A Proposed Data Warehouse Framework to Enhance Decisions of Distribution System in Pharmaceutical Sector

Noura Mahmoud Abd Elazeem, Nevine Makram Labib, Aliaa Kamal Abdella

Computer and Information Systems Department
Sadat Academy for Management Sciences, Cairo, Egypt
nouramahmoud1391@gmail.com, nevmakram@gmail.com, aliaaabdella@yahoo.com

Abstract

Noncommunicable diseases, including cardiovascular diseases, diabetes, cancer, hypertension, and chronic respiratory diseases, are currently the leading national cause of death in Egypt. There are many problems facing the pharmaceutical companies such as increasing the cost of inventory and expired medicines. Recent trends of Business Intelligent have taken attention of both communities and research in pharmaceutical sector. Data warehousing techniques used to enhance decisions of pharmaceutical distribution department by predicting the sales of medicines. Auto regressive moving average model time series and Neural Networks are used to predict sales based on historical data.

Keywords: *Business Intelligence, Data Warehouse, Pharmaceutical Distribution, Decision Support, Auto Regressive Moving Average, Neural Networks.*

1. Introduction

Noncommunicable diseases (NCDs) are estimated to account for 82% of all deaths in Egypt and 67% of premature deaths. The 2011/2012 STEPwise survey, conducted by the Ministry of Health and Population, in collaboration with World Health Organization(WHO), revealed a significantly high prevalence of risk factors for NCDs among the adult population, including: a 24% prevalence of smoking and a growing use of shisha tobacco, one of the most overweight populations in the world, with 66% of women overweight , 42% obese and almost three quarters of the population not involved in vigorous activity, 17% prevalence of diabetes and 40% prevalence of hypertension. Egyptians have an average daily salt intake of 9 grams, nearly double the recommended allowance [1]

Data storage and information retrieval is a very important topic nowadays and affects a large number of people and economic agents, being a valuable source for decision making or increasing business. The efficient and effective use of information is particularly important in Business Intelligence (BI). It is important to understand the relationships between different aspects of the company to be derived towards specific objectives such as increasing the market share and improving customer satisfaction. Business Intelligence is critical in supporting decisions. This type of solution is due to the fact that companies are drowning in data that record in operational databases: payroll data, financial data, customer data, vendor data, and so on. These databases are typically tuned for each operation, such as retrieving a single customer order, or for specific batch jobs, such as processing payroll at the end of each month. These databases are not designed to communicate with one another, allowing users to explore data in an unusual way, or to provide high level summary data at once [2].

Data warehousing (DWH) is popular in business environments and encapsulates the process of transforming and aggregating operational data and bringing it to a platform optimized for efficient storage and advanced analysis [3]. Data warehouse has offered an excellent solution towards right decisions in pharmaceutical companies. A Data Warehouse offers a solution designed to enable these enterprises to easily obtain relevant, accurate and up-to date information about prescribers, managed care organizations, wholesalers, distributors and consumers. Pharmaceutical companies have a growing need to combine the large amounts of their downstream data such as supply chain or inventory information with their abundance of sale data. This data must be combined for a clear picture of the supply chain to be integrated into a data warehouse. A Data Warehouse allows companies to confidently make product and research decisions based on integrated, detailed product and portfolio life cycle data. Integrated data from across the company allows pharmaceutical firms to determine the types of drugs to focus its research initiatives on. The pharmaceutical firm can make "right time" decisions by analyzing data from the DWH. It can also gain insight into its market share as well as those of its competitors [4].

Enhancing sales and operations planning through forecasting analysis and business intelligence is demanded in any industry and business. Sales forecasting, in pharmaceutical distribution companies plays a major role for enterprises in making business plans more accurate and gaining competitive advantage. Data mining (DM) methods are used to analyze large observational data sets, find unsuspected relationships, and discover patterns and trends [5].

The researcher collects from Egyptian Company for Medicine Trade, historical data for ten years about diabetic and hypertension drugs to build databases (DBs) in DWH. Data mining techniques such as Neural Networks (NN) and time series were applied for sales prediction to enhance decisions in pharmaceutical sector.

2. Objective of The Study

The main objective of the study is to propose a data warehouse framework to enhance decisions of distribution systems in pharmaceutical companies to decrease the medicine industry cost and increase the productivity.

3. Previous Studies

Egypt was ranked 8th highest country in the world in terms of diabetes rates in 2013. The prevalence of diabetic in Egypt was found in around 15.6% of all adults aged 20 to 79 in 2015. The World Bank reported an even higher percentage (16.7%). Today, more than 7.8 million Egyptians suffer from diabetic and this number is expected to double by 2035 [6].

A number of people with diabetes from 2005 to 2014 shown in Figure1:

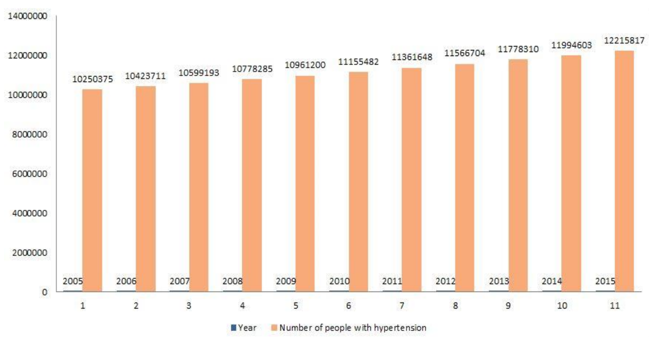


Figure1. Diabetes Rates in Egypt

Hypertension is a chief public health care in both developing and developed countries. Hypertension affects approximately 1 billion individuals worldwide. Egypt was ranked 4th highest country in the world and the first one in Africa in terms of Hypertension rates.

Hypertension is an urgent health problem in Egypt with prevalence rate of 26.3% among the adult population. Its incidence increases with aging, around 50% of Egyptians over the age of 60 years have hypertension [7].

A number of people with hypertension from 2005 to 2015 shown in Figure2:

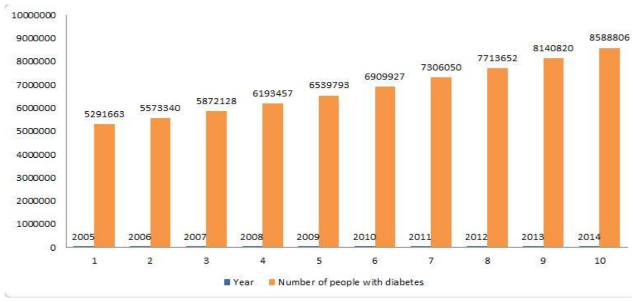


Figure2. Hypertension Rates in Egypt

The amount of expired medicines in 2017 was about 600 million EGP for diabetic, hypertension and cardiovascular medicines in Egypt, according to the Ministry of Health [8].

The techniques of data mining and data warehouse used for prediction in pharmaceutical and healthcare industry that shown in Table 1:

Table1.Techniques Used in Pharmaceutical and Health Care Sector

Domain	Name of Paper	Techniques
Predict sales of pharmaceutical distribution companies	Intelligent Sales Prediction for Pharmaceutical Distribution Companies: A Data Mining Based Approach [9]	Network analysis tools and time series forecasting methods
Drug consumption forecasting in pharmaceutical industry production planning	Application of Data Mining Techniques in Drug Consumption Forecasting to Help Pharmaceutical Industry Production Planning [10]	Artificial Neural Networks (ANN) and Decision Tree (DT)
Healthcare domain to diagnosis diabetes disease	An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques [11]	Decision Tree and K-Nearest Neighbor (KNN)
Healthcare domain to analyze diagnosis and treatment of Breast Cancer disease	Data Mining Techniques in Health Informatics: A Case Study from Breast Cancer Research [12]	Decision Tree
Healthcare domain to predict kidney diseases	Data Mining Techniques for the Prediction of Kidney Diseases and Treatment [13]	Decision Tree
Healthcare domain to predict diabetes	Performance Analysis of Data Mining Classification Techniques to Predict Diabetes [14]	Decision Tree
Healthcare domain in cancer diseases.	The Technology of Using A Data Warehouse to Support Decision-Making In Health Care [15]	DWH, OLAP
Healthcare domain in influenza diseases.	Health Care Data Warehouse System Architecture for Influenza (Flu)Diseases [16]	DWH, OLAP

There are many techniques used in healthcare and pharmaceutical sector that show in previous studies table, such as: OLAP, Decision tree, ANN, Time series, K-NN, Classification and Clustering. Most of the studies used decision tree, ANN and time series because they proved that these techniques give the best results in prediction. So, the researcher found that ANN and time series are efficient techniques for the historical data of Egyptian Company for Medicine Trade because it has no correlation between its variables and it has non-linear trend.

4. Data Warehouses in Healthcare

A framework is a conceptual or actual structure prepared to serve as a conductor or support for building of something that extends the structure into something useful [17]. The main goal of the proposed framework was to enhance decisions in the distributed pharmaceutical company based on sales prediction. The prediction techniques were chosen based on performance evaluation.

4.1. Framework Description

The framework can be described in four phases that shown in Figure3. Phase one is consisted of data preparation phase which has four steps (data collection, building DBs, DWH and data cleaning). Phase two is consisted of training phase which is applying time series to three types of Neural Networks techniques (levenberg marquardt, Bayesian regularized, and Scaled conjugate gradient).Phase three is testing the performance based on mean square error (MSE). Phase four is consisted about evaluating the performance of the best prediction model.

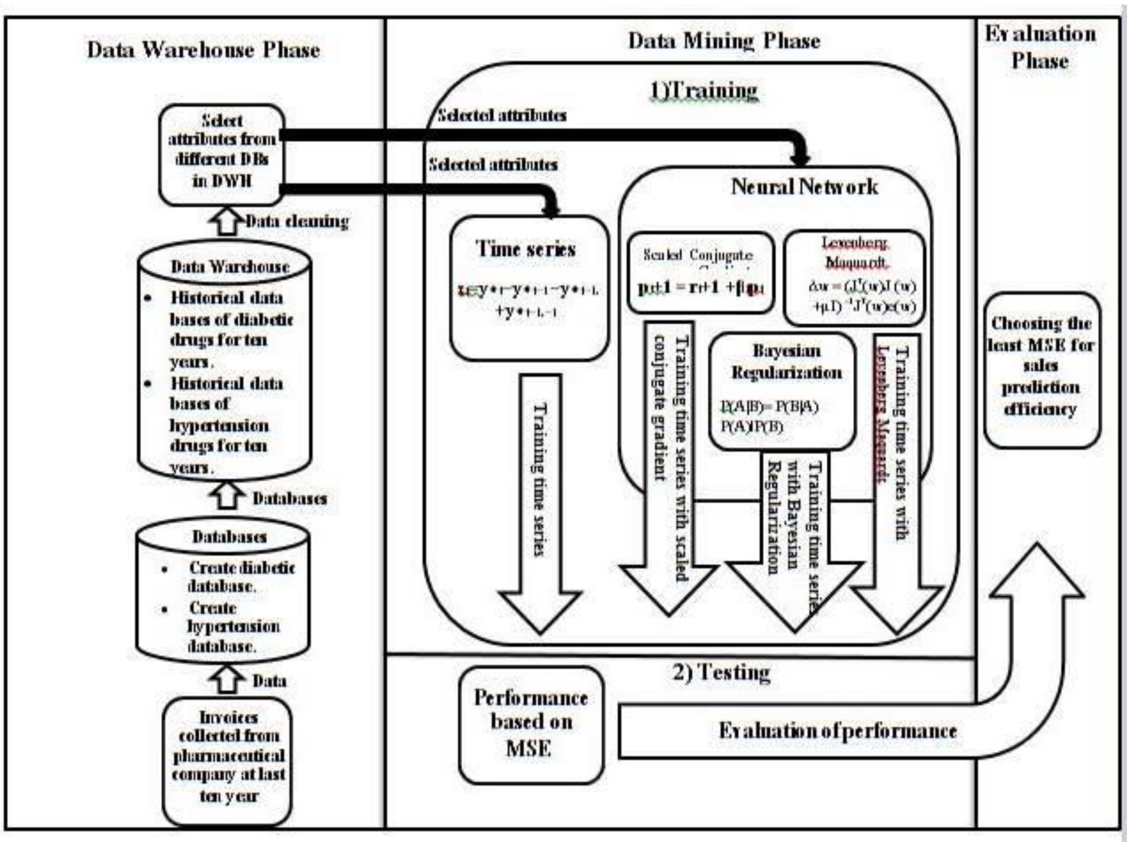


Figure3. The Proposed Framework of Sales Prediction

4.2. Data Warehousing Phase

There are four steps to design and implement DWH that can be shown as the following:

4.2.1. Data Collection

Data were collected from Egyptian Company for Medicines Trade by gathering invoices about Diabetic and Hypertension drugs for the last ten years from 2008 to 2017. There are twenty different brands of Diabetic drugs with different amount of sales for each year and thirty different brands of Hypertension drugs with different amount of sales for each year.

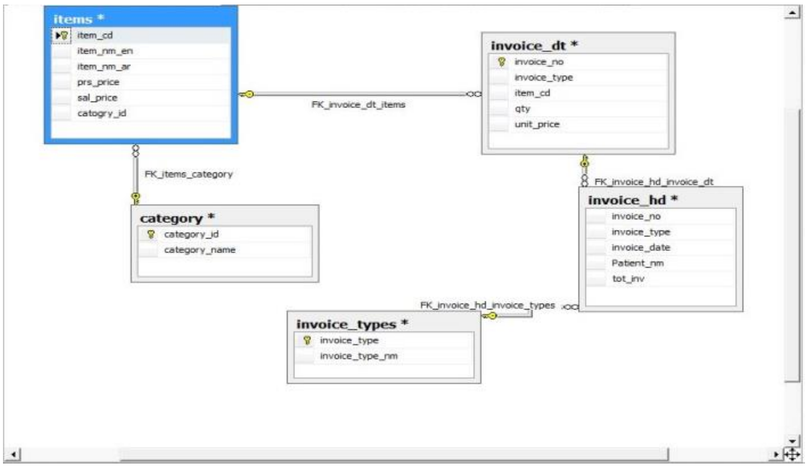
4.2.2. Building Databases

Database is one of the main components of the Information system (IS). The main goal of IS is to transform data into information which used to produce knowledge needed for decision making. The information system DB's purpose should be able to take data and provide tools for aggregation and analysis to help for decision making. Database is an organized collection, because in a database, all data is described and associated with other data [18]. Databases organize in many different ways and take many forms. The most popular form of DB today is the relational DB. A relational DB is a set of described tables from which data can be accessed or grouped in many different ways without having to reorganize the DB tables [18].

The relational DB for pharmaceutical company can be showed in five tables with their relationships in Figure4 as the following:

Figure 4. Relational Database

Invoice_dt table has attributes:



- Invoice_no (with a primary key)
- Invoice_type
- Item_cd
- Qty
- Unit_price

This table has "one to many" relationship with table called Item that has attributes:

- Item_cd (with a primary key)
- Item_nm_en
- Item_nm_ar
- Prs_price
- Sal_price
- Category_id

And Invoice_dt table has indirect relationship with Invoice_type table that has attributes:

- Invoice_type(with a primary key)
- Invoice_type_nm

And Invoice_hd table has attributes:

- Invoice_no
- Invoice type
- Invoice_date
- Tot_invoice

Finally, Category table has “one to many “relationship with Item table and has two attributes:

- Category_id (with a primary key)
- Category_name

4.2.3. Data Warehouse

The purpose of DWH is to take large data from heterogeneous sources and prepare them in known formats that helps in understanding and for making decisions. Data warehouse is providing direct access by using graphical tools for querying and reporting [19, 20]. The data about diabetic and hypertension drugs was collected together which relating to various DBS of individual years at last ten years as shown in Figure5:

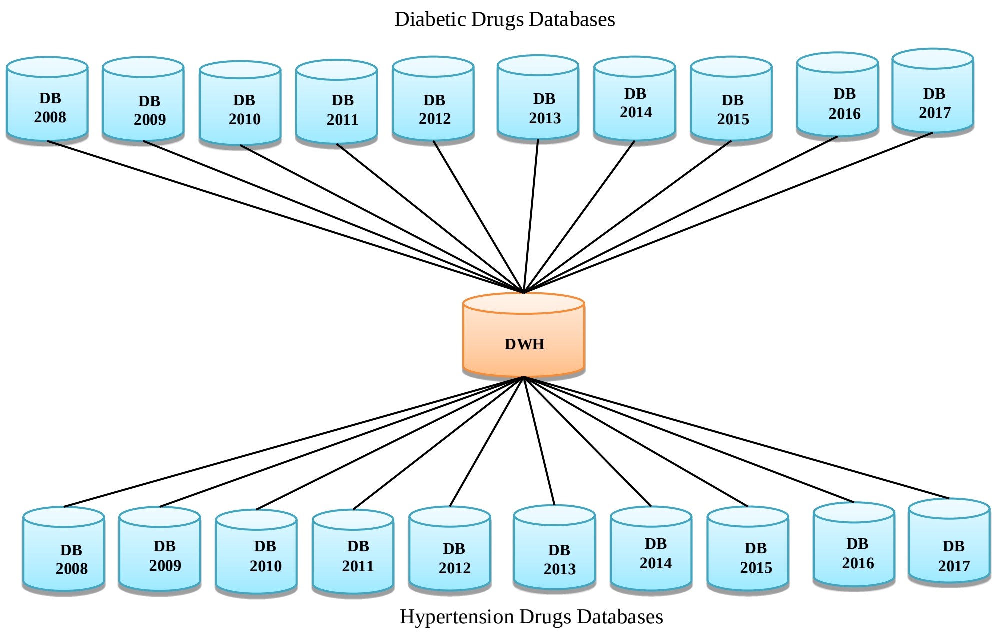


Figure5. Schematic Diagram for Data Warehouse

4.2.4. Data Cleaning

Data cleanup is the process of looking for and fixing inconsistencies to ensure that data is accurate and complete [21]. The main purpose of data cleaning is to detect, correct errors and inconsistencies from data to develop data with its characteristics. The characteristic of data is to be complete, accurate, accessible, economical, flexible, reliable, relevant, simple, timely, verifiable, and secure. It is confirmed to be a difficult but unavoidable task for any IS [22].

The Items, Invoice_dt and Category tables are determined to select most significant attributes (quantity, category_name, and item_name_en) for prediction in data mining techniques.

4.3. Applying Time Series Model

A time series is one of predictive data mining techniques. It is a set of numbers that measures the status of some activity over time. It is also a collection of data recorded over a period of time, weekly, monthly, quarterly, or yearly. Time series equation ($zt=y*t-y*t-1-y*t-L+y*t-L-1$) [23].

4.3.1. Time Series Forecasting in Diabetic Drugs

Mean squared error is calculated as the average of the forecast error values [24]. The results of applying time series in diabetic drugs data at the last ten years show in Table 2 and Figure6:

Table2. Time Series Forecasting for Diabetic Drugs

Years	Mean Squared Error
2008	470074.75
2009	647949.75
2010	645752.25
2011	617818.0625
2012	626656
2013	1139090.04
2014	1773367.4375
2015	2159805.25
2016	2815866
2017	2886640.54

The results showed that the best predicated year was in 2008 that have the lowest mean squared error.

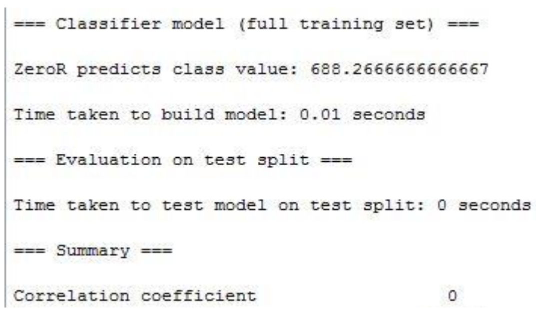


Figure 6. Classifier Model in Diabetic Drugs Data

There is no correlation coefficient between variables as shown in this figure.

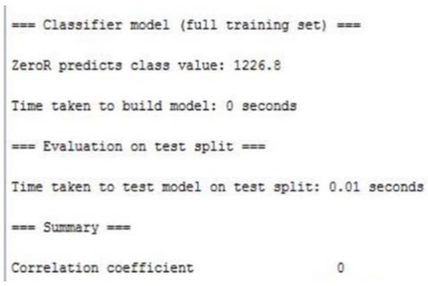
4.3.2. Time Series Forecasting in Hypertension Drugs

The results of applying time series in hypertension drugs data at the last ten years show in Table 3 and Figure7:

Table3.Time Series Forecasting for Hypertension Drugs

Years	Mean Squared Error
2008	226693.5378
2009	218138.2233
2010	211518.037
2011	184188.3363
2012	194010.8619
2013	204505.3611
2014	211581.9167
2015	216821.1389
2016	215795.5285
2017	221755.4563

The results showed that the best predicated year was in 2011 that have the lowest mean squared error.



```
=== Classifier model (full training set) ===

ZeroR predicts class value: 1226.8

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correlation coefficient                  0
```

Figure7. Classifier Model in Hypertension Drugs Data

There is no correlation coefficient between variables as shown in this figure.

There are observations from the previous results that can be conducted as follows:

1. The future prediction of trained data is nonlinear for both diabetic and hypertension drugs as shown in Figure8 and Figure9:

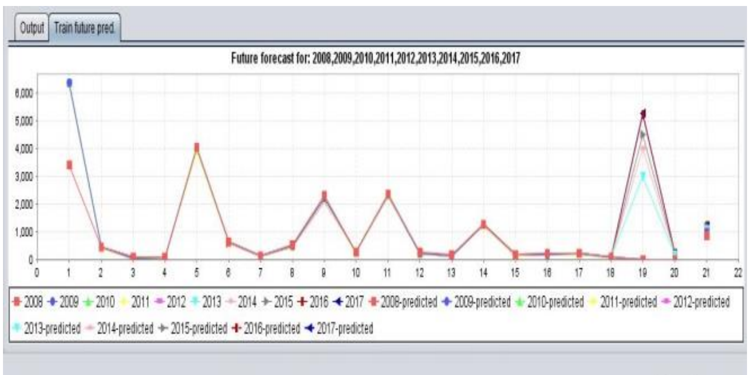


Figure8. Train Future Prediction in Diabetic Drugs Data

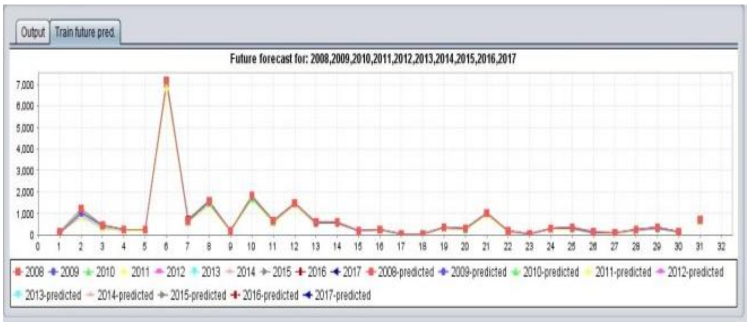


Fig.9. Train Future Prediction in Hypertension Drugs Data

2. Time series are very complex because each observation is dependent upon the previous observation [25].
3. Random error is influential from one observation to another. These influences are called autocorrelation dependent relationships between successive observations of the same variable [25].
4. Focus on univariate data with linear [25].

4.4. Applying Neural Networks

Neural Network is known as a computing system that contained of a number of simple, highly interconnected processing elements, which it can processing information by their dynamic state to external inputs. The Neural Networks using a dynamic network and helps users to selecting data, training, validation, and testing sets, and training the network [26].

4.4.1. Neural Networks Models

There are three models to a high order Neural Networks at MATLAB:

1. The Levenberg-Marquardt (LM) is a numerical least-squares non-linear function minimization technique. The LM computes the weight change according to:$\Delta w = (JT(w)J(w) + \mu I) -1JT(w)e(w)$ [27].
2. Bayesian regularized Neural Networks (BRNNs) restricts the magnitude of the weights by using the equation: $P(A|B) = P(B|A)\ P(A)/P(B)$ [28].
3. Scaled Conjugate Gradient adds to the complexity of the training procedure by performing a line search in each iteration to find the best step size along the conjugate direction. Instead of using a line search, the scaled conjugate gradient method uses a Levenberg-Marquardt approach to determine the optimal step size at each iteration by using the equation: $pt+1 = rt+1 + \beta tpt$ [29].

4.4.2. Applying Neural Networks in Diabetic Drugs

1. Applying LM Technique in Diabetic Drugs

Levenberg-Marquardt is a Neural Networks training function that updates weight and bias values of the Neural Network. LM technique is often the fastest backpropagation technique, and is highly recommended as a first choice supervised technique for training moderate sized (up to several hundred weights) feed-forward Neural Network [27].

- Training LM technique in diabetic drugs data was shown in Figure 10:

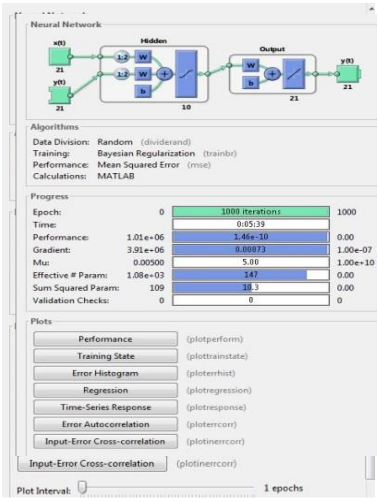


Figure10. Levenberg-Marquardt of Neural Networks in Diabetic Drugs

The results of implementing LM technique declared that the performance (mean squared error) of Diabetic drugs is (149582.28294).

- Regression(R) is values measure the correlation between outputs and targets. An R value of 1 means a close relationship, 0 a random relationship [29]. Regression of LM technique in Diabetic drugs was shown in Fig.11:

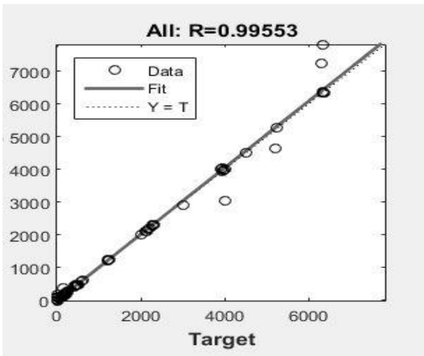


Figure11. Levenberg-Marquardt of Neural Networks in Diabetic Drugs about Regression

The regression between values is (0.99553) which mean random relationship between values.

2. Applying Bayesian Regularization Technique in Diabetic Drugs:

This technique typically takes more time, but can result in good generalization for difficult, small or noisy datasets. Training stops according to adaptive weight minimization (regularization) [28].

- Training Bayesian Regularization technique in diabetic drugs data was shown in Figure12:

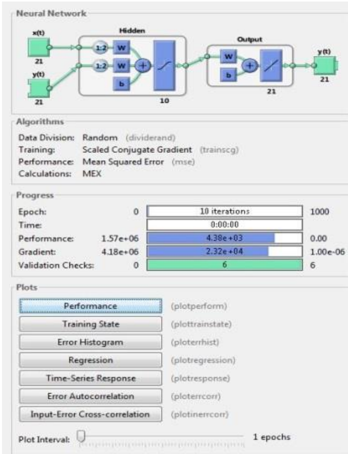


Figure12. Bayesian Regularization of Neural Networks in Diabetic Drugs

The results of implementing Bayesian Regularization technique declared that the performance (mean squared error) of Diabetic drugs after training is (306.10614).

- Regression of Bayesian Regularization technique in diabetic drugs data was shown in Figure13:

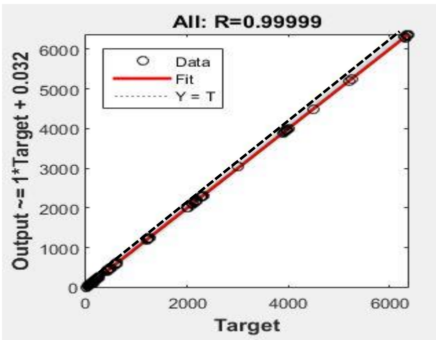


Figure13. Bayesian Regularization of Neural Networks in Diabetes Drugs about Regression

The regression between values is (0.99999) which mean random relationship between values.

3. Scaled Conjugate Gradient Technique in Diabetic drugs:

This technique takes less memory. Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples [30].

- Training Scaled Conjugate Gradient technique in diabetic drugs data was shown in Figure14:

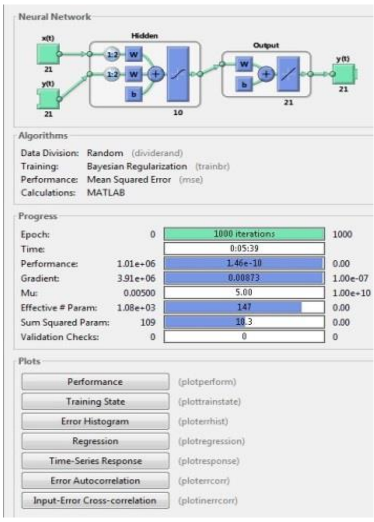


Figure14. Scaled Conjugate Gradient of Neural Networks in Diabetic Drug

The results of implementing Scaled Conjugate Gradient technique declared that the performance (mean squared error) of Diabetic drugs after training is (260490.75382).

- Regression of Scaled Conjugate Gradient technique in diabetic drugs data was shown in Fig.15:

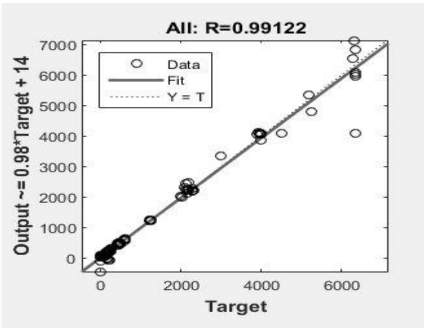


Figure15. Scaled Conjugate Gradient of Neural Networks in Diabetic Drugs about Regression

The regression between values is (0.99122) which mean random relationship between values.

3.4.2. Applying Neural Networks in Hypertension Drugs

1. Applying LM Technique in Hypertension Drugs Data

- Training LM technique in hypertension drugs data was shown in Figure16:

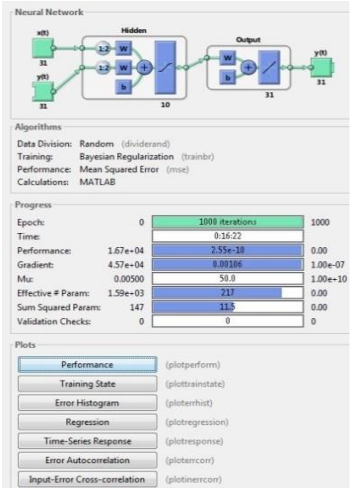


Figure 16. Levenberg-Marquardt of Neural Networks in Hypertension Drugs

The results of implementing LM technique declared that the performance (mean squared error) of Diabetic drugs is (5806.70695).

- Regression of LM technique in hypertension drugs data was shown in Figure17:

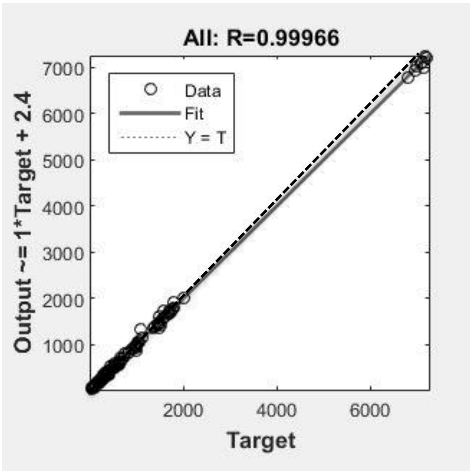


Figure17. Levenberg-Marquardt of Neural Networks in Hypertension Drugs about Regression

The regression between values is (0.99966) which mean random relationship between values.

2. Applying Bayesian Regularization Technique in Hypertension Drugs Data:

- Training Bayesian Regularization technique in hypertension drugs data was shown in Figure18:

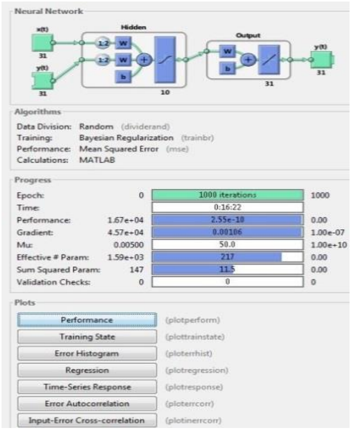


Figure18. Bayesian Regularization of Neural Networks in Hypertension Drugs

The results of implementing Bayesian Regularization technique declared that the performance (mean squared error) of Diabetic drugs is (426.43748).

- Regression of Bayesian Regularization technique in hypertension drugs data was shown in Figure19:

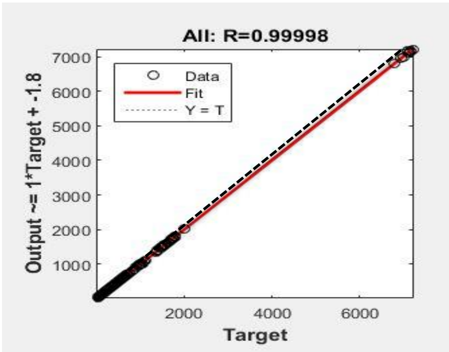


Figure19. Bayesian Regularization of Neural Networks in Hypertension Drugs about Regression

The regression between values is (0.99998) which mean random relationship between values.

3. Applying Scaled Conjugate Gradient Technique in Hypertension Drugs Data:

- Training Scaled Conjugate Gradient technique in pressure drugs data was shown in Figure20:

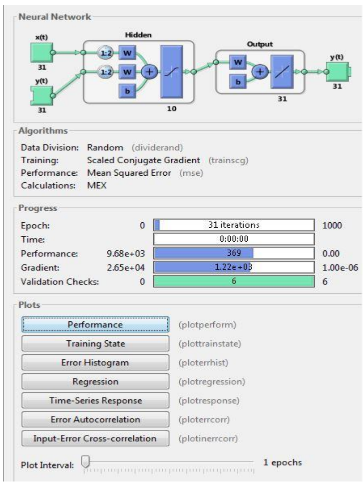


Figure20.Scaled Conjugate Gradient of Neural Networks in Hypertension Drugs

The results of implementing Scaled Conjugate Gradient technique declared that the performance (mean squared error) of Diabetic drugs is (6515.51252).

- Regression of Scaled Conjugate Gradient technique in pressure drugs data was shown in Figure21:

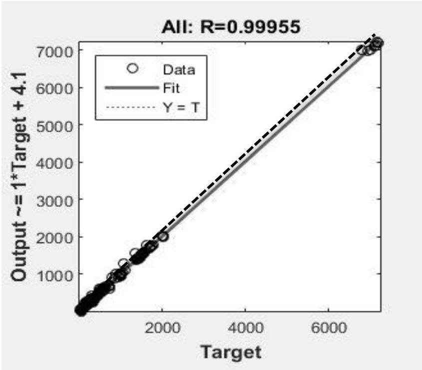


Figure 21.Scaled Conjugate Gradient of Neural Network in Hypertension Drugs about Regression

The regression between values is (0.99955) which mean random relationship between values.

4.4 Comparison between Applied Techniques

The researcher made a comparison between all the applied techniques to focus on the technique with mean square error that shown in Table 4.

Table4. Comparison between Applied Techniques

Applying neural time series	Neural network			Time series
	Levenberg Marquaradt	Bayesian Regularization	Scaled conjugate gradient	
Diabetic drugs	149582.28294	**306.10614**	260490.75382	470074.75
Hypertension drugs	5806.70695	**426.43748**	6515.51252	184188.3363

The Neural Networks and time series techniques are applying in Diabetic drugs data to know the MSE for each technique. First, the results of applying Neural Networks techniques present the MSE of LM technique was (149582.28294) ,the MSE of Bayesian Regularization technique was (306.10614) and the MSE of Scaled conjugate gradient technique was (260490.75382). Second, the MSE of time series technique was (470074.75). By applying Neural Networks techniques and time series technique in Hypertension drugs data to knowing the MSE for each technique, there are many results. First, the results of applying Neural Networks techniques present the MSE of LM technique was (5806.70695) ,the MSE of Bayesian Regularization technique was (426.43748) and the MSE of Scaled conjugate gradient technique was (6515.51252). Second, the MSE of time series technique was (184188.3363).

The data showed that time series technique was not best technique in prediction with Diabetic and Hypertension drugs data because the data was nonlinear and non-smoothly. So time series is preserved under a smooth change of coordinates is trajectory crossing (or lack thereof). As shown in results of implementation, a deterministic system once enough lags are used the dynamics will lie on a single-valued surface. Forecasting in the case of smooth data reduces to modeling the shape of the surface so it is not effective with non-smooth data.

In addition to the Bayesian Regularization was the best technique in Diabetic and Hypertension drugs data because it is searching for hidden correlation relationship that makes data smooth. BR training aims to minimize the sum of mean squared errors although it takes time to get results. So Neural Networks are capable of fitting linear and nonlinear functions without the need for knowing the shape of the underlying function so that it is more suitable for nonlinear time series prediction.

5. Conclusion ,Recommendation and Future Work

This paper proposed a framework to enhance decisions of distributed system in Egyptian company for medicine trade. The selection of an effective prediction technique may be based on comparative tests which cover many forecasting techniques such as Neural Networks and time series.

The conclusions can be summarized as follows:

1. Tests and comparisons between a number of techniques for both Neural Networks and time series model resulted in identifying the Neural Networks is the best sales prediction technique for both drugs based on the least mean square error.
2. Applying the technique of time series for data mining in diabetic and hypertension drugs data showed that the time series was not efficient with nonlinear and non-smoothly data .
3. Adopting the techniques of Neural Networks for data mining is more efficient with nonlinear data that have no correlation with no smooth data. By implementing diabetic and hypertension drugs data, the results shown the best performance for both drugs data in Bayesian Regularization technique which gives best prediction of sales.

Accordingly, it is recommended to apply the proposed model so as to enhance decisions of distribution systems in pharmaceutical companies to decrease the medicine industry cost and increase the productivity.

The researcher recommends the following as future work:

1. Build forecasting model by using NNs technique in Egypt to decrease the expired medicines so it will also help in the sustainable development of the pharmaceutical sector.
2. Developing a proposed web based framework for feedback evaluation.

References

[1]. "World Health Organization," 2018. [Online]. Available: http://www.emro.who.int/egy/programmes/noncommunicable-diseases.html.

[2]. M. D. B. M. P. Crescenzio Gallo, Data Warehouse Design and Management:Theory and Practice, Italy, 2010.

[3]. A. N. a. S. Al-Shammari, "The Gmuware Biomedical Data Warehouse System for Integrating Heterogeneous Biomedical Data Sources," *Egyptian Computer Science Journal,* vol. 31, no. 1, Jun 2009.

[4]. L. Campbell, "An EDW remedies headaches in the pharmaceutical industry," 2015. [Online]. Available: http://apps.teradata.com/tdmo/v08n02/Viewpoints/IndustryInsights/PrescriptionForSuccess.aspx.

[5]. V. H. H. a. D.-R. L. Hani Omar, "A Hybrid Neural Networks Model for Sales Forecasting Based on ARIMA and Search Popularity of Article Titles," *Computational Intelligence and Neuroscience journal,* 2016.

[6]. T. W. B. Group, " The World Bank," 2017. [Online]. Available: https://data.worldbank.org/indicator/SH.STA.DIAB.ZS.

[7]. D. M. Hasan, "Hypertension in Egypt: A Systematic Review," vol. 10, 2014.

[8]. "The Ministry of Health," [Online]. Available: http://www.mohp.gov.eg/.

[9]. M. M. S. H. F. Neda Khalil Zadeh, "Intelligent Sales Prediction for Pharmaceutical Distribution Companies: A Data Mining Based Approach," *HINDAWI,* 2014.

[10]. S. M. a. M. M. Rouzbeh Ghousi, "Application of Data Mining Techniques in Drug Consumption Forecasting to Help Pharmaceutical Industry Production Planning," in *International Conference on Industrial Engineering and Operations Management*, Istanbul,Turky, 2012.

[11]. M. S. U. Z. a. M. R. H. Emrana Kabir Hashi, "An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques," in *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Bangladesh, 2017.

[12]. A. H. D. R. a. M. K. Jing Lu, "Data Mining Techniques in Health Informatics: A Case Study from Breast Cancer Research," in *International Conference on Information Technology in Bio- and Medical Informatics*, United Kindom, 2015.

[13]. M. A. A. a. M. K. Jain, "Data Mining Techniques for the Prediction of Kidney Diseases and Treatment," *International Journal Of Engineering And Computer Science,* vol. 6, no. 2, pp. 20376-20378, 2017.

[14]. M. S. ,. A. G. a. K. K. Sajida Perveena, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," *ELSEVIER,* 2016.

[15]. D. O. E. a. A. N. Eldeen, "The Technology Of Using A Data Warehouse To Support Decision-Making In Health Care," *International Journal of Database Management Systems,* 2013.

[16]. R. Dutta, "HEALTH CARE DATA WAREHOUSE SYSTEM ARCHITECTURE FOR INFLUENZA (FLU)DISEASES," *Department of Computer Science & Engineering, Global Institute of Management & Technology,* 2013.

[17]. Jeremiah, "BRAND DEVELOPMENT PROCESS VS. FRAMEWORK," [Online]. Available: http://jeremiahgardner.com/blog/brand-development-process-vs-framework/.

[18]. D. B. a. D. T. Bourgeois, "Chapter 4: Data and Databases," in *Information Systems For Business and Beyond*, 2013.

[19]. U. S. B. R. M. A. S. A. K. A. A. a. Q. J. Muhammad Bilal Shahid, "Application of Data Warehouse in Real Life: State-ofthe-art Survey from User Preferences' Perspective," *International Journal of Advanced Computer Science and Applications,* 2016.

[20]. R. M. H. a. M. I. M. Yehia M. K. Helmy, "A Survey on Data Warehouse Quality," *Egyptian Computer Science Journal,* vol. 26, no. 2, Jul 2004.

[21]. S. Greengard, "Organizing data and information," in *Information Technology Concept*, 2000.

[22]. T. Peng, "A FRAMEWORK FOR DATA CLEANING IN DATA WAREHOUSES," 2003.

[23]. A. T. J. a. L. Tay, "Introduction to Time Series Analysis for Organizational Research: Methods for Longitudinal Analyses," *SAGE,* 2017.

[24]. J. Moreno, "Artificial neural networks applied to forecasting time series," *Psicothema,* vol. 23, 2011.

[25]. S. S. a. S. R. G. Mahalakshmi, "A survey on forecasting of time series data," in *International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*, Kovilpatti, India, 2016.

[26]. B. S. a. P. K. Venugopalan, "Comparison of Neural Networks Training Functions for Hematoma Classification in Brain CT Images," *IOSR Journal of Computer Engineering,* vol. 16, 2013.

[27]. H. P. Gavin, "the Levenberg-Marquardt method for nonlinear least squares curve-fitting problems," 2017.

[28]. H. Okut, "Bayesian Regularized Neural Networks for Small and Big Data," *INTECH,* 2016.

[29]. S. E. M. A. a. H. F. Alaa F. Sheta, "A Comparison between Regression, Artificial Neural Networks and Support Vector Machines for Predicting Stock Market Index," *International Journal of Advanced Research in Artificial Intelligence,* vol. 4, 2015.

[30]. H. L. J. W. &. Y. T. Zhongbo Sun, "Two modified spectral conjugate gradient methods and their global convergence for unconstrained optimization," *International Journal of Computer Mathematics,* 2015.