Clustering of Chemical Compounds using Unsupervised Neural Networks Algorithms: a comparison

Jehan Zeb Shah[1], Naomie bt Salim[2]
[1, 2] *Faculty of Computer Science & Information Systems, UTM, Malaysia*
[1] zeb@scientist.com, [2] naomie@fsksm.utm.my

***Abstract*-.Clustering of chemical databases has tremendous significance in the process of compound selection, virtual screening and in the drug designing and discovery process as a whole. Traditionally, hierarchical methods like Ward's and Group Average (Gave) and nonhierarchical methods like Jarvis Patrick's and k-means are preferred methods to cluster a diverse set of compounds for a number of drug targets (using fingerprints based descriptors). In this work the applications of a number of self-organizing map (SOM) neural network algorithms to the clustering of chemical datasets are investigated. The results of the SOM neural networks, Wards and Group-Average methods are evaluated for the clustering of different biologically active chemical molecules that can be used as drug like compounds based on topological descriptors. The results show that the Wards and Group Average methods are equally good; however, the performance of Kohonen neural self-organizing maps (SOM) is also important due to its almost similar performance as the hierarchical clustering methods with the advantage of its efficiency.**

Key words: *cluster analysis, chemoinformatics, Neural Gas network, Kohonen self-organizing map.*

I. INTRODUCTION

Drug discovery is a complex and costly process, with the main bottlenecks being the time and costs of finding, making and testing new chemical entities (NCE). The average cost of creating a NCE in a major pharmaceutical company was estimated at around $7,500/compound [1]. And for every 10,000 drug candidate NCE synthesized, probably only one will prove to be a commercial success and there may be 10-12 years after it is first synthesized before it reaches the market [2]. In order to reduce costs, pharmaceutical companies have had to find new technologies to replace the old "hand-crafted" synthesis and testing of NCE approaches.

Currently, many solution- and solid- phase combinatorial chemistry (CC) strategies are well developed [3]. Millions of new compounds can be created by these CC based technologies but these procedures have failed to yield many drug candidates. Enhancing the chemical diversity of compound libraries would enhance the drug discovery. A diverse set of compounds can increase the chances of discovering various drug leads and optimization of these leads can lead to better drugs. In order to obtain a library of great chemical diversity, a number of structural processing technologies such as diversified compound selections, classification and clustering algorithms have been developed.

The term cluster analysis was first used by Tryon in 1939 that encompasses a number of methods and algorithms for grouping objects of similar kinds into respective categories [4]. The main objective of clustering is to organize a collection of data items into some meaningful clusters, so that items within a cluster are more similar to each other than they are to items in the other clusters. This notion of similarity and dissimilarity may be based on the purpose of the study or domain specific knowledge. There is no pre-notion about the groups present in the data set.

The clustering process for chemical structures is outlined in [5] as follows. (1) Select a set of attributes on which to base the comparison of the structures. These may be structural features and/or physicochemical properties. (2) Characterize every structure in the dataset in terms of the attributes selected in step one. (3) Calculate a coefficient of similarity, dissimilarity, or distance between every pair of structures in the dataset, based on their attributes. (4) Use a clustering method to group together similar structures based on the coefficients calculated in step 3. Some clustering methods may require the calculation of similarity values between the new objects formed and the existing objects. (5) Analyze the resultant clusters or classification hierarchy to determine which of the possible sets of clusters should be chosen.

In chemical information system, most of the clustering methods are hierarchical like Single and complete linkage algorithms, Wards and Group Average algorithms. The traditional non hierarchical methods like k-means and Jarvis-Patrick's nearest neighbor methods are not very popular because of their inferior cluster quality.

Willett [6] has found that, among the hierarchical methods, the best result was produced by Ward's hierarchical agglomerative method and Jarvis-Patrick produced the best results compared to the other non-hierarchical methods

tested. They had evaluated almost 30 hierarchical and non hierarchical methods on a smaller dataset of around 14 biological groups, where 2D fingerprints been used as compound descriptors. In another study [7], Barnard and Downs have further investigated Jarvis-Patrick method in more detail using a small dataset of 750 diverse set of compounds from the ECDIN database using 29 physiochemical and toxicological information.

In [8] Downs and Willet have analyzed the performance of Wards, Group Average, Minimum Diameter and Jarvis Patrick methods on two datasets: a small subset of 500 molecules from chemical abstract service [9] database and another one of 6000 molecules. They have incorporated the same 29 physiochemical properties. The performance of Jarvis Patrick's method was very poor. The Minimum diameter method was found to be the most expensive. And the performance of the Wards method was the best. They have used the reciprocal nearest neighbor implementation which is more efficient.

The best principle work on the clustering of chemical dataset was reported by Broewn and martin [5] where Wards, Jarvis-Patrick,s (fixed and variable length nearest neighbor lists), Group Average and Minimum Diameter (fixed and variable diameter) methods had been evaluated on a large dataset of 21000 compounds comprised of four biologically active groups.They have employed a number of descriptors like MACCS 2D structural keys, Unity 2D keys, Unity 2D fingerprints, Daylight 2D fingerprints and Unity 3D pharmacophore screens. The performance of wards was found to be the best across all the descriptors and datasets, Group average and minimum diameter methods were slightly inferior. The performance of Jarvis Patrick method was very poor for fixed as well as for variable length nearest neighbor lists.

Recently fuzzy clustering methods have been applied for clustering of chemical datasets. In [10] Rodgers et al have evaluated the performance of fuzzy k-means algorithm in comparison with hard k-mean Wards methods using a medium size compound dataset from Starlist database for which LogP values were available. Their results show that fuzzy k-means is better than Wards and k-means. They have used simulated property prediction method [11] as performance measure, where the property of the cluster is determined by the average property of all the molecules contained in the cluster. This average property of the cluster is called the simulated property of each of the structure in the cluster. The simulated property of each molecule is correlated with the actual property of the compound to find the performance. In [12], Shah and Salim have used fuzzy Gustafson-Kessel, fuzzy k-means, Wards and Group average methods to cluster a medium size dataset from MDL's MDDR database containing about seven biologically active groups. Instead of using simulated property prediction method, the active cluster subset method where the proportion of active compounds in active clusters is used as performance measures, was employed. Their results show that the performance of Gustafson-Kessel algorithm is the best for optimal number of clusters. The Wards, fuzzy k-mean and group average methods are almost the same for optimal number of clusters.

With the advent of neural network theory, almost every field of science and technology has undergone revolutionary changes. Neural networks theories had been applied to various problems of control, system identification, image processing, pattern recognition, and data analysis. Researchers have devised a number of methods for the resolution of clustering problem; the self organizing map proposed by Kohonen [13], adaptive resonance theory based networks and neural gas network, being some of the important neural algorithms

Self organizing map (SOM) neural network have already been applied to various problems in chemoinformatics like protein classification [14-18], clustering [19], secondary structure mapping [46], and molecular surface unfolding [19-21]. In [22], SOM was used to discriminate dopamine from benzodiazepine agonists out of a training set of 172 compounds by using autocorrelation descriptors.

In this work, three neural network methods namely the Kohonen self-organizing map, neural gas and enhanced version of the neural gas algorithm were employed to cluster the chemical data space. This study uses two types of descriptors; the topological descriptors that are based on 2D topologies of atoms in a molecule and the BCI bitstring have that are strings of bits zero or one representing the presence or absence of an atom, ion or a group in a molecule, been used.

II. DESCRIPTORS

In order to classify compounds structures based on their biological activity, we need structural features with good ability to represent the structure-activity relationship. The stronger the structure activity relationship, the better the results will be in terms of practical importance.

A. Topological Indices (TI)

Topological indices are a set of features that characterize the arrangement and composition of the vertices, edges and their interconnections in a molecular bonding topology. These indices are calculated from the matrix information of the molecular structure using some mathematical formula. Among hundreds of possible descriptors, a few have been found useful in characterization of molecular properties and activities, such as the Weiner's index [23], Harary index [24], MTI index [25, 26], Balaban index [27, 28], and Zagreb group indices [29, 30]. For instance, TI has been used to predict the heats of formation for 60 hydrocarbons and the result show satisfactory predictions [31]. In [32] a QSAR study was carried out for modeling the DNA modeling affinity with help of distance matrix based TI.

Here we have calculated 62 topological descriptors using Melano chemometrics' Dragon software [33] for our MDL's [34] MDDR dataset. Principal component analysis (PCA) was then carried out to reduce the unnecessary linearly

dependent features so that each molecule is characterized by only the first 10 principle components that accounts for 98% variance in our selected dataset. PCA was carried out using the MVSP 3.13 [35]. A list of the 10 features selected is shown in Table 1.

B. BCI Bit Strings

These are strings of bits either zero or one and almost 1052 bits long developed by Barnard Chemical Inc [36]. Each bit describes the presence or absence of a particular atom, ion or a group of atoms in a molecule. In this way large databases of molecules can be represented with the help of such descriptors.

III. METHODS

Although there are available a large number of methods for clustering chemical dataset but a few are more important because of their good performance and efficiency. As discussed in the introduction Wards and Group average are the two hierarchical methods that outperform all the remaining classical methods. So, the Group average, Wards and self organizing map neural network methods that we have used in this work are described briefly.

A. Group Average Method

In all the agglomerative hierarchical methods the dataset is divided into the same number of clusters as the data elements in the dataset and so each individual compound structure is considered as a cluster of its own. In each successive level of the hierarchy only two clusters can be combined together based on similarity measure until a desired hierarchical level is obtained.

The group average clustering method [37] is an agglomerative hierarchical method that merges two clusters in a hierarchy if the distance between the two clusters is the minimum among all the clusters. This distance is an average of all the distances of the elements of one cluster to the elements of the other cluster in a pair of clusters.

The algorithm of the Group Average method is given in figure 1.

B. Ward's Clustering Method

The Ward's clustering method was suggested by Ward [38] in 1963 which is based on the minimization of the information loss associated with the merging of two groups in a hierarchy. According to Ward's those pairs of cluster should be merged which result in the minimum amount of loss of information. Ward defined the information loss in terms of an error sum of squares criterion given as:

$$ESS = \sum_{i=1}^{n} (x - x')^2$$

Wards method has proved to be an extremely powerful grouping mechanism, and is considered the best of hierarchic methods. However, it is criticized for its circular cluster shapes.

The algorithm is the same as for the Group average method shown in figure 1 except the distance is replaced with the error sum square value.

C. Self-Organizing Kohonen Neural Networks

Self organizing map (SOM) is an unsupervised neural network proposed by Kohonen [13] which consists of only two layers of neurons. The input layer and the output Kohonen layer, which is usually designed as 2 or 3 dimensional map of neurons.

It is basically, a competitive network with the characteristics of self organization where similar submaps of output layer neurons designate a class or group of the input dataset. If $X = [x_1, x_2, \ldots x_p]$ is a *P* dimensional data object then $W_l = [w_{l1}, w_{l2}, \ldots w_{lp}]$ is the weight vector associated with the neuron *l* of the output layer. The objective of the Kohonen learning law is to find the winning neuron or a neighborhood of neurons closest to the object presented as input stimulus. The winning neuron is required to be moved closer to the input object by some proportion of the distance between the input and the winning neuron (or neurons).

```
For I=1 to N-1Do
   For J=I+1 to N Do
      Calculte the distance D[I][J] between
          cluster I and cluster J.
   End
End
Search the distance matrix D to identify the
closest pairs of clusters.
Merge the closest pair and set N=N-1
And REPEAT the Algorithm until N==1
```

Fig 1: Group Average Algorithm

For each compound *I* of the dataset, the distance d_i between the weight vector W of each neuron and the input compound X is computed. The neuron (or neurons) having very small distance from the input is the winner. The weights of the winner neuron are then updated using some learning rule. Usually, Euclidean distance is used for distance computations.

The index *q* of the winning neuron (neurons) is determined as follows:

$$q = \arg \min_{l} \|W_l(t) - X_i\|$$

The weights for the winning neurons are updated as follows:

$$W_q(t+1) = W_q(t) + \alpha(t)\|W_q(t) - X_i\|$$

Where α(t) is the learning parameter which is a function of time and is decreased continuously with each iteration.
The value of this function is some positive value between 0 and 1 inside the map neighborhood. If the neuron is not inside the neighborhood of the winning neuron the learning rate parameter is 0. It means that in Kohonen learning rule only the neuron inside the neighborhood are learning neurons.

D. Neural Gas Clustering Method

The neural gas algorithm is an important neural network, first introduced by Martinez [39] for the prediction of time series and then applied successfully to the clustering of various databases[40], vector quantization [40], pattern recognition [41, 42], and topology representation [43] etc.
According to Martinez [39] the neural gas algorithm has a number of advantages like, 1- converges quickly to low distortion errors, 2- reaches a distortion error E lower than that resulting from k-means clustering and maximum entropy clustering (for practically feasible number of iterations), and from Kohonen feature map and 3- at the same time obeys a gradient descent on an energy surface (like the maximum entropy clustering, in contrast to Kohonen's feature map).
The neural gas algorithm generates a list of the ranks of weight vectors $\boldsymbol{W_k}$ corresponding to each input pattern $\boldsymbol{Z_k}$ which gives the weights a descending order based on the closeness to the input pattern with W_{k0} being the closest weight vector to the input pattern, W_{k1} as the second close weight vector and $\boldsymbol{W_{ki}}$, i = 1, 2, …, c-1 being the weight vector for which there are i vectors $\boldsymbol{W_j}$ with $\|Z_k - W_j\| < \|Z_k - W_{ki}\|$. If the ranking index number $\boldsymbol{k}$ associated with each vector $\boldsymbol{W_k}$ is denoted by $\boldsymbol{R(Z_k, W_{ki})}$, which depends on $\boldsymbol{Z_k}$ and the whole set $\boldsymbol{W_{ki}} = (\boldsymbol{W_{k0}, W_{k1}, \ldots, W_{kc-1}})$ of weight vectors, then the adaptation step we employ for updating the $\boldsymbol{W_k}$'s is given by

$$W_{ki}(t+1) = W_{ki}(t) - \eta(t) h_\lambda(R(Z_k, W_{ki}))\left[Z_k(t) - W_{ki}(t)\right]$$

The learning rate parameter $\eta(t) \in [0,1]$ describes the overall extent of modification and usually is taken as an exponentially decreasing function of time

$$\eta(t) = \eta_i(\eta_f / \eta_i)^{t/Maxiteration}$$

Where η_i and η_f are the initial and final values of the learning rate, respectively, which are initialized in advance. The *Maxiteration* is also a constant specifies the number of maximum *t* steps, also initialized at the start of the algorithm.
As already stated the ranking index $\boldsymbol{R(Z_k, W_{ki})}$ which depend on the input pattern $\boldsymbol{Z_k}$ and the whole set of the weight vectors $\boldsymbol{W_{ki}} = (\boldsymbol{W_{k0}, W_{k1}, \ldots, W_{kc-1}})$ of weight vectors which also serve as the prototypes of the dataset $\boldsymbol{Z}$. The value of $\boldsymbol{R(Z_k, W_{ki})}$ is zero for the closest weight vector and is the maximum for the farthest weight vector. The ranking adaptation parameter $\boldsymbol{h_\lambda(t)}$ is a function of the ranking index $\boldsymbol{R}$ and lies between 0 and 1. Martinez [39] has suggested an exponential decreasing function

$$h_\lambda(R) = \exp(-R/\lambda)$$

the value of $\boldsymbol{h_\lambda(t)}$ thus depends on the rank of the weight vector, as the rank increases the updating rate decreases and vice versa.

E. Enhanced Neural Gas Clustering Method

The neural gas (NG) algorithm as described in the previous chapter has a number of advantages like faster convergence to low distortion errors, lower distortion errors than k-means, maximum entropy kohnen's self organizing maps yet the updating formula is highly fragile in an environment having noise and large number of outliers and is also sensitive to the order of input vectors [44].

$$\Delta W_i = \eta(t) h_\lambda(R(Z_k, W_i)).\|Z_k - W_i\|.\frac{(Z_k - W_i)}{\|Z_k - W_i\|}, \quad i = 1,2,\ldots,c$$

It is obvious from neural gas updating formula in the above equation, if an outlier Z_0 is presented to update all the prototypes, the amplitude $\|Z_0 - W_i\|$ generated along the unit direction $\frac{(Z_k - W_i)}{\|Z_k - W_i\|}$ will be considerably large such that the prototypes will be dragged towards the outliers. Moreover, if the outliers are highly scattered around the dataset, the training process will not be smooth and there will be lot of oscillation. To overcome these problems Qin et al [44] suggested the following rule which is called the enhanced neural gas:

$$\Delta W_i = \eta(t) h_\lambda(R(Z_k, W_i)).\exp\left(-\frac{\|Z_k - W_i\|}{\beta d_i^{\,m}(0)}\right).\sigma_i(iter).\frac{(Z_k - W_i)}{\|Z_k - W_i\|}, \quad i = 1,2,\ldots,c$$

Compared with the previous equation, this formula also does obey the stochastic gradient descent rule and heuristically adjust the amplitude $\sigma_i(iter)$ of the gradient descent. The gradient descent amplitude $\sigma_i(iter)$ is given as:

$$\sigma_i(iter) = \sigma_i^{\,m}(t) = \begin{cases} d^m{}_i(t) & if \quad \|Z^m{}_t - W_i^{iter}\| \geq d^m{}_i(t-1) \\ \|Z^m{}_t - W_i^{iter}\| & if \quad \|Z^m{}_t - W_i^{iter}\| < d^m{}_i(t-1) \end{cases}$$

with

$$d_i^{\,m}(t) = \begin{cases} \left\{1/2\left[1/d^m{}_i(t-1) + 1/\|Z^m{}_t - W_i^{iter}\|\right]\right\}^{-1} & if \quad \|Z^m{}_t - W_i^{iter}\| \geq d^m{}_i(t-1) \\ 1/2\left[d^m{}_i(t-1) + \|Z^m{}_t - W_i^{iter}\|\right] & if \quad \|Z^m{}_t - W_i^{iter}\| < d^m{}_i(t-1) \end{cases}$$

and

$$d_i^{\,m}(0) = \left[\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\|Z_j - W_i^{mN}\|}\right]^{-1}$$

where N, Z_t^m, and W_i^{iter} represent the number of input vectors, the input vector given at iteration t of the training epoch m and the prototype vector i at the total iteration step *iter*, respectively. The term $d^m{}_i(t)$ is the restricting distance for the prototype W_i , which includes both historical and current distance information and is used here to limit the large absolute distance due to the outliers. If the absolute distance of an input vector at any time is greater or equal to the historical distance in the previous iteration, they are averaged using the harmonic mean and if absolute distance is less in magnitude than the historical distance, the averaging is done using arithmetic mean.

The objective of the training process of any neural network is to decrease gradually the average absolute distance. Whenever, this absolute distance becomes larger than the historical mean distance, the input to the machine must be an outlier. So, this is the scheme, used to detect the outliers and decrease their influence on the clustering process in general.

IV. RESULTS AND DISCUSSION

Our objective is to investigate the performance of SOM Kohonen neural network, Neural gas and enhanced neural gas algorithms for clustering the drug dataset based on topological indices and BCI bit strings against that of the Ward's and Group Average methods which are considered the industry standard. These methods should cluster together biologically similar structures and separate actives from inactive structures into different clusters. This experiment used the 1388-molecule from the MDL Drug Data Report (MDDR) database, containing molecules of drugs launched or under development, as referenced in the patent literature, conference proceedings, and other sources [34]. Out of these over 100000 molecules, we have chosen seven biologically active groups. The main groups, their subgroups and their aggregate activity are summarized in Table 2. Each of the three clustering methods was repeated ten times for 10, 20... 100 sets of clusters. The purpose of the analysis carried out on each of these sets of clusters is to determine the extent to which the actives have been separated from the inactives. For each of the seven groups in our data, we have taken the group as the active group and the rest of compounds belonging to the remaining six groups are taken as inactives. An active cluster is defined as one in which at least one member of the cluster is active. A subset of the dataset is termed as active cluster subset containing all the compounds in active clusters [13]. The active cluster subset must not contain the structures in the active singletons, as this will give rise to the proportion of actives incorrectly. For example if all the clusters are singletons then the proportion of actives will be 100 %, which contradicts the clustering objective to combine actives with actives and inactives with inactives in a multi member clusters rather than having too many singletons. If a singleton is active it does not give any clue about any other structure to be active or inactive. The proportion of actives in the active cluster subset gives us the degree of separation between actives and inactives.

In SOM neural network clustering, some parameters have to be initialized in advance. These parameters are the learning rate and the neighborhood map shape. The learning rate parameters can be linear, inverted or non linear (power) where the learning rate decreases exponentially. The shape of the neighborhood can be square, hexagonal, or Gaussian. Here, we have used the linear as well as exponential learning rate type and for the neighborhood we kept Gaussian.

Figure 2 shows the results of all the three methods: the Kohonen SOM (with linear and exponential learning rates), Neural Gas, and Enhanced neural gas in comparison with Wards and group average methods. The descriptors used in this experiment were the topological indices. The data is sampled for every 10 clusters. The x-axis plots the number of structures in active cluster subset, whereas the y-axis plots the proportion of actives in active cluster subset. Looking at the four curves, it is very easy to notice that the Wards and Group Average methods are the best, but the performance of Kohonen SOM and other neural networks is slightly inferior. The SOM with the exponential learning rate is better than the one with linear learning rate.

It has been observed that as we increase the number of clusters the proportion of actives in active cluster subsets increases. The best proportion obtained for the SOM, SOM-exp, neural gas, enhanced neural gas, Group average and Wards was respectively 16.39037, 17.03327, 16.54251, 15.13585, 17.92719 and 17.59631.

The results obtained using the BCI bitstring is worth discussion. Since the Bitstring length is very large i.e. 1052 bits which can not be considered a good dimension for computational methods like neural and when the data size is also large. So, the bits were divided into chunks of 8, 10, and 12 bits length and then converted to decimal real numbers and thus the dimensionality was reduced to 132, 105 and 88 variables respectively. The performance of this scheme was evaluated with Wards method as is shown in figure 3, and found that the 10 bit chunks perform better. So, the 10 bit chunk method was adopted for further analysis. It should be noted that all 105 variables accounted only for 71% of the variance in the data.

The performance of Ward's and Group average clustering methods is very brilliant for the BCI bitstrings as is shown in figure 4, due to their agglomerative hierarchical nature. In agglomerative clustering initially the whole data set is converted into as many clusters as the number of data elements. Then each cluster is compared with the rest of the cluster for similarity and each of two clusters is combined to make one larger cluster. On the other side divisive methods divides the whole dataset into a number of clusters and assign each compound to a cluster whose distance is minimum from the cluster center.

It has been noted that it is difficult to differentiate among precise variations than combining larger similarities. That is the reason that when the data is highly correlated and the variances are small, the divisive methods have to face almost failures whereas the agglomerative methods are more successful in such a situation as the similarities among the compounds are higher than their dissimilarities.

V. CONCLUSION

The Wards and Group Average methods which are considered to be the industry standard, once again have performed better than the neural network based methods Although the results of SOM and simple neural gas are not better than Wards and Group Average but still it has the ability to cluster chemical datasets.

REFERENCES

[1] P. Hecht, "High-throughput screening: beating the odds with informatics-driven chemistry", *Current Drug Discovery*, January 2002, 21-24.

[2] W. A. Warr, "High-Throughput Chemistry", Handbook of Chemoinformatics, Wiley-VCH, germany,2003.

[3] D. G. Hall, S. Manku, F. Wang, "Solution- and Solid-Phase Strategies for the Design, Synthesis, and Screening of Libraries Based on Natural Product Templates: A Comprehensive Survey", *Journal of combinatorial Chemistry*, 2001, 3, 125-150.

[4] (a) Troy R. C., *Cluster Analysis*, ANN Arbor, MI: Edwards Brothers, 1939.
(b) http://www.statsoftinc.com/text-book/stcluan.html

[5] R. D. Brown, and Y. C. Martin, "Use of structure-Activity data to compare structure based clustering methods and descriptors for use in compound selection", *Journal of chemical Information and computer science*, 36, 572-584. 1996.

[6] Willett P., *Similarity And Clustering In Chemical Information Systems*. Research Studies Press, Letchworth. 1987.

[7] G. M. Downs and J. M. Barnard, "Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures", Journal of chemical information and computer science, 1992.

[8] G. .M. Downs, P. Willett, and W. Fisanick, "Similarity searching and clustering of chemical structure databases using molecular property data", *Journal of Chemical Information and Computer Science*. 1994. 34:1094-1102.

[9] W. Fisanick, K. P. Cross, A. Rusinko, "characteristics of computer generated 3D and related molecular property data for CAS registery substances", Tetrahedron Computational Methodologies, 3, 635-652, 1990.

[10] J. D. Holliday, SW. L. Rodgers, P. Willet, et all, "Clustering Files of chemical Structures Using the Fuzzy k-means Clustering Method", *Journal of chemical Information and computer science*, 44, 894-902. 2004.

[11] G. W. Adamson and J. A. Bush, "A comparison of some similarity and dissimilarity measures in the classification of chemical structures", *Journal of chemical Information and computer science, 15, 55-58, 1975.*

[12] J. Z. Shah and N. Salim, "FCM and G-K clustering of chemical dataset using topological indices", Proc. of First International Symposium on Bio-Inspired Computing, Johor Bahru, Malaysia, 2005.

[13] T. Kohonen, "The self organizing map", IEEE Proc. Vol. 78, no. 9, 1990.

[14] T. F. Betham, T. Engan, J. Krane, and D. Axelson,"Analysisi and classification of proton NMR spectra of lipoprotein from healthy volunteers and patients with cancer or CHD", AntiCancer Research, 20(4), 2393-2408, 2000.

[15] E. A. Ferran, "Self organized neural maps of human protein sequences", Protein Science, 3(3), 507-521, 1994.

[16] W. Huichun, J. Dopazo, L. G. de la Fraga, Y.P. Zhu and J. M. Carazo, "Self organizing tree growing network for classification of protein sequences", Protein Science, 7(12), 2613-2622, 1998.

[17] J. Kaartinen, Y. Hiltunen, P. T. Kovanen, and M. Ala-Korpela, "Application of self organizing maps for the detection and classification of human blood plasma lipoprotein lipid profiles on the basis of 1H NMR spectroscopy data", NMR in Biomedecine,11(4-5), 168-76, 1998.

[18] J. Schuchhardt, "Local structural motif of protein backbones are classified by self organizing neural networks", Protein Engineering, 9(10), 833-842, 1996.

[19] P. Somervuo, T. Kohonen, "clustering and visualization of large protein sequence databases by means of an extension of the self organizing map", In Discovery Science, 3rd International Conference, DS 2000, Berlin, Germany, pages 76-85, 2000.

[20] P. Unneberg, J. J. Merelo, P. Chacon, and F. Moran, "SOMCD: Method for evaluating protein secondary structure from UV circular dichroism spectra", Protein-structure function and genetics, 42(4), 460-470, 2001.

[21] J. gasteiger, and A. Uschold, "The beauty of molecular surfaces as revealed by self organizing neural networks", Journal of Molecular Graphics, 12, 90-97, 1994.

[22] J. Polanski, and b. Walczak, "comparative molecular surface analysis (COMSA): A novel tool for molecular design", computers and chemistry, 24(5), 615-625, 2000.

[23] H. Wiener, *Journal of American Chemical Society* 1947, 69,17; 1947, 69, 2336; *Journal of Chemical Physics*,1947, 15,766; *Journal of Physical Chemistry*, 1948, 52, 425; 1948, 52, 1082.

[24] D. Plavsic, S. Nikolic, N. Trinajstic, Z. Mihalic, *Journal of Mathematical Chemistry*, 1993, 12, 235.

[25] H. P. Schultz, *of chemical Information and computer science.* 1989, *29*, 237.

[26] W. R .Mueller, K. Szymanski, J. V. Knop, N. Trinajstic, *of chemical Information and computer science* , 30, 169, 1990.

[27] M. Randic, *Journal of American Chemical Society.*1975, 97, 6609.

[28] A. T. Balaban, *Journal of Chemical Physics,* 89, 399, 1982.

[29] I. Gutman, N. Trinajstic, *Chemical Physics Letters* 1972*, 17*, 535.

[30] I. Gutman, B. Ruscic, N. Trinajstic, Jr. C. F. Wilcox, *Journal of Chemical Physics,* 1975, *62*, 3399.

[31] A. Mercader, E. A. Castro, and A. A. Toropov, "Maximum Topological Distances Based Indices as Molecular Descriptors for QSPR. 4. Modeling the Enthalpy of Formation of Hydrocarbons from Elements", *International Journal of Molecular Science*, 2001, 2, 121-132.

[32] A. Thakur M. Thakur, N. Kakani, A. Joshi, S. Thakur and A. Gupta "Application of topological and

physicochemical descriptors: QSAR study of phenylamino-acridine derivatives", *ARKIVOC* (xiv), 2004.
[33] Dragon, melano chemoinformatics: http://www.talete.mi.it
[34] MDL's Drug Data Report: http://www.mdli.com/products/knowledge/drug_data_report/index.jsp
[35] MVSP 3.13, Kovach computing services: http://www.kovcomp.com/
[36] Bernard Chemical Inc
[37] B. S. Everitt, "Cluster Analysis, 3rd Edition", Edward Arnold, London, 1993.
[38] J. H. Ward, "Heirarchical grouping to optimize an objective function", *Journal of American Statistical Association*, 58, 236-244.
[39] T.T. Martinez, S.G. Berkovich, and K. J. Schulten, "Neural gas network for vector quantization and its application to time series prediction", IEEE Tran. On Neural Networks, 4, vol. 4, 1993.
[40] A. K. Qin, and P. N. Suganthan, "Robust growing neural gas algorithm with application in cluster analysis", Journal of Neural Networks, 1135-1148, 17, 2004.
[41] B. L. Zhang, M. Y. Fu, and H. Yan, "hand written character verification based on neural gas based vector quantization", IEEE proc., 1998.
[42] F. Camastra, and A. Vinciarelli, "combining neural gas and vector quantization for cursive character recognition", Journal of Neurocomputing, 147-159, 51, 2003.
[43] Z. Cselenyi, "Mapping the dimensionality, density and topology of data: the growing adaptive neural gas", computer methods and programs in biomedicine, 78,141-156, 2005.
[44] A.K. Qin and P.N. Suganthan, "Enhanced neural gas network for prototype-based clustering", Pattern recognition, 1275-1288, 38, 2005.

Table 1
TOPOLOGICAL INDICES

Sr. No.	TI	Description
1	Gnar	Narumi geometric topological index
2	Xt	Total structure connectivity index
3	Dz	Pogliani index
4	SMTI	Schultz Molecular Topological Index (MTI)
5	PW3	path/walk 3 – Randic shape index
6	PW4	path/walk 4 – Randic shape index
7	PW5	path/walk 5 – Randic shape index
8	PJI2	2D Petitjean shape index
9	CSI	eccentric connectivity index
10	D/Dr05	distance/detour ring index of order 5

Table 2
Groups and some characteristics of the Dataset

S.No	Activity	No. molecules
1	Interacting on 5HT receptor Potentially useful in the treatment of depression, anxiety, hypertension, eating disorders, obesity, drug abuse, cluster headache, migraine, obsessive compulsive, and associated vascular disorders, panic attacks, agoraphobia eating, urinary incontinence and impotence.	
	5HT Antagonists	48
	5HT1 agonists	66
	5HT1C agonists	57
	5HT1D agonists	100
2	Antidepressants Potentially useful as an antiepileptic, antiparkinsonian, neuroprotective, antidepressant, antispastic and/or hypnotic agent. Some of the compounds may be useful in the treatment of dopamine-related CNS disorders such as Parkinson's disease and schizophrenia.	
	Mao A inhibitors	84
	Mao B inhibitors	174
3	Antiparkinsonians Potentially useful in the treatment of septic shock, congestive heart failure and hypertension and in the prevention of acute renal failure.	
	Dopamine (D1) agonists	32
	Dopamine (D2) agonists	104
4	Antiallergic/antiasthmatic Most of these are used as antiinflammatory, antiasthmatic and antiischemic agents. However, adenosine (A3) antagonists are useful as a tool for the pharmacological characterization of the human A3 receptor.	
	Adenosine A3 antagonists	73
	Leukotine B4 antagonists	150
5	Agents for Heart Failure Potentially useful as a bronchodilator, smooth muscle relaxant or cardiotonic agent, accelerator of hormone secretion, platelet aggregation inhibitor, etc.	
	Phosphodiesterase inhibitors	100
6	AntiArrythmics Most of the Potassium channel blockers block the cardiac ion channel carrying the rapid component of the delayed rectifier potassium current.	
	Potassium channel blockers	100
	Calcium channel blockers	100
7	Antihypertensives	
	ACE inhibitors	100
	Adrenergic (alpha 2) blockers	100
	Total molecules	1388

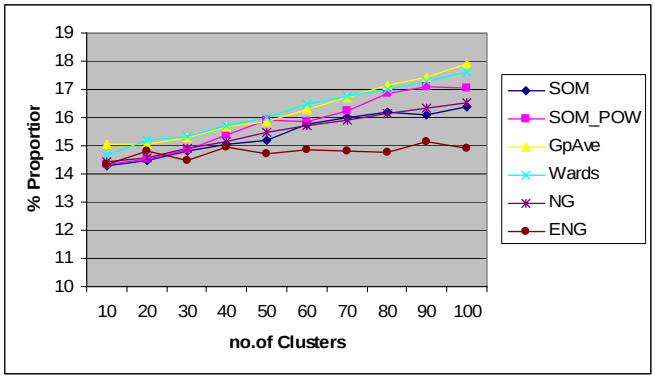

Fig2: Performance of Neural networks methods like Kohonen SOM, Neural Gas, and Enhanced Neural Gas in comparsion with Ward,s and Group Average. Here the BCI 1052 bitstrings were used.

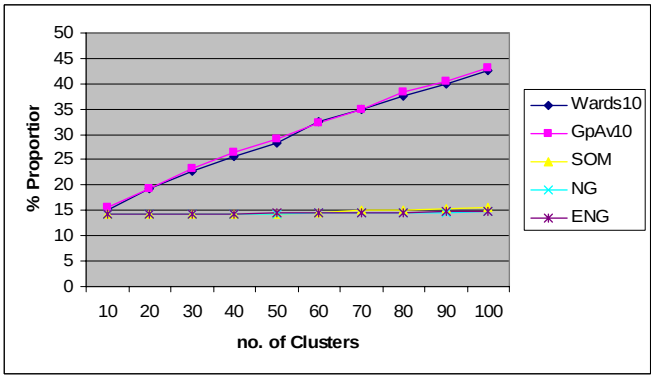

Fig 4: Performance of neural clustering when BCI bits strings have been used as descriptors

Figure 3. Performance of combinations of 8, 10 and 10 bits BCI bitstrings as decimal real numbers using Ward's clustering method.