J. Appl. Environ. Biol. Sci., 6(12)16-26, 2016
© 2016, TextRoad Publication
ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Drug Design and Discovery using Differential Evolution

[1]Waseem Shahzad, [2]Ahsan Yawar, [3]Ejaz Ahmed

Department of Computer Science
National University of Computer and Emerging Sciences, Islamabad, Pakistan

Received: July 25, 2016
Accepted: October 7, 2016

ABSTRACT

The fields such as sciences, biology, chemistry and medicine are facing many challenges due to large scale of data information and persistently continuous expansion of data. In order to utilize persistent information, there is a need for the understanding such data. Although, current Computational Intelligence (CI) techniques such as artificial neural networks, fuzzy systems, evolutionary computation algorithms and other approaches have shown great success in various kinds of model designs problems. The introduction of Differential Evolution (DE) is relatively a new challenge that has attracted new attentions. In this paper, we develop a new method for the drug design and discovery (DD&D) using DE, which is an evolutionary algorithm. We have conducted experiments for the DD&D using DE, and compare it with state-of-the-art CI techniques, and DD&D techniques. In this paper, our focus is on the most important drug targets findings and comparisons for deadly diseases such as Polio, Typhoid and Rabies. The experimental results show that our proposed method for DD&D achieves better performance in terms of accuracy, consistency and computational complexity.

KEYWORDS: Evolutionary Computing, Differential Evolution, Computational Intelligence, Artificial Neural Networks, Genetic Algorithms

1. INTRODUCTION

Recent developments in science and bio informatics depend on the access of large data using computational techniques for finding process of new drugs. This has created an immense opportunity for DD&D, and for macro scale knowledge discovery. A drug is a chemical substance used in the treatment, cure, prevention, or diagnosis of disease; otherwise used to enhance physical or mental well-being [1]. Drug discovery is the process of finding potential new drugs. Drug discovery has evolved from early serendipitous discovery from natural sources such as morphine from poppy seeds. In the current era, the modern DD&D initiates therapeutic effect, induced by modulating the disease or biological target [2]. Searching for potential drugs, compounds are normally tested for their ability to modulate the target. Screening of libraries of millions of compounds is managed with the help of technologies such as high-throughput screening (HTS) and combinatorial chemistry. Screening is continuously growing and number of possible molecules synthesized, and tested infinitely [3].

Drug design or rational drug design means inventing the process of finding new medications or new bioactive compounds with favorable properties. Ligands mean a molecule that binds to another (usually larger) molecule and serve for biological purpose. Design process originates ligands but can also stem from virtual screening of compound libraries. Virtual screening is a computational technique; represent the computational counterpart of HTS screening. Rational drug designing has two approaches such as structure based and ligand based approaches. Structure based drug designing involves designing of small molecules based on the 3D structure of the reported protein; a 3D model can be built by using a suitable template [4]. CI techniques are then applied to find new ligands, or modify existing ones that fit into the target structure. For example, structure-based drug design is managed by De Novo design.

After the drug molecule is virtually designed, the next process is to design a manufacturing process. This includes creation of drug follow clinical trials. If the drug passes the clinical tests, quality standards, and manufacturing resources then it is introduced to the market [6].In target identification work, we choose a molecule in order to target with a drug. If the cause of the disease is known, in other words how the disease is spread out then pharmaceutics selects target for the new drug. Particular target is one molecule, for example, a gene or a protein that relates to a particular disease. Pharmaceutics have to make sure that in the early stage of drug design they choose a target to make up possible [7]. In other words the target must be "drugable" that has the capability to interact and be influenced by a molecule. The target validation phase is to test the identified target and confirm its role in the disease. After a possible target has been selected, the researchers should illustrate that it has role in the disease and it can act upon by the drug [8]. The researchers suggest that a target is significant to a particular disease when it is studied or by carrying out experiments in

***Corresponding Author:** Waseem Shahzad, Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan. waseem.shahzad@nu.edu.pk

living organism's cells as well as in animal models of the disease. From programming perspective, we have shown a chromosome of a drug pattern e.g. *Chromosome*: 0010101011011, where 0's means inactive molecule, no role in the disease and 1's means active molecule have a role in the disease. If the researchers are able to find the nature of the disease, researchers and scientist may immediately start finding drugs with safety tests of promising drugs and lead compound [9, 10]. This leads to preclinical testing. This means main focus of scientist is on the extensive testing strategies to find that whether they should allow the drug testing on humans or not.

In this paper, we have also indicates the latest techniques that are currently being used in DD&D domain by most of the pharmaceutics. These techniques have been implemented in order to bring new medicines to the market for cure/treatment of many deadly diseases such as Polio, Typhoid and Rabies. We develop a new method for the DD&D using DE which is an evolutionary algorithm. It provides an optimal solution for new medicine after comparisons of results.

The rest of the paper is organized as follows; section 2 provides foundation for our research motivation towards designing drug for new medicine using CI and De Novo design, section 3 provides an optimal solution of DD&D using DE. Section 4explains the findings and experimentation comparisons of proposed technique with other techniques.

2. Drug Design and Discovery using Computational Intelligence

The CI techniques will contribute in the research in DD&D till pre-clinical research. The idea of drug discovery using CI techniques is introduced as a novel technology intended to increase efficiency of the DD&D [11]. Although, CI approaches are used to address complex real world problems. Artificial Intelligence (AI) involves many branches of machine learning, statistical pattern recognition and clustering technique. It also includes nature-inspired approaches, such as Artificial Neural Networks (ANN), evolutionary computing, or fuzzy logic modeling, collectively represented as CI. Similarly, state-of-the-art CI or AI approaches such as ANN, fuzzy logic, genetic algorithms (GA), genetic programming, evolutionary programming, evolutionary strategies, Particle Swarm Intelligence (PSO), Ant Colony Optimization (ACO), and Support Vector Machine (SVM) are playing significant role in the research, which are used frequently to fight drug development and discovery issues [12, 16, 17].Similarly, ACO is also used in information system's distance related problems such as traveling salesman problem [14]. De Novo technique is widely used in the context of DD&D. We include selection of data modeling, relevant information, regression, classification, optimization and prediction for types of applications of AI methods [15].Various techniques of machine learning include SVM, clustering, classification, and statistical-based methods are also used in research community [15].

Today, we have tremendous CI techniques and machine learning methodologies but it is hard to decide which technique suits perfectly for the drug design scenario. However, all the existing techniques work well on certain conditions it is designed for. To further improve the optimization of CI techniques enhancing the optimization capabilities in discovering the drug and design problems. DD&D approach will engage in solving global optimization problems. This can be supplemented by an example, if the samples conformational space is going to look for the most optimal molecular structure of the drug [13].

De Novo design is useful in designing lead compounds for structure-based drug design [8]. We have mentioned it earlier, a drug is a key molecule involved in a particular metabolic or signaling pathway specific to a disease. De Novoimproves significantly with customization, combining multi-objective evolutionary algorithms with local search techniques and expands evolutionary algorithms by incorporating novel features such as self-adaptation capabilities. That is why we have chosen DE technique. Through results, we will show how DD&D will advance with DE technique. However, it has created its space in the list of evolutionary algorithms but it is proved to be 30-40% more efficient compared with most evolutionary algorithms.

3.Drug Design and Discovery using Differential Evolution

Evolutionary computation has been introduced a new efficient algorithm called differential algorithm by Price and Storn in early 1995 [18]. It got popularity because of its extraordinary qualities including its efficiency, robustness and handled with less control parameters which make it a simple global optimization technique. As compared to the other evolutionary algorithms, it is considered as the most powerful algorithm in view of solving real world optimization problems. DE is better in convergence speed and robustness than other population based optimizing functions used in real world problem solving [19]. DE works excellent in exceptions; it means that it is not affected by lengthy computation time due to its stochastic nature.

DE has explored a new dimension to the direct global search optimization. DE has a lot of success stories, when it was applied in finding global minimizations on problems. It can be quoted with the comparison of results with the direct search algorithms.

3.1 Robustness of Differential Evolution

However, the robustness of DE decreases with the increase of dimensions. We have used the DE for the problem of DD&D of large dimensions. We will show some experimentation using DE for drug design in experimentation section

for which transitions are rarely publically accessible. The localization of search is integrated with the global exploration which is a prime feature of DE [20]. We will show the experimentation in which the outputs of the technique reflect the strengths of DE.

3.2 Basics of Differential Evolution

As earlier mentioned DE is a stochastic natured algorithm. Now we will see the mathematical formulation of the DE:

Objective function f: $X \subseteq R_D \rightarrow R$ with the possible region X!=∅, the minimization problem is to find best optimal solution as shown in equation 5.1:

$$x^* \in X \text{ such that } f(x^*) \leq f(x)\ \forall x \in X \qquad \text{eq(5.1)}$$

Where $f(x^*)$ is called fitness of the bests solution and x represents elements of sample or solution space.
Finding optimized value is mandatory in the field of engineering, statistics, finance and commerce. Similarly the real world problem such as DD&D has an objective functions that are non-linear, noisy, non-continuous and non-differentiable. It has to search the optimal solution through the multi-dimensional space. These kinds of problems are very sensitive and difficult to find the solutions. DE is better than GA as the nature of the problem is that it may be stuck in the local minimum solution [21]. If it does not solve the problem analytically, DE may get the solution through the search approximation technique.

It generally generates a trial vectors (v) using the following formula of equation 5.2:

$$v_i = x_{r1} + F \cdot (x_{r2} - x_{r3}) \qquad \text{eq(5.2)}$$

It elegantly replaces the two operations that are Crossover and Mutation. DE technique uses less parameter to be tuned and has self-organizing ability. The solution space with two out of three operations is shown in figure 4.

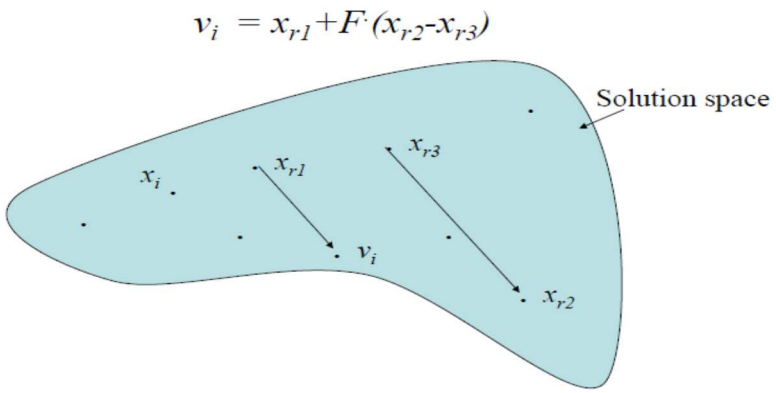

Figure 4: Solution Space [22]

Traditional optimization algorithms may find the optimal solution but the question is that there can be more than one single optimal solution [23]; in that case the traditional optimization algorithm will fail to find a group of possible optimal solution. In that case DE may come up as an effective solution to the problem.

The pros of DE are Simple and easy to implement, widely use, extensions and variations are easily available, less parameters to be tuned, and only one DE parameter needs to be set. Similarly, cons of DE are no convergence proofs, better result is dependent upon the set of population size, and complexity is high O(N^2).

3.3 Pseudo code of Differential Evolution

As mentioned earlier, DE is a population-based search technique alike other standard evolutionary algorithms. The uniqueness lies in the reproduction stage where DE gives edge working efficiently on flat surfaces. This code can is modified with the case of drug designing and discovery. Its implementation has been applied successfully in finding the righteous drug against a disease. In a nutshell, DE operates better on fitness surfaces which are flat:

- Let $g = 0$, initializes the population pr and ƛ.
- Initialize random population Cg of N numbers of candidates.
- For each compound (individual), Cg, n, (n=1,…,N):
  select$c1$, $c2$, $c3 \sim U$ (1,…, N), with $c1 \neq c2 \neq c3 \neq c$

select $i \sim U(1,\ldots, I)$
for j =1…, I
if ($U(0,1) <$ pr or $j=i$)
Og, nj = C$g, n3j$ + ƛ (C$g, n1j$ – C$g, n2j$)
else
Og, nj = Cg, nj

- Now select the new population in next generation C$g+1$ of N candidates (*individuals*):
  Evaluate C[*fitness*]
  if (C[*fitness*] is better than P[*fitness*])
  $P = C$ // *replace parent with child*
  else
  P[i] = P // *keep previous survived parent to next iteration*
- Run iterations for convergence, yet algorithm is not converge, repeat the above step till stopping criteria is reached or desired fitness is achieved.
- According to fitness of candidate of the drugs merged out, it can be named as selected drug against particular target (disease).

We have seen the detail working of DE optimization technique. The finding of best values based on decision variables is the main objective of optimization algorithm. Other objective includes minimization of the cost and it increases the reliability or any other objective function [24]. DE has earned a reputation as a global optimizer over other evolutionary algorithms used during a recent decade. As we have seen in pros of DE algorithm, it is highly considered to use on optimization problem because of its simple structure, easy to understand and use, and its robustness [25].

3.4 Flowchart of Proposed Technique

The flow of the DE algorithm in case of drug discovery is simple and effective. In figure 5, the algorithm starts by initializing the random population of drug candidates. In the next step of the program, fitness of each individual (drug candidate/ molecule) is evaluated. We have added a check loop in our code, which says run till the generation count is less than the total number of count mentioned. Here it is important to mention that the generations are run iteratively and in iteration specific number of generations occurs. It is important to tune number of generations in order to find the right drug against a disease i.e. find drug that is most effective, fittest. Next the children are created by selecting three random children that are not same; again calculate the fitness of children, by evaluating the fitness of the children. We can decide whether to add children to new population or to keep previous parents in the new generation. In our drug design case, we are looking for most effective individuals, we have pruned out the candidates (molecules or compounds) that have fitness value less than 0.5. We only move effective individuals to next generation that has value higher than 0.5. This helps us in removing less fit individuals instantly. The algorithm will run till the stopping criterion is met. In our algorithm if the generations count matches the total counts then the algorithm terminates giving out the output in form of molecules with their fitness.

Our approach aims is to introduce DE as an alternative more effective approach towards DD&D. The DE technique is measured as fast and robust optimization method. The design has an effective global optimization capability as shown in figure 5. The application behavior is fast and its architecture is not complex just like other evolutionary algorithms. One of the key features of our approach is its high precision calculation and removal of less fit individuals promptly. The structure of this approach has reduced tremendous steps, few parameters require to setup and providing space for individuals to keep promoting evolution within less time.

Randomization of initial population equalizes the individuals which creates an opportunity for each candidate to come forward. The generation can be defined as the evolution step towards searching for better candidate(s)/ molecule(s). Fitness function defined afterwards, plays a key role contribution towards finding effective drug. In our design, fitness function is the heart of the algorithm which is tuned for maximization of productivity/effectiveness. Our fitness function gives high rank to those solutions which discover better drugs.

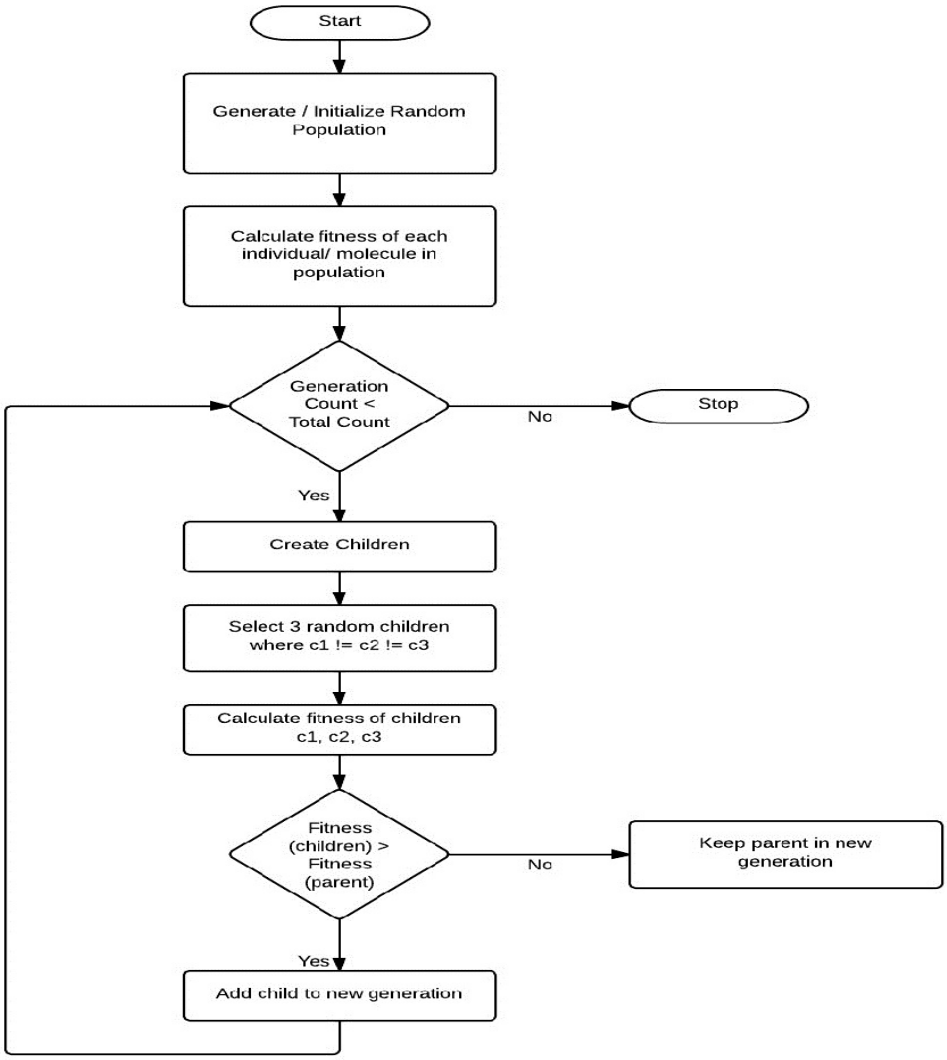

Figure 5: Flowchart diagram of proposed technique

The chromosome modeled in our design is of fix length i.e. 12 gene chromosome as shown in figure 6. The chromosome contains active and inactive states molecules, denoted by 0's and 1's.

0	1	1	0	0	0	1	0	1	1	0	0	0

The 1's are active contribution of molecule in a disease and 0's are inactive molecules in a particular disease. When the parents are initialized, chromosome comes out from the dataset evolved which is discussed in detail in experimentation. It is completely on principals of evolutionary algorithms that randomization is ensured. Later on, against each active gene, molecular weights count on the total weight of the chromosome.

In next section, we test our model over different diseases and observe our model behavior by comparing the results with cases applied with other techniques. How the weak candidates are pruned out, fit candidates more to next generations to form an effective drug and at a certain point, the fitness stops improving further.

3.5 Datasets for Experimentation

For experimental purpose, we have used Drug Bank databases, widely used in bioinformatics and cheminformatics [30]. It is an open data drug and drug target database freely available resource to the public. The datasets of Drug Bank contains large amount of pharmaceutical information which includes drug target, sequence, structure and pathway. It comprises of 6811 drug entries including FDA (Food and Drug Administration USA), approved small molecules drugs. Around 150 drugs, 87 nutraceuticals and 5080 drugs can be used for experiments. Drug Bank information has been

intended to use for educational and scientific research [26]. We have picked out a specific portion of the dataset in order to use our algorithm in case of polio disease.

The datasets of Drug Bank consists of various attributes i.e. large amount of information is available about a drug. This databank has ability to be used for vast purpose research. A focused research will be concerned with this related information and may ignore others. This is how the databank sample record looks like for (Refludan) a single compound drug. The drug is in v card format which can be used to import as a file on various applications.

#BEGIN_DRUGCARD DB00001	
# Absorption:	Bioavailability is 100% following injection.
# Biotransformation:	Lepirudin is thought to be metabolized by release of amino acids via catabolic hydrolysis of the parent drug. However, con-clusive data are not available. About 48% of the administration dose is excreted in the urine which consists of unchanged drug (35%) and other fragments of the parent drug.
# Brand_Names:	Refludan
# Drug_Registry_Number:	120993-33-5
# Chemical_Formula:	C287H440N80O110S6
# Chemical_Structure:	>DB00001 sequence LVYTDCTESGQNLCLCEGSNVCGQGNKCILGSDGEKNQCVTGEGTPKPQSHNDGDFEEIP EEYLQ
# Creation_Date:	2005-06-13 13:24:05 UTC
# DPD_Drug_ID_Number:	02240996
# Description:	Lepirudin is identical to natural hirudin except for substitution of leucine for isoleucine at the N-terminal end of the molecule and the absence of a sulfate group on the tyrosine at position 63. It is produced via yeast cells.
# Dosage_Forms:	Powder, for solution Intravenous
# Drug_Category:	Anticoagulants Antithrombotic Agents Fibrinolytic Agents
# Drug_Type:	Approved Biotech
# FDA_Label_Files:	1997-09-01 /drugs/DB00001/fda_labels/402
# Generic_Name:	Lepirudin
# Half_Life:	Approximately 1.3 hours
# Indication:	For the treatment of heparin-induced thrombocytopenia
# Mechanism_Of_Action:	Lepirudin forms a stable non-covalent complex with alpha-thrombin, thereby abolishing its ability to cleave fibrinogen and initiate the clotting cascade.
# Melting_Point:	65 oC (Otto, A. & Seckler, R. Eur. J. Biochem. 202:67-73 (1991))
# Molecular_Weight_Avg:	6963.425
# Molecular_Weight_Mono:	6958.962
# Organisms_Affected:	Humans and other mammals
# Pharmacology:	Lepirudin is used to break up clots and to reduce thrombocytopenia. It binds to thrombin and prevents thrombus or clot formation. It is a highly potent, selective, and essentially irreversible inhibitor of thrombin and clot-bond thrombin. Lepirudin requires no cofactor for its anticoagulant action.
# RxList_Link:	http://www.rxlist.com/cgi/generic/lepirudin.htm
# State:	Liquid
# Toxicity:	In case of overdose (eg, suggested by excessively high aPTT values) the risk of bleeding is increased.
# Wikipedia_Link:	http://en.wikipedia.org/wiki/Lepirudin
# Drug_Target_1_Chromosome_Location:	11
# Drug_Target_1_GO_Classification:	>>> Function: thrombin activity Function: binding Function: ion binding Function: cation binding Function: calcium ion binding Function: catalytic activity Function: hydrolase activity Function: peptidase activity Function: endopeptidase activity Function: serine-type endopeptidase activity || >>> Process: organismal physiological process Process: regulation of body fluids Process: hemostasis Process: blood coagulation

	Process: physiological process Process: metabolism Process: macromolecule metabolism Process: protein metabolism Process: cellular protein metabolism Process: proteolysis || >>> Component: extracellular region
# Drug_Target_1_General_Function:	Involved in blood clotting cascade
# Drug_Target_1_Reaction:	Selective cleavage of Arg!Gly bonds in fibrinogen to form fibrin and release fibrinopeptides A and B INHIBITOR Benzamidine; D-Phe-Pro-Arg-CH2Cl; Nalpha-(2-naphthyl-sulfonyl-glycyl)-D-p-amidinopheyl-alanylpiperadin e; Argatroban
# Drug_Target_1_Specific_Function:	Thrombin, which cleaves bonds after Arg and Lys, converts fibrinogen to fibrin and activates factors V, VII, VIII, XIII, and, in complex with thrombomodulin, protein C
#END_DRUGCARD DB00001	

Table 1: Sample Dataset of Polio disease

In table 1, we have selected required information which is irrelevant in our case. We have only used drug name, drug chromosome, molecular weight, molecular average weight and molecular structure in order to draw its molecular structure. We have picked these parameters in order to establish a fitness function on the basics of which drug fitness will be evaluated.

In next section, we will see the outputs of our algorithm with the dataset in order to attain a drug against deadly disease i.e. polio [28]. DE evolves during each generation to have global best solution which means the best combination of drug that is most effective against the disease. This evolvement depends on parent and its fitness in the iteration.

4. Findings and Experimental Results

4.1 Case of Polio Disease

In the first run of an algorithm, random population has been initialized, and therefore random parents are generated. The algorithm generates random population with little jumps. In first generation, parents are evolved. For each parent, its fitness is calculated via fitness function. In the table 1, we see the best parent is 1. The fitness of parents below 0.5, the algorithm considered those parents that are weak. Thus in this iteration, the effectiveness is unsatisfied (un-effective drug yet). In second generation, parents are evolved. For each parent, its fitness is calculated via fitness function. In the table 1, we see the best parent is 2. The fitness of parents keeps on improving and moving towards global optimal solution; the algorithm ignores parents that are weak and does not include them in the next iteration. Thus in this iteration, the effectiveness is improving and is effective thus better combination of drug is required. In third generation, parents are evolved. For each parent, its fitness is calculated via fitness function. In the above table 1, we see the best parent is 6. The fitness of parents keeps on improving and moving toward global optimal solution; the algorithm ignores parents that are weak and does not include them in the next iteration. Thus in this iteration, the effectiveness kept the same as previous generation best fit. Though there is need to run more generations. In fourth generation, the fitness has improved tremendously and reached to 0.916368. In the above table 1, we see the best parent is 3. The fitness of parents keeps on improving and moving toward global optimal solution. The combination of drug is effective on polio disease.

The fitness has stopped improving in fifth generation and is now constant. In table 1, we see the best parent is 3. The combination of drug is effective on polio disease. It produces same results on more generations which means till generation 4, 5; DE got the best drug out.

4.2 Comparison of Drug Design using DE vs Drug Design using GA (Case of Polio)

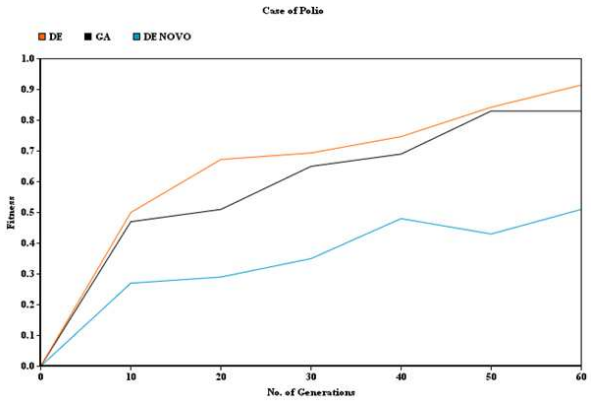

Figure 7: Comparison GA/ DE/ De Novo (Polio)

In graph of figure 7, we have shown the comparison of GA and De Novo with DE in case of drug designing, the red line denotes DE whereas the black line represents GA. We have compared it over 60+ generations. The DE convergence speed and robustness are same as GA in first few iterations but the fitness does not goes decreases with more iteration and search through the search space [29, 30]. It moves towards the global best solution in minimum number of iterations and remains same as number of generation's increased. De Novo has initial boost to fitness however it remains ineffective as the fitness only moves up to 0.2+. The GA also gives good fitness results. It lacks in robustness, convergence speed and consistency. Though DE outperforms GA and De Novo as seen in graph figure 10, as mentioned earlier, the results based on fitness criteria difference is too small. The fitness evaluation using DE is 0.91 and using GA the fitness is 0.83.
Prior to this, GA has overcome another technique, i.e. de Novo drug design. Since the size of the population is fixed, user defined. So it is difficult to find the drug out of the limited chemistry space. In GA, the smooth line shows the drastic increase in fitness and moving towards the ideal fitness i.e. 1. After certain number of generations, the graph is stable and later on a slow increase in its fitness occurs. When it traverses over more population, it gradually increases the fitness and up to 60+ iterations, we see the fitness of GA landed on approx. 0.8.

4.3 Comparison of Drug Design using DE, GA, De Novo Design (Case of Rabies)

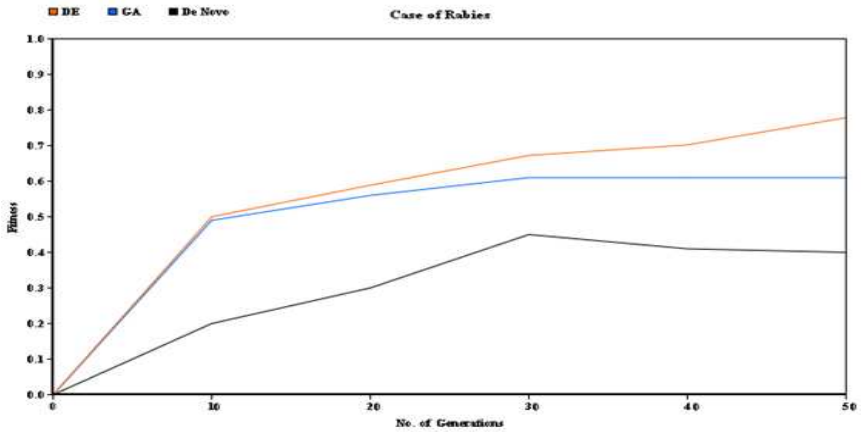

Figure8: Comparison GA vs DE (Rabies)

The initial randomization of DE, GA and De Novo is equivalent. The fitness in start is improving rapidly of all techniques in graph of figure 8. The DE and GA techniques have a check that the compound having fitness less than 0.5 is pruned out. In the first ten generations, DE and GA both succeeded in moving up to 0.5. In second generation DE has slightly improved up to 0.5+, GA has also improved up to 0.5+ but not as DE. De Novo design technique shows signs of less fit values because of its static non evolutionary technique nature. It has fixed length search capability. In third level (30) generations, DE fitness goes to 0.6, GA 0.58 and De Novo 0.37. Fourth one (40) shows better result of DE, drastic increase in fitness to 0.8. GA shows static behavior moving slight up to 0.5. De Novo indicates that it should be out of the game. After fifty generations, DE reaches to 0.8+, GA 0.6+ and De Novo 0.37. DE shows better results than other techniques.

4.4 Comparison of Drug Design using DE, GA, De Novo Design (Case of Typhoid)

The fitness of DE and GA techniques is improving rapidly early in figure 9; De Novo fitness is also adorable. The DE and GA techniques have a check that the compound having fitness less than 0.5 which is pruned out. In the first ten generations, DE and GA both are succeeded in moving up to fitness of 0.5. After twenty generations, DE has improved up to 0.5+ very minor improvement as compared to GA, GA has also improved up to 0.5+. De Novo gradually increases its fitness over generations but slow growth can be noted. After thirty generations, DE fitness goes to 0.6, GA 0.58 and De Novo 0.41. After forty, it shows better result of DE, drastic increase in fitness to 0.77. GA shows static behavior moving slight up to 0.5. De Novo indicates that it should be out of the game. After fifty generations, DE reaches to 0.8+, GA 0.6+ and De Novo 0.41. DE shows better results than other techniques.

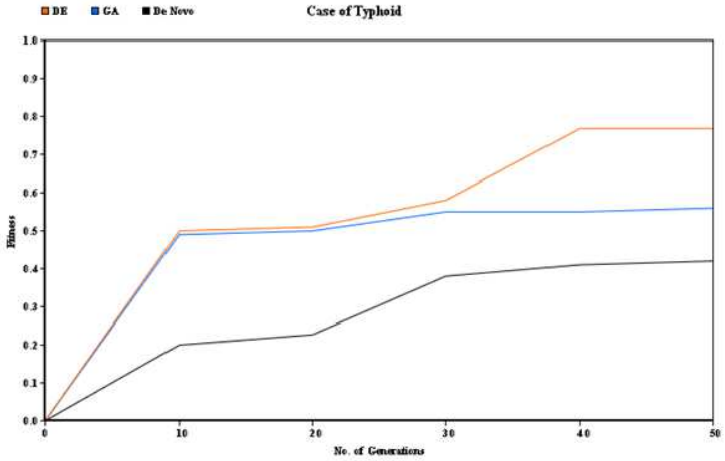

Figure 9: Comparison GA vs DE (Typhoid)

4.5 Performance Measure & Effectiveness of Drug Molecules

We are taking run time combination generations of the chromosomes. A technique can be used by pharmaceutical companies to get best combination of drug to cure a disease.

The giant companies have their datasets in millions, containing molecular structures in millions [15]. There is a high possibility to have high quality drug. In our project, we have defined certain criterion that calculates the quality of a drug. A similar criterion has been used by pharmaceuticals to evaluate drug effectiveness. We have selected properties of molecular weight (MW). We have used Chem Axon Ltd's Calculator Plugins [30] to calculate the molecular weights of the compounds and their properties. Here is an algorithm of the calculation of the performance measure and effectiveness of a molecule:

Following algorithm in terms of an Equation 5.3is used for measuring effectiveness.

$$f_{MW} = \begin{cases} \text{Un-effective} & \text{if } MW \leq 0.7 \\ \text{Less-effective} & \text{if } 0.7 \leq MW \leq 0.9 \\ \text{Effective} & \text{if } 0.9 \leq MW \leq 1 \end{cases} \quad \text{eq(5.3)}$$

5. Conclusion and Future Work

Nowadays the pharmaceuticals and bioinformatics made a big advancement in research and development. There is a race among the industry to find out new drug targets with the confidential availability of drug data. Evolutionary algorithms can be helpful in DD&D process. Various state-of-the-art techniques and designs have been previously used. We have offered a new approach DE towards DD&D which is efficient and promising in identification of drug candidate. We have run the algorithm on three different cases, Polio, Typhoid and Rabies using drug bank as bio-dataset and compared the results. In our comparison results, we observed that DE is producing better results as compared to other techniques. Our holistic approach is a pre-clinical technique which leads to practical testing in lab i.e. practicing it on animal or human under observation of medical specialists. In future, we will try to modify or customize DE to evolve and get more optimal values. As our research area is the evolutionary algorithm, we are more focused on the algorithm performance rather than operational processes. So we can plan to work with the algorithm to reduce its computation cost even though it is better than the previous evolutionary algorithm such as GA. Although, GA produces adequate results in solving operational based optimization problems such as production system produces the demanded product with higher quality and lower amounts of time and cost [31]. Furthermore, a survey paper on evolutionary techniques can help in solving drug design problem.

REFERENCES

1. Margarita Sordo, Sachin Vaidya and Lakhmi C. Jain, "Advanced Computational Intelligence Paradigms in Healthcare", *Springer-Verlag Berlin Heidelberg*, 2008, pp. 2–15.
2. Eelke Van Der Horst, "Drugs, Structures, Fragments", Substructure-based approaches to GPCR Drug Discovery and Design, pp. 11–21.
3. Jorgensen W.L., "The many roles of computation in drug discovery", Science 303, 2004, pp. 1813–1818.
4. Donald G. Truhlar et al, "Rational Drug Design, Springer-Verlag", Volume 108, 1999, pp. 93–101.
5. Klebe G, "Recent developments in structure-based drug design", Journal of Molecular Medicine, 2000 pp. 2-5.
6. Ecemis, M. Ihsan, et al. "A drug candidate design environment using evolutionary computation." Evolutionary Computation, IEEE Transactions on12.5 (2008): pp. 591-603.
7. Al-Globus, John Lawton, and Todd Wipke, **"**Automatic molecular design using evolutionary techniques"6th Foresight Conference on Molecular Nanotechnology. Conference of Nanotechnology: 10 (1999) pp. 290-299.
8. Shimada J., Ekins S., Elkin C., Shaknovich E.I., Wery J.-P.: Integrating computer-based De Novo drug design and multidimensional filtering for desirable drugs. Targets 1, 196, (2002).
9. Lu Chen, John K. Morrow, Hoang T. Tran, Sharangdhar S. Phatak, Lei Du-Cuny and Shuxing Zhang*****, "From Laptop to Benchtop to Bedside: Structure-based Drug Design on Protein Targets", *Current Pharmaceutical Design, 18,* pp. 1217-1239, 2012.
10. Koza, JR., "Genetic programming: a paradigm for genetically breeding populations of computer programs to solve problems", Technical Report STANCS-90- 1314, Department of Computer Science, Stanford University, June 1990.
11. David Hecht, Gary B. Fogel, "A Novel *In Silico* Approach to Drug Discovery via Computational Intelligence", *J. Chem. Inf. Model.* 2009,*49,*pp.1105–1121, American Chemical Society*,* 2009.
12. Gary B. Fogel, "Computational intelligence approaches for pattern discovery in biological systems", Briefings in Bioinformatics Advance Access published May 5, 2008.
13. Nikil Wale, "Machine Learning in Drug Discovery and Development", Drug Development Research 72, Publisher: Wiley-Liss, Inc, pp.33–65, 2011.
14. Hashim Ali, Muhammad Haris, FazlHadi, Ahmadullah, Salman and Yasir Shah, Solving Traveling Salesman Problem through Optimization Techniques Using Genetic Algorithm and Ant Colony Optimization, Journal of Applied Environmental & Biological Sciences, ISSN: 2090-4274, Vol. 6(4S), pp. 55-62, March 2016.
15. Konar A. Computational Intelligence; principles, techniques and applications. Berlin: Springer 2005.
16. Gary B. Fogel, "Computational intelligence approaches for pattern discovery in biological systems", Briefings in Bioinformatics Advance Access published May 5, 2008.
17. David Hecht, "Applications of Machine Learning and Computational Intelligence to Drug Discovery and Development", Drug Development Research 72: pp.33–65 (2011), Publisher: Wiley-Liss, Inc, 2010.
18. Carlos Andre´s, Pena Reyes, Moshe Sipper," Evolutionary computation in medicine: an overview", Artificial Intelligence in Medicine 19 (2000), pp. 1–23.

19. Fogel L.J., Owens A.J., and Walsh M.J., 'Artificial Intelligence through Simulated Evolution', Wiley, New York, 1966 (Book).
20. Back, Hoffmeister and Schwefel. "A Survey of Evolution Strategies", in Proceedings of the Fourth International Conference on Genetic Algorithms. Morgan Kaufmann Publ., San Mateo, California, 1992, pp. 2-9.
21. Price, V. Kenneth, and Storn, M. Rainer. (1997). "Differential evolution - A simple evolution strategy for fast optimization." Dr. Dobb's Journal, 22, pp. 18-24 and 78.
22. Rainer Storn, Kenneth Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces", Journal of Global Optimization 11: pp. 341–359, 1997.
23. Price, V. Kenneth, Storn, M. Rainer., and Lampinen, A. Jouni. (2005). Differential evolution: A practical approach to global optimization. Springer-Verlag Berlin, Heidelberg.
24. DE. "Differential Evolution Homepage." http://www.icsi.berkeley.edu/~storn/code.html.
25. Differential Evolution: a stochastic nonlinear optimization algorithm by Storn and Price, 1996. Journal of Global Optimization 11, 1997, pp. 341-359.
26. Peter Ertl, "Cheminformatics and its Role in the Modern Drug Discovery Process", Cheminformatics in Modern Drug Discovery Process, Novortis Institute for Biomedical Research, 2002.
27. Arunachalam, Vasan, "Optimization Using Differential Evolution" (2008). Water Resources Research Report. Book 22
28. R. Storn and K. Price, "Differential evolution—A simple and efficient adaptive scheme for global optimization over continuous spaces," Berkeley, CA, Tech. Rep. TR-95-012, 1995.
29. DrugBank, Open Data Drug & Drug Target Database, http://www.drugbank.ca.
30. ChemAxon, "Cheminformatics platforms & desktop applications", http://www.chemaxon.com.
31. MahsaRahimi, Optimal Workflow Scheduling In Production Line of a Dairy Factory Using Petri Net and Genetic Algorithm, Journal of Applied Environmental & Biological Sciences, ISSN: 2090-4274, Vol. 6(3), pp. 88-94, February 2016.