

Research

Artificial Neural Network as Data Mining ToolAnita Yadav^{1*}, R. Y. Chaudhari², S. B. Bhise³¹Pharmaceutical chemistry, Dr. L. H. Hiranandani college of Pharmacy, Ulhasnagar- 3, India.²Loksevak Honourable Madhukarrao Chaudhari College of Pharmacy, Faizpur, India.³LNB Chhabada Institute of Pharmacy, Raigaon, India.**Received** May 05, 2018; **Accepted** August 01, 2018; **Published** August 03, 2018**Copyright:** © 2018, Anita Yadav et al

***Corresponding author:** Anita Yadav, Assistant Professor in Pharmaceutical chemistry, Dr. L. H. Hiranandani college of Pharmacy, Ulhasnagar- 3, India. Email: anita.yadav9@gmail.com, Mobile: 9869190383 / 9820201077

Abstract

Data mining is the term used to describe the process of extracting value from a database. By using Neural networks information can be extracted from the databases. This review summarizes how Artificial Neuronal Networks are used as data mining tool in drug design

Keywords: Artificial neural network (ANN); Quantitative structure activity relationships (QSAR).

Introduction

Artificial Neural networks as data analysis method have their application in drug design to form quantitative structure activity relationships (QSAR) [1]. Artificial Neural Networks are one of the best Artificial Intelligence (AI) techniques. Complex nonlinear data can be solved by them. They are broadly used in process which involves complex mathematical modelling like primary virtual screening of compounds, quantitative structure activity relationship studies, receptor modelling, formulation development, Pharmacokinetics [2].

Artificial Neural Network

ANNs gains knowledge from standard data. Neural network technology mimics the brain's own problem solving process. Just as humans apply knowledge gained from past experience to new problems or situations, a neural network takes previously solved examples to build a system of "neurons" that makes new decisions, classifications, and forecasts.

ANN techniques are the mathematical models which functions similar to human brain. All ANN methods consist of "neurons" ("hidden units") in their architecture [3]. Processing Element (PE) is the artificial analogue of biological neuron. A number of PEs (ten to thousands) are organized into groups called layers. There are three types of layers-Input, hidden and Output layer .The input layer is one which collects the information obtained from the surrounding and the output layer then generates a response to a given input . The layers between the input and output layers are called hidden layers. PEs in hidden layers is joined with all the PEs in the input and output layers as shown in figure [1]

Various ANN used are

Feed forward networks, Multilayer Perceptron's, Feedback networks, Kohonen Neural network /Self organizing Maps (SOM), Bayesian Regularized Neural network and K-Nearest Neighbor method

Applications

Based on literature survey the information of all the available numerous drugs under specific categories which have given

significant results in the treatment of diseases and disorders are collected and are used for training the neural networks which will be further used for generating output for the test set of compounds.

ANN Workflow Overview

Initially, a compound set of 313 AChE inhibitors was collected from literature. Descriptors were selected by manually dividing the compound set into a training set and a test set (1:3, respectively.). Then the whole dataset together with the previously determined descriptors was used for training models for prediction of activity, here counter propagation neural networks and novelty detection method were used. Later after model validation, an external dataset - the natural products database DIOS - was screened. The resulting hit lists contained basic, neutral, and acidic compounds. Because of the aromatic environment of the AChE inhibitor binding site, binding of neutral or basic compounds is more favorable than binding of acidic ones. So accordingly Virtual hits were filtered in order to remove acidic compounds and then docked to confirm their potential to fit into the AChE active site. From the resulting hit list, promising natural compounds were selected for biological testing. A graphical workflow is given in Figure (2) [4].

From the available sources i.e literature and other chemical databases like Pub Chem., Pubchem Bioassay and ZINC, etc A Antimicrobial drug database (AMDD) was created. AMDD current version consists of 2900 antibacterial and 1200 antifungal compounds. The details of molecules are given with the following properties such as description, target, format, bioassay, molecular weight, hydrogen bond donor, hydrogen bond acceptor and rotatable bond. These available anti-microbial agents, provides useful information and facilitate the virtual screening process, by saving time and overcoming difficulties in selecting specific type of inhibitors for the specific targets. The information regarding all those compounds are freely available [5]. The information obtained from this database can be utilized as a training set in neural network.

DTP Human Tumor Cell Line Screen

The Anti-cancer Agent Mechanism Database is a set of 122 compounds with anti-cancer activity and reasonably well known mechanism of action. The compound set was assembled as a training

set for neural network analysis of mechanism of action of drug .It was based on the Standard Anti-cancer Agent Database.

Anti-Cancer Databases

In order to find more effective anticancer drugs, the U.S. National Cancer Institute (NCI) screens a large number of compounds in vitro against 60 human cancer cell lines from different organs of origin. Various statistical and artificial intelligence methods including neural network modelling can be used to analyse this large activity database. Mining the database can provide useful information: (a) for the development of anticancer drugs; (b) for a better understanding of the molecular pharmacology of cancer; and (c) for improvement of the drug discovery process [6].

Similarly TB databases

Bio Health Base is now incorporated into PATRIC , The Collaborative Drug Discovery Tuberculosis Database, GenoMycDB, Tbrowse, TDR targets database, Tuberculosis Drug Resistance Mutation Database, TubercuList, The Tuberculosis Database and WebTB [7].

These databases give the knowledge related to the physicochemical property profiles of the drugs, their targets ,mechanism of action, Pharmacophores, descriptors, Pharmacokinetic properties (ADMET).

The inability to understand and control the ADMET properties of molecules is an important reason why many candidate drugs fail late in the development pathway. These failures are expensive and they contribute to the diminishing efficiency of the pharmaceutical industry. In silicon models of ADMET properties allow these properties to be considered at an early, less costly stage, and should reduce the number of late-stage development candidates which fail. Neural network approach provides useful solutions to this problem [8]. Kohonen Artificial neural network and Counter Propagation Neural network is used in Molecular structure -Toxicity studies [9].

The neural networks are trained by utilizing these databases and are prepared for giving solutions/output for new test sets of compounds. If a new set of compounds (test set) is given to the neural network then it will generate the output based on the data contained in the system which was previously used for training the neural network system .Suppose a compound in a test set is having a structural

similarity to the ant tubercular agents then the neural network will predict its QSAR data accordingly Table 1 [10].

So a trained neural network gives the information which can be further utilized for QSAR studies and in drug design it is used for

- Comparison of combinatorial libraries
- Search for new lead structures
- Establishment of structure–activity relationships
- High-throughput screening data analysis
- Optimization of a lead structure

- Exploration of conformational flexibility
- Analysis of ADME (absorption, distribution, metabolism and excretion) data [11].

Analysis of the similarity and diversity of combinatorial libraries

Kohonen Mapping-Kohonen neural networks have been used to show that the combinatorial libraries derived from 1 and 2 are very similar, whereas those obtained with 3 are distinctly different. The depicted kohonen map indicates that the libraries from 1 and 2 are so similar that they need not both be synthesized for screening figure 3.R1 –R4 are one of 18 possible aminoacid residues [12].

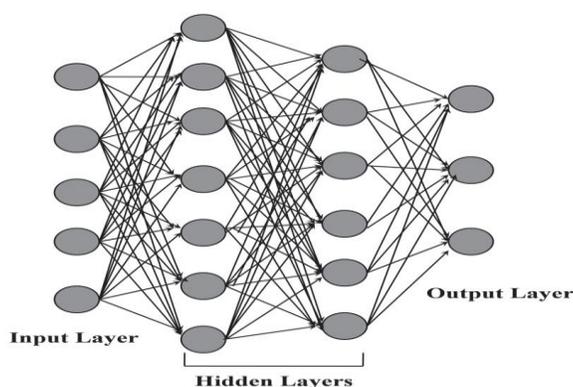


Figure 1: Structure of Artificial Neural Network

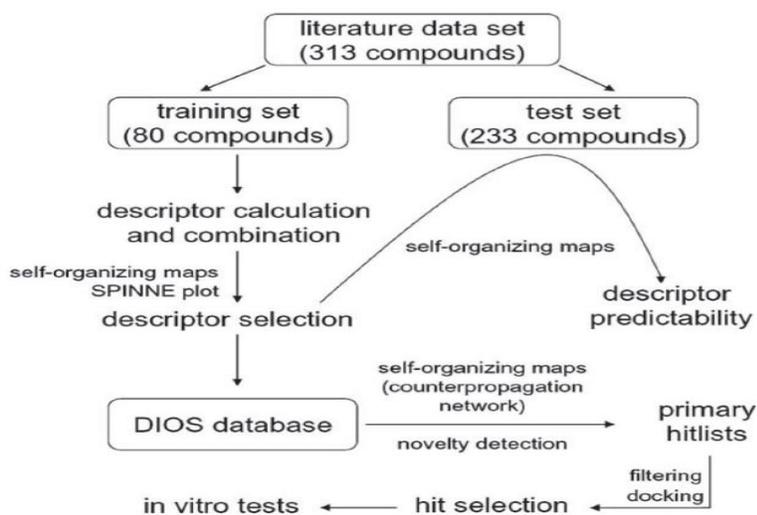


Figure 2: Work Flow followed for the identification of novel AChE inhibitors.

Biological properties	Chemical properties
Bioconcentration	Boiling point
Biodegradation	Chromatographic retention time
Carcinogenicity	Dielectric constant
Drug metabolism and clearance	Diffusion coefficient
Inhibitor constant	Dissociation constant
Mutagenicity	Melting point
Permeability	Reactivity
Blood brain barrier	Solubility
Skin	Stability
Pharmacokinetics	Thermodynamic properties
Receptor binding	Viscosity

Table 1: Summary of biological and chemical properties explored in QSAR studies.

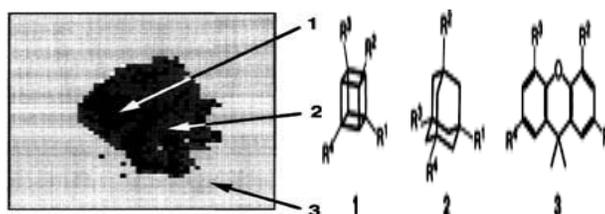


Figure 3: Analysis of the similarity and diversity of combinatorial libraries

Comparison of compound libraries

By using Kohonen two-dimensional maps of molecular surface properties can be generated and used for the comparison of set of molecules [13].

Search for new lead structures-Lead Discovery

Kohonen neural network/SOM is used in drug discovery. To demonstrate the utility of the technique SOM method has been applied to several series of compounds. A classic set of 31 steroids are used for validating novel QSAR and from the resulting Kohonen maps it was a simple task to classify whether a molecule was active or inactive. Potentially these maps could be used to search databases for new lead structures that possess the biological activity of interest [14].

Analysis of multidimensional data

It is done by using Kohonen neural network. This network performs a nonlinear mapping of a high dimensionality data space transforming

it in a low dimensional space. The SOM technique is employed to cluster and extrapolate the data set keeping the original topology [3].

Analysis of data from high-throughput screening (HTS)

To find out active compounds from a large dataset, ANN is trained using High throughput screening (HTS) data. The HTS data acquired from methionine amino peptidases inhibition activity considered of library of 43,347 compounds and the ratio of active to nonactive compounds, R(A/N) was 0.0321. Back propagation ANNs were trained and validated using principal components derived from the physicochemical features of compounds. The results showed that only 10% of all available compounds were needed for training and validation and the rest of the data set was screened with more than 10 fold gain of original R(A/N) value. Thus, ANN trained with limited HTS data might become useful in recovering active compounds from large datasets [15].

Prediction of biological activity and ADME –Tox (absorption, distribution, metabolism and excretion-toxicity) data

ANNs are introduced as robust and versatile tools in QSAR modelling. Feed forward neural networks are mostly utilized in QSAR. K-nearest neighbour QSAR method is used for modelling adverse liver effects of drugs [16].

ANN has become efficient tool for modelling and variety of data analysis in drug discovery. Neural networks are powerful data mining tools with a wide range of applications in drug design. Neural networks are used in Predicting HIV Drug Resistance [17]. 93% correct separation of compounds with and without antibacterial activity (in set of 657) is obtained using 34 “inductive QSAR descriptors”. The elaborated QSAR model based on the Artificial Neural Networks approach has been extensively validated and has confidently assigned antibacterial character to a number of trial antibiotics from the literature [18]. Overlooked visualization methods based on nonlinear dimensionality reduction and highlighted applications in drug discovery with emphasis on ANN technique [19]. Demonstrated importance of quantum chemical calculations in predicting the specificity of tetrahydroisoquinoline derivatives using ANN [20]. Developed ANN model for dopamine receptor subtype (D1 like and D2 like) affinity and selectivity and showed that secondary amines and other nitrogen containing moieties were shown to be important for the D1 like receptor selectivity [21], whereas molecular size, volume and tertiary and quaternary carbons were found to be of significant importance for the D2 like receptor selectivity [2].

The available Neural Network Data Mining tools are

STATISTICA Data miner- It provides information about Stat Soft's complete line of data analysis software.

Bayesia Lab Bayesian network laboratory including a set of data mining and machine learning tools.

Knowledge Miner Self-Organizing Data Mining for your Mac. It works using three advanced self-organizing modeling technologies: Group Method of Data Handling (GMDH Neural Networks), Analog Complexing and Fuzzy Rule Induction. This is the first time that all of these algorithms have been available in one place on any computer platform.

Alyuda Research Inc This provides neural network software for data mining and forecasting as well as consulting and research services in neural networks and data mining

Management Intelligenter Technologien GmbH It is a data mining and clustering software for numerical and textual data. Main product is DataEngine. Interfaces to Lab View and Bridgeview. Data Engine is a software tool for data analysis in which fuzzy rules, fuzzy clustering, neural networks and fuzzy neural systems are offered in combination with mathematics, statistics and signal processing.

Partech Incorporated Pattern recognition software used in life sciences and engineering for gene expression (microarray) data

analysis, high throughput screening, and drug design including SAR and ADME prediction

Smart Research BV Statistical Modeling and Artificial Reasoning Technology for solving complex problems using intelligent techniques such as neural networks and graphical models.

The Chi-Square Works. Inc This provides multi-window, dynamic data mining systems (e.g., Panmo) that use graphical direct manipulation as the main user interface. Data can be specified, retrieved, and passed to analytical functions (e.g., SOM and CART) graphically.

Inforsense Beyond intelligence predictive analytics data mining-Kensington Discovery Edition platform includes data integration, transformation, visualization, mining and discovery processes creation. Vertical application in life science and general discovery informatics.

STATISTICA Automated Neural Networks [22].

Neuroshell Classifier [23]

Neuroshell Predictor [24]

NeuroDimension -Neural networks and Intelligent Software solutions

PrologP 7.0 Software

SONNIA –Self organizing Neural network Package

Molecular descriptor Package **ADRIANA code**

Kohonen and CPANN toolbox

MATLAB Neural network Toolbox

Conclusion

So Neural network serves as data mining tool in drug design by its use in variety of analysis like analysis of multidimensional data, analysis of similarity and diversity of combinatorial libraries, comparison of combinatorial libraries, analysis of data from HTS, lead discovery and prediction of biological activity and ADME-Tox properties.

References

1. Ross D. King, Jonathan D. Hirst, Michael J.E.Sternberg (1993). New approaches to QSAR: Neural networks and machine learning. *Perspect. Drug Discovery Des* 1: 279-290.
2. Jignesh kumar Patel (2013) Science of the science, Drug Discovery and Artificial Neural Networks. *Curr. Drug Discovery Technol* 10:2-7.

3. Vinicius Goncalves Maltarollo, Kathia Maria Honorio, Alberico Borges Ferreira da Silva (2013) Applications of Artificial Neural Networks in Chemical Problems. *INTECH* 10: 203-223.
4. Daniela Schuster, Lisa Kern, Dimitar P. Hristozov, Lothar Terfloth et al. (2010) Applications of Integrated Data Mining Methods to Exploring Natural Product Space for Acetyl cholinesterase Inhibitors. *Comb. Chem. High Throughput Screening* 13: 54-66.
5. Mohd Danishuddin, Lalima Kaushal, Mohd Hassan Baig, Asad U. Khan (2012) AMDD Antimicrobial Drug Database. *Genomics Proteomics Bioinformatics* 10: 360-363.
6. Leming M. Shi , Yi Fan , Jae K. Lee , Mark Waltham , Darren T. Andrews (2000) Mining and Visualizing Large Anticancer Drug Discovery Databases. *J. Chem. Inf. Comput. Sci* 40:367–379.
7. Sean Ekins, Joel S. Freundlich, Inhee Choi, Malabika Sarker, Carolyn Talcott (2011) Computational databases, pathway and chemo informatics tools for tuberculosis drug discovery. *NIH Public Access .Trends Microbial* 19:65-74.
8. Winkler, D.A (2004) Neural networks in ADME and toxicity prediction. *Drugs Fut* 29:1043.
9. Marjan Vracko, Kohonen (2005) Artificial Neural Network and Counter Propagation Neural Network in Molecular Structure-Toxicity Studies. *Curr. Comput.-Aided Drug Des* 1:73-78.
10. Chanin Nantasenamat, Chartchalerm Isarankura-Na-Ayudhya, Thanakorn Naenna, Virapong Prachayasittikul (2009). A practical Overview of Quantitative structure activity relationship. *EXCLI Journal* 8: 1611-2156.
11. Johann Gasteiger, Andreas Teckentrup, Lothar Terfloth, Simon Spycher (2003) Neural networks as data mining tools in drug design. *J. Phys. Org. Chem* 16: 232–245.
12. Jens sadowski, Margus Wagener, Johann Gasteiger (1996) Assessing similarity and Diversity of combinatorial libraries by Spatial Autocorrelation Functions and Neural network. *Angew. Chem. Int. Ed* 34: 2674–2677.
13. Soheila Anzali, Gerhard Barnicke, Michael Krug, Markus Wagener, Johann Gasteiger (2007) Kohonen Neural Network: A Novel Approach to Search for Bioisosteric Groups. *Computer-Assisted Lead Finding and Optimization. WILEY-VCH* 7: 95-106.
14. David T. Manallack, David J. Livingstone (1999) Neural networks in drug discovery have they lived up to their promise, *Eur. J. Med. Chem* 34:195-208.
15. Swapan Chakrabarti, Stan R. Svojanovsky, Romana Slavik, Gunda L.Georg, Georg S. Wilson et al. (2009) Artificial Neural Network-Based Analysis of High-Throughput Screening Data for Improved Prediction of Active compounds,
16. *J BIOMOL SCREEN* 14:1236-1244.
17. Amie Danielle Rodgers (2009) Modeling Adverse Liver Effects of Drugs using KNN QSAR method.
18. Sorin Draghici , R. Brian Potter (2003) Predicting HIV drug resistance with neural networks. *Bioinformatics* 19: 98-107.
19. Artem Cherkasov (2005) Inductive QSAR Descriptors. Distinguishing Compounds with Antibacterial Activity by Artificial Neural Networks. *Int. J. Mol. Sci* 6:63-86.
20. Reutlinger M, Schneider G (2012) Nonlinear Dimensionality reduction and mapping of compound libraries for drug discovery. *J. Mol Graph Model* 34:108-117.
21. Uesawa Y, Mohri K, Kawase M, Ishihara M, Sakagami H (2011) Quantitative structure-activity relationship(QSAR) analysis of tumor-specificity of 1, 2, 3, 4 –tetra hydro isoquinoline derivatives. *Anticancer Res* 31:4231-4238.
22. Karolidis DA, Agatonovic-Kustrin S, Morton DW (2010) Artificial neural network (ANN) based modeling for D1 like and D2 like dopamine receptor affinity and selectivity. *Med Chem.* 6:259-70.
23. Statistical Data Miner.
24. NeuroShell Classifier.
25. NeuroShell Predictor.