Deep Neural Network for Enhancing Drug-Utilization Clustering

HadabKhalid Obayes*[1], Nabeel Al – A'araji[2], EmanAL-Shamery[3]

[1,3]College of Information Technology, University of Babylon, Iraq
[2]Ministry of Higher Education, Iraq
*Corresponding Author Email: hedhab@uobabylon.edu.iq

Abstract

Drug consumption data needs to be linked to the disease. The process of analyzing quantities consumed based on drug name and brand is complex. It needs to be accurate because it is involved in the provision, manufacture,and marketing of medicines. The aim of this paper is to obtain optimal clustersof the drugs according to utilization. Anew model is proposed for the clustering process, specifically the Disease_Drugs_Clustering_Deep_Nural_network (DDC_DNN) as a type of deep neural network. This model consists of four layers. In the first layer,the features have been adapted to the network weights. The normalization and standardization are satisfied in the second layer. The main contributions are concerned in the forming primal clusters according to neighbors' proximity and distance. In the third layer, the final clustering is organized by re-forming clusters depends on the calculation of cluster centers and merging of the nearest clusters according to a carefully selected threshold. Three diseases have been linked with their drugs to be the research data set (diabetes, leukemia,and allergy). The final clusters are optimal clusters. Silhouette validity score has been used to validate the quality of clusters. The result of the proposed model has been compared with the traditional method K-means. Silhouette score of the proposed model result was better than the result of the K-means for the data set.

Keywords: Deep Neural Network, Drug Utilization, Disease_Drugs_Clustering_Deep_Nural_network.

1. Introduction

The development of the field of healthcare and the pharmaceutical industry significantly led to the availability of data there for the area of healthcare is a fertile environment for scientific research. This field differs from other sectors because it is connected with human life. Furthermore, it takes priority since it should be under a high level of medical services regardless of cost and consumes a large proportion of the budget for each country. Pharmacy is perhaps one important field of healthcare. In the pharmaceutical industry, it relies on quantitative analyzes of clinical studies and drug marketing[1].

Data mining is a tool to discover the hidden information in the big dataset. On the other words, it is a collection of techniques used to identify the patterns that state the dataset behavior. It is used to extract important knowledge that helps to predict future events .DM is used in making a decision in the pharmacy industry and identifies the manufactory plans and drugs distribution for the doctors or the customer[2].

Machine learning methods are considered a format of DM. Machine learning is a branch of Artificial Intelligent (AI), which has two classes supervised and unsupervised method. Support vector machine, neural network and deep learning are examples of the machine learning [3].

Recently, deep learning has been interested from all research fields in medical and pharmacy application particularly [1]. Deep learning uses multiple layers of a neural network, it can treat with big data in the training process. According, The health care contains big data there for the DM is considered suit technique for this application which needs to continued development [4].

The research problem includes data set analysis of drug utilization, which represents consumption of drugs in the United States over several years. Drug utilization will be analyzed by the name of the product field representing the drug brand name. This dataset has been includingfields representing the number of prescription, units reimbursed, quarter, year, state, utilization type, and other important fields.

In this work, therefore, a data mining technique has been developing for processing the clustering of the drugs. The paper is structured as follows. Section 2, explain the basic idea and assumptions. Reviewing some related work is illustrated in Section 3. In Section 4, formulate the research problem and discuss their design requirements. Section 5 presents the results and discussions, and conclusions of the paper in Section 6.

2. Related Works

This section summarizes the most noteworthy studies related to this study. In [5] the authors have dealt with the demographic information, wealth distribution, and the drug industry. Therefore, they analyze big data related to the health and drug prices to predict a substitute drug to the custom. Data mining using to find the same drug to the customer with the same affection but with a lower price. ANN and decision tree can be utilized to that task.

OraluckPattanaprateep et. al, [6] employed in this article technique of data mining in the medical field, especially in the use of medicines. In this work, the researchers have extracted a set of a feature of the big data set in the hospital. Hence, they show the

use of the patient's anti-steroidal (analgesics) affecting the digestive system and to reduce the phenomenon of ulceration device Digestive oxidative inhibitors are used. This study applied a rule-of-link model (association rule model) for estimating reasonable no steroidal anti-inflammatory drugs and protective agents in the Outpatient DataSet.

Marjia Sultana et. al, [7] have developed data mining techniques to predict heart disease. Researchers have identified a set of features that are a strong indicator of heart attack risk. For that, they are using predictive techniques in order to build a model for predicting heart attack using Decision Tree, J48 and Bayes Net.

Continuously with this research scope, BasmaBoukenze et. al, [8] have used data mining techniques to predict the occurrence of Kidney failure through the use of multi-purpose techniques to compare the performance of these techniques and the best of them in the prediction process. This can be Support Vector Machine (SVM), Multilayer Perceptron (MLP), Decision Tree (C4.5), Bayesian Network (BN) and K-Nearest Neighbor (K-NN).

SajidaPerveen et. al, [8]also in their work used data mining techniques for the purpose of predicting diabetes by using big dataset recorded by hospitals. Also, in this study, they have used the j48 technique as well as the decision tree for the purpose of forecasting diabetes.

3. Basic Idea and Assumptions

In this section, we explain the main components as a technique which is related to our proposed model.

3.1. Drugs utilization analysis

Analysis of drug utilization is based on drug prescriptions and marketing or distribution in hospitals, pharmacies, clients, and doctors. It relies on databases, electronic records and exploratory questions [2]. The process of drug utilization analysis is used for a number of purposes, for example, it is utilized to track after-sales services for medicines by generic name and multiple brands (substituted brand drug) for the same drug. Thus, it is ensured that the substituted has the exact same properties as the generic drug[9]. Also, detect and determine the illicit drug use or taking the drug over-the-counter medicines [10]. Analysis of drug use over a long historical period can provide support in the planning and marketing of pharmaceutical production, by discovering consumption patterns and predicting production quantity required for future periods [11].

3.2. Deep Learning

Deep learning is defined as a field of machine learning. It is based on multiple computational layers that used for identifying the dataset obstructively. Deep learning has been applied in many areas such as image classification, drug discovery, and speech recognition. The methods of deep learning are a multi-representation learning method. They are obtained by a series of simple nonlinear units that transform representation from raw data into more complex functions. Hence, they can be trained and learn to perform functions such as classification, prediction, and recognition [12]. On the other hand, deep learning techniques deal with a big dataset. In modern computers, by dealing with big data, it can be reduced the probability of over-fitting and the ability to make it is easy to accomplish complex computation of neural networks.

After deep learning went to overcome the problems of neural networks, it became a focus of attention to use in the field of healthcare. Since the healthcare field is an appropriate environment for the application of deep learning. The availability of a large dataset and the existence of many research fields can use deep learning to solve them. Furthermore, based on many researchers (i.e., [12], [4], and[13]), deep learning is used to analyze the medical dataset to extract knowledge from it. There are many areas in the field of healthcare that have succeeded in using deep learning to solve problems: IBM has adopted this technique in the construction of Dr. Watson's application; Medical analysis of images provides a good level of accuracy. It is used in examining and detecting cancer and diseases that require monitoring to be detected by relying on various sources of screening such as X-Ray; CT and MRI scans. Also, Pathology was also based on the use of deep learning to monitor the impact of new drugs through the analysis of a large dataset. Genome representation which contains vast amounts of data needs to be processed, tracked and categorized also relied on deep learning [1]. Drug utilization analysis is the area of the proposed study. It is dealing with the medical industry, drug marketing and brands for distribution.

3.3 Feature Extraction

Feature extraction can be defined as a process of transformation from the space of data entered into a valid value space using a specific function. Features can be an input vector for learning algorithms. The characteristic of the present time is the availability of data is very large and very diverse, which can be millions of cases and examples of the data type. Thus, Image recognition, natural language processing, bioinformatics and medical are an example of applications that have a huge data set. The advantage of extracting the feature is to improve the accuracy of learning models by extracting features with values that are understood by the algorithm. This process eliminates noise and redundant data [14].

Moreover, It also aims to reduce the dimensions of the input data and reduce the required storage space in addition to increasing inference and training speed[12][15]. furthermore, this field represents very importantly for highlighting, since there are a set of terms used in it, such as extract feature, feature selection, feature normalization, feature construction. Finally, neural networks can be used to extract features from the big dataset. In addition, extraction of features from input data is referred to as learning of representations, which is related to the extraction of features across multi-layer neural networks. Deep neural network attains great success in this area. It uses multiple layers to represent data and usually, the upper layer consists of the representation of abstraction and nonlinearity that is hidden within the raw data [14].

3.4. Statistical Features

Statistical features are used to describe a dataset with a few numbers of values. So, when the values have a good central disposition, that means, the values of input data have the tendency to cluster around a given value[16],[17]
- Mean (μ) of the values (x1, x2,…, xN):

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{1}$$

Where (N) is the number of input data.
- The Standard Deviation formulated as follows

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)} \tag{2}$$

Also, the Skewness (Skew) is formulated as follows:

$$Skew(x_1, x_1, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{x_i - \mu}{\sigma} \right]^3 \tag{3}$$

3.5. Evaluation Process

The evaluation process for the proposed model has been done in two ways: qualitative and quantitative.

3.5.1. Qualitative Evaluation

This evaluation was performed using the scatter diagram for the visual comparison purpose of the resulting clusters.

3.5.2. Quantitative Evaluation

Silhouette Validity score: Silhouette depends on the average score for each point in the data set. Each individual point depends on the difference between the average distance of that point and each other point in its group. The minimum distance between that point and the other points in each other group. The process of division is then made on this difference by the term normalization, which is greater than the average:[18], [19].

$$\frac{1}{N}\sum_{i=0}^{N} S_{xi} \tag{4}$$

where N is the number of points in the data set and:

$$S_{xi} = (b_{q,i} - a_{p,i})/\max\{a_{p,i}, b_{q,i}\} \tag{5}$$

If x_i is a point in cluster p, then,

$$b_{q,i} = \min d_{q,i} \quad \text{where } d_{q,i}$$

is the average distance between point x_i and every point of cluster q. On the other hand, $a_{p,i}$ is the average distance between point x_i. and every other point of cluster p. The Silhouette measurement relates separation to compactness that is using subtraction instead of dividing. when the clustering improves, the result is close to 1.

4. The Proposed System

The proposed system involves four layers (the feature extraction layer, the normalization and standardization layer, the initial clustering layer and Re-forming clustering layer), as shown in Figure (1).
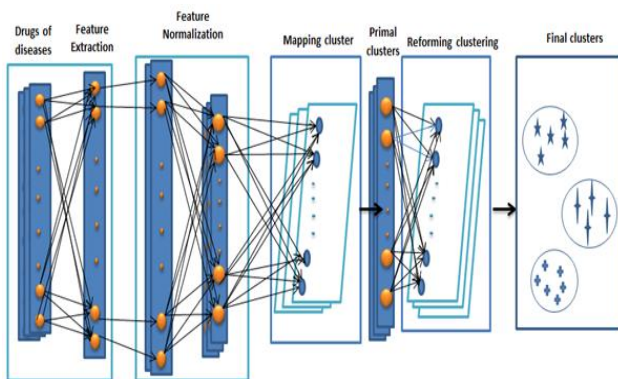

Fig. 1: Disease- Drugs clustering Deep Neural Network

4.1. Data Preparing

Deep learning needs a big dataset for training. USA official website https://www.data.gov provides abig dataset for the UAS drugs utilization over many years. Type of the dataset is a CSV file, which contains (20) attributes and many millions of records. Dataset attributes such as utilization type, state, labeler code, product code, year, quarter, product name, units reimbursed, number of prescriptions and other attributes representing the use of drugs for one year. A file "Medicaid State Drug Utilization Data Description " that contains a detailed explanation of the data attributes in the dataset. In this study, a number of attributes were

selected to describe the drugs utilization. The " product name " used to store a brand name of drugs. Also, "units reimbursed" used to store the "The total number of units (based on Unit Type) of the drug", and " number of prescriptions". Data preparing for this study can be listed as follows:

✓ Identify the diseases for study- three diseases (diabetes, leukemia, and allergy).

✓ Website- collect the brand's name of the drugs for each disease: (diabetes: 48 brands of different drugs),(leukemia: 35 brands of different drugs) and (allergy: 22 brands of different drugs).

✓ Filtering the data set- all the records of each brand names are select from it and store in a list.

✓ All the NaN or empty values have been removed from the list.

✓ Process a large number of prescriptions to extract it from the data set as a float number.

4.2. Disease-Drugs Clustering Deep Neural Network

DDC-DNN system is dealing with the drugs as a brand for certain diseases to apply the clustering process. The aim of the clustering process is to place drug brands in a group of clusters according to the quantity consumed by the United States of America within one year.

4.2.1 The topology of Proposed System
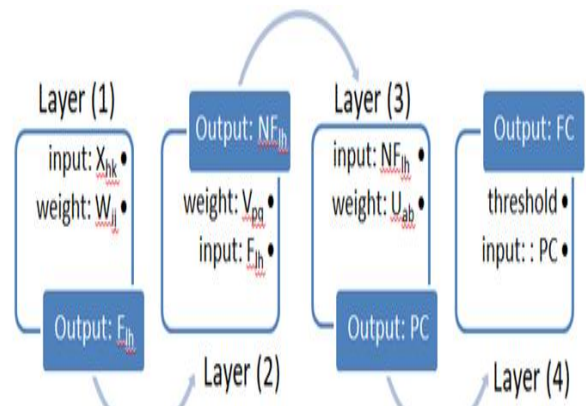
Figure (2) shows the topology of proposed system

Fig. 2: Topology of DDC-DNN

where

X_{hk}: where h number of brands, k is the number of records.

W_{ij}: weight of the 1st layer.

F_{lh}: Features of brands where l is no. of the feature, h is no. of brands.

V_{pq}: weight of the 2nd layer.

NF_{lh}: Features of brands after normalization where l is no. of the feature, h is no. of brands.

U_{ab}: weight of the 3rd layer.

IC_g: Initial clusters where g is no. of result cluster.

Ө: threshold

PC: the Primal clusters of drugs brand.

4.2.1.1. First Layer: Feature Extraction

The input of layer(1) is the raw data of the data set, the important attributes are select from the data set to represent the brands of drugs, Units Reimbursed and Number of Prescriptions for each brand. Thes attributes can give a very suitable indicator for the consuming of the drug's brand. The extracted features from this layer are mean, standard deviation and Skewness. Feature extraction layer generates six feature for each drug brand as shown in the algorithm (1).

Algorithm (1): feature extraction layer

Input:
An array of the selected attributes (X_{hk}) where h: is the number of attributes, and k: is the numbers of records.
Output:
An array of output neurons (O_{jk}), where j is the number of features
Begin:
Step 1: the input neurons are formed in two array two dimensions according to the number of attributes represent the columns and records of data represent rows, the initial weights (W) are set to a constant value ($1/N$)
Step 2:for all disease belongs to dataset diseases
Step 3: for all drug brand belong to drugs for one disease
Step 3.1: Calculate Mean

$$I_{1i} \leftarrow \sum_{i=1}^{no.of\ records} X_{1i}$$
$$O_{1li} = W_1 * I_{1i}$$
$$I_{2i} \leftarrow \sum_{i=1}^{no.of\ records} X_{2i}$$
$$O_{2li} = W_2 * I_{1i}$$

Step 3.2: Calculate Standard Division

$$\mu = (\sum_{i=1}^{no.of\ records} (X_{1n})/n$$
$$I_{3i} = \sum_{i=1}^{no.of\ records} (X_1 - \mu_n)^2$$
$$O_{3i} = SQRT(W_3 * I_{3i})$$
$$\mu = (\sum_{i=1}^{no.of\ records} (X_{1n})/n$$
$$I_{4i} = \sum_{i=1}^{no.of\ records} (X_2 - \mu_n)^2$$
$$O_{4i} = SQRT(W_2 * I_{2i})$$

Step 3.3: Calculate Skewness

$$\mu = (\sum_{i=1}^{no.of\ records} (X_{1n})/n, \quad SD = SQRT(\sum_{i=1}^{no.of\ records} (X_1 - \mu_n)^2/n)$$
$$I_{5i} = \sum_{i=1\ to\ n} [(x_i - \mu)/\ SD]^3$$
$$O_{3i} = W_5 * I_{5i}$$
$$\mu = (\sum_{i=1}^{no.of\ records} (X_{2n})/n, \quad SD = SQRT(\sum_{i=1}^{no.of\ records} (X_2 - \mu_n)^2/n)$$
$$I_{6i} = \sum_{i=1\ to\ n} [(x_i - \mu)/\ SD]^3$$
$$O_{6i} = W_6 * I_{6i}$$

end_loop
end_loop
End_algorithm

Second Layer: Normalization

The values in all feature are normalized, the input of the secondlayer is the output of the first layer. The output of this layer is (6) features for each brand. The aim of this layer is to avoid the large values of any variable to control the calculation results. A theoretical description of the algorithm is below.

Algorithm (2): normalization is the second layer

Input:
Array (Y_{vk}) is the technical indicators (output of the first layer).
Output:
Array of output neurons (O_{2vk})
Begin:
Step1: the inputs are a tow dimension array represent the number of technical indicators as columns and input records as arrows. The weights are set to be constant value equals to ($1/\ (max_v - min_v)$).
Step 2: for g=1 to v
  2:1: compute Mean of each input neurons of $g(\mu_g)$
2:2: find Max and Min value for each input neurons of $g(max_g - min_g)$
end_loop
Step3: for each output neuron:
3:1: compute the weighs of each neurons. It is computed as
$$T_{2g} = 1/(max_g - min_g)$$
3:2: the final value of output neuron is compute as:
$$O_{2gk} = T_{2g} * (Y_{gk} - \mu_g)$$
end_loop
End_algorithm

4.2.1.2.   Third Layer: Amapping Cluster

The output of the normalization layer forming as tow dimension array represent the features of the dataset. this array will be representing the input of the third layer. Patterns are the input of the map layer and the output is a tow dimension map represent the initial clustering result.

Algorithm (3): Mapping cluster layer (3)

Input:
Patters, epochs^{max}, learning_rate^{init}, radius^{init}, map_width, map_hight
Output:
pattern_co_ordinate
Begin:
for i=1 to epochs^{max}
Learning_rate^i ← calculate_learning_rate(i, learning_rate^{init})
        radius^i ← calculate_radius(i, radius^{init})
        Pattern$_i$ ← Select_Input_Pattern(Patterns)
BMU$_i$ ← Select_Best_Matching_Unit(Pattern$_i$, pattern_co_ordinate)
        neighbor ← BMU$_i$
neighbor ← Select_Neighbors(BMU$_i$, pattern_co_ordinate, radius^i)

    for (Vector$_i$ ∈ neighbor):
        for (Vector_Attribute$_i$ ∈ Vector$_i$):
Vector_Attribute ← Vector_Attribute + Learning_rate^i
*(pattern_attribut$_i$ - Vector_Attribute$_i$)
end_loop
end_loop
end_loop
End_algorithm

4.1.1.1.   Fourth Layer: Re-forming Cluster

The output of the mapping layer(3) is the patterns_co_ordinate (Primal cluster). Where the brands of drags mapping as a set of the cluster. The output of the map layer is the input of the Re-forming layer (4).

Algorithm (4): Re-forming layer (4)

Input:
The Primal cluster, the threshold. The threshold has been selected in the method trial and error.

Output:
The Final cluster
Begin*:*
    For i=1 to the number of clusters
center_ of_ cluster ← Calculate mean for all features of cluster
Euclidean_ distance$_{i,i}$ ← find Euclidean_ distancebetween each tow clusters
//Compare the Euclidean_ distance$_{i,i}$ with the threshold //
    If Euclidean_ distance$_{i,i}$ < threshold
            Meregcluster$_i$ with cluster$_i$
end_if
end_loop
End_algorithm

# 5. Experimental Result and Discussion

In this section, the results of the proposed system have been present, the DDC-DNN clustering of drug brands.
Case (1): the diabetic drug the number of drugs is (48), a number of records are (50855). Map dimensionof the third layer has effects on the convergence, small convergence is the best result, as shown in the Table (1).

Table 1: Training for diabetic drug

Map dimension	Number of epochs	primal clusters	convergence criteria
5*5	500	21	0.009
4*4	1000	14	0.017
3*3	1000	9	0.010

Table 3: Training for Allergies drug

Map dimension	Number of epochs	primal clusters	convergence criteria
5*5	1000	17	0.001
4*4	1000	13	0.004
3*3	1000	7	0.010

Figure (3) shown clusters result for diabetic drug brands. The primal cluster number was (21) and the result of the specific clustering layer result was only (4) cluster represent the final cluster with a threshold (0.4), Silhouette validity score was0.610.
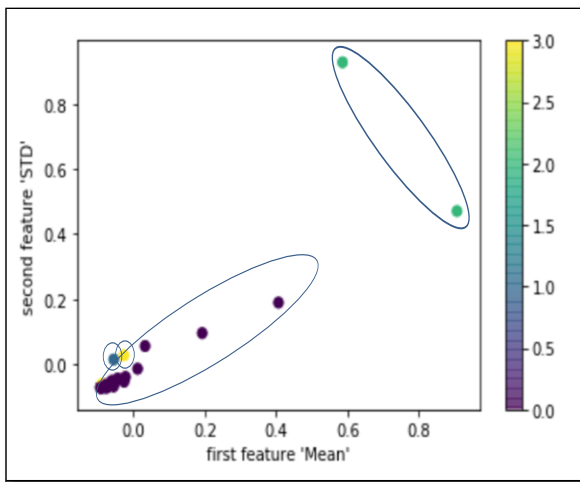
Figure (5) shown Cluster result for Allergies drug brands. The primal cluster number was (18) and the result of specific clustering layer was only (4) cluster represent the final cluster. 0.682 is the silhouette validity score
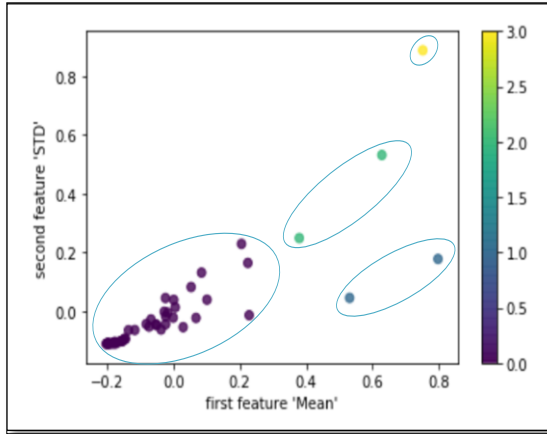
Fig. 3: DDC-DNN clusters of diabetic

Fig. 5: DDC-DNN clusters of Allergies

Case (2): the Leukemia disease has(35) number of drugs, number of records is (3204) where the training of the third layer to find the primal cluster depend on the map dimension, convergence is the difference between the previous and current map, a small convergence value means a better result, as shown in Table (2).

5.1. Qualitative Evaluation of the DDC-DNN result clusters

At the evaluation step, the K-means algorithm has been used to evaluate the results of the proposed model. by comparing them with results of traditional K-means. The number of cluster of K-means has been used the same number of the final cluster number.

The result has been visualized to show the difference between tow method.
At the Figure(6) the result clusters of k-means suffering from overlapping, while the result of DDC-DNN as shown in Figure (3) has the same number of K for the diabetic drug but the resulting cluster can be said a good cluster.

Table 2: Training for leukemia

Map dimension	Number of epochs	Number of primal clusters	convergence criteria
5*5	1000	18	0.008
5*4	1000	14	0.013
3*3	1000	7	0.014

Figure (4) shown Final clusters of leukemia drugs are (4). The primal clusters are (18) and the threshold is (0.32). silhouette validity score is 0.610.
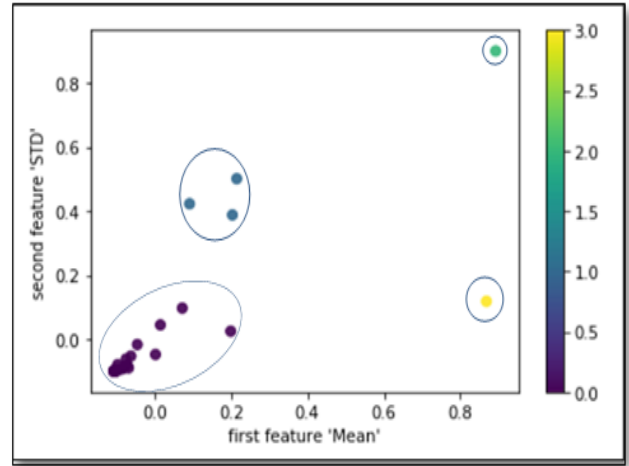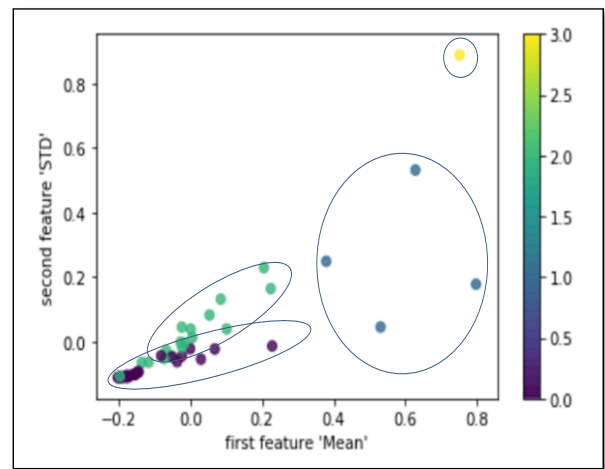

Fig. 4: DDC-DNN clusters of leukemia

Fig. 6: K-means with k=4

Case (3): the allergies number of drugs is (35). A number of records are (41908). So, the best convergence is (0.001) of the Third Layer of the DDC-DNN for Cluster result of Allergies drug brands. As shown in the Table (3)

In Figure (7) the result of leukemia drugs clustering of k-means with k=4 appears with an overlapping cluster. The result of DDC_DNN for leukemia drugs clustering shown in figure (4) when the cluster number is (4)
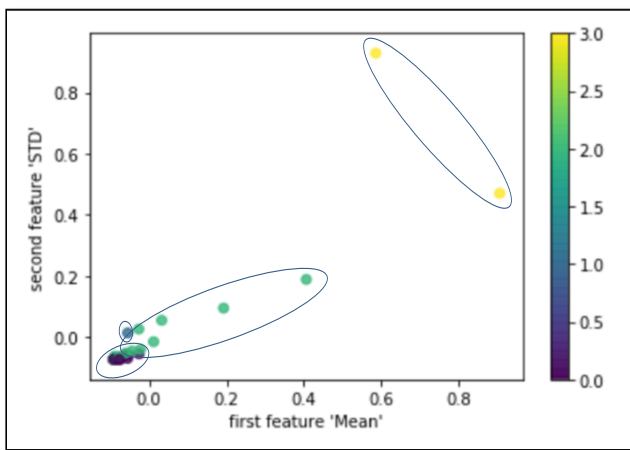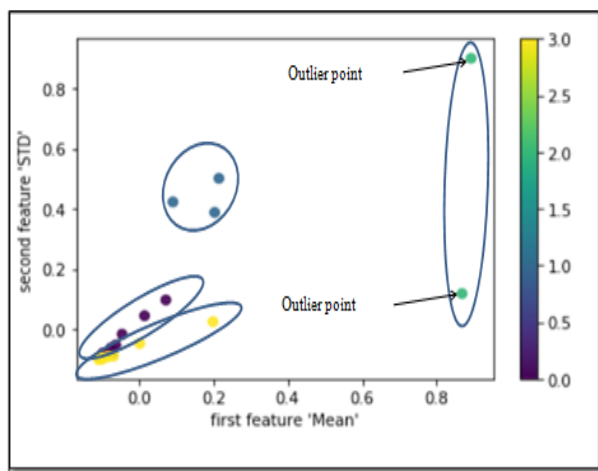
Fig. 7: K-means with k=4

Fig. 8: K-means with k=4

In the Figure (8) the result of the k-means cluster for allergic drugs. The result has two weak points in the form of the resulting clusters. The first is the overlap in the clusters. The second is sensitive to outliers points**.** Where in the figure (4) the resulting cluster of DDC_DNN for allergic drugs.

5.2. Qualitative Evaluation

In the qualitativeevaluation, the Silhouette score has been used toevaluating the result of the DDC_DNN. The result of the proposed model has been compared with the result of k-means for the same data and the same number of cluster (k) the Table (4) represent the Silhouette score for the two result clusters.

Table 4: The Silhouette Score

Drugs of disease	Silhouette score	
	(K-means)	**(DDC_DNN)**
Diabetic drugs 3-clusters	0.464	**0.610**
Leukemia drugs 3-clusters	0.449	**0.563**
Allergies drugs 4-cluster	0.593	**0.682**

6. Conclusions

The paper presents a proposed model of DDC_DNN. The proposed model is fully automated, adaptive clustering method and based on deep neural network concept. It has been used for obtaining the best distribution in the primal clusters for drugs brand depending on the best convergence of the third layer. primal clusters have been re-formed by using a carefully selected threshold in order to find the final clusters. The result of DDC_DNN has been evaluated by Silhouette score. In accordance

with the characteristics of the proposed model referred to earlier the obtained results are promising results and this model can be relied upon in the clustering process. In order to increase accuracy, the result of the proposed model has been compared with the result of k-means on the same data set. the result's of the proposed model was best in all cases**.**

References

[1] M. I. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and Future," CoRR, vol. 1704.06825, pp. 1–30, 2017.

[2] C. Company, "Chapter 5 Application of Data Mining Techniques in Data Mining Needs of Global," pp. 138–180, 1990.

[3] S. Madge, "Predicting Stock Price Direction using Support Vector Machines," 2015.

[4] J.-G. Lee et al., "Deep Learning in Medical Imaging: General Overview," Korean J. Radiol., vol. 18, no. 4, p. 570, 2017.

[5] Zhaojin Zhang, Cunlu Xu, and Wei Feng, "Road vehicle detection and classification based on Deep Neural Network," 2016 7th IEEE Int. Conf. Softw. Eng. Serv. Sci., no. 60903101, pp. 675–678, 2016.

[6] O. Pattanaprateep, M. McEvoy, J. Attia, and A. Thakkinstian, "Evaluation of rational nonsteroidal anti-inflammatory drugs and gastro-protective agents use; Association rule data mining using outpatient prescription patterns," BMC Med. Inform. Decis. Mak., vol. 17, no. 1, pp. 1–7, 2017.

[7] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2016, 2017.

[8] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," Procedia Comput. Sci., vol. 82, no. March, pp. 115–121, 2016.

[9] O. G. D. Fy, R. Science, I. Effective, T. Office, and G. Drugs, "FYs 2013 - 2017 Regulatory Science Report : Analysis of Generic Drug Utilization and Substitution," no. Figure 1, pp. 1–10, 2017.

[10] Y. Zhou, N. Sani, C.-K. Lee, and J. Luo, "Understanding Illicit Drug Use Behaviors by Mining Social Media," 2016 IEEE Int. Conf. Big Data (Big Data), 2016.

[11] R. Ghousi, S. Mehrani, and M. Momeni, "Application of Data Mining Techniques in Drug Consumption Forecasting to Help Pharmaceutical Industry Production Planning," Proc. 2012 Int. Conf. Ind. Eng. Oper. Manag., pp. 1162–1167, 2012.

[12] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[13] E. AL-Shamery and A. AL-haq, "An Optimized Feed Forward Neural Network for Reducing Error Based Stoch Market Prediction," J. Eng. Appl. Sci., vol. 13, no. Special Issue 5, pp. 4616–4621, 2018.

[14] S. Storcheus, Dmitry; Rostamizadeh, Afshin; Kumar, "A Survey of Modern Questions and Challenges in Feature Extraction," 1st Int. "Feature Extr. Mod. Quest. Challenges," vol. 44, pp. 1–18, 2015.

[15] E. AL-Shamery and A. AL-haq, "Enhancing Prediction of NASDAQ Stock Market Based on Technical Indicators," J. Eng. Appl. Sci., vol. 13, no. Special Issue 5, pp. 4630–4636, 2018.

[16] A. Abebe, J. Daniels, W. McKean, and A. Kapenga, Statistics and Data Analysis. Kalamazoo, MI: Western Michigan University, 2001.

[17] H. Khalid Obayes, "Suggested Approach to Embedded Playfair Cipher Message in Digital Image," vol. 3, no. 5, pp. 710–714, 2013.

[18] Z. Ansari, M. F. Azeem, W. Ahmed, and A. V. Babu, "Quantitative Evaluation of Performance and Validity Indices for Clustering the Web Navigational Sessions," vol. 1, no. 5, pp. 217–226, 2015.

[19] J. Baarsh and M. E. Celebi, "Investigation of Internal Validity Measures for K-Means Clustering," Proc. Interational MultiConference Eng. Comput. Sci., vol. I, pp. 14–16, 2012.