Review Article

Bashaer Abdurahman Mousa and Belal Al-Khateeb*

Predicting medicine demand using deep learning techniques: A review

https://doi.org/10.1515/jisys-2022-0297
received December 09, 2022; accepted February 15, 2023

Abstract

The supply and storage of drugs are critical components of the medical industry and distribution. The shelf life of most medications is predetermined. When medicines are supplied in large quantities it is exceeding actual need, and long-term drug storage results. If demand is lower than necessary, this has an impact on consumer happiness and medicine 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. Artificial intelligence applications and predictive modeling have used machine learning (ML) and deep learning algorithms to build prediction models. This model allows for the optimization of inventory levels, thus reducing costs and potentially increasing sales. Various measures, such as mean squared error, mean absolute squared error, root mean squared error, and others, are used to evaluate the prediction model. This study aims to review ML and deep learning approaches of forecasting to obtain the highest accuracy in the process of forecasting future demand for pharmaceuticals. Because of the lack of data, they could not use complex models for prediction. Even when there is a long history of accessible demand data, these problems still exist because the old data may not be very useful when it changes the market climate.

Keywords: forecasting, deep learning, machine learning, prediction

1 Introduction

The world's medical demand is currently on an upward trend, and as a result, the pharmaceutical industry's focus is also growing. Businesses must manage demand to some extent to maintain a competitive market with rising market demand. A situation where forecasting is a crucialcomponent of demand management [1]. With the dual goals of matching the supply with any potential increase or decrease in the demand for the products and maintaining inventory as little as possible, the demand forecasting model is a key machine learning (ML) technique utilized by the majority of international pharmaceutical businesses. If such a system is successfully implemented, it would help manage their supply chain properly [2]. One of the most popular sales management techniques nowadays is forecasting. It enables the capacity, strategy, and location of the plant to be changed, plans spending for the coming year, shows how easily the business process can be controlled, and lowers external dangers by giving an idea of future sales for inventory

*** Corresponding author: Belal Al-Khateeb,** College of Computer Science and Information Technology, University of Anbar, Ramadi 31001, Iraq, e-mail: belal-alkhateeb@uoanbar.edu.iq
Bashaer Abdurahman Mousa: Iraqi Commission for Computer and Informatics, Informatics Institute for Postgraduate Studies, 10066 Baghdad, Iraq, e-mail: ms202120660@iips.icci.edu.iq

management [3]. Obtaining a reasonably accurate estimate of future demand for a good or service given historical data and the current political, social, and economic environment is the objective of forecasting processes across various industries, despite their complexity and execution: to plan and organize businesses accordingly. Accurate forecasting is a significant difficulty for the pharmaceutical sector [4]. Supply chain efficiency is based on accurate demand forecasting because it fundamentally informs all crucial operational choices, including those pertaining to inventory management, production planning, raw material supply, and financial objectives. Demand forecasting can be especially important for pharma manufacturers because any mismatch between supply and demand could have an impact on patients and the drug distribution chain, sometimes even putting their lives in danger [5]. Numerous techniques and studies have been done in the field of demand forecasting, as ML algorithms such as linear regression, decision trees, support vector machines (SVM), random forests (RF), recurrent neural networks (RNNs), and long-short-term memory (LSTM) are used [1,2,6]. A different level of accuracy was obtained, and this can be improved. Time-series forecasting techniques are the foundation of conventional forecasting techniques. These forecasting techniques provide predictions about future demand using historical data. A collection of data points that were gathered at regular periods of time is called a series. These methods only use alimited number of historical time-series data on demand [4]. The objective of the current study is to present and examine various approaches for forecasting the pharmaceutical supply chain. The following are the contributions to this work:

- This review covered the previous methods that are used in the forecasting process.
- This article presents the most important statistical methods, including ML and deep learning methods, that are used in the prediction process.
- This article presents the datasets that are used in the demand for medicine prescriptions.
- The current study shows the importance of artificial intelligence approaches, such as ML and deep learning, that are used in prediction processes. This study focuses on the development of a mechanism using deep learning algorithms for forecasting the pharmaceutical supply chain.

The rest of this article is organized as follows: Section 2 presented pharmaceutical sales forecasting. Section 3 presented the datasets that are used in the prediction process. The ML methods that are used in the prediction process are described in Section 4. Section 5 presented the deep learning methods that are used in the prediction process. Section 6 presented a discussion containing an analysis of this review article results. Finally, conclusions have been provided in Section 7.

2 Pharmaceutical sales forecasting

Accurate sales forecasting is a crucial and affordable strategy to boost earnings, which will also lessen the impact of risks on key business operations. Modeling the ideal environment for the manufacture and distribution of goods is made possible based on the outcomes of the constructed forecast. The forecasting model serves as both the foundation for calculating the precise values that represent the anticipated result, or breakeven sales, and a functional and adequate representation of the process under investigation [1]. Figure 1 shows the stages of the prediction model.

Forecasting can be categorized into two types: qualitative and quantitative. Quantitative forecasting makes predictions from numerical data and historical events using probability calculations. There are two different types of forecasting: qualitative and quantitative. Using probability calculations, quantitative forecasting generates forecasts based on numerical data and past events, impacts and has long-term effects. The foundation for the models' methodology is a mathematical model and an objective analysis. In contrast to quantitative forecasting, which only uses numbers, qualitative forecasting uses expert judgment. It is crucial to have knowledge of highly skilled professionals in this industry. Comparing this to qualitative forecasting is very different. When an abnormality can be anticipated, this strategy is employed. It works well when it is anticipated that future results will differ from those of the past [2].

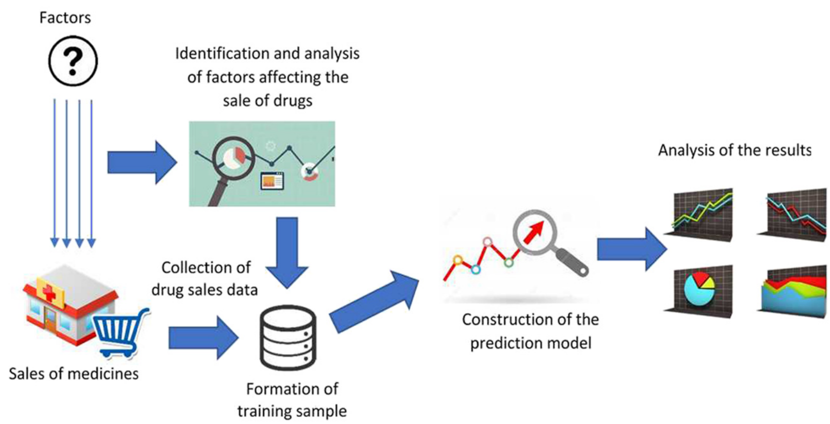

Figure 1: Stages of the prediction model [1].

3 Dataset

A dataset consists of many different pieces of data, which can be used to train an algorithm to search for expected patterns in the entire dataset. Data are a critical component of any AI model and are, essentially, the only reason for the current surge in the popularity of ML. Due to the availability of data, scalable ML algorithms are now possible as standalone solutions that can enhance a company's worth instead of serving as a result of its core operations [6]. Some of the data are not globally available. Table 1 shows the available datasets that are employed in the pharmaceutical sector.

4 ML algorithms

ML, a branch of AI that focuses on finding patterns in massive volumes of data, allows prediction and classification models to be created using training data [7]. Supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the four main divisions of ML algorithms. There are various types of ML techniques.

If several algorithms, like polynomial regression, linear regression, and support vector regression (SVR), which are widely used for prediction purposes, are constructed based on various parameters, it will be simpler to select one. The linear regression approach is used to represent the connection between a scalar dependent variable, Y, and one or more independent variables, X. A linear model can be applied to a linear dataset in the same manner as simple linear regression to get a respectable result; however, when the same model is applied to a non-linear dataset without any modifications, an impressive result is created. The larger loss of function will cause the accuracy to fall and the error rate to climb. Linear regression is a line equation that is used to define a linear relationship between two variables. If one of the values is known, it can be used to estimate the value of the other. To make an accurate estimation from the data, the best truth for the data must be created. The area that is closest to all the points should be chosen when drawing the optimum line. Linear regression requires the operation of two variables: a dependent variable and an independent variable.

A statistical approach for classification and prediction is the Gauss process. Bayesian Gaussian processes are used for nonlinear regression when learning the Gaussian processes method. The Bayesian Gauss method Using all possible probability sequences of the inputs, a smooth and continuous function f assigns a direct distribution to the space of nonlinear functions, correlates the data with the function outputs, and provides a strong correlation between the outputs based on the proximity of the inputs. The final deduction from functions prior to Gauss processes is that they allow supervised learning [13].

Table 1: Pharmaceutical products datasets

Dataset name	Description	Used by	Reference
Pharma sales data	**Data:** 57 drugs **Time period:** 2014–2019 **Note:** This data can be downloaded easily and contains drug sales data for an hour, day, week, and month	[3,4]	https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data" https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data
Rossman Store Sales	**Data:** 1,115 stores **Time period:** 2013–2015 **Note:** This data can be downloaded easily and contains store sales data	[5,8,9]	https://www.kaggle.com/datasets/pratyushakar/rossmann-store-sales" https://www.kaggle.com/datasets/pratyushakar/rossmann-store-sales
AKRIKHIN company	**Data:** 200 drugs **Time period:** 2016–2021 **Note:** There is no link to download this data, which makes it difficult to be publicly available	[10,11]	The data are not available
Hydroxyzine	**Data:** 200 drugs **Time period:** 2014–2017 **Note:** There is no link to download this data, which makes it difficult to be publicly available	[1]	The data are not available
Drugs in Rwanda	**Data:** 500 drugs **Time period:** 2015–2019 **Note:** It is difficult to download this data	[11]	https://www.kaggle.com/datasets/tamilsel/healthcare-providers-data" https://www.kaggle.com/datasets/tamilsel/healthcare-providers-data
Coop service	**Data:** 42,753 drugs **Time period:** 2018 **Note:** It is difficult to download this data	[12]	https://github.com/search?q=Lead+Time+Forecasting+with+Machine+Learning+Techniques+for+a+Pharmaceutical+Supply+Chain&type=code" https://github.com/search?q=Lead+Time+Forecasting+with+Machine+Learning+Techniques+for+a+Pharmaceutical+Supply+Chain&type=code
USA	**Data:** 135 drugs **Time period:** 2001–2016 **Note:** There is no link to download this data, which makes it difficult to be publicly available	[13]	The data are not available
ACT0002UZ01	**Data:** 41 data points **Time period:** 13 weeks **Note:** There is no link to download this data, which makes it difficult to be publicly available	[14]	The data are not available
TRx data	**Data:** 27 products **Time period:** 2010–2019 **Note:** There is no link to download this data, which makes it difficult to be publicly available	[15]	The data are not available
Moroccan industrial company	**Data:** Many drugs **Time period:** 2012–2020 **Note:** There is no link to download this data, which makes it difficult to be publicly available	[16]	The data are not available

The SVR algorithm for supervised ML primarily divides data into groups. It is an offshoot of the well-known classification algorithm SVM. In contrast to other algorithms, SVM employs a decision boundary called a hyperplane to divide the groups. With the help of SVM, it is possible to divide the data into segments, each containing a single type of data. The efficiency of the model may be determined by comparing the prediction functions for linear regression, polynomial regression, and SVR together with estimated accuracy and error values [8]. Figure 2 shows the SVR architecture.

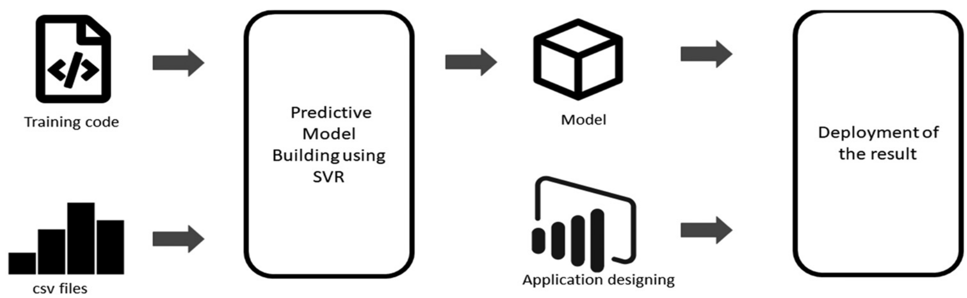

Figure 2: SVR architecture [4].

RF is a common ensemble technique for tackling classification and regression-related ML learning issues. A collection of decision trees makes up an RF. Each decision tree uses the forest's input and then produces an estimated value. The training procedure produces the tree's structure. A judgment is made for each tree by evaluating the input against a predetermined threshold. A fresh comparison is conducted with a different input and threshold based on the comparison's outcome, which establishes the tree's prediction.

Each decision tree output is combined to increase accuracy. The average of all the trees output is frequently used as the aggregation approach in regression [7,11,19] as shown in Figure 3.

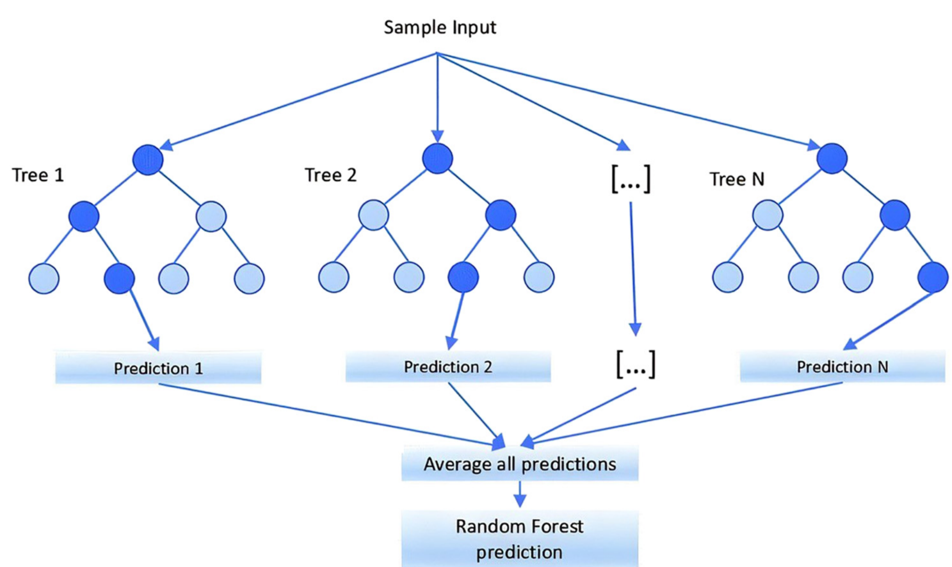

Figure 3: RF prediction scheme [19].

Table 2 shows the strengths and weaknesses of the above-mentioned algorithms.

The recent and well-known studies worthy of note that have been conducted on this issue are summarized. To forecast future consumption, Kravets et al. [1] used models to forecast the volume of drug sales

Table 2: Strengths and weaknesses of ML algorithms [20]

Algorithm	Strengths	Weaknesses
Linear regression	– Simple and fast technology – Easy to use as it requires minimal adjustment	– Limited to linear relationships – Sensitive to the direction of outliers – Unable to elicit relationships in some complex cases – The data must be independent – Limited to predicting digital output
Decision tree	– Easy to learn – Useful in data exploration – Does not require a lot of data elaboration – Not restricted to one data type – Enable predictive models with high accuracy, stability, and ease of interpretation	– Overfitting – The more decisions in the tree, the less accurate the results
Support vector machine (SVM)	– Effective in high-dimensional spaces – Best-rated performance in training data – Effective when the number of dimensions exceeds the number of samples – No strong assumptions are made about the data	– Does not perform well with a large dataset, because the required training time is high – When the dataset has a lot of noise, the target groups are overlapping – Does not provide direct probabilistic estimates
Random forest	– Can solve both types of problems: classification and prediction – The power of dealing with a large set of data with larger dimensions – Maintains accuracy even when a significant portion of the data is lost thanks to an effective mechanism for predicting lost data – Possesses techniques for correcting inaccuracies in data sets with unbalanced categories	– It categorizes data well, but the prediction problem is where it falls short because it cannot provide precise, continuous normal forecasts – Like a black box to statistical modelers. It means less control over what the model does

in Volgograd and compared their accuracy. The RF approach was used to build the model, which was exclusively based on historical data and took into consideration the elements that influenced it. The model was effective, using a small set of records, for predicting drug sales, considering the factors of influence. The used data covered the period from the beginning of 2014 to the end of 2017. The volume of the sample was 40 records. The dataset used in this study is not widely available, and the accuracy could be improved further. Merkuryeva et al. [14] proposed the basic moving average method, multiple linear regression, and symbolic regression with genetic programming as three experimental scenarios based on the use of various forecasting techniques that are examined. The following variables have been considered for regression scenarios: a distributor pricing list, the product's discounted selling price, the number of sales per week in a month, and the weekly average exchange rate. Analysis of forecasting results and forecast errors for each scenario is done, and application viability and consequences are given. The experimental examination of three forecasting scenarios reveals that the symbolic regression-based forecasting model offers the best curve fitting to historical demand data, reduced error estimates across all scenarios and experiments, and the capacity to forecast peak sales more accurately in the research. In addition, the flaws in accurate forecasting are a significant difficulty for the pharmaceutical sector. Historical data sets are scarce since they cover a shorter time frame than a year. On the other hand, the environment in emerging markets is volatile. A periodical demand fluctuation's clear correspondence with past demand data is obscured by the data's complexity. The dataset is not accessible locally [17]. The RF algorithm was proposed by Kumari et al. [5], and the XGBoost began with the direct application of ML algorithms on unprocessed data and then naturally developed by enhancing the data and employing reliable methods. The exercise underlined the significance of conducting data engineering and analysis prior to applying the algorithms to the data. Due to the time-series structure of the data, Rossman Store Sales were used to gather the data. It has been amply demonstrated that using series modeling techniques can considerably improve the accuracy and significance of data analysis and data engineering before applying the algorithms to the data. Gradient boosting and bagging ensemble regression approaches helped the researchers achieve better outcomes (RF). Using time-series modeling methods like ARIMA can greatly improve accuracy. Due to renovations, very few stores were closed for longer than 6 months. This information had to be removed from the final training set since it was affecting the model. Also, the prediction accuracy can befurther improved [5] proposed linear, Gaussian, m5rules, multilayer perceptron's smoreg, M5P, and RF algorithms by comparing the mean absolute error (MAPE) of these algorithms with the objective of estimating drugstore purchase data and subsequent sales quantities. It attempts to use the findings of the sales prediction in accordance with the algorithm that produces the best outcomes when applying the seven ML techniques to analyze sales data pertaining to previous years. The information organized from 5 years of medicine sales data between December 2019 and December 2020 had the best result of the Gaussian operations algorithm in terms of the mean absolute error rate. The obtained accuracy can be improved [13]. Tugay et al. [21] proposed linear regression, RF regression, gradient boosting, and decision tree regression. The fundamental implementation of stacked generalization consists of two steps. All learning algorithms are taught by utilizing the available data in the initial step. Use gradient boosting, decision tree regression, linear regression, and RF regression as the first level regressors at this point. In the second stage, a combiner algorithm is used to compile all of the predictions made by the learning algorithms used in the first stage into a final forecast. Choose the model that best fits this issue at this level using the same regression techniques as in the first phase. When more data are used, the real-world dataset acquired from the e-commerce company technique will predict significantly better than other single classifiers. Due to its accuracy with less data and the lack of statistical significance between the suggested model and RF, the proposed method can be utilized to forecast demand [22].

Nowadly and Jung [15] proposed RF, ANN, LR, and SVM. They used both direct and recursive forecasting models to create multi-step, long-range systems. In the one-step models, RF, ANN, and LR generated results that were reasonably accurate. However, for the accuracy of data scarcity in modeling and model evaluation, new elements, such as external factors that could affect demand, were not included in the model since the data were not accessible when the forecasting horizon was extended using a multi-step prediction. The dataset is not globally available [6]. Oliveira et al. [12] compared three straightforward non-linear approaches – RFs, k-nearest neighbors, multilayer perceptions, and linear regression – against two effective linear approaches – linear regression and linear SVM Averaging less than two days' error, the best method produced the best overall performance. The outcomes from the Coop Service Group dataset are quite positive, demonstrating how the purchasing lead time can be predicted with great accuracy, particularly for linear SVR. Simple non-linear techniques, in particular, do not appear to significantly improve prediction [12]. Keny et al.'s [4] compared the LR, Polynomial Regression, and SVR prediction functions and estimated accuracy and error values, the model's efficacy can be determined. The suggested model demonstrates the application of SVR to sales forecasting. It is possible to draw the conclusion from the study and application done during the work that the Web application's performance when using SVR and other data visualizing and analyzing methods was favorable in forecasting product sales for individual products, obtaining results that were more accurate and dependable for the company. The datasets include sales information for the drugs on an hourly, daily, weekly, and daily basis. It was created using the original dataset, which contained 600,000 transactions amassed over the course of 6 years (period 2014–2019). The data can be found on the Kaggle website. Performance development centers can more easily identify trends and seasonality thanks to the suggested model's improved sales data presentation (PDCS). Accuracy could be increased. When a linear model is applied to a linear dataset, such as in simple linear regression, a respectable result can be produced. However, when the same model is applied to a non-linear dataset without any adjustments, a striking result is obtained. The error rate will be considered a result of the enhanced loss function, but accuracy will be reduced. In these situations, when the organization of the data points is non-linear, the polynomial regression model is necessary [4]. In 2021, Mbonyinshuti et al. [11] developed three types of models, including LR, RF, and ANN models. The goal of this study is to see whether ML techniques might be used to improve forecasting demand accuracy and therefore optimize the availability of essential medicines. The dataset, which was referred to, is based on consumption. Data from 2015 to 2019 included approximately 500 in Rwanda. Realistic and evidence-based predictive modeling has the potential to provide optimal results. The management of essential medicine stock levels and storage facilities, as well as the reduction of safety inventories and waste, are constantly improving. RF predicted data with an 88% accuracy. It passed 76% of the tests and thus can be used to predict future consumption. The complex, multi-staged technique involves estimating the quantity of necessary medications and acts as a direction for future study. as a basis for choosing the right quantity to purchase [14]. Mbonyinshut et al. [11] focused on ML applications. An RF model using essential drug consumption data from 2015 to 2019 for about 500 medical items was used to forecast demand. With a 78% accuracy rate for the training set and a 71% accuracy rate for the testing set, the RF model accurately predicted the trend in demand for 17 critical drugs for NCDs. RF is an ML model that is well suited for difficult jobs like managing the supply chain for medical care because it can swiftly handle a variety of variables. Because the RF model does not provide reliable continuous normal predictions, data sets can be automatically balanced if a class is less prevalent in the data than other classes. This model has an R-square accuracy of 0.78 on a training set and 0.71 on a testing set [23]. Table 3 shows the previous work in which ML algorithms were used.

5 Deep learning methods

The next generation of conventional neural network technology is called deep learning. Similar to a neural network with numerous layers [3]. Significant advancements in computer science have aided in the creation of multilayered neural networks for deep learning. Therefore, one of the most crucial deep learning capacities is the finding of nonlinear patterns. The autonomous learning function of enormous datasets is where deep learning gets its power. It enables the best learning processes and makes it easier to extract attributes. Convolutional neural networks, deep belief networks, deep neural networks, and RNNs are the primary deep learning algorithms utilized in many ML applications [7,16].

RNNs are created to distinguish repeated characteristics and predict the next most probable event. Unlike traditional ANNs, they employ feedback loops. Although the structure of traditional RNNs theoretically permits the management of dependencies in long-term memory, their practical influence is minimal. As a result, the short-term sequences are better matched to the memory storage capacity of RNNs. The hidden layer built on conventional RNNs now has cell states and a gate mechanism, which means that its control gates can largely solve the gradient vanishing problem. Furthermore, a range of control gates with various functions is utilized to handle the past and present information if the historical message is transmitted to the neurons in the buried layer. RNNs are members of the family of RNNs when prior outputs are used as new inputs; they become recurrent, meaning that they are used once or more when determining new outputs. Through recurrent connections, a simple RNN accepts both the current input and the previous input as input. This emphasizes the significance of earlier events [18], as shown in Figure 4.

LSTM networks are RNNs that can manage both long-term and short-term dependencies. Hochreiter and Schmidhube initially showed them in 1997 [16,26]. Figure 5 shows the LSTM architecture.

Table 4 shows the strengths and weaknesses of the above-mentioned algorithms

The recent and well-known studies worthy of note that have been conducted on this issue are summarized. To forecast future consumption, Amalnick et al. [23] proposed a neural network. The major goal of the suggested approach is to combine neural network techniques with grouping and classification techniques. Data are grouped to do this before being entered into neural network models. The optimum structure is then chosen after many have been produced for each cluster. The information gathered from a pharmaceutical factory in Iran was applied. The prediction error has greatly decreased. The findings showed that clustering products not only improves prediction accuracy but also makes it possible to evaluate anticipated values more accurately for each individual cluster. The neural network's training process is slowed, and its accuracy suffers as a result. The dataset is not globally available [23]. Zdravkovi et al. used three forecasting techniques: LSTM, Prophet, and ARIMA, and the train-test split technique (52 weeks of test data). Mean squared error (MSE) was used as a loss of function for LSTM as well as a performance metric. To

Table 3: Previous works based on ML algorithms

Author(s)/year	Methods/techniques	Brief description	Dataset	Strengths	Weaknesses
Kravets et al. (2018) [1]	RF	The accuracy of models used to forecast the volume of drug sales in the Volgograd region is compared in this study	Sales of hydroxyzine in the Volgograd area from the start of 2014 to the end of 2017	In light of the influencing elements, compared to the identical models that solely use historical data, are more accurate	Accuracy can be improved The dataset is not globally available
Merkuryeva et al. (2019) [14]	Model SMA, Genetic programming combined with multiple linear regression and symbolic regression	Three experimental scenarios based on the use of various forecasting techniques are looked at	The historical weekly sales ACT0002UZ01product contains 41 data points	The findings demonstrate that the symbolic regression-based forecasting model offers the most accurate curve fitting to historical demand data as well as reduced error estimates for all scenarios and experiments	Forecasting accuracy is still big challenge. The dataset is not globally available
Kumari and Bohra (2021) [5]	RF and XGboost	Applying ML algorithms on raw data	Rossman Store Sales	Series modeling techniques Enhanced the obtained accuracy	Accuracy can be improved
Poyraz et al. (2020) [24]	LR, Gaussian process, m5rule, MLP, smoreg, M5P, RF	It aims to utilize the findings of the algorithm that generates a sales forecast the finest outcomes	Weka Arff pharmacy in Turkey, 5 years of pharmaceutical sales data are used	Gaussian processes achieved the best mean absolute percentage error	Accuracy can be improved
Tugay and Ögüdücü (2017) [22]	LR, RF, Gradient Boosting, and Decision Tree	All algorithms are taught in the initial step. Then, a combiner algorithm is employed	Real world dataset obtained from the e-commerce company	Demand predictions performed better with far less training data (just 20% of the dataset)	Accuracy can be improved
Nowadly and Jung (2020) [15]	RF, ANN, LR, and SVM	The creation of multi-step long-range forecasting models involves both direct and recursive methods.	Data were made by multiple manufacturers	In the one-step models, RF, ANN, and LR generated results that were reasonably accurate. However, the accuracy decreases as the forecasting horizon is extended using a multi-step forecast	Exogenous factors are not available The dataset is not globally available
Oliveira et al. (2021) [12]	KNN, SVM, RF, LR, and MLP	Compare LR and SVM against RF, KNN, and MLP	Coop service group	SVM shows how to predict the length of the purchasing lead time using high precision	Accuracy can be improved
Keny et al. (2021) [4]	SVR	The suggested model demonstrates the application of SVR to sales forecasting	The collection included 600,000 transactional data points collected over a 6-year period (2014–2019)	The suggested model makes it easier for performance development centers to recognize trends and seasonality by better presenting the sales data	Accuracy can be improved

(*Continued*)

Table 3: *Continued*

Author(s)/year	Methods/techniques	Brief description	Dataset	Strengths	Weaknesses
Mbonyinshuti et al. (2022) [11]	LR, RF, and ANN	The goal of this study is to see whether ML techniques might be used to improve demand forecasting accuracy	The dataset is based on data consumption from 2015 to 2019 in Rwanda	The accuracy of RF's data predictions is 88% with the train set and 76% with the test set	Accuracy can be improved
Mbonyinshuti et al. (2022) [25]	RF	By using the usage of vital drugs, RF was used to forecast demand	Top 17 non-communicable diseases are selected from medicines consumption data from 2015 to 2019	On a training set, RF's accuracy was 0.78, while on a testing set, it was 0.71	Accuracy can be improved

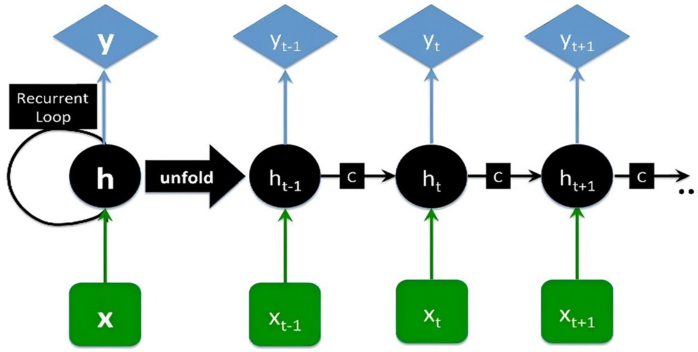

Figure 4: RNN architecture [10].

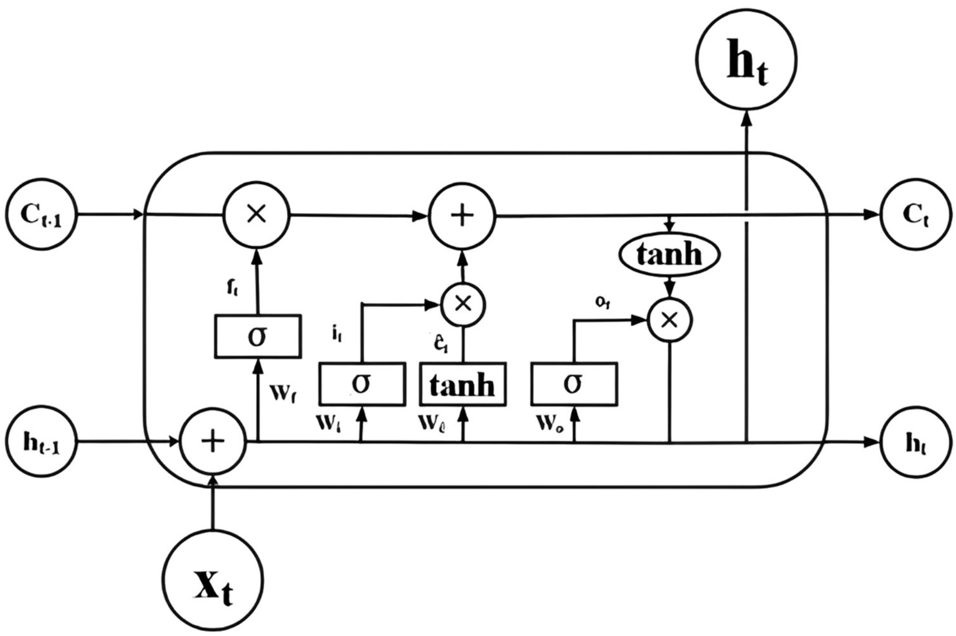

Figure 5: LSTM network's architecture [27].

establish the baseline, there were three tests. While the Nave and seasonal Nave were utilized for rolling predictions, the average technique served as a basis for long-term forecasting. Different methods and approaches related to the preparation, analysis, and forecasting of sales time-series data with the aim of facilitating recommended sales and marketing are validated. Sales of eight distinct categories of pharmaceutical products are forecasted based on forecasting strategies based on trend and seasonality characterized by a variety of traits, including stationarity, seasonality, residual amount, and sales data volatility. For a single distributor, a pharmacy chain, or even just one store, all of these evaluations and projections are conducted on a small scale. The initial dataset consisted of 600,000 transactional records collected over 6 years (2014–2019). Making potentially valuable deductions and suggestions for enhancing sales and marketing tactics is possible by this purpose. Then, the time-series in each group were examined for stationarity, autocorrelation, and predictability to derive the initial set of parameters for using the forecasting techniques. The limitation in this study was the lack of data and low accuracy [3].

Manjunath et al. [29] proposed RNN with LSTM. The proposed methodology is based on the seasonal number of prescriptions required with the number of quarters as an indicator. The aim of the research is to forecast the drug amounts needed for 1 year. An auto-encoder is used as a data abstraction for the shallow

Table 4: Strengths and weaknesses of deep learning algorithms [10,16,28]

Algorithm	Strengths	Weaknesses
Neural network	– Simple and fast technology – The availability of multiple training algorithms – The ability to implicitly reveal complex nonlinear relationships between independent variables – The ability to detect most possible interactions between predictor variables	– Need an extensive calculations and fast hardware – Determining the appropriate network architecture is a challenge – Sometimes the network has unexplained behavior, and it is difficult to monitor and debug the network
Deep neural networks	– The latest science that has come to it its architecture (the number and structure of layers) can be adapted to many kinds of problems – Hidden layers reduce the need for distinct geometry	– Usually not suitable as general-purpose algorithms because they require a large amount of data – Requires extensive calculations and time for training – Requires fast and sophisticated computers
RNN	– Very useful for knowing the sequence dependence of a given dataset (predicting the next value of the sequence)	– When the time gap is too long, it is highly challenging to transmit all of the information carried by the RNN. The network becomes untrainable when it has a lot of deep levels. This issue is known as the "vanishing gradient problem."
LSTM	– It can learn how to produce a multi-step prediction with a single step as well as learn long sequences, which can be useful for time-series forecasting	– The difficulty with LSTM is that it can be difficult to configure and may require setup to obtain data in the correct format for learning

neural network. The abstracts are used to train the second shallow network. The system consists of five steps, each of which has successfully accomplished the intended objective. The data were initially derived from actual world data. The gathering of data in a format suitable for time-series is the state's most crucial step. The test lasted 6 years, during which the input data were used for 5 years, and the anticipated data were used for the final year. In the second stage, visualization is used to make the data more understandable. This stage is crucial for understanding the distribution of the data and its behavior. The division of data into training and test sets is the third step. Fourth, a deep network of the LSTM type was created, and it had positive outcomes. After testing more than the deep network's design, the researchers arrived at the lowest error value. The outcomes of deep network analysis and real-world data are contrasted through visualization to illustrate network performance. To predict the future, the system has produced integrated and encouraging findings. Over the course of 6 years, the information was extracted for 135 medications, separated into four quarters (2011–2016). The data for the years (2011–2015) are used as the input for a training set. The numbers in 2016 were the expected or desired values. After being proven to be an efficient method of forecasting the handling of nonlinear input and output variables, neural networks are a highly well-known technology for predicting time-series, with the ability to estimate any function under certain conditions. In other words, the RNN network has demonstrated outstanding performance in a wide range of applications, including speech recognition, translation, and picture captioning. The application of LSTMs to forecasting time-series is what makes this network successful. If the data exhibit nonlinear behavior, the error's magnitude grows with time as uncertainty rises in line with expectations. Nonlinear latent connections cannot be modeled using linear approaches. Recently, datasets were not globally available [29]. Zhu et al.'s [20] proposed exploratory RNN analysis indicates that different drugs may have some similar demand patterns. Therefore, a new paradigm framework is designed to take advantage of data from various sources to train the model to capture these patterns. To further improve the prediction model across chains, the model selected (VAR and various models of ML) is being used, and it is suggested to use an RNN model with big data from major pharmaceutical manufacturers. The proposed model enables the cross-series training to be further optimized using a variety of grouping strategies and addresses the issue of a shortage

Table 5: Previous work based on deep learning algorithms

Author(s)/year	Methods/ techniques	Brief description	Dataset	Strengths	Weaknesses
Amalnick et al. (2020) [23]	Neural network	The major goal of the suggested approach is to combine neural network techniques with grouping and classification techniques	Data collected from an Iranian pharmaceutical factory	The prediction error has decreased significantly	Accuracy can be improved The dataset is not globally available
Zdravković et al. (2020) [3]	ARIMA, Prophet, and LSTM	Validate various techniques and strategies in relation to time-series for sales	The dataset included 600,000 transactional data points collected over a 6-year period (2014–2019)	Make findings and recommendations that could be useful for enhancing sales and marketing tactics	Lack of data Accuracy can be improved
Manjunath et al. (2021) [29]	RNN with LSTM, an auto-encoder	The system consists of five steps, each of which has successfully accomplished the intended objective	The data collected for 135 medications over a 6-year period was separated into four quarters (2011–2016)	RNN with LSTM produced good accuracy	The magnitude of the error increases over time The dataset is not globally available
Zhu et al. (2021) [20]	RNN	A new paradigm framework is designed to take advantage of data across various sources	Big data from major pharmaceutical manufacturers	Solves the problem of lack of data	Accuracy can be improved The dataset is not globally available
Galkin et al. (2022) [10]	Neural network	The study examines the principal forecasting techniques, alternative neural network topologies for time-series prediction, and neural network training techniques	A study of the pharmaceutical business AKRIKHIN JSC's income from 2016 to 2021 was done	The issue of under-learning was resolved by the choice of measures The number of MAE fell from 118 to 23 million	High time complexity The data have not been preprocessed
El Filali et al. (2022) [16]	LSTM	The goal of this endeavor was to create an LSTM model that could deliver the most accurate forecasts	The information used comes from a Moroccan pharmaceutical company's monthly sales figures for the period from 2012 to 2020	The model's prediction accuracy for a given time-series is improved. To use forecasts over extended periods of time and make long-term predictions, this kind can extract the most pertinent Using the data knowledge, divide the time-series signal into what is important in the long term and what is important in the short term	Accuracy can be improved The dataset is not globally available

of data for more advanced ML models [26]. Galkin et al. [10] showed the primary sales of a pharmaceutical company using neural network forecasting. The study analyzes the primary forecasting techniques, numerous neural network architectures that are employed to predict time-series, as well as neural network training techniques. An analysis of the revenue of the pharmaceutical company AKRIKHIN JSC was carried out from 2016 to 2021. They selected LSTM networks, a unique kind of neural network that can analyze lengthy sequences of numbers. There were more layers (six instead of four), and there were more neurons in each LSTM module (from 50 to 90). The problem of under-learning (where the model repeats the values in the training sample badly and predicts the values in the test sample poorly) was resolved by the choice of metrics. From 118 to 23 million, MAE fell. There are some weaknesses in this study to account for the occurrence of various demand conditions and the variety of the product selection. Due to seasonality and consumer tendencies to buy drugs in the future, the demand for medicines is highly unpredictable [12].

The objective of the work of El Filali et al. [16] was to create an LSTM model utilizing real data from the sales history of a pharmaceutical product of a Moroccan company, capable of giving the best forecasts possible. Comparing the proposed multilayer LSTM with the Grid search method to conventional approaches, multilayer RNNs, and monolayer LSTM models, the findings demonstrate that this approach has the lowest error measures, with an root mean squared error (RMSE) of 4487.32 and a symmetric mean absolute percentageerror of 0.026. This demonstrates that the fitted LSTM model can generate predictions that are more precise and effective. The data used come from a pharmaceutical company's monthly product sales from 2012 to 2020. 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. To use forecasts over extended periods of time and make long-term predictions, this kind of analysis may separate the time-series signal into what is significant in the near term and what is relevant in the long term, extracting the most pertinent information from the data. The complexity of non-linear data can be managed by the deep learning model, which can also automatically extract its features. Because of this, even with arbitrary parameters, it is simpler to get good prediction results on the first training run. Accuracy can be improved. The dataset is not locally available ref. [16]. Table 5 shows previous work in which deep learning algorithms were used.

6 Discussion

By analyzing the results for reviewed works, it was found that ML can use small amounts of data to make predictions. It requires features to be precisely defined and created by users. It divides the learning process into smaller steps and then combines the results from each step into a single output. It takes relatively little time for training, from a few seconds to a few hours. The output is usually a numeric value, such as a score or rating, whereas, when using deep learning, it needs to use large amounts of training data to make predictions. It learns high-level features from the data and creates new features by itself. It moves through the learning process by solving the problem on an end-to-end basis. Training usually takes a long time because the deep learning algorithm includes many layers. The output can have multiple formats, such as text, pitch, or audio. It was noticed that when using traditional forecasting methods such as ARIMA and exponential smoothing, the error values are larger compared to ML methods because they do not have the ability to deal with non-linear data, unlike deep learning algorithms, such as RNN, and LSTM, which have the ability to automatically determine the best model for forecasting demand as studies have shown its ability to capture the non-linear features found in time-series data. The results of measuring the error rate in previous works ranged between 2.9 and 0.029.

7 Conclusions

Since the pharmaceutical supply chain (pharma) has some distinctive characteristics, demand forecasting is essential for effective supply chains. Demand forecasting can be especially important for pharmaceutical (pharma) manufacturers for two reasons: (1) any mismatch between demand and supply could have an impact on patients and sometimes even put their lives in danger and (2) any unfulfilled demand could potentially result in a permanent loss of sales. Demand forecasting is given a lot of attention in the supply chain to fulfill customer demands, prevent any out-of-stock situations, and conserve resources. Previous literature on artificial intelligence has been studied to create models to predict demand for pharmaceutical products. They have used different ML algorithms such as SVM, RF, RNN, and others. In some works, more than one algorithm was used at the same time. The error was measured using various metrics, such as MSE, MAE, RMSE, and others. Because of the lack of data, they could not use complex models for prediction. Even when there is a long history of accessible demand data, these problems still exist because the old data may not be very useful when it changes market climate. Despite the favorable outcomes, there are still many challenges, such as the issues brought on by COVID-19, which triggered a serious health and economic crisis. Future work on time-series forecasting includes increasing the number of data points used, using deep learning algorithms such as LSTM, and bidirectional LSTM and measuring the error rate for each algorithm when using the same dataset to obtain the best model with the highest accuracy to predict demand for pharmaceutical products. Also, considering the influencing factors can be another future direction.

Author Contributions: The authors are equally contributed to the paper. Authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare that there is no conflict of interest.

Ethical approval: The conducted research is not related to either human or animals use.

Data availability statement: Data sharing is not applicable to this article since no datasets were generated or analyzed during this study.

References

[1] Kravets AG, Al-Gunaid MA, Loshmanov VI, Rasulov SS, Lempert LB. Model of medicines sales forecasting taking into account factors of influence. In Journal of Physics: Conference Series. Vol. 1015. No. 3. May 2018. p. 032073. doi: 10.1088/1742-6596/1015/3/032073.

[2] Hafiz M, Sabir Ibna Sazzad M, Hasan KI, Hasnat J, Mishu MC. Predicting the demand of prescribed medicines in Bangladesh using artificial intelligent (AI) based long short-term memory (LSTM) model. ACM International Conference Proceedings Series; 2020. p. 1–5. doi: 10.1145/3377049.3377056.

[3] Zdravković M, Đorđević J, Đorđević AC, Pavlović S. Case study: univariate time series analysis and forecasting of pharmaceutical products' sales data at small scale. ICIST. 2020:1–4.

[4] Keny S, Nair S, Nandi S, Khachane D. Sales prediction for a pharmaceutical distribution company. Int J Eng Appl Phys. 2021;1(2):186–91. https://ijeap.org/.

[5] Kumari A, Bohra N. Prediction of drug sales by using neural network algorithm. 2021:235–42. doi: 10.4108/eai.27-2-2020.2303124.

[6] Kumari A, Bohra N. Prediction of Drug Sales by Using Neural Network Algorithm. Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 27–28 February 2020, Jamia Hamdard, New Delhi, India. doi: 10.4108/eai.27-2-2020.2303124.

[7] Punia S, Nikolopoulos K, Singh SP, Madaan JK, Litsiou K. Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. Int J Prod Res. Aug. 2020;58(16):4964–79. doi: 10.1080/00207543.2020.1735666.

[8] Pavlyshenko BM. Sales time series analytics using deep Q-learning. Int J Comput. 2020;19(3):434–41. doi: 10.47839/IJC.19.3.1892.

[9] Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Taieb SB, et al. Forecasting: theory and practice. Int J Forecast. Jul. 2022;38(3):705–871. doi: 10.1016/j.ijforecast.2021.11.001.

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

[11] Mbonyinshuti F, Nkurunziza J, Niyobuhungiro J, Kayitare E. The prediction of essential medicines demand: A machine learning approach using consumption data in rwanda. Processes. Jan. 2022;10(1):4282–7. doi: 10.3390/pr10010026.
[12] Oliveira MB, Zucchi G, Lippi M, Cordeiro DF, Silva NR, Iori M. Lead time forecasting with machine learning techniques for a pharmaceutical supply Chain. 2021;1(no. Iceis):634–41. doi: 10.5220/0010434406340641.
[13] Pornima P, Boyapati S. Prediction of water consumption using machine learning algorithm. Lect Notes Electr Eng. 2021;698(3):891–908. doi: 10.1007/978-981-15-7961-5_84.
[14] Merkuryeva G, Valberga A, Smirnov A. Demand forecasting in pharmaceutical supply chains: A case study. Procedia Comput Sci. 2019;149:3–10. doi: 10.1016/j.procs.2019.01.100.
[15] Nowadly K, Jung S. Using machine learning approaches to improve long-range demand forecasting; 2020. https://dspace.mit.edu/.
[16] El Filali A, Ben Lahmer EH, El Filali S, Kasbouya M, Ajouary MA, Akantous S. Machine learning applications in supply chain management: A deep learning model using an optimized LSTM network for demand forecasting. Int J Intell Eng Syst. Apr. 2022;15(2):464–78. doi: 10.22266/ijies2022.0430.42.
[17] Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci. May. 2021;2(3):160. doi: 10.1007/s42979-021-00592-x.
[18] H. Saleh and J. A. Layous, Machine Learning – Regression, PhD thesis, Jan. 2022. doi: 10.13140/RG.2.2.35768.67842.
[19] Segura D, Khatib EJ, Barco R. Dynamic packet duplication for industrial URLLC. Sensors. Jan. 2022;22(2):587. doi: 10.3390/s22020587.
[20] Zhu X, Ninh A, Zhao H, Liu Z. Cross-series demand forecasting using machine learning: Evidence in the pharmaceutical industry. Prod Oper Manag. no. March 2019. 2021;1–54.
[21] Tugay R, Ögüdücü ŞG. Demand prediction using machine learning methods and stacked generalization. Sep. 2020, [Online]. http://arxiv.org/abs/2009.09756.
[22] Tugay R, Ögüdücü ŞG. Demand prediction using machine learning methods and stacked generalization. DATA 2017 - Proceedings of the 6th International Conference on Data Science, Technologies and Applications; 2017. p. 216–22. doi: 10.5220/0006431602160222.
[23] Amalnick MS, Habibifar N, Hamid M, Bastan M. An intelligent algorithm for final product demand forecasting in pharmaceutical units. Int J Syst Assur Eng Manag. Apr. 2020;11(2):481–93. doi: 10.1007/s13198-019-00879-6.
[24] Guerrero E, Poyraz İ, Gürhanlı A. Demand forecasting with time series analysis using drug sales data IJERA journal mark last JOURNAL OF APPLIED QUANT ITAT IVE MET HODS JOURNAL OF APPLIED QUANT ITAT IVE MET HODS Demand Forecasting with Time Series Analysis Using Drug Sales Data. Int J Eng Res Appl. www.ijera.com. 2020;10:51–4. doi: 10.9790/9622-1007035154.
[25] Mbonyinshuti F, Nkurunziza J, Niyobuhungiro J, Kayitare E. Application of random forest model to predict the demand of essential medicines for non-communicable diseases management in public health facilities. Pan Afr Med J. 2022;42. doi: 10.11604/pamj.2022.42.89.33833.
[26] Han Y. A forecasting method of pharmaceutical sales based on ARIMA-LSTM model. Proceedings - 2020 5th International Conference on Information Science, Computer Technology and Transportation ISCTT 2020; 2020. p. 336–9. doi: 10.1109/ISCTT51595.2020.00064.
[27] Li W, Wu H, Zhu N, Jiang Y, Tan J, Guo Y. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). Inf Process Agric. 2021;8(1):185–93. doi: 10.1016/j.inpa.2020.02.002.
[28] Zhu X, Ninh A, Zhao H, Liu Z. Demand forecasting with supply-chain information and machine learning: Evidence in the pharmaceutical industry. Prod Oper Manag. Sep. 2021;30(9):3231–52. doi: 10.1111/poms.13426.
[29] Manjunath CR, Rony SI, Nandi S, Ghosh RC, Pasha S. Prediction of medicine consumption using machine learning. J Emerg Technol Innov Res. 2021;8(5). www.ijcrt.org.