Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/140218

Prediction of Chemical Toxicity for Drug Design Using AIRS Algorithms and Hybrid Classifiers

Amena Mahmoud[1,*], *Taher Hamza*[2] *and M. Z. Rashad*[2]

[1]Department of Computer Science, Faculty of Computers and Information, Kafr El Sheikh University, Egypt
[2]Department of Computer Science, Faculty of Computers and Information, Mansoura University, Egypt

Received: 2 Jan. 2020, Revised: 12 Feb. 2020, Accepted: 22 Feb. 2020
Published online: 1 Mar. 2020

Abstract: Studying toxicology and its relationship to drugs is expected to provide substantial human benefits; however, such benefits demand the ability to recognize and understand drug side effects and prevent them from happening. The proposed model aims to improve toxicity prediction by classifying chemical synthesis using an Artificial Immune Recognition System (AIRS) algorithm. The core of the current approach is its emphasis on constructing a hybrid classification system that achieves an effective performance. This system is achieved by merging three different types of artificial immune recognition system algorithms with a detector-based classifier in a hybrid model and optimizing the final output to improve the overall system performance.

Keywords: Computational drug design, Biological Immune System, Artificial Immune System, Hybrid Computational Classifiers System.

1 Introduction

Utilizing intelligent approaches in medication disclosure and development procedures is quickly gaining in popularity, appreciation, and implementations. Furthermore, rapid progress at developing the computational drug design field has been achieved by improving software computational power and protein compound structure databases and by an increasing understanding of chemical molecular targets [1].

Computational drug design approaches are used to identify active drug candidates by choosing the most promising candidates for valuation and improving them, such as converting biologically active synthesis to appropriate medications by reinforcing their pharmaceutical properties. This approach is intended to minimize the size of the chemical area and therefore allow drug detection and optimization efforts to concentrate on more promising candidates [2]. Thus, these approaches are applied to effectively reduce resource requirements and minimize the time required for chemical synthesis and laboratory testing. Computational approaches are considered to be extremely promising techniques in the drug discovery field; they improve the prediction of drug properties by recognizing chemical compounds and substantially reducing laboratory testing and traditional resource demands [3,4].

Over the past decade, Artificial Immune Recognition System (AIRS) have attracted the concerns of researchers aiming to improve immune-based paradigms and mechanisms that involve complicated computational problems. The AIRS-based algorithms discussed in this paper are improved based upon the tradition of ”immunity principle” sets.

The contribution of the current research is to briefly introduce the most common toxicological experiences and evaluate the progress made by computational toxicity prediction to recognize toxic drug entities.

2 Literature Review

Computational toxicology—the recognition and prediction of toxicity through computer designed techniques is considered a novel field of applied science that integrates standards from biology and chemistry with computer science in an effort to determine the relations between chemical synthesis and toxicological endpoints that can be exploited to predict the toxicity of new syntheses that have not yet been tested in vivo. The

* Corresponding author e-mail: Amena_mahmoud@fci.kfs.edu.eg

fundamental inroads in this subject began early in the twentieth century with research into the interaction of molecules with biological systems.

An extensive volume of subsequent experiments expanded the level of proof for a receptor-mediated mode of action for the plurality of pharmaceuticals and toxicants and resulted in reasonably specifications of the fundamental structure-activity relationships. The thermodynamic foundation of the parametric techniques characterizing one property using the parameters that represent molecular structure was given by the linear free-energy relationships developed in the 1930s by Hammet [5] through his seminal research to quantify the effects of constituents on ester hydrolysis.

As the accessibility and power of computers have increased, academics [6] have begun to identify the roles of diverse molecular structures by associating linear-free energy relationship essentials with the available toxicity data. The fusion of toxicology, computer science, and statistical analysis is considered to be the beginning of the science of computational toxicology.

Another step was taken by HazardExpert, which is considered the beginning of computational toxicity research. HazardExpert was developed by CompuDrug Chemistry Ltd. and introduced a toxicity predictor expert system [7]. HazardExpert provides a prediction for the extent of toxicity endpoints, including immunotoxicity, mutagenicity, teratogenicity, irritation, neurotoxicity, and carcinogenicity. The proposed expert system includes a toxic partial synthesis knowledge base extracted from US EPA statements and literature on structure-toxicity relationships. The prophetic rules in HazardExpert are based on the influences of these fragments on biological systems and combine expert judgement with fuzzy logic.

3 Artificial Immune Recognition Algorithms

AIRS are computational intelligence models that were inspired by and simulate natural immune systems. During the previous decade, many computer science researchers have aimed to produce immune-based paradigms to solve complicated computational issues by using the distinctive characteristics of natural immunity, such as learning, memory, pattern recognition, and self-organization [8].

In this study, three basic AIRS algorithms are selected to first evaluate their efficiency at recognition and classification cases; then, they were employed to build the AIRS-based system. The algorithms are chosen based on the variations in the learning procedures used by each algorithm. All three algorithms are considered unsupervised learners [9].

3.1 ClonalG Algorithm

The ClonlG selection algorithm was proposed as a simulation of the Clonl selection hypothesis of obtained immunity, and it represents the characteristics and behaviours of antibodies in the immune system [10]. Its hypothesis suggests that when selecting B and T-cells (antigens for lymphocytes) and connecting them to a specific antigenic, each cell divides to make duplicates of itself and differentiates to form other cell types such as memory cells or plasma. Memory cells remain alive for an extended duration in the host to anticipate future recognition of specific antigens, while plasma cells stay alive for a lifetime and create enormous numbers of antibody molecules. A significant characteristic of that hypothesis is considered when choosing a cell that subsequently increases; the cell is subject to tiny duplication errors that change the form of the expressed receptors, and thus, the subsequent specific recognition capabilities of the antibodies that plasma cells create and the antibodies that bind to the surfaces of lymphocytes cells [10].

3.2 V-Detector Algorithm

The V-detector algorithm [11] stems from the negative selection mechanism, which is based on self-nonself differentiation behaviour in the immune system. The negative selection process identifies and ejects cells that are self-reactive during cell production and division. This procedure has been observed in the preparation of T-lymphocytes, naive versions of which mature using both positive and negative selection processes in the thymus [11].

3.3 aiNet Algorithm

The iNet algorithm is based on the immune network theory of the immune system [12], which suggests that any antibody has idiotypes (surface features) to which the receptors of another antibody can connect. These idiotypes are dynamic because of receptor interactions, in which receptors continually inhibit and excite each other through complicated receptor regulatory chains. In addition to the surface characteristics of the pathogen, the hypothesis underlying this algorithm assumes that the ClonlG selection procedure may also be triggered by the idiotypes of other immune cells and molecules and that the maturation procedures apply both to the receptors themselves and the idiotypes that they expose.

4 Hybrid Computational Classifiers Approach

The hybrid computational model is a well-known way to gain more accurate classifiers by merging different

algorithms. The hybrid learning approach forms a machine learning system composed of a collection of particular learner paradigms and a mechanism for incorporating their decisions by combining outputs; this mechanism produces single answers for input problems. Our proposed approach aims at merging a blend of effective experts to benefit from the outcomes created by the paradigm's constituent expert parts. By merging all the outputs from each expert paradigm, a final product with better performance can be obtained.

4.1 AIRS Based Hybrid Model

Figure 1 introduces the designed structure of the model proposed in this paper. The effective learning AIRS schema is constructed to allow the dynamic decisions of multiple classifiers to be incorporated. In this schema, the proposed accumulation procedure is able to adjust the modifications in both the input and output to improve the final decision. Using this architecture, the idea is to be able to understand the adjustments to the input set and their influences on the base classifiers.

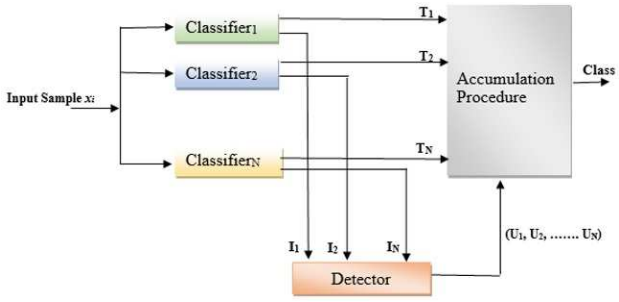

Fig. 1: The proposed Hybrid Classifier model

The proposed hybrid system is composed of the previously-mentioned AIRS algorithms (ClonlG, V-Detector, and iNet) fused into a classifier model using a combination of weighted averaging and majority voting.

4.2 Detector-Based Classifier

The proposed classifier model aims to extend a dynamic procedure reinforced by incorporating the decision of hybrid classifier models to improve the final output of the AIRS classifier models. This result is achievable by designing a detector ingredient that obtains characteristics from the connected classifiers to assist in effectively implementing the accumulation procedure. Our proposed detector permits the classifier paradigm to recognize modifications to the input data and determine their effect on classifier efficiency; then, according to those adjustments, formulate the decision of the entire model. The learning procedure extends the paradigm by adding the desired adaptability to change the classifiers' input/output to enhance the aggregate classification process.

Considering that purpose, two primary components are seriously considered when planning the detector-based schema. Initially, the proposed detectors design basically relies on token characteristics from the proposed classifiers. The detector then requests new characteristics for the current problem that differ from the previous characteristics used for classification purposes. However, characteristic extraction is not found to be an easy mission because the characteristics are not typically obtainable. The second component is base classifier planning, in which the mechanism used for training is a continuous challenge.

Efficiency diversity among the adopted classifiers is a result of several factors such as the internal randomness from the training algorithm itself, the selection of training and testing sets, and random classification error [13]. To achieve an enhanced classification performance, the base classifiers are trained sufficiently and appropriately.

The proposed detector-based schema presents a novel procedure for calculating the confidence level of the upcoming output $T = [T_1, T_2, ..., T_n]$ in the classifier model by extracting characteristics after each training phase and using them to derive a confidence value allocated to each subsequent output. The confidence levels for the proposed classifiers are demonstrated by vector $I = [I_1, I_2, ..., I_n]$, where n_p is the number of classifiers. Then, the subsequent result with the confidence measurement is utilized in the accumulation procedure implementation to obtain the overall decision.

4.3 Proposed Classifier Ingredients

At the previously-introduced schema, the classifier ingredients clarify the particular classification mechanisms utilized to compose the hybrid paradigm. The function of each selected classifier is to divide characteristic space into class-specific decision zones, which are then applied to classify the novel input patterns into the classes established in advance. First, the proposed classifiers are trained sufficiently using the same training datasets with the same characteristics spaces, allowing each specific classifier address the same classification issue utilizing its specific procedures. The suggested schema proposes that each classifier results in a predicted class identifier $T = [T_1, T_2, \ldots, T_n]$ and the confidence level of a classifier concerning its predicted output I =

$[I_1, I_2, \ldots, I_n]$, where n_p is the number of classifiers. The overall result of the hybrid model classifier is obtained by fusing the outputs of all of the specific classifiers.

The Detector Ingredients. The primary role of the proposed detector is to create a weighting factor for every classifier. The generated weights are identified by the vector U = $[U_1, U_2, \ldots, U_n]$, where n is the number of classifiers. The weights present the confidence level for each selected classifier related to its classification precision. The assigned weights are then applied to join these several classifiers by applying a standard multiple-classification resolution mechanism, such as majority voting or weighted averaging.

The Accumulation Procedure. The accumulation procedure refers to the integration technique applied to merge several outputs created by the proposed classifiers to reach a final decision by the proposed hybrid paradigm. The accumulation procedure utilizes the weights returned from the detector that specifies its confidence in the result by each classifier (as shown in Figure 1). The various classification outputs are accumulated, resulting in an eligible decision using a basis-joining technique. Various joining techniques can be utilized in the incorporation layer; this study concentrates on the weighted averaging technique.

The proposed adaptive learning AIRS-based classifier model has several major processing stages:

1. Initialization: the first step is training the introduced base classifiers (for instance, the iNet, ClonlG and V-Detector algorithms) on the selected training dataset to generate a trained model of all the algorithms, which are utilized in the testing step.

2. Experts Decision: for each sample x_j in the testing dataset, the predicted class for x_j is determined by each AIRS classifier identified by T = $[T_1, T_2, \ldots, T_n]$, where T_j is the portended result of classifier number j.

3. Confidence Measurement: acquire the confidence degree vector I = $[I_1, I_2, \ldots, I_n]$ on the predicted result to every classifier by:

• Calculate the Euclidean distance between sample x_j and all detectors in detectors set.

• If distance ¡ threshold, then set confidence level I_j is equal to 100 % and $T_j = 1$. Otherwise, set T_j= 0 and divide the area around x_j into bins and then calculate I_j by:

$$I_j = 1 - \frac{\sum_{j=1}^{10} n_j \times k_j}{\sum_{j=1}^{10} n_j} \quad (1)$$

Where nj is the no. of detectors in every bin and kj is bin index.

The above stages are executed iteratively for all the test samples by the hybrid AIRS classifier model. In this case, the adaptive learning AIRS-based classifier paradigm functions according to the following scenario. First, the diverse elements of the hybrid paradigm are trained on the entire training dataset. Then, the trained models are used to produce a decision on an input pattern x_j from the testing dataset. The proposed schema uses the V-Detector, ClonlG and iNet AIRS algorithms as base classifiers. The predictions from the AIRS classifiers are identified by the vector T = $[T_1, T_2, \ldots, T_n]$. According to that, the confidence values symbolized by vector I = $[I_1, I_2, I_3]$ are calculated, with a confidence value assigned to each of the base classifiers to adjust their first-level precisions concerning the class to be predicted.

5 Results

An experimental study was performed to confirm the efficiency of the proposed AIRS hybrid classifier paradigm using various anticancer drugs synthesis datasets. The major goals in this experiment are to clarify the effectiveness of the proposed adaptive learning AIRS paradigm compared with traditional classifiers and to evaluate the effectiveness of various joining techniques, such as weighted averaging and majority voting.

In this study, the datasets used for the experiment are analysed initially before applying the traditional classifiers and the hybrid model. The datasets are divided into two sets: 90% of the samples are used as a training dataset and 10% as a testing dataset. In the experiment, 10 runs (each containing 10 datasets) are executed to ensure that diverse training patterns and testing patterns are selected randomly for every run.

Before estimating the proposed hybrid model, the V-Detector, ClonlG and iNet AIRS algorithms are selected through an objective experiment to evaluate the performances of each individual classifier for this classification task. The results are shown in Table 2 and in Figures 2, 3 and 4 (where the x-axis represents the number of runs and the y-axis represents the criteria level). The performance evaluations comparing the adopted algorithms are listed in Table 1. The results show that the V-Detector algorithm has achieved better classification precision than the ClonlG and iNet algorithms in seven of the ten runs. ClonlG is best in three of the ten runs, while the third algorithm iNet has achieved the minimum precision. However, in the sensitivity criteria calculation, iNet is sufficiently consistent that it achieved the optimum results over the ten rounds, with a 100% rate. The V-Detector algorithm is second-base, with results that are approximately 68% while the ClonlG algorithm has achieved poor results on this criterion. Regarding accuracy, the Clonal G algorithm has consistently achieved the optimum results across all ten runs; the V-Detector has achieved second place with its results (approximately 77%); while the iNet algorithm

results has achieved the highest accuracy in 0% of the overall runs.

Table 1: Performance evaluations

Performance evaluation	Definition
Precision	Measures the ratio of the correctly classified samples.
Sensitivity	Measures the concrete positive samples that were classified
Accuracy	Measures the concrete negative samples that were classified

Table 2: Performance properties' results for the AIRS Algorithms

Run #	Precision			Sensitivity			Accuracy		
	Clonal G	V-Detector	aiNet	Clonal G	V-Detector	aiNet	Clonal G	V-Detector	aiNet
1	0.743	0.771	0.286	0.3	0.9	1	0.92	0.72	0
2	0.559	0.765	0.294	0	0.7	1	0.792	0.792	0
3	0.686	0.743	0.371	0.154	0.692	1	1	0.773	0
4	0.8	0.743	0.2	0.143	0.714	1	0.964	0.75	0
5	0.8	0.744	0.2	0.43	0.714	1	0.893	0.75	0
6	0.657	0.743	0.343	0.25	0.667	1	0.87	0.783	0
7	0.704	0.796	0.235	0.125	0.75	1	1	0.692	0
8	0.75	0.719	0.188	0	0.667	1	0.923	0.731	0
9	0.676	0.735	0.412	0.357	0.429	1	0.9	0.95	0
10	0.618	0.706	0.471	0.313	0.625	1	0.889	0.778	0

To create the proposed hybrid model, a combination of the V-Detector, ClonlG, and iNet AIRS algorithms were selected through an objective evaluation of their performances on the current classification task. The experiential performance results (shown in Table 3) describe the three criteria of precision, sensitivity, and accuracy the algorithms achieved during the 100 sample runs.

6 Conclusion

To validate the proposed adaptive learning model, an experimental study is conducted on a combination of the three AIRS algorithms, which are applied to anticancer drug synthesis using actual drug datasets. The selected AIRS algorithms for the experimental study are inspired by the fundamental V-Detector, ClonlG, and iNet immunology algorithms.

Table 3: The total performance of the AIRS Algorithms

Run #	Precision	Sensitivity	Accuracy
1	0.891	0.949	0.946
2	0.885	0.933	0.844
3	0.895	0.942	0.923
4	0.867	0.932	0.850
5	0.834	0.922	0.936
6	0.857	0.950	0.855
7	0.866	0.956	0.833
8	0.889	0.928	0.947
9	0.875	0.978	0.869
10	0.899	0.966	0.820

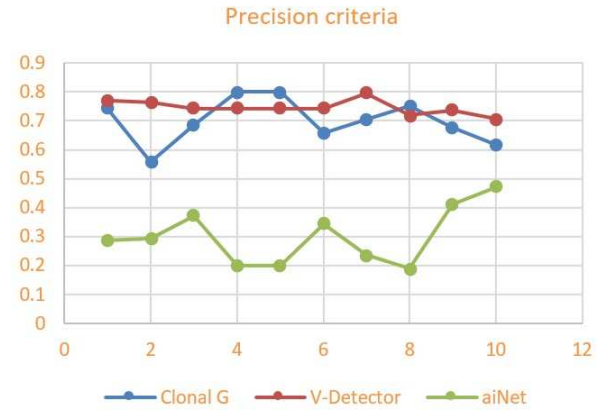

Fig. 2: Precision performance for AIRS Algorithms

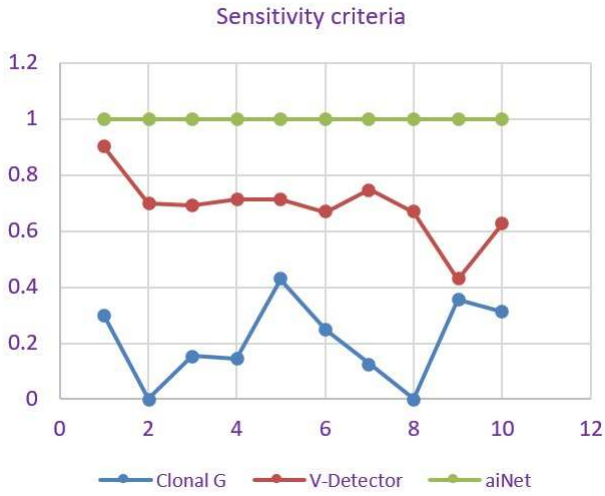

Fig. 3: Sensitivity performance for AIRS Algorithms

The presented experimentation results are reported to estimate the performance of the proposed hybrid model for the current classification task. The experimental results are obtained by measuring the precision, sensitivity, and accuracy that the three algorithms have achieved over 10 sample runs each on 10 datasets. The results suggest that the hybrid AIRS algorithms are more efficient for the drug development industry than the previous approaches.

Different mechanisms and schemes are presented to address these challenges, and a literature review is included as well. A new AIRS-based classifier model that fuses the three proposed AIRS algorithms is presented as a first step toward achieving the research objectives. A novel mechanism to adjust the confidence degree of the

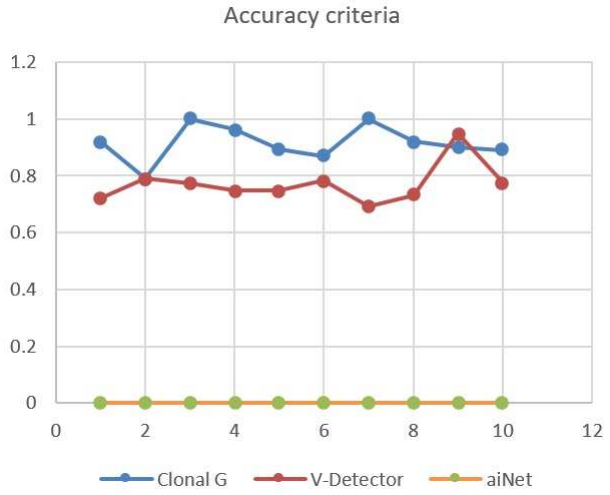

Fig. 4: Accuracy performance for AIRS Algorithms

proposed hybrid model classifiers is presented in this research. The process focuses on setting weights for each classifier in the hybrid classifier model based on its classification efficiency and ability to obtain the best performance.

References

[1] Green DV., Virtual screening of virtual libraries, *Prog Md Chm*, 41(2003).
[2] E. Pozzan A., Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery, Curr Pharm Des 2006; 12(17): 2099–2110.
[3] Stahl M, Guba W, Kansy M., Integrating molecular design resources within modern drug discovery research: the Roche experience, Drug Discovery Today 2006;11(7–8): 326–333.
[4] Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA, Applying computational modeling to drug discovery and development, Drug Discovery Today 2006;11(17–18):806–811.
[5] Hammett LP., Physical organic chemistry: Reaction rates, equilibria and mechanism, 2nd edition. New York: McGraw-Hill, 1970.
[6] Enslein K, Craig PN., A toxicity prediction system, *J Environ Toxicol*, 2, 115–21, (1978).
[7] Smithing MP, Daravas F. HazardExpert: An expert system for predicting chemical toxicity, *In: Finlay, JW, Armstrong DJ, Robinson*, SF, editors, Food safety assessment. Washington, DC: American Chemical Society, 191-200, (1992).
[8] Delahunty A. and Callaghan D., Artificial Immune Systems for the Prediction of Corporate Failure and Classification of Corporate Bond Ratings, *Master of Management Science*, University College Dublin, August, (2004).
[9] Timmis J., Hone A., Stibor T. and Clark E., Theoretical Advances in Artificial Immune Systems, *Theoretical Computer Science*, 1(403), 11-32, (2008).
[10] F. M. Burnet, The clonal selection theory of acquired immunity, *Vanderbilt University Press*, (1959).
[11] Z. Ji and D. Dasgupta, Revisiting Negative Selection Algorithms, *Evolutionary Computation*, (2007).
[12] De Castro L. and Zuben F. J., aiNET: An artificial immune network for data analysis, *Idea Group Publishing*, Chapter XII, 231-259, (2001).
[13] Küncheva L. I., Combining Pattern Classifiers: Methods and Algorithms, John Wiley and Sons, Inc, 2004.

Amena Mahmoud is an assistant lecturer of computer science dept. at Kafr El Sheikh University, Egypt. She is a Ph.D. researcher in Computer Science dept. at the Faculty of Computers and Information Mansoura University, Egypt and is interested in BioInformatics and Artificial Intelligence research topics like: Pattern Recognition, Machine Learning and Image Processing.

Taher Tawfiq Hamza is a professor of computer science at Mansoura University, Egypt. Professor Taher holds a Ph.D. in Computer Science from Faculty of Science Mansoura University in Egypt. He has served as a head of computer science dept. and a vice dean of faculty of computers and information systems Mansoura University.

Magdi Zakaria Rashad is a professor of computer science at Mansoura University, Egypt. Professor Magdi holds a Ph.D. in Computer Science from Faculty of Engineering Cairo University in Egypt and is the author of over 160 papers published in refereed international journals. He has served as a head of computer science dept. and a vice dean of faculty of computers and information systems Mansoura University. He has also served as a reviewer for various international journals, such as IJCSIT, MJCSIS and he is interested in the following fields: Artificial Intelligence, Pattern Recognition, Machine Learning, Image Processing and Cloud Computing.