PREDICTING MEDICINE DEMAND USING DEEP LEARNING TECHNIQUES

Bashaer Abdurahman Mousa[1]

[1]Informatics Institute for Postgraduate Studies,

Baghdad, Iraq.

Belal Al-Khateeb[2*]

[2]College of Computer Science and Information Technology, University of Anbar,

Ramadi, Iraq

Abstract **- Medication supply and storage are essential components of the medical industry and distribution. Most medications have a predetermined expiration date. When the demand is met in large quantities that exceed the actual need, this leads to the accumulation of medicines in the stores, and this leads to the expiration of the materials. If demand is too low, this will have an impact on consumer happiness and drug marketing.**
Therefore, it is necessary to find a way to predict the actual quantity required for the organization's needs to avoid material spoilage and storage problems. A mathematical prediction model is required to assist any management in achieving the required availability of medicines for customers and safe storage of medicines. The research question is to design a system based on deep learning that can predict the amount of drugs required with high efficiency and accuracy based on the chronology of previous years.Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional LSTM, and Gated Recurrent Unit (GRU) are used to build prediction models. Those models allow for the optimization of inventory levels, thus reducing costs and potentially increasing sales. Various measures such as mean squared error (MSE), mean absolute squared error (MASE), root mean squared error (RMSE), and others are used to evaluate the prediction models. RNN model achieved the best result with MSE: 0.019 MAE: 0.102, RMSE: 0.0.

Index Terms – RNN; LSTM; Bidirectional LSTM; GRU; Prediction medication needs.

I. INTRODUCTION

Predicting the actual need for medicines can be an effective approach to avoid storage problems caused by overstocking or understocking of medications. By accurately predicting the demand for medications, healthcare providers can order and stock medications in the right quantities, ensuring that medications are available when they are needed without creating excess inventory that may expire before it can be used.

There are several methods that can be used to predict the demand for medications, including time series analysis, machine learning, and other statistical methods [1].Time series analysis involves analyzing historical data on medication usage to identify trends and patterns that can be used to predict future demand [2]. Machine learning algorithms can be trained on historical data to identify patterns and make predictions about future demand based on a variety of factors, such as patient demographics, disease prevalence, and seasonal variations [3]. Other statistical methods can also be used to predict medication demand, such as regression analysis, which involves identifying the relationship between medication usage and various demographic or clinical factors that may influence demand. By using these methods (time series analysis, machine learning, and other statistical methods to predict medication demand, healthcare providers can adjust their ordering and stocking practices to ensure that they have the right amount of medication on hand to meet patient needs without creating excess inventory that may expire before it can be used. This can help minimize waste, reduce costs, and ensure that patients have access to the medications they need when they need them [4]. Pharma sales data is collected for the period (2014-2019) on sales of drugs predicted in 8 ATC categories contain information about the sales of drugs. This paper aims to develop a system that utilizes deep learning algorithms to forecast the necessary quantity of drugs by analyzing time series data. To achieve this goal, the following objectives have been identified:
1. To construct a model based on one of the deep learning algorithms that can accurately predict the demand for drugs using globally available data.
2. To determine the best deep learning architecture that yields the lowest error rate in predicting drug needs.
The rest of this article is organized as follows: Section 2 presented related works. Section 3 presented deep learning algorithms that are used in the prediction process. The proposed method that are used in the prediction process are described in Section 4. Section 5 presented results and discussion. Finally, conclusions have been provided in Section 6.

II. RELATED WORKS

This section discusses the literature on predicting medicine demand that employs deep learning in its research. The method is based on deep learning, a subset of machine learning that has emerged as a semi-pioneer in a number of scientific, security, and other domains, including healthcare, due to its potent powers and high information requirements, below are some studies related to this work:

In 2020, Zdravković et.al. tested the Autoregressive Integrated Moving Average (ARIMA), Prophet, and LSTM forecasting techniquesThe data was divided into training data and test data (52 weeks of data were used for the test).

MSE was used as a loss function for LSTM as well as a performance metric. Three tests were conducted to establish the baseline. The Nave and seasonal Nave were used for rolling predictions, while the average approach served as the baseline for long-term forecasting. In order to facilitate recommended

sales and marketing, several methods and processes linked for collecting, analyzing, and forecasting sales time series data are validated. Using forecasting techniques based on trend and seasonality, the sales of eight different classes of pharmaceutical items are predicted. With a variety of traits, including stationarity, seasonality, residual amount, and variation in the sales data. The study only used mean squared error (MSE) as a loss function and performance metric. Also, the data are limited data and the testing period was only for 52 weeks [5]. **Hafiz et.al.** [6] Splitted the dataset into two groups with a 70/30 ratio for training and testing, using the LSTM algorithm. After neural network training. Basis Model creation begins with developing the model. Following training, the model will be applied to forecast future sales data for analgesic medications for the following LSTM-related issues in Dhaka: The issue with RNNs long-term reliance, and the explosion and diminishing gradient for greater accuracy, the model requires a large amount of training data.

[1]aimed to estimate a pharmaceutical company's major sales using neural network forecasting. The researchers analyzed basic forecasting techniques, different neural network topologies, and training techniques to predict time series. They studied the pharmaceutical business's income from 2016 to 2021 and selected LSTM networks due to their ability to handle long value sequences. They added more layers (six instead of four) and more neurons in each LSTM module (from fifty to ninety) to improve the model's performance.
One of the major limitations of the Galkin et.al study is the high degree of time complexity involved in using neural network models. The researchers also did not apply any data preprocessing techniques to the dataset, which could have resulted in inaccurate predictions. The model's performance is affected by the heterogeneity in product variety and demand conditions. Additionally, the demand for medicines is highly unpredictable due to seasonality and consumer tendencies to buy drugs in the future. The study provides valuable insights into the use of neural network models to predict pharmaceutical sales. However, further research is needed to address the limitations and improve the accuracy of the predictions [7]. The objective of the work of El Filali et.al. (2022) was to build an LSTM model capable of producing the best possible forecasts using real data from the sales history of a pharmaceutical product of a Moroccan company. The results show that the proposed multilayer LSTM with the Grid search method has the lowest error measures: RMSE = 4487.32 and SMAPE = 0.026, compared to the traditional approaches and the multilayer RNN and monolayer LSTM models. This proves that the fitted LSTM model is capable of producing more accurate and efficient predictions. The data used is the monthly amount of product sales from 2012 to 2020 of a pharmaceutical company The suggested approach has the capacity to automatically configure by iterating over various combinations of LSTM hyperparameters, which enables improving the model's prediction accuracy for a given time series. In order to use forecasts over extended periods of time and make long-term predictions, this kind can extract the most pertinent information from the data and divide the time series signal between what is significant in the short term and what is important in the long run. The deep learning model is able to handle the complexity of non-linear data and automatically extract its features. For this reason, it is easier to obtain good prediction results on the first training run, even with arbitrary parameters. In this work the accuracy can be improved, and the dataset is not globally available [8].
Rathipriya et. al. (2022) used suggested ARIMA, Radial suggested ARIMA, Radial Basis Function Neural Network (RBF NN(, Generalized Regression Neural Network)GR NN(, Probabilistic Neural Network (PNN), and LSTM. The used dataset included 600,000 transactional records amassed over a six-year period (2014–2019). The dataset is divided into a training set (70%) and a validation set (the remaining 30%). Spiking Neural Networks (SNN)models should be trained using a variety of hyperparameter values and training data. The validation dataset is used to assess SNN models' performance. The best model should have the lowest RMSE value. The optimal parameter values should be returned by the best NN model. The optimal SNN net model should be trained and simulated with the best parameter settings. The accuracy of the forecast should be evaluated using test data. Forecast future timestamp values and evaluate forecasting accuracy. The ARIMA model worked well for three drug classes, while DFM-based neural network models performed well in five of the eight drug classes. It has been demonstrated through empirical investigation that an individual demand prediction model is not the best choice for all drug classes. Accuracy can be increased [9].

III. DEEP LEARNING ALGORITHMS

Deep learning algorithms such as RNN, LSTM, bidirectional LSTM, and GRU are all used for sequential data analysis, such as time-series data. Here is a simplified explanation of each algorithm:

RNN is a type of neural network that is capable of handling sequential data by remembering the previous inputs in the sequence. In RNN, the output from the previous step is fed as input to the current step, and this process is repeated until the end of the sequence. However, standard RNNs can suffer from the vanishing gradient problem, which means that the gradients

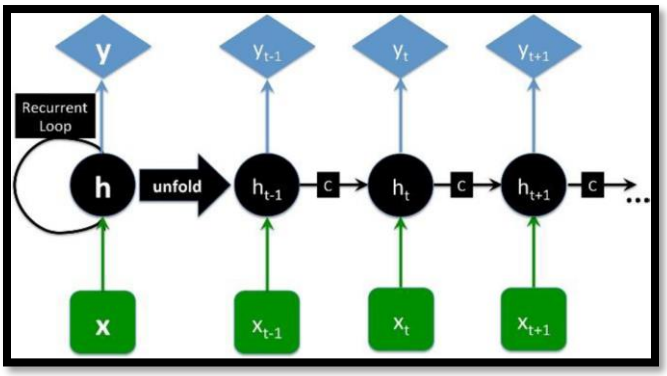

Figure 1: RNN architecture[1]

can become too small, making the model difficult to train on long sequences [8] [10] [11]. The figure 1 shows the RNN architecture [12].

The problem of disappearing gradients in typical RNNs, which can happen when the network tries to propagate incorrect gradients through numerous time steps, was addressed by the development of LSTM networks. Input gates, output gates, forget gates, memory cells, and other components make up the unique architecture of LSTM networks, which enables the network to selectively remember or forget data from earlier time steps. This makes it possible for the network to successfully propagate gradients through numerous time steps and discover long-term dependencies in the input sequence. [8]. Figure 2 shows the LSTM architecture.

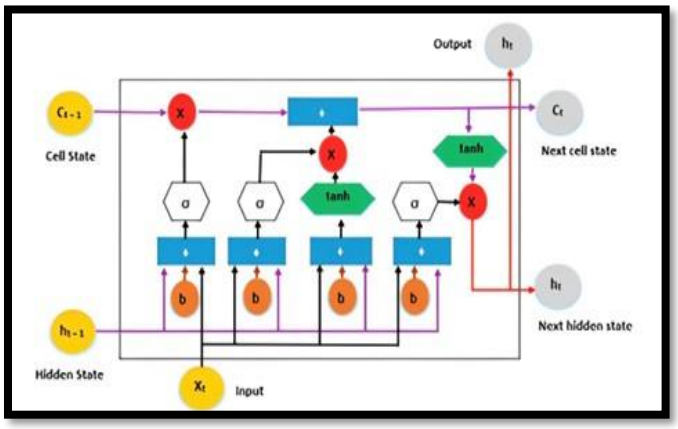

Figure 2: LSTM network’ s architecture[2]

A version of LSTM known as bidirectional LSTM processes the sequence both forward and backward. This enables the model to consider the past and the future forecasts that take into account context, which is important in some situations [12] [13].The figure 3 shows the bidirectional LSTM architecture.

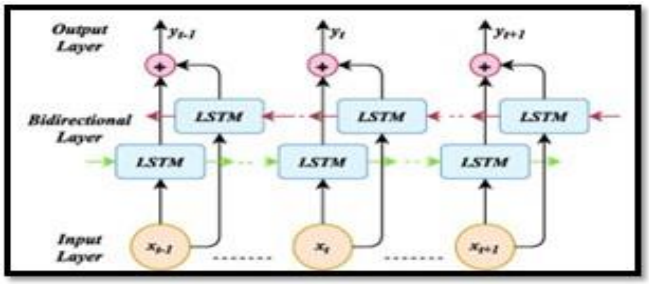

Figure 3 Bidirectional LSTM Network's Architecture[3]

Another LSTM variation called the Gated Recurrent Unit (GRU) streamlines the architecture by fusing the input and forget gates into a single update gate. As a result, the model has fewer parameters and can be trained more quickly [14] [15]. Figure 4 shows the GRU architecture.

IV. PROPOSED METHOD

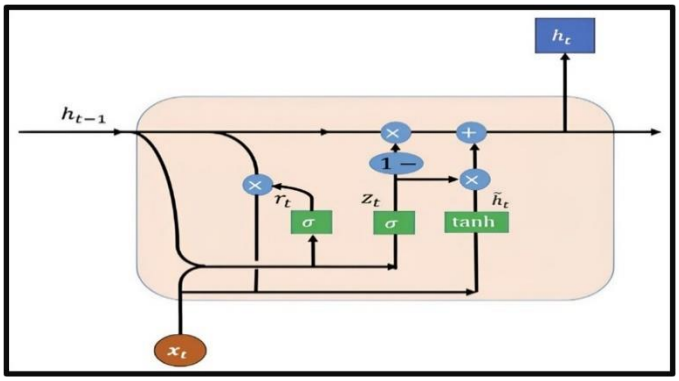

Figure 4: Gated recurrent unit (GRU)[4]

The proposed model is to predict the need for the actual required quantity of drugs to ensure their permanent availability and at the same time avoid out-of-date drugs. The methodology includes five steps.

- Load the dataset (time series).
- data preprocessing
- Divide the dataset into training data and test data.
- Apply a group of deep learning algorithms to obtain a model that predicts the required need for medicines.
- Evaluate all models to determine which one has the highest prediction accuracy.
- Choose the best model.

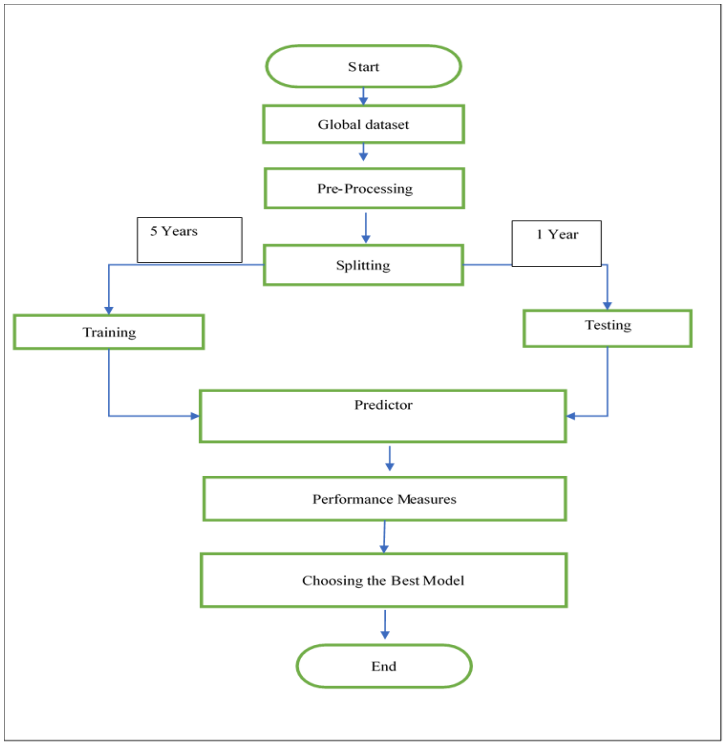

Figure 5: The General Structure of the predicting medicine.

A. Dataset

In this work, the pharma sales dataset is chosen due to its popularity. The dataset can be downloaded from the reference [7]

Pharma sales data is collected for the period (2014-2019) on sales of drugs predcited in 8 ATC categories contain information about the sales of drugs in eight different categories over a period of six years. It includes information about the quantity and value of drugs sold in each category, as well as the number of units sold, the average price per unit, and other relevant information. The dataset is a subset of a larger dataset consisting of 2106 transactional records collected over a period of six years (2014-2019) from a point-of-sale system in individual pharmacies. The ATC categories included in the dataset are used for market analysis, forecasting, and decision-making in the pharmaceutical industry as shown figure 1. The ATC categories included in the dataset are:

- M01AB: Anti-inflammatory and antirheumatic products, non-steroids, acetic acid derivatives and related substances.
- M01AE: Anti-inflammatory and antirheumatic products, non-steroids, propionic acid derivatives.
- N02BA: Other analgesics and antipyretics, salicylic acid, and derivatives.
- N02BE/B: Other analgesics and antipyretics, pyrazolones and anilides.
- N05B: Psycholeptics drugs, anxiolytic drugs.
- N05C: Psycholeptics drugs, hypnotics, and sedatives drugs.
- R03: Drugs for obstructive airway diseases.
- R06: Antihistamines for systemic use.

	datum	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06	Year
0	1/2/2014	0.00	3.670	3.40	32.40	7.0	0.0	0.0	2.00	2014
1	1/3/2014	8.00	4.000	4.40	50.60	16.0	0.0	20.0	4.00	2014
2	1/4/2014	2.00	1.000	6.50	61.85	10.0	0.0	9.0	1.00	2014
3	1/5/2014	4.00	3.000	7.00	41.10	8.0	0.0	3.0	0.00	2014
4	1/6/2014	5.00	1.000	4.50	21.70	16.0	2.0	6.0	2.00	2014
...	...	...	...	...	...	...	...	...	...	...
2101	10/4/2019	7.34	5.683	2.25	22.45	13.0	0.0	1.0	1.00	2019
2102	10/5/2019	3.84	5.010	6.00	25.40	7.0	0.0	0.0	0.33	2019
2103	10/6/2019	4.00	11.690	2.00	34.60	6.0	0.0	5.0	4.20	2019
2104	10/7/2019	7.34	4.507	3.00	50.80	6.0	0.0	10.0	1.00	2019
2105	10/8/2019	0.33	1.730	0.50	44.30	20.0	2.0	2.0	0.00	2019

Figure 6. Sample of pharma sales data.

B. Data preprocessing

The dataset was preprocessed using the following operations:

1- Data cleaning: process involves removing duplicates, missing values, and irrelevant columns to ensure the accuracy and suitability of the dataset for analysis working with machine learning models.

```
Before data cleaning: (2106, 13)
Before removing irrelevant columns:
      datum  M01AB  M01AE  N02BA  N02BE  N05B  N05C   R03  R06  Year  Month  Hour  Weekday Name
0  1/2/2014    0.0   3.67    3.4  32.40   7.0   0.0   0.0  2.0  2014      1   248    Thursday
1  1/3/2014    8.0   4.00    4.4  50.60  16.0   0.0  20.0  4.0  2014      1   276      Friday
2  1/4/2014    2.0   1.00    6.5  61.85  10.0   0.0   9.0  1.0  2014      1   276    Saturday
3  1/5/2014    4.0   3.00    7.0  41.10   8.0   0.0   3.0  0.0  2014      1   276      Sunday
4  1/6/2014    5.0   1.00    4.5  21.70  16.0   2.0   6.0  2.0  2014      1   276      Monday

********************************************************************************
After data cleaning: (2106, 8)
After removing irrelevant columns:
   M01AB  M01AE  N02BA  N02BE  N05B  N05C   R03  R06
0    0.0   3.67    3.4  32.40   7.0   0.0   0.0  2.0
1    8.0   4.00    4.4  50.60  16.0   0.0  20.0  4.0
2    2.0   1.00    6.5  61.85  10.0   0.0   9.0  1.0
3    4.0   3.00    7.0  41.10   8.0   0.0   3.0  0.0
4    5.0   1.00    4.5  21.70  16.0   2.0   6.0  2.0
```

Figure 7. Removing irrelevant columns.

In figure 2, the Data Frame is printed before removing irrelevant columns. The irrelevant columns, namely datum, Year, Month, Hour, and Weekday Name, are removed. The shape of the cleaned data and the Data Frame after removing irrelevant columns are printed, showing the difference in the number of columns. This is a common preprocessing step in data analysis and machine learning.

-2- Data Normalization: This process is used to put the extracted data of drugs in a range of [0-1] that promotes better model performance, fairness in feature importance, and avoids potential numerical problems in certain algorithms.as shown in figure. The equation(1) is used to normalize the dataset [5].

$$\text{Normalize} = \frac{X(i) - MIN(X)}{MAX(X) - MIN(X)} \qquad (1)$$

Where: X represents the observed values in the set. Min and max are the minimum and maximum values of x.

```
Before feature scaling:
   M01AB  M01AL  N02BA  N02BE  N05B  N05C  R03  R06
0    0.0   3.67    3.4  32.40   7.0   0.0  0.0  2.0
1    8.0   4.00    4.4  50.60  16.0   0.0 20.0  4.0
2    2.0   1.00    6.5  61.85  10.0   0.0  9.0  1.0
3    4.0   3.00    7.0  41.10   8.0   0.0  3.0  0.0
4    5.0   1.00    4.5  21.70  16.0   2.0  6.0  2.0

After feature scaling:
   M01AB    M0IAE    N02BA    NO2BE    NOSB    NO5C    R03    R06
0  0.000000  0.253751  0.21250  0.201242  0.127668  0.000000  0.000000  0.133333
1  0.461361  0.276568  0.27508  0.314286  0.291793  0.000000  0.444444  0.266667
2  0.115340  0.069142  0.40625  0.384161  0.182371  0.000000  0.200008  0.066667
3  8.230681  0.207426  0.43758  0.255288  0.145897  0.000000  0.066667  0.008888
4  0.288351  0.069142  0.28125  0.134783  0.291793  0.222222  0.133333  0.133333
```

Figure 8. Data normalization

C. Splitting

Dividing a time series into a training set and a test set typically involves selecting a portion of the data to be used for training the model and another portion to be used for evaluating its

performance. The training set is used to train the model, the test set is used to evaluate the final performance of the model.

The proposed study used a 5-year training phase and a 1-year testing period to divide a time series into a training set and a test set.

D. Experimental Setup

Many experiments were conducted using deep learning algorithms (RNN, LSTM, Bidirectional LSTM, and GRU) using different architectures, where each algorithm used a different number of layers (two layers, three layers, four layers), and the prediction error percentage was measured for each architecture. Algorithm1 shows the steps of predicting drug need.

Algorithm (1): Prediction of drug need using RNN algorithm

Input: Pharma sales data (Daily Sales).

Output: Forecasting next year's sales.

Step 1: Load the Dataset of pharmaceutical sales for 6 years.
Step 2: Data preprocessing
Remove the outliers value.
Normalize all values to be between 0 and 1.
Step 3: Split the dataset to five years training and one year testing.
Step 4: Bulid the architecture of model by:
Determine the number of layers.
Determine the hyperparameters such as number of epochs, number of batch size, activation function, and optimizer.
Step 5: Train the model by:
Compute current input use the equation $H_{t+1}=f(x_t,h_t,w_x,w_h,b_h)=f(w_x x_t+w_h h_t+b_h)$
Compute the output from the recent past use the equation $Y_t=f(h_t,w_y)=f(w_y*h_t+b_y)$).
Compute the error use the equation $e=y_{t+k}-p_{t+k}$.
Backpropagate the error across the unfolded network and update the weight.
Step 6: Test the model.
Step 7: Evaluate the model by computing MSE, MAE, and RMSE.

Algorithm (2): Prediction of drug need using LSTM algorithm

Input: Pharma sales data (Daily Sales).

Output: Forecasting next year's sales.

Step 1: Load the Dataset of pharmaceutical sales for 6 years.
Step 2: Data preprocessing
Remove the outliers value.
Normalize all values to be between 0 and 1.
Step 3: Split the dataset to five years training and one year testing.

Algorithm (2): Prediction of drug need using LSTM algorithm

Input: Pharma sales data (Daily Sales).

Output: Forecasting next year's sales.

Step 4: Build the architecture of model by:
Determine the number of layers.
Determine the hyperparameters such as number of epochs, number of batch size, activation function, and optimizer.
Step 5: Train the model by:
Calculate forget gate activation vector (ft) using equation $f_t=\delta(W_f*[h_{t-1},x_t]+b_f)$.
Calculate input gate's activation vector (It) using equation $I_t=\delta(W_i*[h_{t-1},x_t]+b_i)$.
Calculate new memory vector ($\hat{C}t$) using equation $\hat{C}_t=\tanh(W_c*[h_{t-1},x_t]+b_c)$.
Calculate cell stat (Ct) using equation $C_t=F_t*C_{t-1}+I_t*\hat{C}_t$.
Calculate output gate's activation vector (Ot) using equation $O_t=\delta(W_o[h_{t-1},x_t]+b_o)$.
Calculate hidden state vector ht (final new hidden state), Ot using equation $h_t=O_t*\tanh(C_t)$.
Calculate error using equation error = $(y-\hat{y})^2$.
Update weights using equation Wi(p+1) =wi(p) + alpha *xi * e(p).
Step 6: Test the model.
Step 7: Evaluate the model by computing MSE, MAE, and RMSE.

Algorithm (3): Prediction of drug need using Bidirectional LSTM algorithm

Input: Pharma sales data (Daily Sales).

Output: Forecasting next year's sales

Step 1: Load the Dataset of pharmaceutical sales for 6 years.
Step 2: Data preprocessing
Remove the outliers value.
Normalize all values to be between 0 and 1.
Step 3: Split the dataset to five years training and one year testing.
Step 4: Build the architecture of model by:
Determine the number of layers.
Determine the hyperparameters such as number of epochs, number of batch size, activation function, and optimizer.
Step 5: Train the model by:
Calculate Forward LSTM using equation $h_{t(\tau)}^{\rightarrow}=\text{LSTM}^{\rightarrow}(x_i\ t), t\in[1,T]$.
Calculate Backward LSTM using equation $h_{t(\tau)}^{\leftarrow}=\text{LSTM}^{\leftarrow}(x_i\ t), t\in[1,T]$.
Step 6: Test the model.
Step 7: Evaluate the model by computing MSE, MAE, and RMSE.

Algorithm (4): Prediction of drug need using GRU algorithm
Input: Pharma sales data (Daily Sales).
Output: Forecasting next year's sales.
Step 1: Load the Dataset of pharmaceutical sales for 6 years. Step 2: Data preprocessing Remove the outliers value. Normalize all values to be between 0 and 1. Step 3: Split the dataset to five years training and one year testing. Step 4: Build the architecture of model by: Determine the number of layers. Determine the hyperparameters such as number of epochs, number of batch size, activation function, and optimizer. Step 5: Train the model by: Calculate Reset gate using equation $R_t = \delta(X_t W_{xr} + H_{t-1} W_{hr} + b_r)$. Calculate Update gate using equation $Z_t = \delta(X_t W_{XZ} + H_{t-1} W_{hz} + b_z)$. Calculate hidden state using equation $\hat{H} = \tanh(X_t W_{xh} + (R_t \Theta H_{t-1}) W_{hh} + b_h)$. Calculate previous time step hidden state using equation $H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \hat{H}_t$. Step 6: Test the model. Step 7: Evaluate the model by computing MSE, MAE, and RMSE.

Below are the parameters that were used for all the experiments:
- Number of epochs= 400.
- Batch size=50.
- Activation Function: Tanh Function and ReLU.
- Optimizer=Adam.

4.4 Error ratio metrics

To determine the accuracy of the work performed by the system employed in the prediction process, the error rate is measured using a variety of methods.

1) Mean Squared Error (MSE)

The average squared variance between actual and expected values is expressed quantitatively by this statistic. The MSE is equal to zero when a model has no errors. The value of a model rises in direct proportion to the amount of mistake it contains [7]. It is calculated by the following equation.

$$MSE=\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y_t})^2 \quad (3)$$

2) Mean absolute error (MAE)

Is a measurement of the typical error in a set of predictions that ignores the direction of the errors. It is measured as the average absolute difference between the predicted values and the actual values and is used to check whether a regression model is functioning appropriately [15].

$$MAE= \frac{1}{T} \left(\sum_{t=1}^{T} |yt - \hat{y}t|\right) \quad (4)$$

V. RESULTS AND DISCUSSION

The results of using RNN, LSTM, Bidirectional LSTM, and GRU to predict medicine demand have shown promise in optimizing inventory levels and reducing costs, potentially leading to increased sales. Evaluation of the prediction model using various measures, such as MSE, MASE, and RMSE, demonstrated the effectiveness of the model in accurately predicting medicine demand. Overall, the use of deep learning techniques in predicting medicine demand has the potential to revolutionize the pharmaceutical industry by providing insights into demand patterns and allowing for more efficient inventory management.

Table 1 includes columns for drug materials and actual expenditures for the year 2019, which will be used as projected expenditures. The ratio between actual and projected data is then measured to find out the percentage of prediction error. Based on Table 1, The data represents daily sales, with each row representing a different day. The minimum and maximum values in each column provide some insight into sales trends for each category over the time period covered by the dataset. Some notes could include:

- Column N05B has the highest maximum value, indicating that this category had the highest sales during the given time period.
- Column N02BE has the highest range (difference between the minimum and maximum values), indicating that there were significant fluctuations in sales for this category over the given time period.
- Column N05C has the lowest maximum and range values, indicating that sales for this category were consistently low over the given time period.
- Overall, there doesn't seem to be any clear trends in the data that suggest significant changes in sales over time for any of the categories.

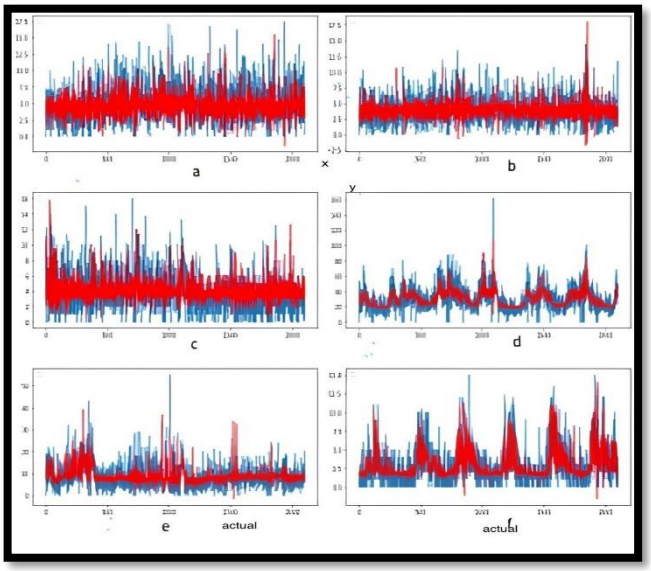

Figure 9. The actual and predicted values for each medication substance a) 'M01AB' b) 'M01AE' c) 'N02BA' d) 'N02BE' e) 'N05B' and f) 'R06

Figure 4 shows the plots of the actual and predicted values for each medication substance. Each plot contains two lines: the blue line (X-access) shows the actual sales values, while the red line (Y-access) shows the predicted sales values. Looking at the plots, we can see that the model is performing well in predicting the sales of some medication substances, while in other cases, the model predictions are not very accurate due to the different data variances. For example, for the medication substance 'M01AE', the actual and predicted values are very close, and the red line follows the trend of the blue line, indicating that the model is performing well in predicting sales for this substance. Overall, the plots show that the model is performing reasonably well in predicting the sales of most of the medication substances.

Table 1. Comparison of the results of the algorithms with normalization

Models	Node	MSE	MAE	RMSE
RNN three layer	(L1:50,L2:60,L3:70)	0.019	0.102	0.0
LSTM three layer	L1:50,L2:60,L3:70)	0.027	0.122	0.0
Bidirectional LSTM Four layer	L1:50,L2:60,L3:70, L4:60)	0.024	0.112	0.0
GRU four layer	L1:50,L2:60,L3:70, L4:60)	0.028	0.123	0.0

Table (1) shows the best results when using the RNN algorithm with three layers, and the figure 10 shows the results. By comparing the results for each algorithm, it was found that the best results for the RNN algorithm were when using three layers, where the MSE scale was 0.019 and the MAE was 0.102, while the best architecture for LSTM is the three layers: MSE: 0.027, MAE: 0.122, and for Bidirectional LSTM, the best results were at using the four layers MSE: 0.024, MAE: 0.112, and finally the GRU had the best results when using four layers, where the results were MSE: 0.028, MAE: 0.123, as shown in Figure 10.

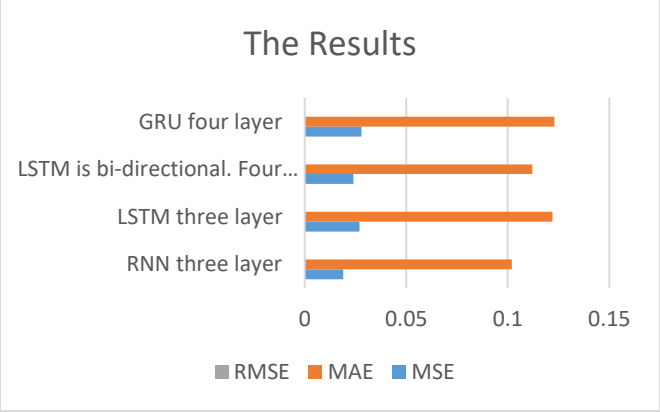

Figure 10. Best obtained results

VI. CONCLUSIONS

The use of deep learning algorithms, (RNN, LSTM, Bidirectional LSTM, and GRU) has shown promising results in analyzing time series and predicting the need for drugs. However, the performance of each algorithm may vary depending on the specific dataset. In general, the choice of the best algorithm for a specific task should be based on a comparison of their performance metrics, such as error measures like MSE, RMSE, and MAE. Comparing the performance of different algorithms can help to identify the one that provides the most accurate predictions. In some cases, the results of different deep learning algorithms may be somewhat close in terms of error measures, and other factors such as computational complexity and ease of implementation may come into play. Therefore, it's important to carefully consider all factors when choosing the best algorithm for a specific task. In summary, the use of deep learning algorithms can provide accurate predictions of the need for medications. The choice of the best algorithm should be based on a comparison of their performance metrics, and other factors should also be taken into consideration when making the final decision. In this study, the three-layer RNN algorithm (with 50 nodes in the first hidden layer, 60 nodes in the second hidden layer, and 70 nodes in the last hidden layer) gave the best results with MSE: 0.019 MAE: 0.102, RMSE: 0.0002. For the future research, we suggest using previously learned models to perform operations on new datasets that are collected locally. Examine several modeling approaches: LSTM, RNN, and GRU models were utilized in this study, but there are a number of other deep learning models that can be investigated, including Transformers and Convolutional Neural Networks (CNN) models.

REFERENCES

[1] D. Galkin, T. Dudkina, and N. Mamedova, “Forecasting time series using neural networks on the example of primary sales of a pharmaceutical company,” SHS Web Conf., vol. 141, p. 01014, 2022, doi: 10.1051/shsconf/202214101014.

[2] S. Punia, K. Nikolopoulos, S. P. Singh, J. K. Madaan, and K. Litsiou, “Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail,” Int. J. Prod. Res., vol. 58, no. 16, pp. 4964–4979, Aug. 2020, doi: 10.1080/00207543.2020.1735666.

[3] M. Ala’raj, M. F. Abbod, and M. Majdalawieh, “Modelling customers credit card behaviour using bidirectional LSTM neural networks,” J. Big Data, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00461-7.

[4] M. J. Hamayel and A. Y. Owda, “A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms,” Ai, vol. 2, no. 4, pp. 477–496, 2021, doi: 10.3390/ai2040030.

[5] C. Badgujar, S. Das, D. M. Figueroa, and D. Flippo, “Application of Computational Intelligence Methods in Agricultural Soil–Machine Interaction: A Review,” Agric., vol. 13, no. 2, pp. 1–39, 2023, doi: 10.3390/agriculture13020357.

[6] W. A. Al-Zaidi, A. J. Al-Karawi, and A. K. Al-Zuhairi, “Role of Demand Forecasting and Lead Time on Waste in Supply Chain: A Case Study in Diyala Health Sector-Iraq,” J. Adv. Manag. Sci., pp. 155–160, 2018, doi: 10.18178/joams.6.3.155-160.

[7] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, “Deep Learning for Time Series Forecasting: A Survey,” Big Data, vol. 9, no. 1. Mary Ann Liebert Inc., pp. 3–21, Feb. 01, 2021. doi: 10.1089/big.2020.0159.

[8] S. J. Maceachern and N. D. Forkert, “Machine learning for precision medicine,” Genome, vol. 64, no. 4. Canadian Science Publishing, pp. 416–425, 2021. doi: 10.1139/gen-2020-0131.
[9] R. Ferreira, M. Braga, and V. Alves, Forecast in the pharmaceutical area – Statistic models vs deep learning, vol. 747. Springer International Publishing, 2018. doi: 10.1007/978-3-319-77700-9_17.
[10] M. Zdravkovi, J. Đorđević, A. C. Đorđević, and S. Pavlović, “Case study : univariate time series analysis and forecasting of pharmaceutical products ’ sales data at small scale,”pp.1-4 2020.
[11] M. Hafiz, M. Sabir Ibna Sazzad, K. I. Hasan, J. Hasnat, and M. C. Mishu, “Predicting the demand of prescribed medicines in Bangladesh using artificial intelligent (AI) based long short-term memory (LSTM) model,” ACM Int. Conf. Proceeding Ser., 2020, doi: 10.1145/3377049.3377056.
[12] https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data.
[13] A. El Filali, E. H. Ben Lahmer, S. El Filali, M. Kasbouya, M. A. Ajouary, and S. Akantous, “Machine Learning Applications in Supply Chain Management: A Deep Learning Model Using an Optimized LSTM Network for Demand Forecasting,” Int. J. Intell. Eng. Syst., vol. 15, no. 2, pp. 464–478, Apr. 2022, doi: 10.22266/ijies2022.0430.42.
[14] R. Rathipriya, A. A. Abdul Rahman, S. Dhamodharavadhani, A. Meero, and G. Yoganandan, “Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model,” Neural Comput. Appl., vol. 35, no. 2, pp. 1945–1957, 2023, doi: 10.1007/s00521-022-07889-9.
[15] R. Vinayakumar, K. P. Soman, P. Poornachandran, and S. Sachin Kumar, “Detecting Android malware using Long Short-term Memory (LSTM),” J. Intell. Fuzzy Syst., vol. 34, no. 3, pp. 1277–1288, 2018, doi: 10.3233/JIFS-169424.
[16] A. Shewalkar, D. nyavanandi, and S. A. Ludwig, “Performance Evaluation of Deep neural networks Applied to Speech Recognition: Rnn, LSTM and GRU,” J. Artif. Intell. Soft Comput. Res., vol. 9, no. 4, pp. 235–245, 2019, doi: 10.2478/jaiscr-2019-0006.
[17] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K. W. Chau, “Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting,” Water (Switzerland), vol. 12, no. 5, pp. 1–18, 2020, doi: 10.3390/w12051500.
[18] R. Chandra, A. Jain, and D. S. Chauhan, “Deep learning via LSTM models for COVID-19 infection forecasting in India,” PLoS One, vol. 17, no. 1 January, pp. 1–28, 2022, doi: 10.1371/journal.pone.0262708.
[19] J. Chen, C. Lu, C. Chen, H. Cheng, and D. Xuan, “An Improved Gated Recurrent Unit Neural Network for State-of-Charge Estimation of Lithium-Ion Battery,” Appl. Sci., vol. 12, no. 5, 2022, doi: 10.3390/app12052305.
[20] Z. Huang, F. Yang, F. Xu, X. Song, and K. L. Tsui, “Convolutional Gated Recurrent Unit-Recurrent Neural Network for State-of-Charge Estimation of Lithium-Ion Batteries,” IEEE Access, vol. 7, pp. 93139–93149, 2019, doi: 10.1109/ACCESS.2019.2928037.
[21] https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data
[22] M. S. Amalnick, N. Habibifar, M. Hamid, and M. Bastan, “An intelligent algorithm for final product demand forecasting in pharmaceutical units,” Int. J. Syst. Assur. Eng. Manag., vol. 11, no. 2, pp. 481–493, 2020, doi: 10.1007/s13198-019-00879-6.