Predicting Hospital Medicine Needs Based on a Multiple Linear Regression Model

Bashar Talib AL Nuaimi[1], Wedyan Habeeb Hameed[1] and Wisam Sami Mohsin[2]
[1] *Department of Computer Science, University of Diyala, 32001 Baqubah, Diyala, Iraq*
[2] *Department of Islamic Banking and Finance, University of Diyala, 32001 Baqubah, Diyala, Iraq*
alnuaimi_bashar@uodiyala.edu.iq, wedyanhabeeb@uodiyala.edu.iq, wissamsami@uodiyala.edu.iq

Keywords: Correlation and Regression Analysis, Forecast, Future Needs, Multifactorial Model, Retrospective Analysis, Time Series Analysis.

Abstract: The study's issue is that hospitals are unable to accurately match the predicted demand and actual consumption for medications due to the use of traditional and ineffective forecasting techniques. The study's objective: 1) Finding a systematic approach based on modern scientific forecasting methods, based on multi-factor mathematical models. 2) Selecting the optimal forecasting model among the methods utilized in this study to predict the future needs of hospitals for medicines in the coming years. 3) Calculating the discrepancy rate for each drug annually to check the forecasting accuracy. Three statistical analyses were performed: correlation and regression, time series, and retrospective forecast between estimated demand and actual consumption. Three accuracy indicators (MAPE, MAD, and MSD) were used as criteria for selecting the optimal model that best describes the pattern of future demands for six pharmaceuticals and five factors influencing drugs' consumption from 2015 to 2019. Three time series forecasting techniques (ALT, SES, and DES) were tested and compared with the MLR model to verify its forecasting accuracy. Time series techniques were compared to each other; DES was selected as the optimal technique among them. The results showed that the mean MAPE for all medications by using MLR and DES models was 15.63 and 24.61, respectively. Therefore, we conclude that MLR is the optimal model for hospital inventory management and forecasting future needs since it has a lower relative error rate compared to DES. This indicates that the influence of independent factors on demand is stronger than the time factor. Therefore, MLR outperformed DES, which relies on the time factor. With the exception of Infliximab 100mg and Tocilizumab 20mg/ml, whose values exceed 25%, the average discrepancy rates between the estimated demand and actual consumption of each medication over a 5-year period are statistically significant and within the acceptable bounds.

1 INTRODUCTION

Hospital inventory management and medication needs assessment are still challenging problems that have an impact on the standard of care given to patients and society [1]. It is assumed that medical facilities must cope with a limited budget and, at the same time, meet the patients' requirements. Optimizing administrative procedures for planning and organizing medical and medication treatment can help solve some issues, particularly those pertaining to predicting health care demands, which is the primary objective of health development in all nations. To guarantee that patients receive the finest medical treatment possible, it is important to increase the precision of the drug needs assessment utilizing contemporary forecasting techniques [2]. The development of a set of mathematical methods for predicting the need for medicines and comparing actual consumption with estimated demand is one of the main ways to rationalize the use of financial resources and protect medicines from spoilage due to poor storage or expiration. Accurate demand forecasting reduces drug overstocking due to expired medications and sudden shortages, preventing hospitals from resorting to purchasing at high prices on the black market. Accurate demand forecasting may also improve medical quality and safety, improving patient outcomes and reducing Mortality Rate, and reduce medication errors [3]. When inventory is organized perfectly, the likelihood of inappropriate substitute medications being dispensed due to shortages is reduced. In the case of healthcare, it is also critical to emphasize that the forecasting process, which takes into account both the creation of budgets and the usage of medications, is predicated on the assessment of historical consumption, a method that compromises the complexity of the industry. In addition, the market share of competing products may be affected by external factors such as expiration, seasonality, new surgical procedures, epidemics, or the introduction of innovative drugs, which significantly impact drug demand forecasts by adjusting historical consumption [4]. Therefore, this study aims to select optimal models for predicting hospital drug needs and identify the main factors influencing the consumption of these drugs, depending on the degree of correlation analysis between drugs and factors on the one hand and the degree of correlation analysis between the factors themselves on the other [5], [6]. To achieve this goal, we needed to complete the following tasks:

1) Using correlation analysis techniques to identify the key variables (factors) influencing the needs for medicines in hospitals and medical facilities.
2) Establishing a multi-factor linear regression (MLR) model and create a predicting equation for each drug separately based on the factors that selected from correlation analysis and compared the forecasting of those equations with three methods of time series forecasting to verify the accuracy of the our model.
3) Calculating the average discrepancy percentages between actual consumption and estimated demand to verify the accuracy of the forecast and select the optimal predictive.

In healthcare data analysis, time series models and multiple linear regression are used to forecast future drug demand, but each has distinct advantages and applications. Time series models such as ARIMA or exponential smoothing are appropriate when data are based on historical values and seasonal trends, making them ideal for forecasting drug demand based on historical consumption patterns [7]. Multiple linear regression is used when there is a clear relationship between drug demand and multiple independent variables such as patient population, age, season, or economic factors. These models were chosen because they offer high analytical accuracy; time series captures temporal and cyclical changes, while linear regression quantifies the influence of external factors on demand [8], [9]. Combining them can enhance the accuracy of forecasts, especially in healthcare systems where drug demand is influenced by both temporal and demographic factors. Therefore, the current study focused on using the multiple linear regression model as a basis for forecasting future drug demand. This model was chosen for several reasons related to the nature of the data, ease of interpretation, and the model's efficiency. Linear regression is considered an ideal choice because it provides accurate and easy-to-understand results due to the linear relationship between the independent variables and the antecedent variable. Linear regression also provides directly interpretable coefficients, allowing for a clear understanding of the effect of each independent variable on drug demand (for example, a 10% increase in the number of patients increases drug demand by X%). Its simplicity and speed of implementation compared to more complex models such as neural networks or random forests are some of the reasons for choosing it for forecasting future demand. Furthermore, it is easy to integrate with time series analysis methods, as the data in the study sample is linked over the time factor, which facilitates the use of multiple linear regression for forecasting based on historical trends.

2 RELATED WORKS

The health sector is one of the most important service sectors for the individual and society, as it is closely related to human life. The process of correctly assessing the need for medicines is one of the most important elements affecting the quality of services provided to patients, so it is necessary to find modern methods or mechanisms for assessing the need for medicines in different health sectors [7]. Health centres around the world face many difficulties and challenges because of the mismatch between actual consumption and expected demand resulting from the use of primitive traditional methods. This section will review the literature related to the processes of estimating the drug needs of health centres and hospitals in order to improve demand estimation and minimize waste of public money.

In 2022, Mbonyinshuti F. et al. [10] the researchers employed a number of data preparation techniques to choose the top 17 critical medications for non-communicable diseases (NCDs), A Random Forest (RF) model was used to anticipate demand using vital medicine consumption data from 2015 to 2019 for roughly 500 medical goods as part of the machine learning applications study. The Random Forest model forecasted the demand trend for 17 key NCD medications with an accuracy rate of 78% for the training set and 71% for the test set. Entering the month, year, region, and name of the necessary NCD medication allowed for this to be accomplished. The Random Forest model can be used to predict demand trends based on past consumption data. According to the study's findings, RF can improve operational management and health supply chain planning by increasing the precision of forecasting the demand trend for critical NCD medications.

In 2022, Mamoon Rashid et al. [11] suggested a medication prediction model that would help people take the right drug to treat a specific illness and act as an educational aid. The authors suggested storing medical datasets pertaining to medications in a Hadoop distributed file system. The suggested model will assist in examining the patient's medical history in order to identify any potential drug adverse effects. In order to make it easier for patients to locate local pharmacies that would have the necessary medication, the idea will also integrate Google Maps and weather.

In 2023, Nabizadeh A. H. et al. [12] showed that predicting the number and types of medicines consumed by each patient is done using different training groups. Proper timing of forecasting makes the predictions accurate, as the hospital management will have enough time to provide the required quantity of the most commonly used drugs. The performance of the algorithms was evaluated using a confusion matrix. The results showed that random forest had promising results in predicting the number of most frequently used drugs per month using two years of data (83.3% accuracy), while its accuracy in predicting drugs was 35.9%. After comparing the researchers' results with some previous studies, they found that they are consistent with those reported in (Mbonyin-Shuti et al., 2022). The authors of this study predicted the need for NCD medications using an RF model. According to their results, the RF model was able to predict medication usage one month in advance.

In 2025, Machado, D.M. et al. [13] researcher employ two methods: one that predictions for specific hospitals and another that integrates data from several hospitals in order to maximize resource allocation and logistics. This work explores machine learning for predicting hospital medication consumption. Based on consumption trends, researchers investigated manual pair clustering as well as K-means clustering. Manual clustering revealed particular combinations of medications with noticeably improved forecast accuracy, however K-means clustering produced no gains (e.g., Medicine 15 at Hospital 1: MAPE fell from 19.70% to 3.30%). However, not all hospitals consistently benefited from the unified approach (e.g., Medicine 9). This emphasizes the necessity of weighing possible losses in certain hospitals against accuracy gains in others. In general, the split approach's manual clustering shows promise if applys on larger datasets.

3 METHODOLOGY

The methodological basis of this study was based on a mathematical and statistical analysis of the field of pharmacy regulation and the work of domestic and foreign scientists in the field of forecasting drug provision in hospitals. The study was conducted using a dataset from the Institute of Fundamental Medicine and Biology in Kazan (Volga Region) Federal University (KFU), which consists of six types of drugs and five factors for the period (2015-2019). The data was analysed using Jupyter Notebook. No pre-processing steps were applied to address missing or inconsistent data by the researchers. Since the Institute of Fundamental Medicine and Biology of Kazan University prepares accurate data for scientific research purposes to support postgraduate students and data mining researchers. We relied on the available data without focusing on specific drugs or factors, as the main purpose of our study is to find an optimal prediction model that can be applied to any type of data in order to reduce the discrepancy between the actual consumption of materials (drugs in our case) and the estimated demand.

Within the framework of the current research program, three main analyses were used, namely:

1) the method of correlation and regression analysis;
2) time series forecasting analysis;
3) the method of retrospective analysis.

Multidimensional modeling is mainly based on correlation and regression analysis, which identifies patterns of demand for drugs and the influence of factors on the consumption of those drugs. Therefore, it is necessary to:

1) identify and analyze which factors have the greatest influence on the dependent variable (medicine in our study);
2) assess the degree of influence of the factors on the dependent variable;
3) determine the form of the relationship between the factors;
4) construct a forecasting equation for each medicine separately.

To determine the degree of influence of various factors on the need for medicines, the degree of correlation between factors and medicines, on the one hand, and the factors themselves, on the other hand. The Pearson correlation coefficient (r) was used, which is the most common way to measure linear correlation [14]. It is a number from -1 to 1 that measures the strength and direction of the relationship between two variables. Below is the formula for calculating the correlation intensity, which is also called the correlation coefficient r*XY*:

$$rXY = cov(X, Y)/\sigma X\sigma Y, \ \ rXY \in [-1;1]. \quad (1)$$

Where, $cov(_{X,Y})$: is the covariance between the two variables; σ_X: is the standard deviation of X; σ_Y- is the standard deviation of Y. Strong correlations between dependent variables (factors) and the independent variable (estimated) are necessary for methods of causal forecasting of demand. The relationship between the dependent variables should be known in order to exclude redundancy, which may affect the quality of the forecast [15]. The purpose of this analysis is to determine whether there is a relationship between the amount of drugs consumed and the selected factors using a correlation test. We assume that a relationship is significant if the significance level, or P-value, is less than 0.05.

Based on the values of the coefficients obtained, regression equations were constructed for each drug individually, which have the following general form:

$$Y = a + b1\ X1 + b2\ X2 + \ldots + bn\ Xn + e. \quad (2)$$

Where, Y: is the output or target vector $b_1 + b_2 + \ldots + b_n$ – are the regression coefficients, $X_1 + X_2 + \ldots + X_n$ – are independent variables or input variables, e: an error term.

Methods of forecasting time series. To obtain a preliminary assessment of the nature of drug consumption, a time-series methodology was used, which included the introduction of only one variable, mathematically expressed as y= f(t). Three time series forecasting methods were selected: linear trend analysis (LTA), simple exponential smoothing (SES), and double exponential smoothing (DES) [16], [17], and to obtain the most appropriate forecasting method among them, they were compared depending on the values of three error indicators (MAPE, MAD, and MSE). The method with the lowest absolute relative error is the most suitable way to more accurately predict future values compared to other methods [18]. The methods and indicators were calculated using the following models:

1) Linear trend analysis (LTA). This method is the simplest form of trend analysis, which assumes that the data follows a straight line with a constant slope and intersection [6]. The prediction equation of this method has the following form (3):

$$Ft = a + bt. \quad (3)$$

Where, Ft: is the predicted time period t, a: y- is the figurative interception, bt: is the slope of the trend. The slope (bt) indicates the amount of subtraction if the value (b) is negative or the amount of addition if the value (b) is positive, from one period to another. This method is suitable for time series in which the trend of the data demonstrates long-term stability.

2) Single exponential smoothing (SES). The method of simple exponential smoothing is the basic form of exponential smoothing [19], [20]. The future values of the time series are a weighted average of the previous values, and the weights decrease significantly as the distance from the present increases. There is only one parameter in the model, α, which controls the rate of weight loss. The lower the alpha value, the more weight is given to the previous values, and the forecast becomes smoother [20], [21]. The simplest form of the exponential smoothing is given by the following (4) [20]:

$$St = St\text{-}1 + \alpha\ (Xt - St\text{-}1). \quad (4)$$

Where, St: is a simple weighted average of the current observation. Xt, St-1: is the previous smoothed statistics, α: is the smoothing coefficient of the data; $0 < \alpha < 1$, t: is a period of time. The choice of the smoothing constant value is important for determining the performance characteristics of exponential smoothing. The lower the (α) value (i.e., close to 0), the slower the response. Large values of α (i.e., close to 1) cause the smoothed value to react quickly not only to real changes but also to random fluctuations [22].

3) Double exponential smoothing: holt's linear exponential smoothing is an extension that includes a directional component. It is assumed that the future values of the time series are a weighted average of the previous values and a linear trend. The model has two parameters, alpha and beta, which control the rate of weight loss and the rate of direction change, respectively [22], the standard form of this model:

$$S'(t) = \alpha Y(t) + (1 - \alpha)\ S'(t - 1). \quad (5)$$

Under simple exponential smoothing, let S(t+1) = S'(t) at this point. Then let S" denote the doubly smoothed series obtained by applying simple exponential smoothing (using the same a) to series S':

$$S''(t) = \alpha S'(t) + (1 - \alpha) S''(t - 1). \quad (6)$$

4) Finally, the forecast Y'(t + 1) is given by:

$$Y'(t + 1) = a(t) + b(t). \quad (7)$$

Where a(t) = 2S'(t) - S''(t) is the calculated level for the period t, b(t) = (a/(1-a)) (S'(t) - S''(t)) is the expected trend for the period t.

4 ANALYSIS OF THE ASSESSMENT OF THE DRUG NEEDS BASED ON STATISTICAL ANALYSES

4.1 Analysis of Correlation and Regression Between Drug Consumption and Factors Affecting It

To build models of drug consumption, we used a mathematical modelling method using multiple linear regression analysis. This method allows us to take into account the degree of influence of each factor on the consumption of each drug. The needs for the analysed drugs were encrypted and designated as "Y," whereas the factors influencing the drug consumption were encoded as "X.". The result of correlation and regression analysis showed that the correlation coefficients (r) between drug consumption and factors ranged from 0.014 to 0.968. These values indicate the presence of both a weak and strong relationship between drug consumption (Y) and factors (X). Correlation coefficients were selected, reflecting a strong correlation between drug consumption and factors affecting their consumption (|r| > 0.7). A direct correlation was established between the total numbers of disabled people among RA patients. (X5) and the need for 5 selected drugs: Rituximab 500 mg/ 50 ml (Y1), Infliximab 100 mg (Y2), Certolizumab 200 mg/ml (Y3), Tocilizumab 20 mg/ml (Y5), and Abatacept 125 mg (Y6), with correlation coefficients (r) of 0.949, 0.761, 0.954, 0.961, and 0.929, respectively. The volume of purchases of genetically engineered biological drugs (GIBP) for the treatment of prostate-specific antigen (PsA) (X3) has a weak relationship with the need for (Y2) (r = 0.130). There is a direct correlation between the amount of funding for the compulsory medical insurance program (CMI) for the so-called clinical and statistical groups (CSG) (X1) and the financial support of the program of provision of necessary medicines (ONLS) (X2) with the need for (Y6) (r = 0.879; r = 0.912), respectively. Factor (X2) also has a significant correlation with the need for Golimumab 50 mg/0.5 ml (Y4) (r = 0.890) and an average correlation with the need for (Y2) (r = 0.523). A weak negative correlation was found between the need for (Y2) and the price level for genetically engineered biological drugs (GIBP) (X4) (r = -0.014). Drug demand models were created based on coefficients obtained from correlation and regression analysis. The regression equations are presented in Table 1.

4.2 Accuracy Indicators

The main purpose of using more than one forecasting method was to compare the obtained estimates and determine which method provides the best prediction results based on three accuracy indicators: Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Deviation (MSD). The most suitable model is the one with the minimum accuracy indicators (MAPE, MAD, and MSD), which will select and use to estimate the needs of medicines in hospitals for the coming years [20].

Table 1: Multiple linear regression models for predicting drug demand and assessment their reliability.

Drugs	Model	Fisher (F)	
		Table_val of F_coeff	empirical_value
Rituximab 500mg/ 50ml	Y1 = 0.367* X5 + 13.643	6.39	1.11
Infliximab 100mg	Y2 = 0.003 *X2 + 0.002 *X4 + 0.159 *X5 - 97.211	6.39	1.00
Certolizumab 200mg/ml	Y3 = 0.513 * X5 - 51.110	6.39	1.10
Golimumab 50mg/0.5ml	Y4 = 0.003 * X2 + 0.0002 * X3 - 30.420	6.39	1.00
Tocilizumab 20mg/ml	Y5 = 0.606 * X5 -50.900	6.39	1.08
Abatacept 125mg	Y6 = 0.002* X1+ 0.270 * X5 - 81.469	6.39	1.00

Mean Absolute Percentage Error (MAPE). This indicator calculates the average ratio of the absolute error to the actual values. The method of its calculation is shown in (8):

$$MAPE = [\sum | (yt - yt^\wedge) | /n] * 100, (yt \neq 0). \quad (8)$$

Where y_t - is the actual value, $y_t^\wedge$- is the predicted value, and n - is the number of observations.

Mean Absolute Deviation (MAD). This indicator measures the accuracy of fitted time series values. It expresses accuracy in the same as the data. This helps to theorize the error also [21]:

$$MAD = \sum| (yt - yt^\wedge)| /n. \quad (9)$$

Where y_t - is the actual value, - $y_t^\wedge$- is the predicted value, and n is the number of observations.

Moving Standard deviation (MSD). This indicator has calculated using the same denominator, n, regardless of the model [23]. This is to enable comparison of MSD values across models. This makes MSD a more sensitive of a usually largest forecast error than MAD:

$$MSD = \sum|(yt - yt^\wedge)|2 /n. \quad (10)$$

Where y_t - is the actual value; $y_t^\wedge$ - is the predicted value, n - is the number of observations [20].

4.3 Evaluation of Forecasting Using Time Series Methods and Multiple Linear Regression Model

To determine the best model for forecasting drug needs, we compared the MLR model with three time series methods: (1) linear trend analysis (LTA), (2) single exponential smoothing (SES), and (3) double exponential smoothing (DES). First, we compared the forecasts of the three time series approaches with each other. Based on the three accuracy measures (MAPE, MAD, and MSD), we selected the method that best fit the data, which turned out to be DES as shown in Table 2. Next, we compared MLR model with DES as shown in Table 3.

Based on the data in Table 2, MAD was selected as the best indicator that describes those data. This indicator was chosen because it has the lowest error values compared with other two indicators.

When comparing the accuracy of the two models based on the data in Table 3, the MAPE model was chosen as the best indicator to describe the data. This indicator was chosen because it had the lowest error values compared to the other two. Based on the MAPE indicator values, the mean MAPE was calculated for all medications by using the MLR model and was 15.63 and 24.61 for DES. Therefore, MLR was selected as the best model for predicting future needs for all medications.

4.4 Retrospective Forecast of the Hospitals Future Drug Needs

The demand predictions produced by the two models MLR and DES were contrasted with actual consumption based on the findings of the two analyses mentioned above for the period 2015–2019. To verify the accuracy of the forecast, the discrepancy rate for each drug was calculated annually.

Table 2: Selecting the optimal time series method for forecasting drug needs according to the criterion of minimum accuracy of indicators.

Linear Trend Analysis (LTA)						
Measuring Accuracy	Y1	Y2	Y3	Y4	Y5	Y6
MAPE	10.9	26.3	47.1	34.4	51.7	8.4
MAD	64.3	18.6	17.2	9.9	33.4	36.5
MSD	88.1	32.9	184.2	26.4	160.6	6.8
Single Exponential Smoothing (SES)						
Measuring Accuracy	Y1	Y2	Y3	Y4	Y5	Y6
MAPE	146.1	146.9	161.2	172.1	131.3	122.8
MAD	24.1	6.8	11.3	7.6	9.2	7.5
MSD	793.3	74.7	244.2	73.9	186.4	80.6
Double Exponential Smoothing (DES)						
Measuring Accuracy	Y1	Y2	Y3	Y4	Y5	Y6
MAPE	15.3	25.4	38.1	16.3	39.7	12.9
MAD	9.95	3.6	20.7	2.0	16.7	3.8
MSD	199.9	6.5	1554.0	32.0	1355.5	223.8

Table 3: Comparison of MLR model with DES method according to the criterion of minimum accuracy of indicators.

Multiple Linear regression (MLR)						
Measuring Accuracy	Y1	Y2	Y3	Y4	Y5	Y6
MAPE	7.2	1.0	16.1	7.7	59.2	2.5
MAD	61.1	18.0	12.7	8.4	28.1	36.3
MSD	26.9	0.03	62.7	1.4	54.0	1.0
Double Exponential Smoothing (DES)						
Measuring Accuracy	Y1	Y2	Y3	Y4	Y5	Y6
MAPE	15.3	25.4	38.1	16.3	39.7	12.9
MAD	9.9	3.6	20.7	2.0	16.7	3.8
MSD	199.9	6.5	1554.0	32.0	1355.5	223.6

Table 4: Analysis of the forecast demand of drug calculated using (MLR and DES) models, in comparison with actual consumption for the period (2015-2019).

Y6	Y5	Y4	Y3	Y2	Y1	Drugs	
13	5	0	0	14	49	2015	Actual consumption
21	8	0	0	11	49	2016	
36	32	8	13	22	63	2017	
54	74	15	59	31	92	2018	
61	55	36	35	18	75	2019	
12.3	15.2	/	/	13.7	53.7	2015	Demand forecast calculated using MLR model
5.4%	204%			2.1%	9.6%	% discrepancies	
21.4	4.3	/	/	11.3	47.1	2016	
2.9%	46.3%			2.7%	3.9%	% discrepancies	
36.5	24.3	8.8	12.5	21.9	59.2	2017	
1.4%	24.06%	10%	3.9%	0.5%	6.03%	% discrepancies	
55.8	67.9	13.2	49.4	31.1	85.7	2018	
3.3%	8.2%	12%	16.3%	0.3%	6.9%	% discrepancies	
58.9	62.4	36.5	44.8	17.9	82.3	2019	
3.4%	13.5%	1.4%	28.0%	0.6%	9.7%	% discrepancies	
14	6	/	/	15	50	2015	Demand forecast calculated using DES
7.7%	20.0%			7.1%	2.0%	% discrepancies	
17	7.1	/	/	18.1	65.1	2016	
19.1%	11.3%			64.6%	32.9%	% discrepancies	
29	9.5	9	14	8.8	66.4	2017	
19.4%	70.3%	12.5%	7.7%	60.0%	5.4%	% discrepancies	
51	31.5	12	18.6	30.9	72.7	2018	
5.6%	57.4%	20.0%	68.5%	0.3%	21.0%	% discrepancies	
72	80	15	39.7	40.7	97.9	2019	
18.0%	45.5%	58.3%	13.4%	126.0%	30.5%	% discrepancies	
8.6	50.0	19.0	23.0	26.4	12.8	The average degree of discrepancy	

In addition, the average degree of discrepancy between actual consumption and estimated forecasts was calculated over the five years, as shown in Table 4. Discrepancy rates have been shown to be statistically significant for all drugs, with the exception of Y2 and Y5, where they actually exceed 25% [24].

5 RESULTS AND DISCUSSIONS

In this paper, we discussed the research methodology for multidimensional forecasting of hospital drug needs, which consists of five main steps related to each other.

In the first stage, a correlation analysis was carried out on a dataset to measure the power of the correlation between the medications and the relevant factors influencing the consumption of these medications. The results of the analysis showed a correlation, often strong (up to more than 70%) and statistically significant, between the amount of the drug consumed and some of the factors selected for the study.

In the second stage, a MLR model has been built to find a prediction equation for each drug separately. According to the forecast's equations derived using MLR model and based on the findings of the correlation and regression analysis (Table 1), it turned out that the percentages of discrepancy between demand forecasts and actual consumption are acceptable. Thus, it can be said that MLR model for predicting demand for medicines are very effective.

At the third stage, three prediction models using time series techniques have been built and compared based on three error indicators (MAPE, MAD, and MSD) to find the most appropriate model for predicting future medication needs. Based on the results of the accuracy indicators analysis, DES method has selected as the best model to describe drugs in our dataset because it has the lowest relative error rate compared to the other time series models (Table 2).

In the fourth stage, DES method was compared with MLR according to the same three error indicators (Table 3). The result showed that MLR is the best model to describe the drug data because it has the lower average error, since the mean MAPE values for MLR and DES were 15.63 and 24.61, respectively. Therefore, MLR can be used to predict future values more precisely than other models. Multiple linear regression's superiority over time series models in predicting drug demand may be due to several factors, including:

- The influence of independent factors on demand is stronger than the time factor; MLR model addresses multiple independent variables, while time series models rely only on historical time data;
- MLR model explains the influence of external factors better than time series models, which rely solely on trends and seasonality;
- The correlation effect among the factors (such as increased demand in winter and flu outbreaks), while time series models do not address these complex relationships;
- Ease of interpretation and applicability. Regression coefficients provide a quantitative indication of the impact of each variable. For example, a 10% increase in the number of patients leads to a 7% increase in drug demand.

In the fifth stage, retrospective forecasting was used to compare the actual consumption with estimated forecasts for the period 2015–2019. To verify the accuracy of the forecast, the discrepancy rate for each drug was calculated annually. In addition, the average degree of discrepancy between actual consumption and estimated forecasts was calculated over the five years, as shown in Table 4. Discrepancy rates have been shown to be statistically significant for all drugs, with the exception of Y2 and Y5, where they actually exceed 25%. This may be due to high discrepancies for some years, resulting from significant discrepancies between actual consumption and projected demand in one or both models. It could also be due to several other factors, both internal to the supply or distribution system and external factors, such as sudden changes in demand.

6 CONCLUSIONS

This study has accomplished the following:

- A systematic approach based on a multi-factor mathematical model has been built;
- An optimal forecasting model among all the methods utilized in this study has been selected to predict the future needs of hospitals for medicines in the coming years;
- A discrepancy rate for each drug annually has been calculated to check the accuracy of the forecast.

As a result, three statistical analyses were performed: correlation and regression analysis, time series analysis, and retrospective forecast analysis between estimated demand and actual consumption. The results have shown the following:

A) The model of MLR is the optimal model for predicting the needs of medicines. Since the accuracy values were lower in MLR compared to the DES method. Thus, a multiple linear regression model can:
   1) Improve hospital medication management if implemented correctly (e.g., inventory management, reducing waste, and minimizing surpluses and shortages);
   2) Enhancing medical quality and safety by ensuring the availability of vital medicines and reducing errors and costs;
   3) Reducing medical errors: when stocks are organized, the likelihood of inappropriate substitute medications being dispensed due to shortages is reduced;
   4) Adapting to crises, especially in emergencies (such as pandemics), where model equations can be adjusted to include new variables (such as disease prevalence).

B) In general, the applied methods provided an adequate assessment for our dataset. Based on the results of the analysis of methods on the dataset, we can make sure that our predictive model and all methods used in our study can be applied to other datasets from different countries if the data is accurate and sufficient;

C) It is also possible that industrial enterprises will benefit from the proposed forecasting model in this study when it comes to predicting spare parts.

The study had some limitations because of the limited historical data period because of the missing data for some prior years, in addition, COVID-19 pandemic made it harder to estimate and predict these events, therefore pharmacological data for years after 2019 were excluded because they do not reflect the new reality.

7 FUTURE WORK

Future prospects for this research include the following:

- Develop a forecasting model by combining multiple linear regression and time series techniques (a hybrid model) to improve forecast accuracy. The hybrid model captures the linear relationships between demand and influencing variables and addresses time trends not explained by regression;
- Test the robustness of forecasts over longer time periods by adding more drugs and independent variables.

REFERENCES

[1] W. H. Hameed, "Forecasting drug needs and quantification tools in healthcare institutions," Quantum Center, pp. 3-12, 2021.

[2] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," Computers & Industrial Engineering, vol. 143, p. 106435, 2020.

[3] E. Belyi, P. J. Giabbanelli, I. Patel, et al., "Combining association rule mining and network analysis for pharmacosurveillance," The Journal of Supercomputing, vol. 72, no. 5, pp. 2014-2024, 2016, [Online]. Available: https://doi.org/10.1007/s11227-016-1714-y.

[4] W. S. M. Al-Lami, "An economic study of some factors affecting the fiscal deficit in the Iraqi budget (2003-2016)," in Economics of Enterprises, Regions, and Countries: Current Issues and Modern Aspects, Penza: Library.Ru, 2021, pp. 19-22.

[5] H. Mohammed, I. Nazeeh, W. Alisawi, Q. Kadhim, and S. Ahmed, "Anomaly detection in human disease: a hybrid approach using GWO-SVM for gene selection," Revue d'Intelligence Artificielle, vol. 37, no. 4, pp. 913-919, 2023, [Online]. Available: https://doi.org/10.18280/ria.370411.

[6] H. A. A. Mohammed, S. T. Ahmed, R. M. H. Zaki, and Q. K. Kadhim, "Detection and localization of wrist fractures in x-ray imagery using deep learning teaching," Review of Computer Engineering Research, vol. 11, no. 3, pp. 85-98, 2024.

[7] A. Z. Kamaeva, "Multivariate forecasting of the need for anti-asthmatic drugs used in hospital settings (on the example of the Republic of Tatarstan)," 2013, p. 340.

[8] M. R. Safiullin, "Methodological approach to multivariate forecasting of hospital need for antibacterial drugs: on the example of destructive pancreatitis," Moscow: Russian State Library (RGB), 2018.

[9] O. V. Zhukova, "Methodology of assessment and forecasting of drug consumption based on clinical and economic analysis taking into account the principles of evidence-based medicine," 2021.

[10] F. Mbonyinshuti et al., "Application of random forest model to predict the demand of essential medicines for non-communicable diseases management in public health facilities," Pan African Medical Journal, vol. 42, p. 89, 2022, [Online]. Available: https://doi.org/10.11604/pamj.2022.42.89.33833.

[11] M. Rashid, V. Goyal, S. A. Parah, and H. Singh, "Drug prediction in healthcare using big data and machine learning," 2022.

[12] A. H. Nabizadeh, M. Ghaemi, and D. Gonçalves, "Predicting the pharmaceutical needs of hospitals," 2023, [Online]. Available: https://doi.org/10.21203/rs.3.rs-3263123/v1.

[13] D. M. Machado, Z. Yahouni, G. Alpan, and D. Koala, "Clustering analysis for forecasting medicine consumption," in Innovative Intelligent Industrial Production and Logistics, IN4PL 2024, Communications in Computer and Information Science, vol. 2372, Springer, Cham, 2025, pp. 10, [Online]. Available: https://doi.org/10.1007/978-3-031-80760-2_10.

[14] D. Koala, Z. Yahouni, G. Alpan, and D. S. Mohand, "Correlation analysis of factors impacting health product consumption in French hospitals," IFAC-PapersOnLine, vol. 55, no. 10, pp. 895-900, 2022, [Online]. Available: https://doi.org/10.1016/j.ifacol.2022.09.415.

[15] T. Falatouri, F. Darbanian, P. Brandtner, and C. Udokwu, "Predictive analytics for demand forecasting–a comparison of SARIMA and LSTM in retail SCM," Procedia Computer Science, vol. 200, pp. 993-1003, 2022.

[16] H. W. Hameed, "Comparing linear trend analysis and double exponential smoothing techniques for predicting the demand for chronic medications," Automation in Industry, vol. 11, 2023, [Online]. Available: https://doi.org/10.25728/avtprom.2023.11.04.

[17] W. H. Hameed and E. M. Jassim, "Efficient time-series clustering of drugs demand to inventory optimization in health centers," Scientific and Educational Journal for Students and Teachers ‘StudNet’, no. 6, pp. 6905-6916, 2022.

[18] D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," Management Science, vol. 66, no. 3, pp. 1025-1044, 2020.

[19] P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 3rd ed., New York: Springer, 2016, pp. 449.

[20] M. Zellner, A. E. Abbas, D. V. Budescu, and A. Galstyan, "A survey of human judgement and quantitative forecasting methods," Royal Society Open Science, vol. 8, no. 2, p. 201187, 2021.

[21] D. Koala, Z. Yahouni, G. Alpan, and Y. Frein, "Factors influencing drug consumption and prediction methods," in CIGI-QUALITA: Conférence Internationale Génie Industriel QUALITA, Grenoble, France, May 5-7, 2021, [Online]. Available: https://doi.org/10.48550/arXiv.2109.11779.

[22] F. Benhamida, O. Kaddouri, T. Ouhrouche, M. Benaichouche, D. Casado-Mansilla, and D. Lopez-de-Ipina, "Demand forecasting tool for inventory control smart systems," Journal of Communications Software and Systems, vol. 17, no. 2, pp. 185-196, June 2021, [Online]. Available: https://doi.org/10.24138/jcomss-2021-0068.

[23] A. I. Taloba, R. M. Abd El-Aziz, H. M. Alshanbari, and A. H. El-Bagoury, "Estimation and prediction of hospitalization and medical care costs using regression in machine learning," Journal of Healthcare Engineering, vol. 2022, p. 7969220, 2022, [Online]. Available: https://doi.org/10.1155/2022/7969220.

[24] L. V. Kobzar and T. A. Safronova, "The study of hormonal drug consumption by the method of expert assessments," Pharmacy, no. 4, pp. 5-9, 1981.