM networks, retail outlets, restaurants, banks, and call centers (Koole, 2013). More recently, its application in the optimizations of cloud-computing resources, data-center operations, and task scheduling in distributed systems has increased rapidly (Harchol-Balter, 2013; Newell, 1982). There will undoubtedly be other possible future applications; the possible fusion of artificial intelligence and deep learning with queueing theory will lead to intelligent queuing systems. The application of distributed queueing systems integrated with blockchain technology is also suggested (Xiong, 2023). In healthcare, this can be applied in emergency-room-service, patient-triage, and appointment systems (Green, 2013; Harki, 2024); thus, the formations of queues exist in day-to-day life. The various applications of queueing theory are summarized in Figure 1 (as speculated by Elalouf & Wachtel, 2021 as well as by Shortle et al., 2018).

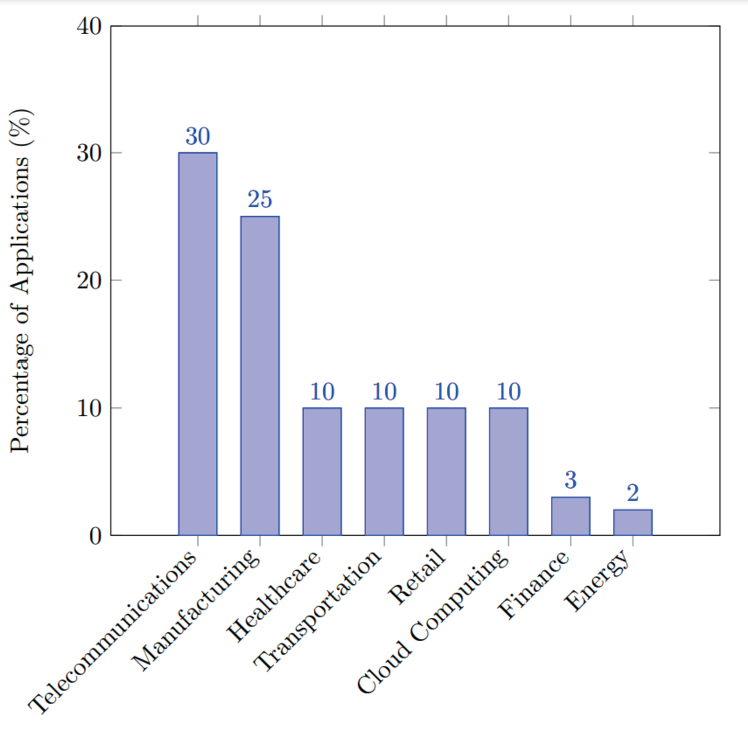

Fig. 1. *Percentages of applications of queueing theory in different fields*

There are factors that affect waiting times; for example, Biya et al. (2022) found the distance that was traveled, the day, and the time of a hospital visit to affect patients' waiting times. Thus, those patients who travel from afar, visit a hospital on a Monday, and arrive in the early morning were found to spend more waiting time than others.

Queueing theory has been extensively utilized in computer science for managing processes efficiently. Ding (2023) provided a comprehensive analysis of computer queueing theory – from its historical background and fundamental concepts to its mathematical models. The paper highlighted key contributions of queueing theory in computer systems; it also examined the characteristics of queues, like arrival and service rates, queue lengths, and waiting times, using different queueing models like single-server queues, multi-server queues, and network queues as well as their relevance and applicability in computer systems. Thus, some of the practical applications of queueing theory in computer systems include performance analyses. The authors of Chakka et al. (2009) introduced and proposed a generalized Markovian queueing model for performance analyses of telecommunications networks. The authors indicated that the model was a potential viable performance predictor, with its applications in the performance analysis of high-speed downlink packet access and optical burst switching nodes. Research by Li (2024) focused on analyzing and optimizing queueing systems in the context of cloud computing, with attention being paid to energy conservation. $M\&M$ queues tend to be widely used in cloud computing in its queueing technology, in physical queueing scenarios like grocery stores, fast food establishments, or the queues at airport terminals before boarding airplanes. The study examined the performance of $M|M|$ queues for developing efficient and expedient queues for cost-effectiveness. The $M|M|C$ queueing models were discovered to be of benefit when handling tasks of varying sizes and greater income levels as compared to the $M|M|1$ models.

In the health sector, queueing theory has been utilized in various departments in helping to identify and analyze existing queues by providing insights on queue management and resource allocation for serving patients with minimal waiting times (Saastamoinen et al., 2023). Elalouf & Wachtel (2021) demonstrated the use of both scientific methodologies like queueing models and simulations as well as managerial approaches like bed management to manage queueing-related problems in hospital emergency departments. Although queueing models have been used in healthcare systems, Peter & Sivasamy (2021) noted the limitations in its applications. For this reason, the authors investigated the applicability of queueing techniques to understand queueing modeling and solution techniques that are useful in applications. Ji (2023) described queueing theory as a mathematical approach that concerned the dynamics of waiting lines (or queues); thus, its applications were pivotal in analyzing and optimizing systems where the timing of customer arrivals and service was of importance in various sectors. Furthermore, the theory provided insights into patient arrival, patient flow, and the ways of optimizing both processes and people to meet demand according to (Johnston et al., 2022).

Waiting times not only differ in countries but even in the centers within these countries. This is a major problem for both developed and developing countries, and some reports on long waiting times have been made (although most date back to as long as 20 years ago).

Some of the studies that have been carried out have noted the management of waiting times as a challenge in hospitals (Mbwogge et al., 2022), where wait times and patient satisfaction were assessed in an eye hospital in Cameroon. The study revealed the lack of a significant association between wait times and satisfaction, with mean pre-intervention waiting, service, and idling times of 449.6, 111.9, and 337.7 minutes, respectively. Through the application of the Plan-Do-Study-Act (PDSA) quality improvement method, a 14.5% reduction in waiting time was achieved. Furthermore, Almusawi et al. (2023) did a study in Saudi Arabia at primary healthcare centers in Riyadh. Excluding emergency cases, the median total waiting time was found to be 23 minutes, whereas the median waiting times before and during the service were 6 and 6.78 minutes, respectively. The total waiting times at urban primary healthcare centers were longer than at rural primary healthcare centers, with a significant difference between both groups ($t = -15.5$, $P < 0.001$). Among the significant factors affecting waiting times on weekdays, they identified a patient's age, marital status, educational level, and occupation as influential ($P < 0.05$). In Nigeria's Ahmadu Bello University, Zaria, Kaduna State Teaching Hospital, Adeniran et al. (2022) considered a multi-server single-channel queueing model and obtained a utilization factor of 13%. From the questionnaires that were distributed, they also discovered dissatisfaction with the service quality in the hospital, where the number of patients outnumbered the hospital staff. Such pressure on a hospital staff forces them to dispose patients without thorough treatment. Still, a multi-server exponential queueing system was adopted by Suleiman et al. (2022) in northwestern hospitals in Nigeria; it showed the busiest and least-busy hospitals having utilization factors of 89 and 10%, respectively. In terms of waiting times, the average times that were spent in queues by patients at the busiest and least-busy hospitals were 0.35 and 0.29 minutes, respectively. On the other hand, Rema and Sikdar (2021) used a Monte Carlo simulation technique to analyze queueing patterns and study patient flow in order to manage queues and minimize delays. Le et al. (2021) sought to improve patient wait times due to overcrowding in the emergency departments (EDs) at public hospitals in Vietnam. Using a lean approach, the reductions in delay and waiting times were 33% for patients who required operations (from 134.4 to 89.4 minutes).

With attempts to reduce patient wait times, Fun et al. (2022) used discrete event simulation (DES) in a Malaysian public hospital. The model was used to evaluate the effects of changing consultation start times and patient arrival times. As per the findings, matching consultation start times and patient arrival times indicated the potential of reducing both waiting times and crowding by 40%, with the number of patients waiting per hour possibly reduced by 10–21% during peak hours. Mohammadi et al. (2022) planned the optimal use of resources and the improvement of service quality to estimate the average length of a stay (LOS), bed occupancy rate (BOR), bed blocking probability (BBP), and the throughput of patients in a cardiac surgery department (CSD) through simulation models at Farshchian Hospital, Hamadan, Iran; they used post-operative ward (POW) and intensive care unit (ICU) patients and beds as servers. With a combination of a Monte Carlo simulation and the use of Python software, the queueing simulation results indicated that, for a fixed number of beds in an ICU, the BOR in a POW decreased as the number of beds in the POW increased, and the LOS in an ICU increased with a decrease in the number of beds in a POW. According to the simulation, the results indicated the problem to be poor queuing-system management rather than insufficient resources; the possibility of reducing the overall average waiting times in the department during business hours was shown (from 37.24 to 29.22 minutes).

On the other hand, Proudlove (2022) applied queueing theory in Greater Manchester, United Kingdom (UK), at a National Health Service (NHS) acute hospital trust for resizing its pediatric inpatient department. The hospital sought to reduce the number of bed occupancies from 54 to 85% using a Monte Carlo simulation. Through the basic application of queueing theory, the recommendation was not to reduce the bed occupancy level, as it was revealed that using a bed occupancy target of 85% would result in a risk of 33% that all beds would be full and that using a very low risk of all beds being full of 0.1% would result in an average bed occupancy of 55%. Other researchers like Palomo et al. (2023) provided insights on queueing theoretic methods when analyzing the time evolution of patients who were hospitalized due to the coronavirus (COVID). The number of COVID patients was modeled as a dynamical system based on the theory of infinite server queues with time inhomogeneous Poisson arrival rates. One of the key findings was the existence of a lag between the time of the peak arrival rates of infected patients not coinciding with the times of the peak numbers of hospitalized patients (or deaths). In addition, Kalwar et al. (2021) analyzed the contribution of queueing theory and discrete event simulation in the improvement of healthcare and noted the queueing system mismanagement of resources and the queuing system to be among the main reason for the low quality of the healthcare service delivery in public sector hospitals of Pakistan. Medical service reports indicated the problems that were faced by patients to be delayed service, long waiting times, and less departmental capacity (at emergency, out-patient departments [OPDs], and laboratories) as well as the inadequate number of doctors. The authors were therefore convinced that applications of queueing theory could largely contribute in the healthcare system.

Concerning predictions, Joseph et al. (2022) ascertained the importance and usefulness of predicting patient waiting times in reducing uncertainty regarding wait times. Their study used machine learning to predict patient waiting times before consultations and throughput times in the outpatient clinics. The predictions made use of four models (random forest, XGBoost, random forest with SMOTE, and XGBoost with SMOTE) regarding gender, the day of a visit, the month of a visit, the time of a visit, a consultation's start time, vital examinations, laboratory visits, pharmacy visits, repeated arrivals, consultation sessions, and weather conditions as some of the input variables. In terms of feature importance, the time of a visit was considered to be the major predictor in determining the throughput time for all of the models, with high feature importance scores of 0.396, 0.321, and 0.266 for random forest, RF with SMOTE, and XGB, respectively. The waiting time before consultation was predicted with an accuracy of 0.86, and the throughput time accuracy was 0.93. The areas under the curve (AUC) for the best models that predicted waiting times before consultations were 0.85 and 0.82 for XGBoost and XGBoost with SMOTE, respectively. The AUC that was obtained for the highest performing XGBoost model for predicting the throughput time was 0.89, and the random forest model received an AUC of 0.87.

In a recent study, Benevento et al. (2023) tested various machine-learning techniques using predictive analytics that were applied to two large data sets from actual emergency departments (EDs). They evaluated the predictive ability of Lasso, random forest, support vector regression, an artificial neural network, and the ensemble method using different error metrics and computational times. For prediction accuracy improvement, new queue-based variables that captured the current states of the EDs were defined as additional predictors. The results indicated the ensemble method to be the most effective at predicting waiting times. Concerning the accuracy and computational efficiency, random forest was a reasonable trade-off. In addition, Walker et al. (2022) sought to validate machine-learning models to predict patient waiting times in various emergency departments. They discovered the best-performing models to be random forest, and linear regression models performed the best in waiting-time predictions. The important variables were the triage category, last$-k$ patient average wait times, and arrival times.

2. METHODS

The study was conducted in Limbe (in the Southwest Region of Cameroon) at one of the local hospitals (Regional Hospital Limbe [RHL], which also serves as the principal referral hospital in the region). For this reason, the hospital is busy; this results in high patient traffic. It is a public government hospital with a capacity of 200 beds and offers a wide range of services, like an out-patient department (OPD), a laboratory, and a pharmacy; these units include general and specialist units (like cardiologists, among others).

The instruments that were used in the data-collection process were questionnaires and observations; in addition, registers were used with the required information from some selected departments. Questionnaires (in French and English) were distributed to both patients and medical personnel in order to obtain opinions and general information on the functioning of the hospital. The survey was carried out over 28 days using the convenience sampling method, with a patient response rate of 84% (a total of 170 respondents and 32 refusal cases). The medical personnel had a 100% response rate, with 26 respondents and no refusal cases.

To simplify the work, the out-patient department (OPD) that was mentioned in the questionnaire was composed of four service segments:

1) almoner (cash payments);
2) triage (parameter observations);
3) screening room (registration of patient details, complaints, and taking of vital parameters);
4) consultation (both general practitioner [GP] and specialist doctors).

To account for different perceptions and to avoid exaggeration and biases, the study had to be complemented by observations. These were performed upon the arrival of the patients in the respective departments for their various services. A stopwatch was used to time and record important variables such as service times, waiting times, and patient arrival rates. In total, the number of days for the observations amounted to 43. Initially, three departments were to be observed (namely, OPD, laboratory, and imaging center); however, the imaging center was omitted due to its complexity. The OPD had four service segments, but only two were selected: the almoner, and the consultation (GPs and specialists). Of the specialists, only five were observed: the internist (only one server), the ear nose throat doctors (ENT) (two servers alternating on different days), and the cardiologist (two servers working in parallel). The reason for not considering the other specialists was that they either had a finite population that was not part of the study or they rarely experienced queues.

For the data-analysis process, the questionnaire and observation data was entered into IBM SPSS software (Statistical Package for Social Sciences – Version 20) and analyzed using Python. The analysis was in terms of general descriptions and visualizations as well as the performances of the correlation tests on some of the observed variables.

The research implemented two methodologies: machine-learning concepts (for the predictions), and queueing theory. From the different machine-learning algorithms that were used for the predictions, this study focused on random forest only because of its flexibility and ease of use; it is widely used among the other related algorithms because of its simplicity and use for both regression and classification. Random forest is a supervised machine-learning algorithm that creates a forest and makes it random; it works by building and merging several decision trees in order to obtain more-accurate and stable forecasts (Abdulkareem & Abdulazeez, 2021). It adds randomness to a model while the trees grow, and it searches for the best feature among a random subset of features instead of searching for the most important feature while splitting a node; as a result, there is a wide variety that usually leads to a better model. Therefore, the algorithm only takes a random subset of characteristics into account to split a node in the random forest; one can even randomize trees by using random thresholds for each function instead of looking for the best possible thresholds (like a normal decision tree). Although random forest has its advantages and disadvantages, one of the important points is its ability to provide a fairly good indicator of the importance that it assigns to features.

The predictions for both the waiting and the total time spent were made using the Jupyter Notebook in Python. The main objective was to check the possibility of predicting the waiting and total times that a patient spends in the hospital on any other day and to determine how accurate this model was. The problem represented a supervised regression machine-learning problem, as both the hospital data and the times to be predicted were available (and also real values). This type of machine learning requires a lot of data, and the model can be trained as well.

The observed data was divided into two-thirds training data and one-third testing data while fitting the model using the random forest regression. Initially, the model had 1000 decision trees, which were further trimmed for simplicity to only 2 levels with 10 decision trees. The study used only two measures to assess prediction accuracy – the mean absolute error (MAE) and the mean absolute percentage error (MAPE). Of these, MAE was considered the more favorable metric for describing the average error and the expected magnitude of errors in the forecast.

Another important step was to use the importance of the characteristics to find the most important or relevant variables to predict these times. In terms of the queueing theory application, we illustrated the queueing system and its various components. The formation of the queues had two important properties: the maximum size (the capacity of a system)/queueing capacity, and the queueing discipline. The size was the population that could either be limited (finite) or unlimited (infinite). The queueing rules for selecting the patients for the services were called the queue discipline, and they were classified as First In/First Out (FIFO), Last In/First Out (LIFO), Priority, and so on. For simple queues, some queueing theory formulas exist, while for complex situations (like this study), computer simulations are needed.

From the two basic approaches that are available for analyzing queue systems, analytical and simulation approaches exist. The former attempts to find formulas (some algorithms) for calculating the steady-state performance measures of a system, while the latter has two types: simulating a known distribution, and simulating a non-specified distribution (bootstrapping). Thus, the simulation method has more flexibility when compared to the analytical approach, which simulates random elements of a system at the same time as keeping track of events as they occur over time. It is based on a simulation model, which is a computer model that mimics a situation in real life. This model is similar to other mathematical models, but it incorporates uncertainty in one or more input variables. The benefit of a computer simulation is the ability to answer what-if questions without actually changing the physical system.

In queueing theory, a queueing concept model known as 'Kendall's notation' exists in its simplest form (A/B/C), which was formalized by Kendall (1953); he was an English mathematician and statistician who was known for contributing to areas like probability, statistical shape analysis, and queueing theory. His notation is used in the description of his queue system; an explanation is given in Figure 2.

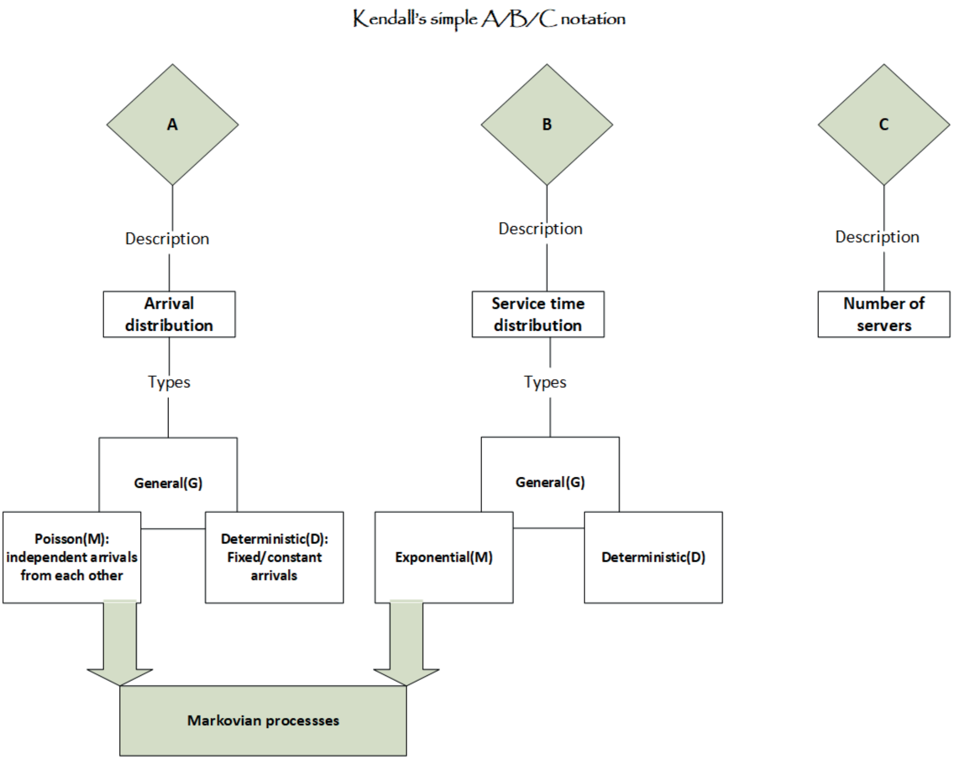

Fig. 2. *Kendall notation*

An extension of 'Kendall's notation' is given by A/B/C/D/E, where D is the queueing capacity and E is the queueing discipline. By the basic assumptions of queueing theory, the following apply using 'Kendall's notation':

A: It is assumed to follow a Poisson distribution and is a Markovian process.

B: It is assumed to follow an exponential distribution and is a Markovian process.

C: It is a positive integer.

D: Is assumed to be ∞, implying that no one is turned away when they come for service.

E: It is assumed to be FIFO/FCFS.

The basic model notation is, thus, given as M/M/C/∞/FIFO/FCFS, where 'M' is the notation for the Markovian processes. The Markov process is a random process in which, given the present, the future is independent of the past and assumes an arrival or service rate. An exponential distribution exhibits an important property of being memoryless; that is, the time for the next arrival is independent of when the last arrival occurred. Its characteristics include an equal mean and standard deviation.

If these assumptions are met, then the analytical approach can be used that makes use of Little's formulas (which are important formulas in queueing theory).

The notation for a Poisson distribution is given as $X \sim \mathrm{Po}(\lambda)$, where λ is the only parameter and signifies the average arrival rate of a set of patients. On the other hand, the notation for an exponential distribution is given as $X \sim \mathrm{Exp}(\mu)$, where μ is the only parameter (and is the average service rate of the patients).

For many queueing situations, arrivals occur randomly, so the occurrence of the next arrival cannot be predicted. Among the distributions that represent the times between successive arrivals, the most important is exponential distribution.

In terms of hospital settings, most arrivals are modeled by a Poisson distribution (where the patients arrive independently one after another) and an exponential service distribution. An example of deterministic arrivals in a hospital setup is when the consultations are by appointment and the patients must arrive at given fixed times. A general distribution is nonspecific and could be any kind of distribution.

We begin by illustrating a simple queue (shown in Figure 3), where queuing behaviors such as reneging and baulking are exhibited.

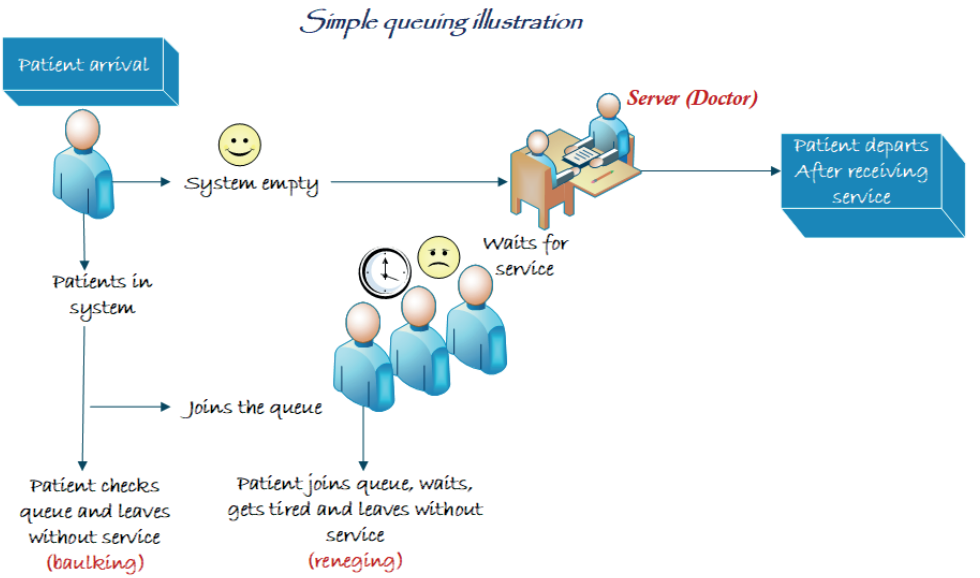

Fig. 3. *Illustration of simple queue*

The illustration simply shows the level of patience from when a patient enters a system up to when he/she is either served or decides to give up and leave. Thus, the queues may be organized in different ways, and they have various types (shown in Figure 4), where only three common types have been illustrated; in this case, the term server refers to a doctor.

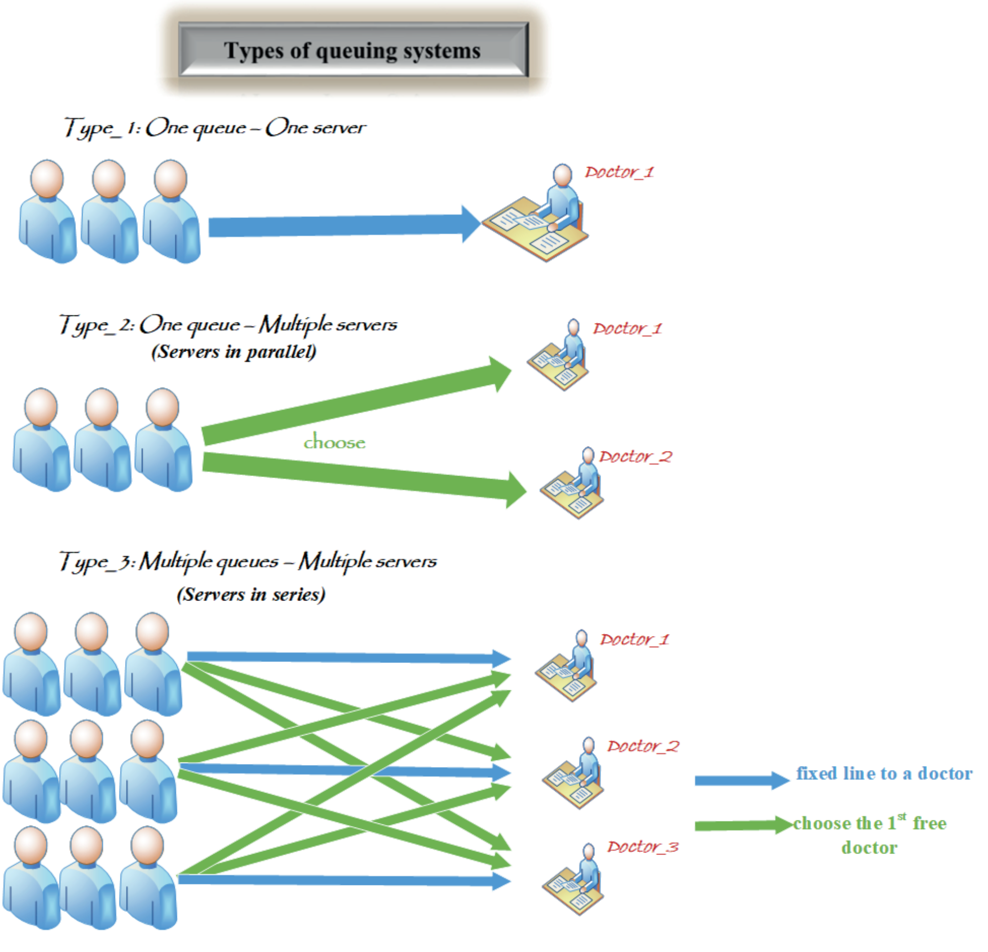

Fig. 4. *Queue types*

In some systems, patients have the option of seeing the first free doctor or simply maintaining the queue to the fixed doctor to whom they have been assigned. The patients are then selected for service by the various disciplines in the queue. In our case, the only department that was chosen was the consultation by a GP; this was mainly because of the time constraint and a lack of personnel. The system belongs to *Type_3* of Figure 4, with three consultation rooms (implying three servers) with patients that maintain a fixed queue. At times, a free doctor can select patients from another queue (which makes the situation complex). Moreover, the patients were from an infinite population, meaning that the system was able to receive all of the patients who came for consultation. The queueing discipline was FIFO, although priority and emergency cases were treated first at times, making the system preemptive.

During the observations, the system was noted to be more complex than was anticipated, and the queueing theory assumptions were violated. For example, the system service time was supposed to exhibit exponential distribution properties by the assumption of healthcare waiting lines; however, it did not since there was no continuity in serving the patients. The arrival process was not constant over time either, and it could not fit the Poisson process; hence, the arrivals and service times were instead categorized as a general distribution (meaning that it could be any kind of a distribution). As a result, the assumption of modeling it as M | M | 1 and using an analytical approach was, therefore, ruled out. The system was observed to be a 3 × G| G | 1 | ∞ | FIFO (or priority), implying a complex system that required a simulation approach.

From the observations, the three observed doctors (named '*Doctor_1*,' '*Doctor_2*,' and '*Doctor_3*') observed 19, 12, and 13 patients, respectively; the patients arrived during the period of 7:30–12:40. The most important parameters that were taken note of were the arrival times, waiting times, service times per hour, and numbers of servers.

Since the system was a bit complex, only the simulation approach on this system (to find the averages) was possible (instead of using the standard queueing theory, with its associated mathematical formulas). The type of simulation that was used was bootstrap simulation, as the service time was not from a specified distribution. The tool that was used for this analysis was Excel, where the observations were used to perform a simulation. The initial step for this approach involved reforming the probability distribution based on random-number generation in reference to the real data, where the random numbers were generated by the *RAND*( ) function. One simulation was performed per doctor for each of the three using the *VLOOKUP*( ) function to simulate 1000 replications, where each replication was an independent replay of the occurring events. The replications were generated using a data table; to do this, the observed data in the spreadsheet was used to construct a typical "prototype" of the simulation. Also, a 95% confidence interval (CI) was estimated. Part of the replication process table that was created for '*Doctor_1*' is shown in Figure 5.

The simulation analysis was also useful for investigating what might have happened if a different policy or strategy had been used. After the observations, the simulation model was used to compare two situations: the first was an observed system "with disruptions" (where the doctors had other things to attend to during their consultations), and the second was termed a system "without disruptions" (for comparison, where the doctors only consulted without any other disturbances).

One of the virtues of the simulation was that it allowed us to experiment with alternatives. Although not used in this study, other alternatives could include simulating a system in which the number of doctors is increased to see the difference that can be made.

No. Replications	Average time in system	Average time in queue	% time server is idle
	22.4	13.72	0.1033058
1	15.32	8.72	0.5657895
2	8.8	2.52	0.6884921
3	8.68	2.08	0.4728435
4	9.52	2.16	0.5523114
5	12.8	5.92	0.4723926
995	14.24	6.76	0.5630841
996	11.2	3.84	0.54
997	18.12	10.4	0.4373178
998	11.12	4.72	0.638009
999	8.84	2.08	0.5517241
1000	9.88	2.16	0.6553571

Fig. 5. *Replication process for 'Doctor_ 1' (supporting software tool output)*

After meticulous data training and simulations, the results are stated in the next section.

3. RESULTS

The data collected for the analysis was both quantitative and qualitative; that is, arrival, waiting, service, and total waiting times in addition to feedback on busy departments and busy days, respectively.

After gathering the opinions from the respondents, making observations, and using the hospital records (registers), the results indicate the waiting times, busy departments, busy days and times, and possible causes of any long queues (where the term 'busy' was associated with a high number of patient arrivals and long queues). The perceptions that were provided by both the patients and the medical personnel helped confirm the findings. All analyses of the patients' movements were performed on an hourly basis, with the times being recorded in minutes.

The arrival rate in the selected departments is presented in Figure 6. Observing the trend, most patients arrived in the morning, and the numbers decreased as the day went by (although the emergency cases tended to increase after hours). Department-wise, the payment section experienced the highest number of arrivals – most likely because it was the starting point. These patients were then distributed to the various segments of the consultation service. Among the specialists, the internist seemed to be burdened the most.

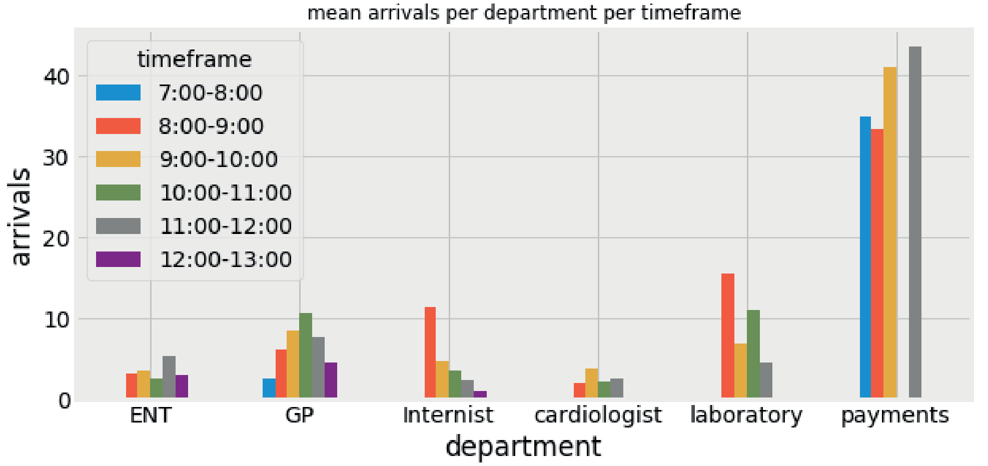

Fig. 6. *Department hourly arrivals*

Figure 7 shows the amount of time that a patient waited for service in the selected departments, given that they found zero patients (no one in the system) up to more than six people. This factor alone triggered waiting times, as other incoming patients were also affected. Normally, if no one is in the system upon arrival, the patient who is first in the queue did not incur any waiting time, with expectations of being served immediately; therefore, this was a concern.

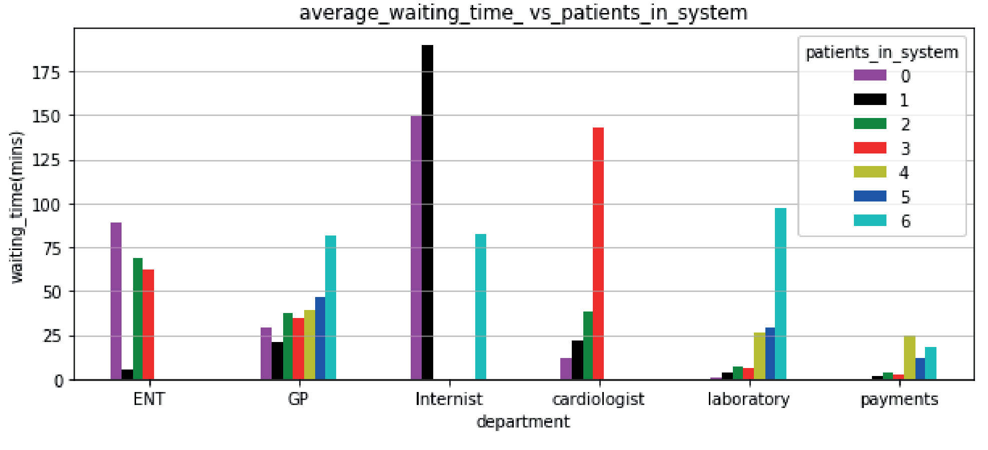

Fig. 7. *Department waiting time when patients are already in system*

The averages for the waiting, service, and total times that were spent in the hospital by the patients in the selected service segments were calculated; of all of the services, the internist had the highest times (illustrated in Figure 8).

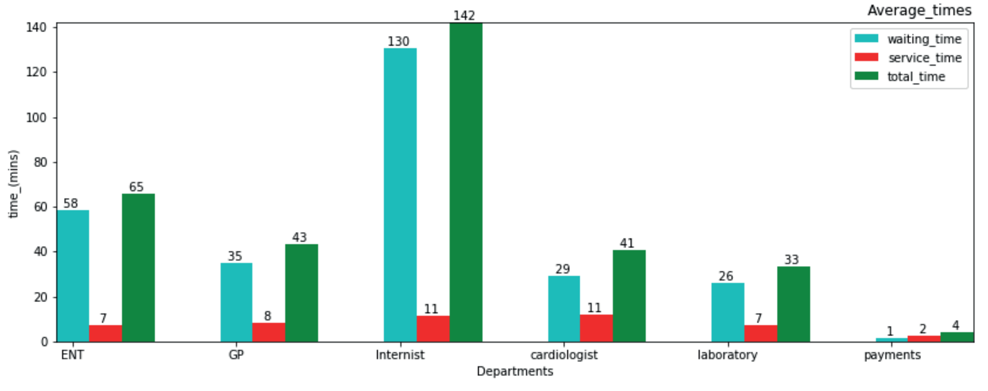

Fig. 8. *Observed times in hospital*

In terms of the laboratory department, the waiting times were in terms of turnaround times (TAT). The averages were calculated by using a sample of five patients for each laboratory test. TAT is defined as the total time that it takes from when a sample is submitted to when the results come out. For the research, only 6 common tests were selected (shown in Table 1); it can be seen that none of the tests were below 56, likely explaining why the patients waited so long.

Table 1. *TATs for RHL lab results*

Laboratory test	TAT (time)
Mp (malaria parasite)	0:56
Hb (hemoglobin)	1:09
FBC (full blood count)	1:45
MS	2:19
FBS (glycaemia)	2:34
Stool analysis	3:50

For correlation clarity, Figure 9 was constructed to check the existence of a relationship among the variables. Of all of the variables, only waiting times and total time spent show a significant relationship (with a strong positive correlation of 0.99).

Concerning the predictions on the waiting and total times that were to be spent by a patient in the hospital tomorrow (referring to the future), a random forest tree was used (where the output of the tree was random each time the notebook was run). Also, the variables that were important for predicting both parameters were calculated and constructed according to their importance. For the prediction of waiting times, the mean absolute error (MAE) was 10.53 minutes, with a standard deviation of 3.54 minutes. The MAE value implied that the model predicted an average of 10.53 minutes of error. To show how accurately the trained model fit the data set, the R-squared value was calculated and found to be 0.64; this showed that the model prediction was almost close to the actual data set.

Fig. 9. *Correlation of variables*

Figure 10 shows that, to predict the waiting times, the most important variables were the times of arrival and the numbers of patients that one would find in the system.

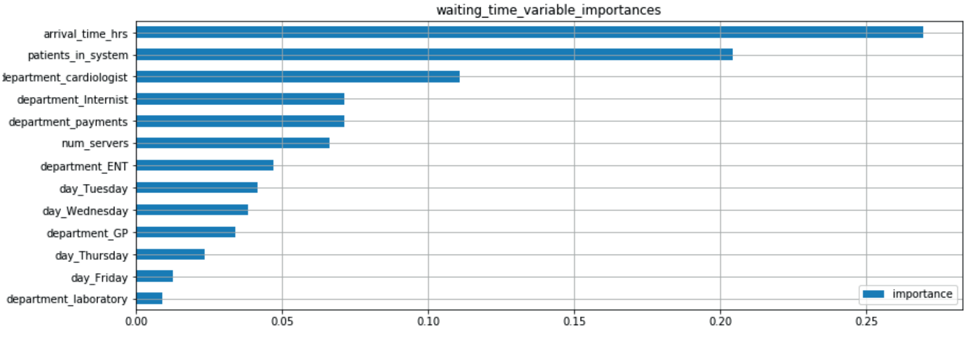

Fig. 10. *Important variables for predicting waiting times*

The trimmed waiting-time tree for prediction is given in Figure 11. Given the variables on the tree, a patient was able to predict how long they could expect to wait in the hospital before receiving service.

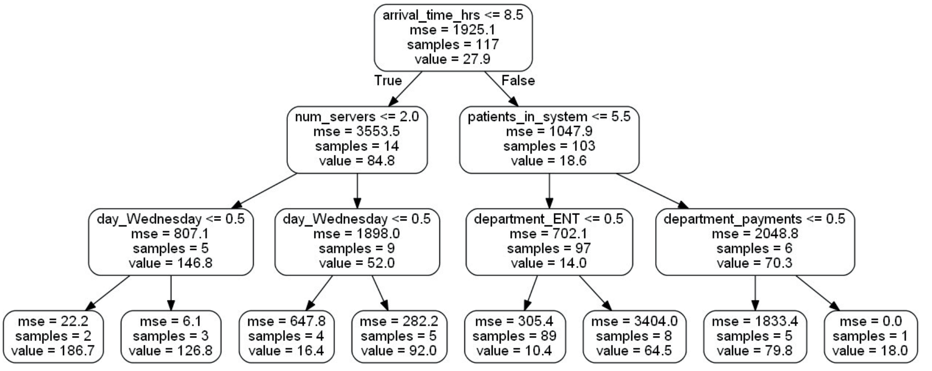

Fig. 11. *Waiting time predictions*

To interpret this random forest in particular, if a patient arrived before 8:50 a.m., they chose the indicated true arrow from the first node; otherwise, they chose the false arrow. If the first node was true, the next question encountered was the number of servers. We moved down the forest answering questions up to the last row, whose nodes indicated the predicted waiting times (which were given as 'value' in minutes). Consider the following example: if a patient arrived before 8:50 a.m. and the service had fewer than two servers on a Wednesday, the waiting time prediction was 186.7 minutes. The 'samples' that were shown on each node were indications of the numbers of samples that were randomly taken for sampling data points, and 'MSE' was the mean squared error.

For the total time prediction, predicting total time that is spent by a patient is somewhat infeasible in reality, as the outcome cannot be predicted without a patient going through the process of waiting and receiving the service first. Therefore, this prediction served as a verifier of the methods rather than a predictor of information beforehand. After training the model, however, the MAE was found to be 1.75 minutes; this implied that the model predicted 1.75 minutes of error on average. The median absolute error (MedAE) was 0.77 minutes, and the mean absolute percentage error (MAPE) was 7.45%.

As highlighted in Figure 12, the most important variables for the total time prediction were the waiting and service times.

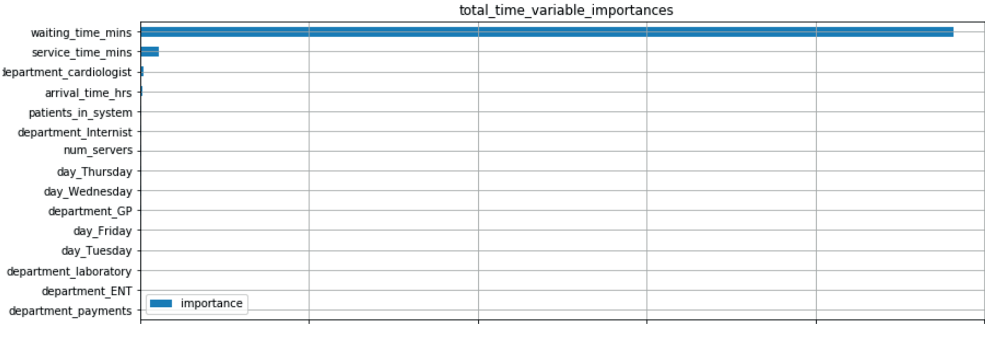

Fig. 12. *Important variables for predicting total time*

Removing all of the other variables and leaving the two important ones, MAE was reduced to 1.4 minutes, while MedAE decreased to 0.39 minutes and MAPE decreased to 5.23%. This implied that all of the other features were not important for the total time prediction after all. Scoring the model on the training data, the R-squared value was 0.99 minutes; this showed that the model prediction was very close to the actual data set. An illustration tree for the total time is shown in Figure 13.

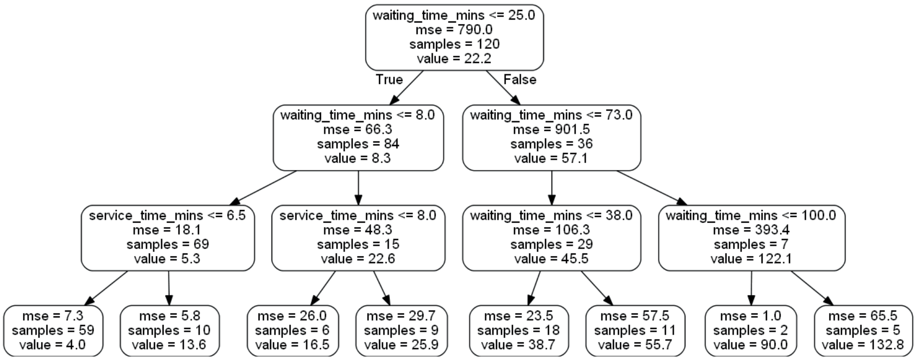

Fig. 13. *Total time predictions*

In order to see the samples from the validation data set and the range within which their errors lay, error plots were constructed for both the waiting and total times. It can be noted from the plots that most of the samples had small minute errors (as seen in Figure 14).

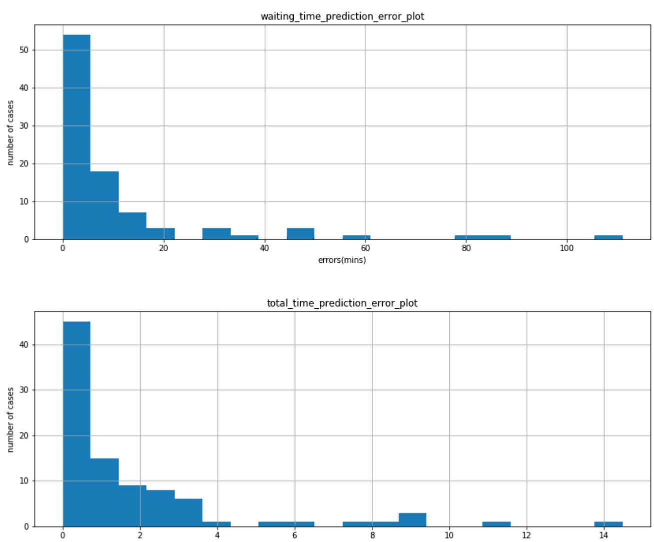

Fig. 14. *Prediction error plots*

For the queueing theory part, each of the three doctors had a summary of the results after a simulation of 1000 replications. The summary measures of performance for the simulations that were calculated included the average time in the system, the average time in a queue, and the percentage of time that a server was idle. The time in the system was the total time that a patient spent from his/her arrival waiting in a queue up to the time that he/she received service (which was simply the waiting time before the service). The time that a server was idle was when a doctor was not busy.

From the results of Figures 15 and 16, 'queueing results with disruptions' was the observed system where the doctors' arrival times were inconsistent and they engaged in other work other than serving the patients in queues. On the other hand, 'queueing results with no disruptions' was the model that it was compared to if the operating strategy was different, with doctors only serving patients without attending to other commitments.

	Average time in system (mins)	Average time in queue (mins)	% time server is idle	Sample Standard deviation for average time in system	Sample Standard deviation for average time in queue	95% Confidence interval for time in system		95% Confidence interval for time in queue	
				54.37	50.64	-40.14	177.34	0.00	156.89
MIN	7.48	1.92	0%						
AVERAGE	68.60	55.61	22.76%	Doctor_1 queueing results with disruptions					
MAXIMUM	378.08	350.08	71.38%						

Fig. 15. *Model for Doctor 1 with disruptions (supporting software tool output)*

	Average time in system (mins)	Average time in queue (mins)	% time server is idle	Sample Standard deviation for average time in system	Sample Standard deviation for average time in queue	95% Confidence interval for time in system		95% Confidence interval for time in queue	
				3.35	2.98	4.28	17.67	0.00	10.00
MIN	5.76	0.16	9.69%						
AVERAGE	10.98	4.03	54.97%	Doctor_1 queueing results with no disruptions					
MAXIMUM	37.76	29.28	78.89%						

Fig. 16. *Model for Doctor 1 with no disruptions (supporting software tool output)*

When comparing the 2 Systems for each doctor, it could be observed that there was a huge difference – especially in the times that a patient spent in both a queue and the system. For example, a patient spent 55 minutes on average queueing in Figure 15, whereas they only spent 4 minutes in Figure 16 because the doctor had no other duties other than consulting. The same concept was applied to the other two doctors, with any differences noted in the systems.

CI play a role in the modeling and simulation, as they were used in the model validation. The wide CI that was outputted in Figure 15 indicated the small sample size that was used; in our case, standard deviation (σ) told us how the collected time was spread out from the average.

Since random numbers were generated, the output was also random; this meant that running the spread-sheet again would have given slightly different results from what could be seen. It is to be noted, however, that a simulation is an approximate method and might not give exact answers; hence, the reason why the ‘with no disruptions’ comparison system might have seemed to be close to perfect. In the following section, we give an in-depth discussion of the results as well as our conclusions.

4. DISCUSSION AND CONCLUSIONS

After performing the analysis, we drew conclusions on the findings and gave recommendations by providing possible solutions. The results and findings of this research made use of registers where Mondays were normal days; therefore, the study assumed the nonexistence of a ghost town. Given the case that a ghost town happened and patient arrivals declined on Mondays, however, then these findings were not applicable and were subject to change only in terms of the busy days and the number of arrivals.

From the hospital registers, it was generally noted that the number of patient arrivals had been reducing over the past years; this was likely because of the crisis or standards being lowering without being noticed.

After observing and analyzing the possible causes of the queues that result in long waiting times, we found that patient waiting times were sensitive to doctors’ arrival times and the time that the doctors spent on other activities. This could be supported by the waiting times that were noted for all of the patients who arrived and found zero people in the system. The primary and secondary data played important roles in our attempt to answer the research questions.

From the combined patient and medical responses, the three main departments were OPD, the laboratory, and the imaging center. As much as these results were obtained from the perceptions of 196 people, we concluded them to be true without further investigations. All of the other observations that were necessary for the research were then carried out in two of these departments.

According to the perception of the questionnaire, the top-three busy days were found to be Monday, Wednesday, and Tuesday; at the same time, our observations indicated that Monday, Wednesday, and Tuesday had the highest numbers of patient arrivals. In combination, this is evidence enough to conclude that Monday, Wednesday, and Tuesday are the top-three busiest days of the week for the hospital. As was stated from the other studies in the literature review, we can also compare our results with Biya et al. (2022) (who found Monday to be the busiest day of the week). Again, when analyzing the times that were spent in the hospital, Wednesday and Tuesday had the highest amounts of time that were spent by the patients. Here, Monday could not be observed because of the crisis; from our secondary data findings, however, we could assume that it featured the highest total time that was spent in the hospital.

We also wanted to find out if there could be a correlation that existed between the waiting lines and the day of the week (or even the time of day). From our assessment, there was an implication that waiting lines and the time that was spent were related to the day of the week. In simple form, we could put it in an equation as follows:

$$\uparrow \text{ in arrivals} \Longrightarrow \text{long queues} = \text{busy day} \Longrightarrow \text{more time spent in hospital} \tag{1}$$

In terms of busy times of the day, most of the medical personnel responded to mornings; again, this matched with our observations, where the hospital was discovered to be more congested and busy during the morning hours.

In response to the predictions that were made (and using the developed models), a patient can predict the amount of time that they can expect to wait the next day in the hospital. Figure 12 complements Figure 9 regarding the fact that the waiting time is the most important variable for total time prediction; to support this, the two were portrayed to have a strong correlation.

In queueing theory, we based our conclusions on the data analysis (including the simulation and assumptions), as the data was not enough to draw strong scientific conclusions. This implied that the value of the essay was not in its actual conclusions but rather in the methodology that was used. Using a what-if analysis, an insight can be gained into the models in understanding and implementing waiting-time strategies. We can only explain what can be done to get a more reliable appreciation of the current situation vs. different possible working policies (which are indicated in the recommendation section). The summarized results are presented in Table 2.

Table 2. *The findings*

Research questions	**Subjective findings**
Main causes of long waiting times	Doctors' arrival times and commitments to other activities
Departments with long waiting times	OPD, laboratory, and imaging center
Busy days	Monday, Wednesday, and Tuesday
Busy times	Morning

Based on our work, an attempt to improve hospital performance, and the general operation, we suggest the following for each department:

1. OPD Department
   – Change the queueing system: use different queueing rules, like adopting "*Type* 2: *One queue- Multiple servers*" queueing from Figure 4 (or any other combination), where a patient is assigned to the first free/available doctor.
   – Increase the number of doctors per shift: the queuing system setup could be changed by having an increase of one or two doctors consulting (especially on the highlighted busy days) to reduce waiting times.
   – Morning meetings: the doctors on duty could either send a representative to any meetings, or they could be shifted to take place at a less-busy times (e.g., in the afternoon).
   – Ward rounds: the doctors on duty at OPD could be exempted from ward rounds while non-consulting doctors do rounds and attend to emergency cases.
   – Classify services: a class of patients that require shorter service times (like those coming for medical certificates) could be dedicated to a different server. Waiting an hour for a signature and a stamp that takes two minutes is not ideal.

- For specialist consultations, appointment times could be introduced:
  - *Triage*: the process from triage to the screening room is more or less repetitive; there is a need to either combine or cut off some unnecessary steps and procedures to reduce the waiting times.
  - *Almoner (Payment section)*: the system is not computerized, and a lot of time is spent invoicing by hand; if this system were changed, many differences could be noted.

2. Laboratory Department
   From the findings concerning the long waiting times in the laboratory, this was due to the fact that results take a long time to come out (and/or handed to the patient) than the time it takes to collect samples. As a result, the solution can only lie within reviewing the way the results are grouped or timed. The results should be produced in two shifts (i.e., 11 a.m. and 2 p.m.).
3. Imaging center
   This was among the departments that were mentioned to have long waiting times; however, no observations could be made because of the complex setup and the sensitivity of the process. We cannot make any conclusions but to recommend a study of this for future work.

In general, the hospital can take the following points into consideration:

1. Staff allocation: in general, the staff allocations could be more concentrated for all of the service segments to suit the busier days and times.
2. Entertainment: providing patients with educational health talks, books, magazines, etc. could also help keep them occupied while they are in queues. This can also help avoid reneging and baulking, as most patients become impatient and bored and leave without receiving service.

For future research and use, the type of operational data that is needed as input for a queueing model could be introduced, as it is often unavailable in the registers. This is necessary for reviewing system performance as well.

Acknowledgements

Funding for the project was provided by The African Institute for Mathematical Sciences (AIMS)-Cameroon and Mastercard Foundation.

REFERENCES

Abdulkareem N.M. & Abdulazeez A.M. (2021). Machine learning classification based on random forest algorithm: a review. *International Journal of Science and Business*, **5**(2), pp. 128–142.

Adeniran D.A., Burodo M.S. & Suleiman S. (2022). Application of qeuing theory and management of waiting time using multiple server model: Empirical evidence from Ahmadu Bello University Teaching Hospital, Zaria, Kaduna State, Nigeria. *International Journal of Scientific and Management Research*, **5**(4), pp. 159–174. **DOI**: https://doi.org/10.37502/IJSMR.2022.5412.

Almusawi M.A., Radwan N.M., Mahmoud N.E., Haji A.M. & Alabdulkareem K.I. (2023). Patient waiting time at primary healthcare centers in Riyadh Region, Saudi Arabia: A cross-sectional study. *Dr. Sulaiman Al Habib Medical Journal*, **5**, pp. 65–69. **DOI**: https://doi.org/10.1007/s44229-023-00032-3.

Benevento E., Aloini D. & Squicciarini N. (2023). Towards a real-time prediction of waiting times in emergency departments: A comparative analysis of machine learning techniques. *International Journal of Forecasting*, **39**(1), pp. 192–208. **DOI**: https://doi.org/10.1016/j.ijforecast.2021.10.006.

Biya M., Gezahagn M., Birhanu B., Yitbarek K., Getachew N. & Beyene W. (2022). Waiting time and its associated factors in patients presenting to outpatient departments at Public Hospitals of Jimma Zone, Southwest Ethiopia. *BMC Health Services Research*, **22**, art. no. 107. **DOI**: https://doi.org/10.1186/s12913-022-07502-8.

Boukas E.-K., Zhang Q. & Yin G. (1995). Robust production and maintenance planning in stochastic manufacturing systems. *IEEE Transactions on Automatic Control*, **40**(6), pp. 1098–1102. **DOI**: https://doi.org/10.1109/9.388692.

Brockmeyer E., Halstrøm H. & Jensen A. (1948). *The life and works of A.K. Erlang*. Copenhagen: Copenhagen Telephone Comp.

Chakka R., Van Do T. & Pandi Z. (2009). A generalized Markovian queue and its applications to performance analysis in telecommunications networks. In: D.D. Kouvatsos (Ed.), *Performance Modelling and Analysis of Heterogeneous Networks*. River Publishers, pp. 371–387. **DOI**: https://doi.org/10.1201/9781003339052-23.

Cooper R.B. (1981). Queueing theory. In: *Proceedings of the ACM'81 conference* (1981), pp. 119–122.

Ding S. (2023). Analysis and application of computer queueing theory. In: P. Kar, J. Li & Y. Qiu (Eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*. Amsterdam: Atlantis Press, pp. 447–453. **DOI**: https://doi.org/10.2991/978-94-6463-300-9_45.

Dshalalow J.H. (Ed.) (1995). *Advances in Queueing Theory, Methods, and Open Problems*. Boca Raton: CRC Press. **DOI**: https://doi.org/10.1201/9781003418283.

Elalouf A. & Wachtel G. (2021). Queueing problems in emergency departments: a review of practical approaches and research methodologies. *Operations Research Forum*, **3**, art. no. 2. **DOI**: https://doi.org/10.1007/s43069-021-00114-8.

Erlang A.K. (1909). The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, **20**, pp. 33–39.

Fun W.H., Tan E.H., Khalid R., Sararaks S., Tang K.F., Ab Rahim I., Md. Sharif S., Jawahir S., Sibert R.M.Y. & Nawawi M.K.M. (2022). Applying discrete event simulation to reduce patient wait times and crowding: The case of a specialist outpatient clinic with dual practice system. *Healthcare*, **10**(2), art. no. 189. **DOI**: https://doi.org/10.3390/healthcare10020189.

Garcia-Corchero J.D. & Jimenez-Rubio D. (2022). Waiting times in healthcare: equal treatment for equal need? *International Journal for Equity in Health*, **21**(1), art. no. 184. **DOI**: https://doi.org/10.1186/s12939-022-01799-x.

Green L. (2013). Queueing analysis in health care. In: R. Hall (Ed.) *Patient Flow: Reducing Delay in Healthcare Delivery*, Second Edition, Springer, pp. 361–384. **DOI**: https://doi.org/10.1007/978-1-4614-9512-3.

Harchol-Balter M. (2013). *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge: Cambridge University Press. **DOI**: https://doi.org/10.1017/CBO9781139226424.

Harki N.A. (2024). Multi-level feedback queue scheduling technique. *Cihan University-Erbil Scientific Journal*, **8**(2), pp. 36–42. **DOI**: https://doi.org/10.24086/cuesj.v8n2y2024.pp36-42.

Ji J. (2023). Application of queuing theory in public service counters. In: P. Kar J., Li & Y. Qiu (Eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Amsterdam: Atlantis Press, pp. 509–517. **DOI**: https://doi.org/10.2991/978-94-6463-300-9_52.

Johnston A., Lang E. & Innes G. (2022). The waiting game: managing flow by applying queuing theory in Canadian emergency departments. *Canadian Journal of Emergency Medicine*, **24**(4), pp. 355–356. **DOI**: https://doi.org/10.1007/s43678-022-00318-9.

Joseph J., Senith S., Kirubaraj A.A. & Ramson J.S. (2022). Machine learning for prediction of wait times in outpatient clinic. *Procedia Computer Science*, **215**, pp. 230–239. **DOI**: https://doi.org/10.1016/j.procs.2022.12.026.

Kalwar M.A., Marri H.B., Khan M.A. & Khaskheli S.A. (2021). Applications of queuing theory and discrete event simulation in health care units of Pakistan. *International Journal of Science and Engineering Investigations*, **10**(109), pp. 6–18. **URL**: https://www.researchgate.net/publication/356980604.

Kendall D.G. (1953). Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*, **24**(3), pp. 338–354.

Koole G. (2013). *Call center optimization*. MG Books.

Lakatos L., Szeidl L. & Telek M. (2019). Queuing networks, In: L. Lakatos, L. Szeidl, M. Telek, *Introduction to Queueing Systems with Telecommunication Applications*, Second Edition, Springer, pp. 421–444. **DOI**: https://doi.org/10.1007/978-3-030-15142-3_10.

Le D.X., Do H.T., Bui K.T., Hoang T.Q., Nguyen G.H., Nguyen A.V., Nguyen Q.T., Gorgui-Naguib H. & Naguib RG.N. (2021). Lean management for improving hospital waiting times – Case study of a Vietnamese public/general hospital emergency department. *The International Journal of Health Planning and Management*, **37**(1), pp. 156–170. **DOI**: https://doi.org/10.1002/hpm.3310.

Li X. (2024). Performance evaluation of M | M | 1 queueing models in cloud computing environments. *Highlights in Science, Engineering and Technology*, **85**, pp. 841–848. **DOI**: https://doi.org/10.54097/1xcw8y74.

Liberopoulos G., Papadopoulos C.T., Tan B., Smith J.M. & Gershwin S.B. (Eds.) (2006). *Stochastic modeling of manufacturing systems*. Springer. **DOI**: https://doi.org/10.1007/3-540-29057-5.

Machangara D.T. (2018) *Predictions and Application of Queuing Analysis at Regional Hospital Limbe* [MSc thesis], African Institute for Mathematical Sciences (AIMS), Cameroone, December 2018. **URL**: https://library.nexteinstein.org/wp-content/uploads/2023/01/Daphne-Essay-Template-en-2017-2018.tex_.pdf [1.10.2024].

Mbwogge M., Astbury N., Nkumbe H.E., Bunce C. & Bascaran C. (2022). Waiting time and patient satisfaction in a subspecialty eye hospital using a mobile data collection kit: pre-post quality improvement intervention. *JMIRx Med* **3**(3), e34263. **DOI**: https://doi.org/10.2196/34263.

Mohammadi T., Roshanaei G., Faradmal J., Sadeghifar M., Manafi B. & Mahjub H. (2022). Improving service efficiency and throughput of cardiac surgery patients using Monte Carlo simulation: a queueing setting. *Scientific Reports*, **12**(1), art. no. 21217. **DOI**: https://doi.org/10.1038/s41598-022-25689-y.

Nair A.M., Sreelatha K. & Ushakumari P. (2021). Application of queuing theory to a railway ticket window. In: *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, pp. 154–158. **DOI**: https://doi.org/10.1109/ICIPTM52218.2021.9388368.

Newell G.F. (1995). *Applications of Queueing Theory*. Second Edition. London – New York: Chapman an Hall. **DOI**: https://doi.org/10.1007/978-94-009-5970-5.

O'Malley M.S., Fletcher S.W., Fletcher R.H. & Earp J.A. (1983). Measuring patient waiting time in a practice setting: A comparison of methods. *The Journal of Ambulatory Care Management*, **6**(3), pp. 20–27. **DOI**: https://doi.org/10.1097/00004479-198308000-00006.

Palm C. (1943). Intensity fluctuations in telephone traffic. *Ericsson Technics*, vol. **44**.

Palomo S., Pender J.J., Massey W.A. & Hampshire R.C. (2023). Flattening the curve: Insights from queueing theory. *PLoS ONE*, **18**(6), e0286501. **DOI**: https://doi.org/10.1371/journal.pone.0286501.

Peter P.O. & Sivasamy R. (2021). Queueing theory techniques and its real applications to health care systems – outpatient visits. *International Journal of Healthcare Management*, **14**(1), pp. 114–122. **DOI**: https://doi.org/10.1080/20479700.2019.1616890.

Proudlove N. (2022). Use and misuse of queueing theory for hospital capacity decisions. In: J. Vissers, S. Elkhuizen, N. Proudlove (Eds.) *Operations Management for Healthcare*. Second Edition, London: Routledge, pp. 197–216. **DOI**: https://doi.org/10.4324/9781003020011.

Rema V. & Sikdar K. (2021). Optimizing patient waiting time in the outpatient department of a multi-specialty Indian hospital: The role of technology adoption and queue-based Monte Carlo simulation. *SN Computer Science*, **2**, art. no. 200. **DOI**: https://doi.org/10.1007/s42979-021-00595-8.

Saastamoinen K., Rissanen A., Suni J., Hyttinen J., Paakkunainen P. & Liakka A. (2023). Simulation of the queuing situation of patients at a health center. *Engineering Proceedings*, **39**(1), art. no. 72. **DOI**: https://doi.org/10.3390/engproc2023039072.

Saaty T.L. (1957). Résumé of useful formulas in queuing theory. *Operations Research*, **5**(2), pp. 161–200.

Shortle J.F., Thompson J.M., Gross D. & Harris C.M. (2018). *Fundamentals of Queueing Theory*, Fifth Edition. Hoboken: John Wiley & Sons.

Suleiman S., Burodo M.S. & Ahmed Z. (2022). An application of single and multi-server exponential queuing model in some selected hospitals of the North-Western Nigeria. *Asian Journal of Probability and Statistics*, **16**(2), no. AJPAS.82945. **DOI**: https://doi.org/10.9734/ajpas/2022/v16i230396.

Walker K., Jiarpakdee J., Loupis A., Tantithamthavorn C., Joe K., Ben-Meir M., Akhlaghi H., Hutton J., Wang W., Stephenson M., Blecher G., Buntine P., Sweeny A. & Burak T. (2022). Emergency medicine patient wait time multivariable prediction models: a multicentre derivation and validation study. *Emergency Medicine Journal*, **39**(5), pp. 386–393. **DOI**: https://doi.org/10.1136/emermed-2020-211000.

Xiong Y. (2023). Research and application analysis of the basic theory of queuing theory. In: P. Kar, J. Li & Y. Qiu (Eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Amsterdam: Atlantis Press, pp. 454–463. **DOI**: https://doi.org/10.2991/978-94-6463-300-9_46.

Yao D.D. (Ed.) (1994). *Stochastic modeling and analysis of manufacturing systems.* Springer Science & Business Media. **DOI**: https://doi.org/10.1007/978-1-4612-2670-3.

Zavanella L., Zanoni S., Ferretti I. & Mazzoldi L. (2015). Energy demand in production systems: A queuing theory perspective. *International Journal of Production Economics* **170**(B), pp. 393–400. **DOI**: https://doi.org/10.1016/j.ijpe.2015.06.019.