31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Drug-like and Non Drug-like Pattern Classification Based on Simple Topology Descriptor Using Hybrid Neural Network

Wan Mohd Fahmi Wan-Mamat [1], Nor Ashidi Mat Isa[2], *Member, IEEE*, Habibah A. Wahab [3], Wan Mohd Fairuz Wan-Mamat [4]

***Abstract*— An intelligent prediction system has been developed to discriminate drug-like and non drug-like molecules pattern. The system is constructed by using the application of advanced version of standard multilayer perceptron (MLP) neural network called Hybrid Multilayer Perceptron (HMLP) neural network and trained using Modified Recursive Prediction Error (MRPE) training algorithm. In this work, a well understood and easy excess Rule of Five + Veber filter properties are selected as the topological descriptor. The main idea behind the selection of this simple descriptor is to assure that the system could be used widely, beneficial and more advantageous regardless at all user level within a drug discovery organization**

I. INTRODUCTION

Over past decade, modern drug discovery philosophy is principally based on the high-throughput screening (HTS) of large compound libraries either taken from the existing compound databases or obtained by chemical synthesis using combinatorial chemistry (CC) techniques [1]. These technologies provided more capacity for making and screening a larger number of compounds in relatively short time. However, it is apparent that the high throughput synthesis and screening paradigm has not delivered the results that were initially anticipated [2]. Therefore the ability to effectively predict if a chemical compound is "drug-like" or "non drug-like" would be a valuable tool in the design, optimization, and selection of drug candidates in order to reduce attrition during expensive clinical development [3].

There are many definitions given to drug-like by various authors [4]. Generally, "drug-like" means molecules which contain functional groups and/or have physical properties consistent with the majority of known drugs [5]. It is based on the assumption that typical drugs have same characteristic in common (pattern) that other compound lack [6]. In additional, drug-like compounds are expected to meet ADME (absorption, distribution, metabolism, excretion and toxicology) profiles [7]. From literature review, numeral of recent methods reported to codify set of "rules" or "filters" which is hoped to help chemist to recognize drug-like properties ragging from simple counting methods, physicochemical filters, functional group filters and chemistry space evaluation methods to artificial intelligence technique such as genetic algorithms, decision trees and neural network [8].

The first widely used drug-like filter was developed by [9] who proposed a surprising simple set of easily calculated physicochemical properties; the so-called "rule of five" (RO5). RO5 have been derived from the 90th percentile of orally drug candidates that have achieved phase II clinical trial [10]. It is an algorithm consisting of four rules in which many of the cut-off numbers are five or multiples of five, thus originating the rule's name [11]. RO5 stated that, designated drug-like candidate should have less than 10 hydrogen bond acceptors (HBA), less than five hydrogen bond donors (HBD), a molecular weight (MW) of less than 500 Dalton, and a partition coefficient log P of less than 5. Nowadays, this method has had a major impact on the daily practice of medicinal chemistry across the pharmaceutical industry and served as very useful guideline for drug discovery [12].

Gaining momentum from Lipinski publication, several other researchers reported analyses of identical drug-like filter. Veber et al. proposed that the number of rotational bonds (<10) and polar surface area (<140Å2) were two important properties to obtain oral bioavailability in rat [13]. Veber filter proved to be a useful descriptor for QSAR analysis [14] and prosperously highlight the importance of pharmacokinetics properties in drug discovery process [10].

Neural networks have been proved to possess high capability to solve such complex classification problems [15], but hardly ever be implemented in pharmaceutical applications. However, [6] and [16] seminal publication come into sight simultaneously in 1998 that described the successful employment of different neural network approach to classify drug-like and non-drug like pattern before followed by [3] and most recently [17]. As shown in Table 1, the prediction systems vary in the type of network architecture, datasets and topological descriptor used. [16] use the so-called ISIS keys with other molecular properties including RO5 as a descriptor. [6] and [3] selected atom Ghose and CONCORD atom types respectively. Meanwhile [17] use various topological indices as descriptor.

Corresponding author phone : Tel: +604-5996051, Fax:+6045941023
Email : [1]wmfahmi@yahoo.com, [2]ashidi@eng.usm.my,
[3]habibahw@usm.my and [4]fiuses@gmail.com
[1,2]School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Penang, MALAYSIA
[3]School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, MALAYSIA
[4]Faculty of Mechanical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA

TABLE 1: A COMPARISON OF VARIOUS METHOD FOR PREDICTION OF DRUG-LIKE MOLECULES

	[16]	[6]	[3]	[17]
Neural Network	Bayesian	SNNS [1]	MLP	MLP
Drug Database	CMC [2]	WDI [3]	MDDR [4]	MDDR [4]
Non-drug Database	ACD [5]	ACD [5]	ACD [5]	ACD [5]
Descriptor	ISIS keys	Ghose Atom Type	Concord atom types	Atom/bond indices
%Correct Drug	90	77	88	76
%Correct Non Drug	90	83	88	70

[1]Stuttgart Neural Network Simulator, [2]Comprehensive Medicinal Chemistry, [3]World Drug Index, [4]MACCS-II Drug Data Report and [5]Available Chemicals Directory

Motivated by the previous investigations, the advanced version of conventional MLP neural network called Hybrid Multilayer Perceptron (HMLP) was considered and used to form a decision making system for drug-like prediction. The intelligent system will be trained using Modified Recursive Prediction Error (MRPE) training algorithm. In this paper, we proposed a well known and easy excess RO5 + Veber filter features as our system topological descriptor (for molecule structure encoding). The main idea behind the selection of this simple descriptor is to assure that this system can be used on a daily basis as widely an audience as possible within a drug discovery organization. In the same time, sustaining the system performance. To validate our claim, we duplicate strictly experiment done by [17] using our proposed hybrid neural network and the simple descriptors for the new system. The performance of both systems will then be compared. [17] was preferred as comparison reference because; there are complete list of experiment dataset (molecules) provided in this paper which is not available in [3,6,16]. Hence, more solid comparison result can be produced.

II. HYBRID MULTILAYER PERCEPTRON NETWORK

Mashor showed in [18] that the MLP network was highly nonlinear and ever modeling a linear model using the standard nonlinear network is never be the best solution. The MLP network consists of a set of input layer, one or more hidden layer and an output layer. The output of the standard MLP network with *m* outputs can be expressed as:

$$\hat{y}_k(t) = \sum_{i=1}^{n_i} w_{jk}^2 F\left[\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t) + b_j^1\right];\ 1 \le k \le m \qquad (1)$$

where w_{ij}^1 and w_{jk}^2 denote the weight of the connection between input and hidden layer, weights of connection between hidden and output layer, respectively. b_j^1 and v_i^0 denote the thresholds in hidden nodes and inputs that are supplied in the input layer, respectively. F(•) is an activation function that is normally selected as a sigmoidal function.

The HMLP network is build as an optimum network of modeling both linear and nonlinear system. The HMLP network allows the network input to be connected directly to the output nodes with some weighted connections to form a linear system (i.e. represented by dotted line connection in Figure 1). This additional linear system is parallel with the original nonlinear system from the standard MLP model (i.e. represented by line connection in Figure 2). HMLP network proved to give significant improvement over standard MLP network and has been successfully applied to solve pattern recognition problem in many field [19].

For m output nodes, the output of the HMLP network is given by:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 u_k + \sum_{i=0}^{n_i} w_{ik}^l v_1^0(t)\ \mathrm{z};1 \le k \le m \qquad (2)$$

where

$$u_k(t) = F\left[\sum\nolimits_{i=1}^{n_i} w_{ij}^1 v_1^0(t) + b_j^1\right] \qquad (3)$$

where w_{ik}^l denotes weights of the extra linear connection between input and output layer, The weights $w_{jk}^2, w_{ik}^l, w_{ij}^1$ and threshold b_j^1 are unknown variables and should be selected to minimize the prediction error, defined as:

$$\varepsilon(t) = y(t) - \hat{y}(t) \qquad (4)$$

where $y(t)$ and $\hat{y}(t)$ are the actual and network outputs respectively.

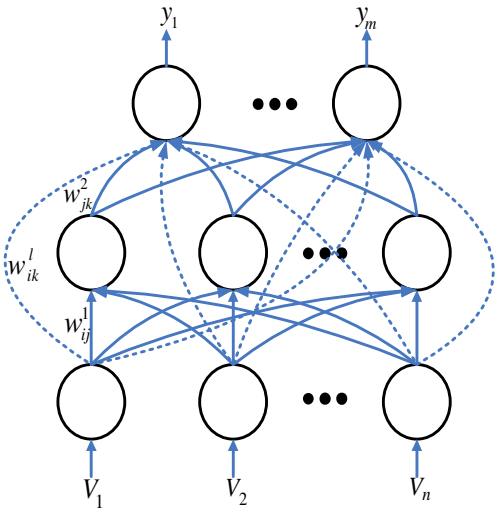

Figure 1: The HMLP network

Learning algorithm for the HMLP network has been proposed in [20] to handle the additional linear connections called Modified Recursive Prediction Error (MRPE). MRPE is a modified version of Recursive Prediction Error that originally derived by [21] and modified by [22] initially to train the MLP network. The MRPE algorithm is able to

improve the convergence rate of the RPE algorithm by using the optimized momentum and learning rate [20].

The RPE algorithm modified by [22] minimizes the following cost function:

$$J\left[\hat{\Theta}\right]=\frac{1}{2N}\sum \varepsilon^{T}\left[t,\hat{\Theta}\right]\Lambda^{-1}\varepsilon\left[t,\hat{\Theta}\right] \quad (5)$$

by updating the estimated parameter vector, $\hat{\Theta}$ (consists of *ws* and *bs*), recursively using Gauss-Newton algorithm:

$$\hat{\Theta}(t)=\hat{\Theta}(t-1)+\mathrm{P}(t)\Delta(t) \quad (6)$$

and

$$\Delta(t)=\alpha_m(t)\Delta(t-1)+\alpha_g(t)\psi(t)\varepsilon(t) \quad (7)$$

where $\varepsilon(t)$ and Λ are the prediction error and an m x m symmetric positive definite matrix, respectively, and m is the number of output nodes; $\alpha_m(t)$ and $\alpha_g(t)$ are the momentum and the learning rate respectively. $\alpha_m(t)$ and $\alpha_g(t)$ can be arbitrarily assigned to some values between 0 and 1, and the typical values of $\alpha_m(t)$ and $\alpha_g(t)$ are closed to 1 and 0, respectively. In [21], $\alpha_m(t)$ and $\alpha_g(t)$ are varied to further improve the convergence rate of RPE algorithm according to:

$$\alpha_m(t)=\alpha_m(t-1)+a \quad (8)$$

and

$$\alpha_g(t)=\alpha_g(0)\left(1-\alpha_m(t)\right) \quad (9)$$

where a is a small constant (typically a=0.01); $\alpha_m(0)$ and $\alpha_g(0)$ are initial values of $\alpha_m(t)$ and $\alpha_g(t)$ that have the typical values of 0 and 0.5 respectively. $\psi(t)$ represents the gradient of the one-step-ahead predicted output with respect to the network parameters:

$$\psi(t,\Theta)=\left[\frac{d\hat{y}(t,\Theta)}{d(\Theta)}\right] \quad (10)$$

The inverse correlation matrix, P*(t)* in Eq. (6) is updated recursively according to :

$$P(t)=\frac{1}{\lambda(t)}\left[P(t-1)-P(t-1)\psi(t)\left(\lambda(t)\mathrm{I}+\psi^{T}(t)P(t-1)\psi(t)\right)^{-1}\psi^{T}(t)P(t-1)\right] \quad (11)$$

where $\lambda(t)$ is the forgetting factor, $0<\lambda(t)<1$, and has been updated using the following scheme:

$$\lambda(t)=\lambda_0\lambda(t-1)+(1-\lambda_0) \quad (12)$$

where λ_0 and the initial forgetting factor $\lambda_0(0)$ are the design values. The initial value of the $\mathrm{P}(t)$ matrix, $\mathrm{P}(0)$ is normally set to $\alpha\mathrm{I}$, where I is the identity matrix and α is a constant, typically between 100 and 10,000.

The gradient matrix, $\psi(t)$ can be modified to accommodate the extra linear connections for one-hidden-layer HMLP network model by differentiating Equation (10) with respect to the parameters, θ_c to yield:

$$\psi_k(k)=\frac{dy_k(t)}{d\theta_c}=\begin{cases} u_k & if\ \theta_c=w_{jk}^2\ \ 1\le j\le n_h \\ v_i^0 & if\ \theta_c=w_{ik}^1\ \ 0\le i\le n_i \\ u_k(1-u_k)w_{jk}^2 & if\ \theta_c=b_j^1\ \ 1\le j\le n_h \\ u_k(1-u_k)w_{jk}^2v_i^0 & if\ \theta_c=w_{ij}^1\ \ 1\le j\le n_h, 1\le i\le n_i \\ 0 & otherwise \end{cases} \quad (13)$$

According to [21], the MRPE algorithm for one-hidden-layer HMLP network can be implemented as follows:

1. Initialize weights, threshold, P(0), a, b, $\alpha_m(0)$, $\alpha_g(0)$, λ_0 and $\lambda(0)$. Where b is a design parameter that has a typical value between 0.8 and 0.9).
2. Insert input to the network and compute the network outputs according to Equation (2)
3. Calculate the error prediction based on Equation (4)
4. Compute matrix $\psi(t)$ according Equation (13). Note that, elements of $\psi(t)$ should be calculated from output layer down to the hidden layer.
5. Compute matrix P(t), and $\lambda(t)$ according to Equation. (11) and (12), respectively
6. If $\alpha_m(t)<b$, update $\alpha_m(t)$ according to Equation. (8).
7. Update $\alpha_g(t)$ and $\Delta(t)$ according to Equation. (9) and (7), respectively.
8. Update parameter vector $\hat{\Theta}(t)$ according to Equation. (6).
9. Repeat Steps (2) to (8) for each training data sample.

III. Intelligent Drug-like Prediction

A. Experiments and result

Making comparison between different systems is a crucial task. In order to make a legit comparison, we followed closely the experiment methodology as mentioned in [17]. Note that, only 86% from original dataset in [17] were used due to limited source to our current database archive. Yet, we emphasize that the comparison still applicable since there are enough information recorded in [17] for a new accuracy recalculation. The accuracy [23] is calculated as follows:

$$\text{System accuracy (\%)} = \frac{\text{Total number of cerrectly predicted cases}}{\text{Total number of cases}}\times 100\% \quad (14)$$

There are a total of 508 data used in the system, which are 364 data for training purpose and 144 data for testing purpose. Among the training data, there are 240 drug samples and remaining as non-drug sample. Meanwhile, there are 98 drug and 64 non-drug samples used in testing phase. The training data set was used to build the system

model in the recognition stage while the testing data set was applied to evaluate the performance of the network.

The dataset had six attributes selected based on Lipinski RO5 and Veber drug-like filter features namely; hydrogen bond acceptors (HBA), hydrogen bond donor (HBD), Molecular Weight (MW), Partition Coefficient log P (Log P), rotatable bonds and polar surface area. Table 2 presents a summary of the molecule data characteristics.

TABLE 2: DESCRIPTION OF MOLECULES DATA

Number of data	508
Training sample size	364 (240 drug & 124 non drug)
Testing sample size	144(98 drug & 64 non drug
Descriptor, v_n	$n = 6$; i.e. RO5 (4) + Veber (2)
Output classes, y_m	$m = 2$; i.e. Drug-like and Non Drug-like Compound

Table 3 tabulates the performance for each network when implemented with their optimum network structure. Despite having trained with fewer data, it is clearly seen that the HMLP network is extensively capable to produce better drug-like prediction performance with 91.84% of accuracy. Furthermore, this result also shows that there is no significant superior classifier that outperforms one another for non drug-like prediction, although, [17] recorder slightly higher accuracy with 73.91 % compared to 71.74% of accuracy obtained by HMLP network.

TABLE 3: CLASSIFICATION ACCURACIES OF THE PROPOSED SYSTEM AND MIGUEL-SOLER ET AL.

Method	Proposed System	Miguel-Soler et al. [17]
% Drug Predicted Correctly	91.84	77.55
% Non Drug Predicted Correctly	71.74	73.91

IV. CONCLUSION

We have developed a simple intelligence drug-like molecule compound prediction system using HMLP network that trained using MRPE by manipulating Lipinski + Veber features as a topological descriptor. The system was designed to discriminate between drug-like and non drug-like chemical compound.

Although the results obtained so far are encouraging, this system development is still in a basic foundation. More investigations on both theoretical and practical aspects are needed to further vindicate the applicability of the proposed system as drug-like chemical compound prediction.

ACKNOWLEDGMENT

This work is partially supported by the Ministry of Science, Technology and Innovation (MOSTI) Malaysia, under ScienceFund grants entitle "Development of a New Architecture and Learning Algorithm of Artificial Neural Network for Determination of Potential Drug in Herbal Medicine".

REFERENCES

[1] Biswas D., Roy S. and Sen S. "A Simple Approach for Indexing the Oral Druglikeness of a Compound: Discriminating Druglike Compounds from Nondruglike Ones". *J. Chem. Inf. Model*, 2006, Vol 46, No 3, pp 1394–1401.

[2] Hann M.M., Leach A.R. and Harper G. "Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery". *J. Chem. Inf. Comput. Sci.*, 2001, No 41, No 3, pp 856–864.

[3] Frimurer T.M, Bywater R., Nærum L., Lauritsen L.N., Brunak S. "Improving the Odds in Discriminating "Drug-like" from "Non Drug-like" Compounds" *J. Chem. Inf. Comput. Sci.*, 2000, Vol 40, No 6, pp 1315–1324.

[4] Clark D.E. and Pickett S.D. "Computational methods for the prediction of 'drug-likeness'". *Drug Discov Today*, 2000, Vol 5, pp. 49–58

[5] Walters W.P. and Murcko, M.A. "Prediction of 'drug-likeness". *Adv. Drug Delivery Rev*. 2002, Vol 54, pp 255–271.

[6] Sadowski J. and Kubinyi H. "A Scoring Scheme for Discriminating Between Drugs and Nondrugs". *J. Med. Chem*, 1998, Vol 41, pp. 3325–3329

[7] Muegge I., Heald S.L. and Brittelli D. "Simple Selection Criteria for Drug-like Chemical Matter". *J. Med. Chem*, 2001, Vol 44, pp. 1841–1846.

[8] Muegge I. "Selection Criteria for Drug-like Compounds". *Med. Res. Rev*. 2003, Vol 23, pp 302–321.

[9] Lipinski C.A., Lombardo F., Dominy B.W and Feeney P.J. "Experimental and Computational Approaches to Estimate Solubility and Permeablity in Drug Discovery". *Adv. Drug Deliv. Rev*, 1997, Vol 23, pp. 3–25.

[10] Keller T.H., Pichota A., Yin Z. "A Practical View of 'druggability'". *Current Opinion in Chemical Biology*, 2006, Vol 10, No 4, pp 357-361.

[11] Quinn R.J., Carroll A.R., Pham N.B., Baron P, Palframan M.E., Suraweera L, Pierens G.K. and Muresan S. "Developing a Drug-like Natural Product Library". *J. Nat. Prod.*, 2008, Vol 71, No 10, pp 1792.

[12] Zhang M-Q., Wilkinson B."Drug Discovery Beyond the 'rule-of-five'". *Current Opinion in Biotechnology*, 2007, Vol 18, No 6, pp 478-488

[13] Veber D.F., Johnson S.R., Cheng, H.-Y., Smith, B.R., Ward K.W., Kopple K.D. "Molecular Properties That Influence the Oral Bioavailability of Drug Candidates". *J. Med. Chem*, 2002, pp 2615−2623.

[14] Hubbard R.E. Structure-based Drug Discovery: An Overview. Royal Society of Chemistry, 2006.

[15] Sadik K. and Mustafa O. "Atrial Fibrillation Classification with Artificial Neural Networks". *Pattern Recognition*, 2007, Vol 40, No 11, pp 2967-2973.

[16] Ajay X., Walters W.P and Murcko M.A, "Can we learn to distinguish between 'drug-like' and 'nondrug-like' molecules?". *J. Med. Chem*, 1998, Vol 41, pp. 3314–3324.

[17] Soler M.M., Giménez F.P., March F.J.G., Salvador M.T.S., Villanueva W.D and Bleda M.J.C. "Drugs and Nondrugs: An Effective Discrimination with Topological Methods and Artificial Neural Networks".*J. Chem. Inf. Comput. Sci.*, 2003, Vol 43, No 5, pp 1688–1702.

[18] Mashor M.Y, "Hybrid Multilayer Perceptron Networks". *International Journal of Systems Sciences*, 2000, Vol 31, No 16, pp. 771-785.

[19] Mat-Isa N.A, Al-Batah, M. S., Zamli, K. Z., Azizli, K. A., Joret, A. and Mat-Noor, N. R. "Suitable Features Selection for The HMLP and MLP Networks to Identify The Shape of Aggregate". *Construction and Building Materials*, 2008, Vol 22, No 3, pp 402-410.

[20] Mashor M.Y, "Modified Recursive Prediction Error Algorithm for Training Layered Neural Network", *International Journal of the Computer, the Internet and Management*, 2003, Vol 11, No 2, pp. 29-36.

[21] Ljung L and Soderstrom T, "Theory and Practice of Recursive Identification". *MIT Press*, Cambridge, MA, 1983.

[22] Chen S, Billings S.A, and Grant P.M, "Recursive Hybrid Algorithm for Non-Linear System Identification Using Radial Basis Function Networks". *International Journal of Control*, 1992, Vol 55, pp. 1051-1070.

[23] Kahramanli H. and Allahverdi. N. "Design of a Hybrid System for the Diabetes and Heart Diseases". *Expert Systems with Applications*, 2008, Vol 35, No 1-2, pp 82-89.