

Examination and Forecasting of Drug consumption Based on Recurrent Deep Learning

Hadab Khalid Obayes, Nabeel Al – A'araji, Eman AL-Shamery

Abstract: *The provision of pharmaceutical drugs in quantities appropriate to consumption is an important point in the pharmaceutical industry and storage of medicines, as the production of large quantities of unnecessary drugs leads to a longer storage of drugs. Meanwhile most medicines have a short shelf life. When the amount of production is less than required, this affects the satisfaction of the customer and the marketing of the drug. Time series analysis is the appropriate solution to this problem. Deep learning has been adapted for the purpose of time series analysis and a prediction of the required quantities drugs. A recurrent neural network with Long-Short Term Memory LSTM has been used by deep learning. The proposed methodology is based on the seasonal number of prescription required quantities with the number of quarters as indicators. The aim of the research is to forecast the drugs amount needed for one year. The proposed method is assessed using two types of evaluation. The first one is based on MSE and the visualization of the actual data and forecasted data. The proposed method has reached a low value of MSE and the visualization graph is semi-identical, whereas the second evaluation method compares the result of the proposed method with traditional forecasting method. Multiple linear regression is a traditional prediction method used with the data set, whose results are relatively good and promising compared to the results of the traditional method.*

Keywords : *Drugs consumption forecasting, DNN, LSTM, Recurrent long-short term memory-deep learning based drug analysis and forecasting, RNN.*

I. INTRODUCTION

Forecasting the quantity of drugs is a process used to determine the amount of product required for the purpose of purchase. Specifically, however, the quantification process involves assessing not only the quantities required for a given item but also the financial means required to purchase the item. Accurate drug identification requires different pieces of information, including the Essential Medicines List (EML), average consumption, total prescriptions, minimum and maximum inventory levels, inventory replication, and epidemiological information. The required quantities of drugs should always be chosen in light of the resources and information available. If no reliable information on prior consumption or morbidity is available, the use of the drug can be extrapolated from data from other facilities, regions or countries [1]. This study employs real-world data, which has been configured and processed to be appropriate for the

forecasting process. Forecasting future observations is one of the most important tasks of time series analysis[2]. The univariate time series is a sequence of observations of a single random variable at different moments over a set of uniform time span. In fact, in forecasting the time series, the magnitude of the error increases over time, as uncertainty increases with expectations[3] Therefore, a suitable technique for the prediction process should be selected in proportion to the available data. There are many ways that can be applied in the prediction process. In the proposed paper, the deep learning technology is used for the purpose of forecasting. Deep neural networks have been extensively studied in related fields and have had exceptional effects on solutions to a wide range of problems. RNN models, for example, have become more popular in natural NLP research. A special case of RNN in particular, namely the Long Short-Term Memory (LSTM)[4] which has provided a distinct solution to many forecasting problems.

II. RELATED WORKS

This section summarizes the recent and popular noteworthy studies related to this problem. The authors employed in [1] methods of time series forecasting to predict the future consumption. The paper focuses on finding the most effective indicators on the quantitative of the drug consumption. Consequently, the authors find 27 indicators for the qualitative and five indicator for the quantitative. Monthly use indicators for the drug, seasonal changes, and data distribution were evaluated. These indicators were the most important ones that affect the provision of the number of medicines required annually. In [5] the researchers present a Smart Sales Prediction Analysis to forecast the requirement of the Pharmaceutical distribution companies with a system that depends on real-time application, the forecasting time series method (ARIMA) and neural network. The hybrid method presented represents a real-time data mining methodology with a combination of a predetermined factor and a new factor. The results of the proposed method demonstrated that it is a real-time approximation forecasting method.

Continuously with this research scope, Oscar Chang and et al, have worked in a deep neural network in [6] to forecast the weekly data of pharmaceutical products. In this paper, an auto-encoder is used as data abstraction for the shallow neural network. The abstracts are used to train the second shallow network, which in its turn, specializes in one week. This recent information is entered with abstractions into the third shallow network for the purpose of predicting its own week.

Revised Manuscript Received on September 25, 2019

Hadab Khalid Obayes, College of Information Technology, University of Babylon, Iraq E-Mail: hedhab@uobabylon.edu.iq

Nabeel Al – A'araji, Ministry of Higher Education, Iraq

Eman AL-Shamery, College of Information Technology, University of Babylon, Iraq

The system can produce a good and stable weekly forecast.

The researchers adopted in [7] a new methodology for pre-evaluating the history of release and evolutionary behavior in a homogeneous category of pharmaceutical drugs. The methodology is based on dynamic meta-analysis, addressing the problem of predicting new products of drugs. The properties of drug spread operate in a clear ranitidine class, whereas the meta-analysis is based on the dynamic evolution study of propagation parameters and the observation of drug sales information. This method is used to predict the behavior of new drug labeling. The authors employed in [8] methods of time series forecasting to predict the future consumption and purchase of the drug RAPILYSIN LYPDINJ 2X1.16G/VIAL (RL). They used Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA). Using a very short time series of drug consumption, the MSE is used for determining the accuracy of the forecasting method.

III. TIME SERIES DATA AND FORECASTING PROCESS

Time series data is recording of observations of a particular subject over multiple time periods. The time series can be described as a set of observable and evenly spaced data recorded at regular intervals. The time series analysis includes methods for analyzing data in order to extract utility and meaning from statistics and give a description for the data. In fact, the time series forecasting process is the use of a model to predict future values based on ones that have been previously observed. Regarding the methods related to this process, forecasting only depends on past values and assumes that the factors that have affected the past and the present will continue to affect the future. Meanwhile, if the future values of a time series can be predicted from their previous values, the time series is inevitable (deterministic). If the future of a time series can be partially determined by previous values, this implies that the time series is stochastic or random[8]. Linear models, such as the (ARMA) and (ARIMA), are successful linear methodologies, but their predictive power is constrained by linear behavior. As a consequence, this will not always be acceptable, especially for application in predicting nonlinear time series. Accordingly, many prediction techniques have been developed over the past decades. Each technique has advantages and disadvantages when compared to other methods. There are factors that have emerged in the last century which led to the development of time-series forecasting. On the one hand, developments in computer science resulted in the use of more complex algorithms. Whereas, on the other hand, the development of machine learning techniques, such as neural networks, has led to significant advances in time series. In fact, neural networks are a very well-known technique in predicting the time series after being confirmed as an effective way of predicting the handling of nonlinear input and output variables, with the ability to estimate any function in specific circumstances. In other words, if the data are affected by nonlinear behavior, linear methods are unable to model nonlinear latent relationships. Recently, neural networks used in the prediction of time series in particular have significantly improved in terms of its performance. The popularity of neural networks is due to their high ability to

emulate a wide range of basic nonlinear behaviors and to cover a wide range of fields. Therefore, many studies have applied neural networks to predict time series and compare them with traditional methods. However, the results were not well-defined for the purpose of adopting a particular methodology[9]. In fact, neural networks have developed rapidly in deep learning, which is considered to be a revolution in machine learning and used in all fields such as speech recognition, natural language processing, prediction, classification, image and video processing[10].

IV. DEEP LEARNING TECHNIQUE

Deep learning is the evolving form of traditional neural network technology. It can be described as a neural network that consists of many layers[11]. The great development of computer science has helped to build the deep learning of multilayered neural networks. Therefore, the discovery of nonlinear patterns is one of the most important deep learning capabilities[12]. Deep learning has been used extensively in speech, signal, image, video, and text exploration. The most recent performance improvement was found to be more than 30%. The power of deep learning lies in the automatic learning feature of massive data sets [13]. It facilitates the extraction of properties and allows for optimal learning processes[14] [10]. However, the main algorithms of deep learning used in many machine learning applications are convolutional neural network, deep belief network, deep neural network, and recurrent neural network[12]. Recurrent neural networks are a class of interconnected models that have an internal cell or short-term memory because of recurrent feedback connections and are suitable for dealing with series problems such as speech classification, prediction, and generation[15]. The use of recurrent neural networks gave promising results in machine learning tasks. In contrast to the results presented through neural networks with feedforward, the network can deal with sequence data in varying lengths by relying on a repeated hidden cell that is activated each time based on the one the one that preceded [16]. In the proposed paper, the RNN (LSTM) is used to predict of the amount of drug brands used based on its past information

STM (Long-Short Term Memory)

RNN network proved very successful in various applications such as speech recognition, translation and image captions. The reason for the success of this network lies in the use of LSTMs: long-term memory, a special type of recurrent neural networks is LSTM. Figure (1) shows the LSTM cell diagram.

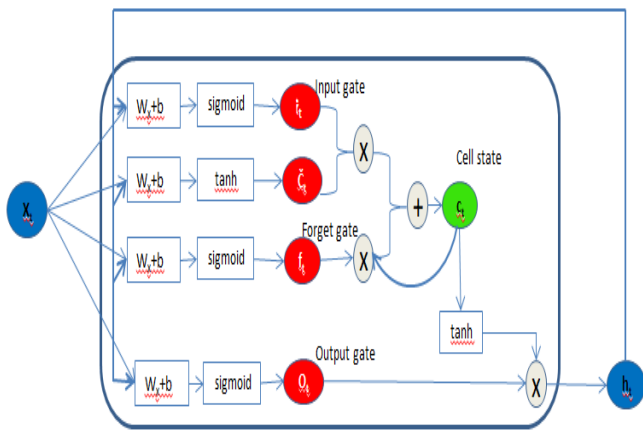


Figure 1: LSTM cell diagram

The LSTM unit differs from the traditional recurrent unit, maintaining the current memory $c_t \in R^n$ at time t . The input at time t is x_t, h_{t-1}, c_{t-1} , and the output is h_t, c_t . These can be updated by means of the following equations [RNN-write]:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$g_t = \tanh(W_g x_t + U_g h_{t-1} + b_g) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Where $\sigma(\cdot)$ the logistic sigmoid function and \odot is the operator denoted to the element-wise vector product. At each time step t , there are

- Input gate i_t ,
- forget gate f_t ,
- output gate o_t ,
- memory cell c_t ,
- hidden unit h_t, h_0

and c_0 can be initialized to 0 and the parameters of the LSTM are W, U, b [LSM]. Figure (2) shows the RNN with LSTM diagram.

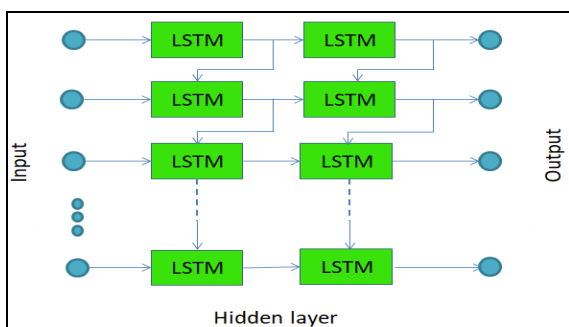
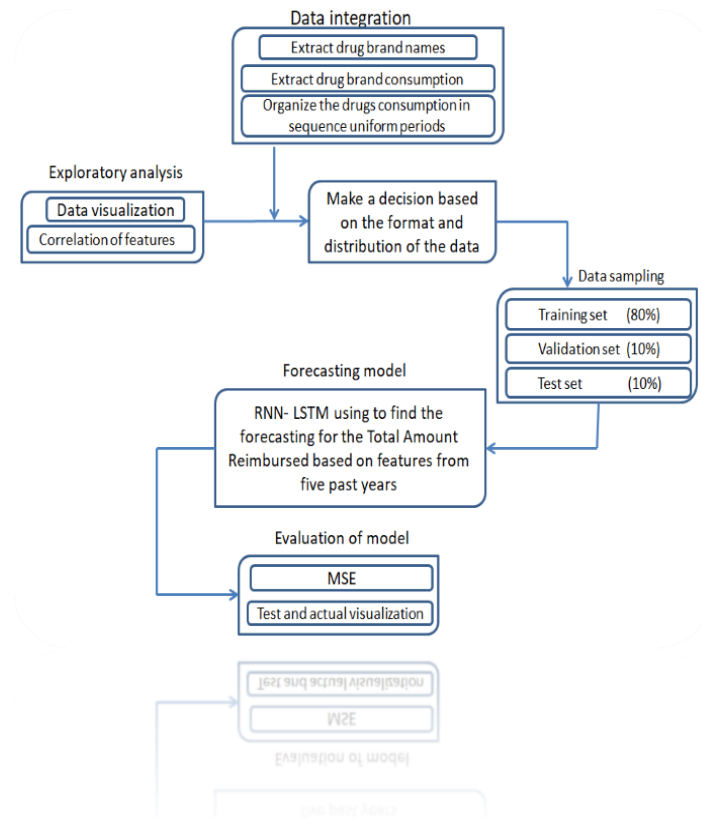


Figure 2: RNN with LSTM

V. THE PROPOSED METHODOLOGY

This section explains the methodology steps in detail: As shown in Figure (3), the methodology steps consists of data

integration, exploratory analysis, data sampling, forecasting model and the evaluation of model.



5.1. Data integration: For the purpose of predicting the future need for drugs, the utilization data for these drugs should be available in advance. The proposed research was treated with a real-world data collection representing the use of drugs in the United States over a period of five years. This step included three sub-steps for configuring data:

Extract drug brand name: A collection of brand names for specific drugs has been extracted from the website specializing in drug and brands used in USA" www.drug.com". 135 drug brand names are extracted.

Extract drug brands consumption: At this point, the real world data collection, which represents the consumption of medicines in the United States, has been loaded for the past five years. The consumption restrictions for a group of drugs extracted at a previous point are extract from this dataset.

Organized the drug consumption in sequence uniform period: The real-world data set is unordered but contains important and valuable data, which makes it an essential step. It is known that the analysis of time series requires past data at fixed time intervals. In this step, the number of prescriptions and quantity required for each drug brand is collected over the quarter. In other words, the extracted data represents the total number of prescriptions and quantity required for each drug over four quarters of the year and for the previous five years. This is the most significant step in the

search as the results of the forecasting process depend heavily on the organizer of the data. Otherwise, the process of prediction will be unsuccessful.

Data Normalization: this process is used to put the extracted data of drugs in a range of (0-1). Equation (7) is used to normalize the data set.

$$normalized = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (7)$$

5.2. Data exploratory analysis: This step is used to understand the nature of the data used in the research, and includes the following steps

Data visualization: Data is visualized at this stage. Figure (4) depicts the data of the required quantities over six years. It is possible to observe that the data does not have a specific format, as it is unstable. For some drugs, the total consumption has decreased, whereas others have increased over the years. Some drugs started to appear gradually over the years, while others disappeared. This seasonal pattern seems to be weak and unstable for drugs.

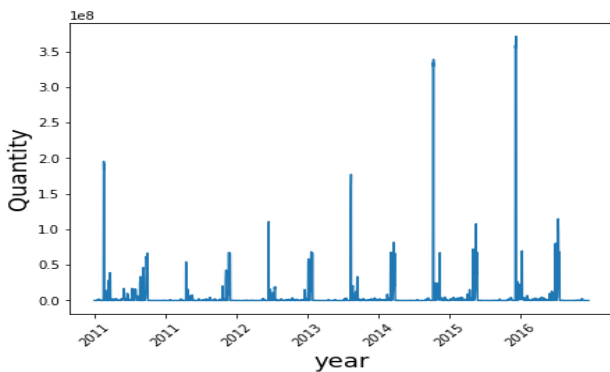


Figure 4: Quantity Required for six year from (2011-2016)

Figure (5) shows the total number of prescription for each drug. The data is organized into the total number of prescriptions: one each quarter, four quarters for each year. The data is integrated for six year from (2011-2016).

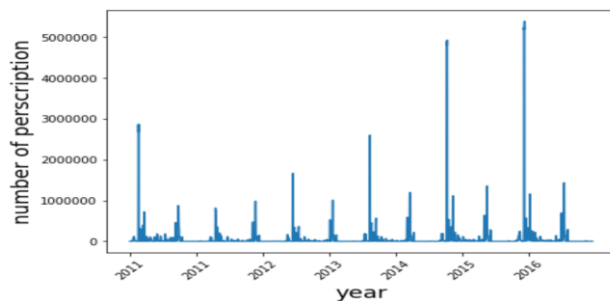


Figure 5: number of prescription for six year from (2011-2016)

Features correlation: With the aim of analyzing available data and after its visualization, the correlation between the number of prescriptions and the quantity required of the drug is calculated, eventually resulting in a correlation between them up to more than 90%, as Table (1) shows. This means

that when the number of prescriptions increases, the required amount of the drug increases accordingly, and vice versa.

Table 1: Correlation of quantity required & number of prescription

| | quantity required | number of prescriptions |
|-------------------------|-------------------|-------------------------|
| quantity required | 1.00 | 0.96 |
| number of prescriptions | 0.96 | 1.00 |

5.3. Data sampling: The training process is one of the most important stages in machine training. In order for the training to be successful, the data should be separated into the training set, the validation set, and the test set. The dataset obtained from the data integration step is as follows (135 drugs; the data for each drug was arranged over four quarters per year for a period of six years (2011-2016)). The training set is (80%), (10%) for the validation set, and (10%) for the test set. The data for the first five years (2011-2015) represent the deep network inputs that form the time series, whereas the data for the last year (2016) represents the desired value for the deep network, as the (LSTM) is a supervised network.

5.4. Forecasting model, Figure (6) shows the deep LSTM network designed for drugs forecasting in the future.

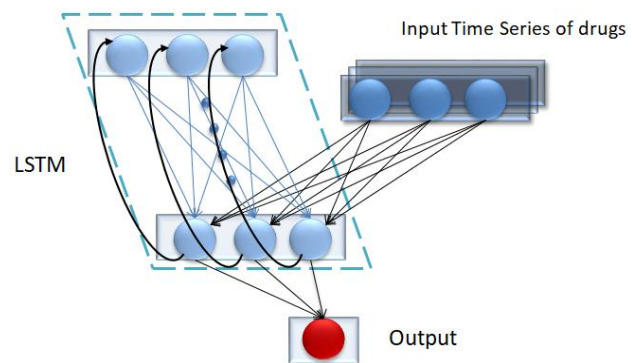


Figure 6: deep neural network (LSTM)

Deep drugs forecasting descriptive algorithm

Let:

t: t=0 to T where T represent the time series (5 previous years).

N: represent the number of LSTM blocks.

M: represent number of inputs (number of drugs is 135).

X_t: represent the input vector of x at time t.

H_{t-1}: previous cell output.

C_{t-1}: previous cell memory.

H_t: current cell output.

C_t: current cell memory.

the following weights for an LSTM layer:

- Input weights: W_i, W_f, W_o ∈ R N×M
- Recurrent weights: U_i, U_f, U_o ∈ R N×N
- Bias weights: b_i, b_f, b_o ∈ R N

The vector formula for LSTM compute as

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$g_t = \tanh(W_g x_t + U_g h_{t-1} + b_g) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Where σ is logistic sigmoid ($\sigma(x) = \frac{1}{1+e^{-x}}$)
And $g(x) = h(x) = \tanh(x)$

VI. EVALUATION OF FORECASTING MODEL

When constructing any model, this model should be evaluated. This is accomplished through:

Mean squared error one of the most famous error measures is used to evaluate the accuracy of model.

$$MSE = \sum_i \frac{(d_i - y_i)^2}{n} \quad (8)$$

Where desired values (d_i), output of the network (y_i) and n is number of input pattern

Visualization of the test set (desired values (d_i)) and the actual values (output of the network (y_i)).

VII. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the result of the proposed system are shown. The data extracted for (135) drugs divided into four quarters over a timespan six years (2011-2016). The input as a training set is the data for the years (2011-2015) The desired or the forecasted values were the data in 2016. Figure (7) shown a sample of this data.

| year | drug_name | Requierd_amount | number_perscription | Quarter |
|------|------------|-----------------|---------------------|---------|
| 2011 | JANUMET | 378075.000 | 6516.0 | 1 |
| 2011 | JANUMET | 402860.000 | 6973.0 | 2 |
| 2011 | JANUMET | 436082.000 | 7525.0 | 3 |
| 2011 | JANUMET | 508824.000 | 8750.0 | 4 |
| 2011 | TOUJEO SOL | 0.000 | 0.0 | 1 |
| 2011 | TOUJEO SOL | 0.000 | 0.0 | 2 |
| 2011 | TOUJEO SOL | 0.000 | 0.0 | 3 |
| 2011 | TOUJEO SOL | 0.000 | 0.0 | 4 |
| 2011 | LANTUS SOL | 1182383.220 | 64643.0 | 1 |
| 2011 | LANTUS SOL | 1396380.667 | 76826.0 | 2 |
| 2011 | LANTUS SOL | 1489961.410 | 81680.0 | 3 |
| 2011 | LANTUS SOL | 1434770.520 | 78586.0 | 4 |
| 2011 | LANTUS | 1623162.610 | 101557.0 | 1 |
| 2011 | LANTUS | 1781418.080 | 106992.0 | 2 |

Figure 7: Sample of extracted data

Influential Parameters: Results and parameters are reviewed in this section.

- *Shuffling the dataset:* The data set has been shuffled before separated into the training set, validation set, and the test set. The shuffling process results in a randomly arranged data set, which prevents the bias state during network training.

Batch Size Selection: In fact, choosing the size of the batch of data to be processed greatly affects the outcome of the

training. The increase in batch size is often more efficient in terms of computing but can lead to less accurate results than choosing a small batch size. On the contrary, the deep training technique usually prefers a small batch size. Processing a small batch of data means that the gradient is very close to its actual value and in turn leads to an increase in computing. Table (2) shows some experiments of the selection of batch size and MSE.

Table (2) show batch size and MSE

| Batch size | MSE | No.of epoch |
|------------|--------------|-------------|
| 10 | 0.002 | 200 |
| 100 | 0.006 | 200 |
| 200 | 0.017 | 200 |

Figure (8) shows the error decay with batch size where it reaches 0.002 with batch size 10.

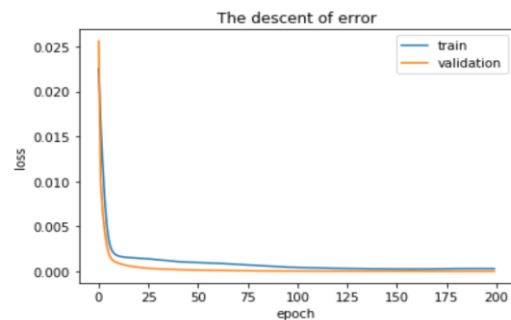


Figure 8: shows the batch size=10 and MSE=0.002

Figure (9) shows the batch size=100 and MSE=0.006

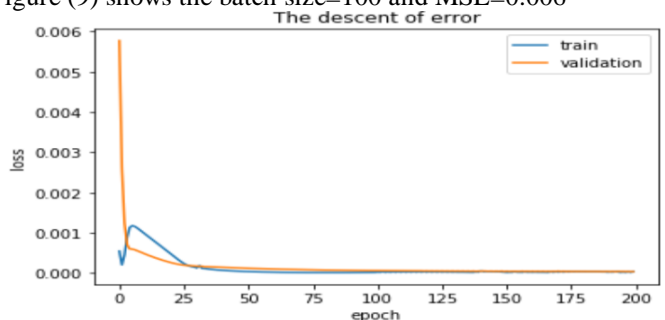


Figure 9:batch size=100 with MSE= 0.006

Figure (10) shows the batch size=200 and MSE=0.017

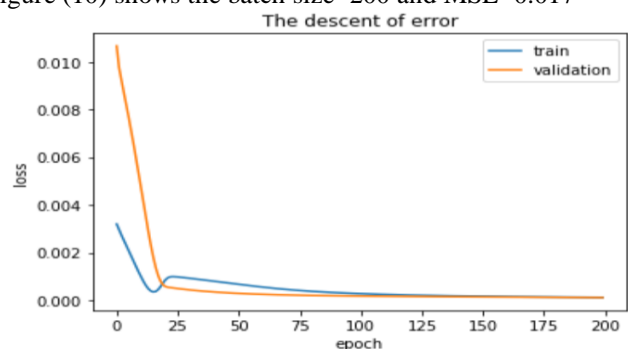


Figure 10: batch size=200 with MSE= 0.017

The structure of LSTM-DNN: Most of the structure of the network has been tested and the following visualization shows the real value (desired) and the prediction value (actual) through the network. Figure (11) presents the performance of the LSTM-DNN with the number of cell (20), a batch size of 10, and an epoch at (200). The resulting MSE was (0.002). This structure is the most preferable one. The error reached the minimum value of (0.002) and the visualization of the forecasted value and the real world value was identical.

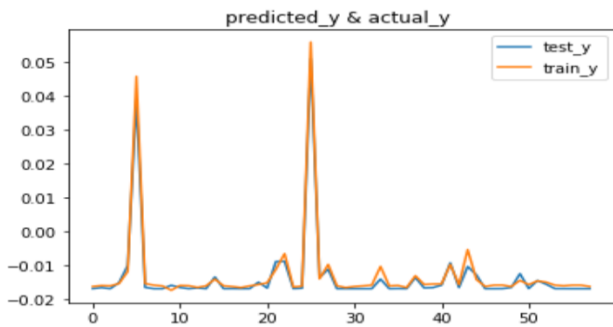


Figure 11: the identical forecasting and real values with MSE=0.002

Figure (12) shows the performance of the DDF with the number of cell (50), a batch size of 100, and an epoch at (200). The resulting MSE is (0.005).

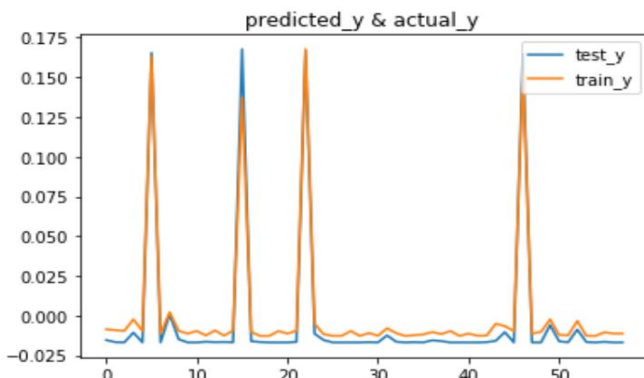


Figure 12: forecasting with error (0.005)

Figure (13) shows the performance of the DDF with the number of cell (70), a batch size of 70, and an epoch at (200). The resulting MSE was (0.017).

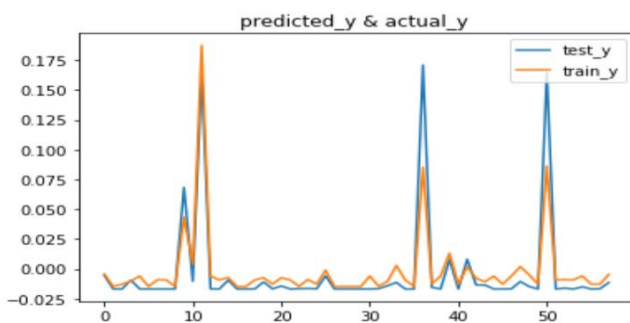


Figure 13: forecasting with error (0.017)

VIII. EVALUATION SYSTEM

For the purpose of evaluating the results obtained through the application of the proposed method, an Applied Linear regression has been used on the data. Regression analysis can be expanded to include more than one independent variable. Regressions involving more than one independent variable are referred to as multiple regression. Multiple regression was applied on the data of drugs consumption of five year (2011-2015) to predict the consumption for the year (2016). The result is shown in Figure (14), with an MSE value of (0.0312).

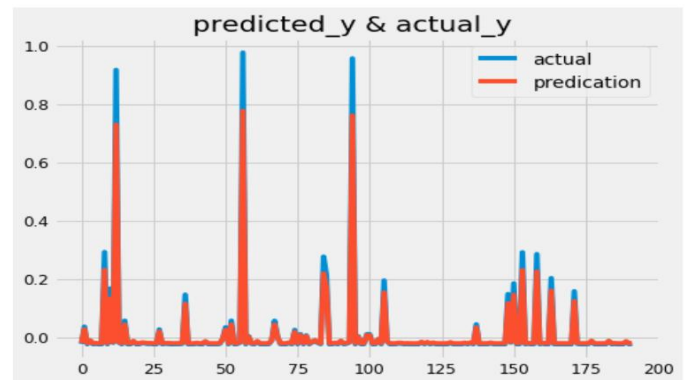


Figure 14: multiple regression forecasting result

IX. CONCLUSION

The proposed paper presented an automated system from the beginning of its work to its end. The system consists of five steps, each of which has achieved the desired goal clearly. First, the data was extracted from real-world data. The most important step at this stage is the collection of data in a format suitable for time series. The duration of the test was six years, five of which represented the input data and the last year represented the forecasted data. In the second stage, the data is clarified by means of visualization. The importance of this stage is to identify the distribution of the data and understand its behavior. Third, the separation of data into training data and test data. Fourth, a deep network of LSTM type is designed, which gave good results. The research reached the lowest error value after testing more than the design of the deep network. Finally, for the purpose of displaying network performance, the result of deep network and real-world data are compared through visualization. The system has provided integrated as well as promising results for the purpose of predicting the future needs of medicines in order to control the quantities of medicines produced according to the supply and demand theory.

REFERENCES

- [1] M. J. Iqbal, M. I. Geer, and P. A. Dar, "Evaluation of Medicines Forecasting and Quantification Practices in Various Evaluation of Medicines Forecasting and Quantification Practices in Various Public Sector Hospitals Using Indicator Based Assessment Tool," J. Appl. Pharm. Sci., vol. 7 (12), no. December 2017, pp. 072–076, 2018.
- [2] P.-A. Cornillon, W. Imam, and E. Matzner-LZber, "Forecasting time series using principal component analysis with respect to instrumental variables Forecasting time series using principal component analysis with respect to instrumental variables," Comput. Stat. Data Anal., vol. 52, no. July, pp. 1269 – 1280, 2008.

- [3] I. A. Gheyas and L. S. Smith, "A Neural Network Approach to Time Series Forecasting," Proc. World Congr. Eng., vol. II, no. 1, pp. 1–5, 2009.
- [4] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks," SIGIR, no. July, 2018.
- [5] K. N. Mahajan, "Business Intelligent Smart Sales Prediction Analysis for Pharmaceutical Distribution and Proposed Generic Model," vol. 8, no. 3, pp. 407–412, 2017.
- [6] Y. Tech, "A Deep Learning Algorithm to Forecast Sales of Pharmaceutical Products," no. September, 2017.
- [7] R. Guseo et al., "Pre-launch forecasting of a pharmaceutical drug," Int. J. Pharm. Healthc. Mark., vol. 11, no. 4, pp. 412–438, 2017.
- [8] A. Papan, D. Folinas, and A. Fotiadis, "Forecasting the consumption and the purchase of a drug," Int. Conf. SUPPLY Chain. Funct., vol. 2, 2016.
- [9] N. K. Zadeh, M. M. Sepehri, and H. Farvaresh, "Intelligent Sales Prediction for Pharmaceutical Distribution Companies : A Data Mining Based Approach," Hindawi Publ. Corp., vol. 2014, 2014.
- [10] T. Pham, T. Tran, and D. Phung, "Predicting healthcare trajectories from medical records : A deep learning approach Predicting healthcare trajectories from medical records : A deep learning approach," no. October, 2017.
- [11] E. AL-Shamery and A. AL-haq, "An Optimized Feed Forward Neural Network for Reducing Error Based Stoch Market Prediction," J. Eng. Appl. Sci., vol. 13, no. Special Issue 5, pp. 4616–4621, 2018.
- [12] F. Jiang et al., "Artificial intelligence in healthcare : past , present and future," stroke Vasc. Neurol., vol. first publ, 2017.
- [13] E. AL-Shamery and A. AL-haq, "Enhancing Prediction of NASDAQ Stock Market Based on Technical Indicators," J. Eng. Appl. Sci., vol. 13, no. Special Issue 5, pp. 4630–4636, 2018.
- [14] H. Khalid Obayes, N. Al – A'araji, and E. AL-Shamery, "Deep Neural Network for Enhancing Drug-Utilization Clustering," Int. J. Eng. Technol., vol. 8, pp. 290–298, 2019.
- [15] J. Koutn and K. Greff, "A Clockwork RNN," arXiv, vol. 1402.3511v, 2014.
- [16] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Aaron, 2016.