

A Review on Optimization Algorithm for Deep Learning Method in Bioinformatics Field

Siti Noorain Mohmad Yousoff*, 'Amirah Baharin, Afnizanfaizal Abdullah*

Synthetic Biology Research Group,

Faculty of Computing,

Universiti Teknologi Malaysia,

81310 UTM, Johor, Malaysia

*ainyousoff@gmail.com, ami.b2525@gmail.com, *afnizanfaizal@utm.my

Abstract—In the past few years, deep learning has been used widely in bioinformatics area to solve common problems such as protein sequence prediction, phylogenetic inferences, multiple sequence alignment and many more. It has been in the spotlight as a powerful approach which makes significant advances in taking care of the issues that haunt artificial intelligence community for many years. However, several weaknesses such as trap at local minima, lower performance and high computational time still occur in deep learning. Therefore, global optimization technique such as differential search algorithm can be used to assist deep learning method in order to get best finding result and data. This review will cover fundamental of deep learning and their involvement in bioinformatics field as well as implementation of differential search algorithm and their involvement in bioinformatics field.

Keywords—*bioinformatics, deep learning, neural network, backpropagation, optimization algorithm, differential search algorithm*

I. INTRODUCTION

Bioinformatics is a research area that is widely known to integrate numerous core subjects such as biology, mathematic, engineering and computer science. With this multidisciplinary field, bioinformatics usually involves in development of new methods and software tools used for better understanding of biological data as the main aim of bioinformatics is to help researchers to extract knowledge encoded in the biological data. As mentioned by Hapudeniya [1], essential issues that occur in bioinformatics such as protein sequence prediction, phylogenetic inferences, multiple sequence alignment, etc are usually hard in nature especially in non-deterministic polynomial-time aspect. Therefore, advanced computational technologies, tools and algorithms need to be used in order to solve these problems [1].

To solve growing problems in bioinformatics field, artificial intelligence methods such as deep learning method offer an efficient and powerful approach to be used. Deep learning is one of the artificial intelligence methods that always been in the spotlight due to their major advances in taking care the problems that haunt artificial intelligence community for a long time [2]. It has turned out that deep learning can be very good at foreseeing the action and activity of potential drug molecules [3], anticipating the impacts of mutations in the non-coding DNA towards gene expression and disease [4-5] and

reconstructing brain circuits [6]. This shows that deep learning can be taken into consideration as a method that can be used to solve multiple problems that rise in bioinformatics field.

Despite successfulness and advantages offered by deep learning, this method still suffers from several other weaknesses such as trap at local minima, lower performance and high computational time [7-9]. Due to these drawbacks, optimization algorithms can be implemented in order to assist deep learning to achieve best finding data and results. There are several optimization algorithms, but this paper will focus on optimization algorithm called Differential Search Algorithm (DSA). Several comparison studies [10-11] have been carried out which proved that DSA is more powerful optimization algorithms compared to Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) and Gravitational Search Algorithm (GSA).

II. DEEP LEARNING

Representation learning is an arrangement of strategies that permits a machine to identify raw data input and allow automatic discovery of representations which is required for classification or detection [2]. Deep learning method is one of the examples of representation learning methods with multiple levels of representation. Deep learning is a branch of machine learning which depends on a set of algorithms that endeavor to model high-level abstraction in data by using model architecture and composed of multiple non-linear modules [12-15]. Deep learning is working by exploiting the idea of the hierarchical concept where the higher level, more abstract concepts are learned from the lower level ones. This architecture often developed with a greedy layer-by-layer strategy and deep learning serves to unravels these abstractions and choose which elements are helpful for training purpose [16].

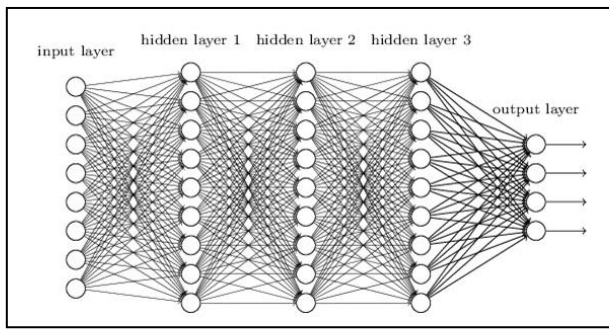


Fig. 1 Deep learning with more than two hidden layer [17].

Fig. 1 above visualizes deep learning architecture with more than two hidden layers. At each layer, the input is changed by a processing unit whose parameters are now learned through the training phase [15]. The main key aspect of deep learning is that it only requires little engineering by hand since deep learning is a unique method in which the layers of features are usually learned from data using general-purpose learning procedure.

The next sub sections will describe briefly fundamental architecture of deep learning for better understanding about deep learning which consists of two main things that make up deep learning. They are neural network and also backpropagation.

A. Neural Network

Neural networks have been around and exist since the 1950s and already contributed in so many things and findings. The most successful deep learning methods usually will involve neural networks. Schmidhuber [15], in his article explains that, neural network basically consists of many simple, connected processors called neurons and each of these neurons produces a sequence of real-valued activation. These neurons are activated through weighted connections from previously active neurons as well as through sensors perceiving the environment. Schmidhuber [15] also stated that neurons' behavior may require long casual chains of computational stages where every stage typically transforms in a non-linear manner and aggregate activation of the network.

Basically, deep learning used neural network concept to operate. Both deep learning and neural network look similar as deep learning applied neural network concept which used neurons and also hidden layers, but actually there are slightly different between these two methods. Firstly, deep learning has more hidden layers if compared to neural network which only have one or two hidden layer and secondly, deep learning can be trained in both unsupervised and supervised manner for both unsupervised and supervised learning task [18-19].

B. Backpropagation

Backpropagation procedure occurs after feedforward procedure happened. By the end of feedforward procedure, it will give actual value which is then be compared with the expected value. This comparison will result in an error value. This will lead to the backpropagation procedure where the connection weights in the network are adjusted and working

backwards from the output layer, to the hidden layer until the correct output or the output that are closely to the expected output is produce.

However, while using backpropagation, there are several problems that might occur such as differentiability to compute the gradient [20]. Therefore, backpropagation cannot handle discontinuous node transfer functions and it cannot handle optimality criteria. The other problem with backpropagation is the scaling problem [20]. Backpropagation works efficiently with simple training problems, however as the problems complexity increases, the performance might fall off rapidly. Backpropagation has an advantage where it can escape local minima. Unfortunately, it leaves them unknowingly whether the next one it finds is better or worse which then end up make the backpropagation bouncing between local minima without much improvement thus making for a very slow training rate.

C. Deep Learning in Bioinformatics

As mentioned before, deep learning is a powerful approach that can be used for learning complex patterns at multiple layers. Due to this characteristic, it also can be used to capture multiple levels of data abstraction and processing within the cells. This feature makes deep learning suitable for genomics studies as mentioned by Park and Kellis [21] and also can be used in pathway analysis purpose.

Alipanahi *et al.* [22], used deep learning strategy in their genomics research to calculate protein-nucleic acid interactions from diverse experimental dataset. They found out that deep learning method performed better than other state-of-the-art methods. Furthermore, Alipanahi *et al.* incorporated deep learning with their own algorithm called DeepBind. This method allowed the prediction of binding affinity of a protein to a DNA or RNA sequence in two steps. These two steps consist of convolution module for representation learning and prediction module for feature combinations. This method is better illustrated in Fig. 2.

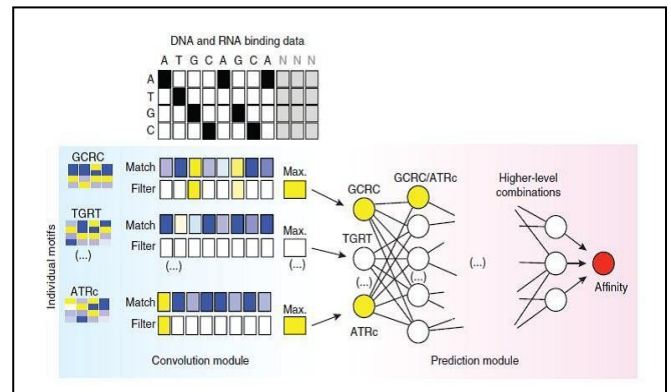


Fig. 2 Deep convolutional neural network designed by Alipanahi *et al.* [22].

III. OPTIMIZATION

Optimization algorithms have been widely used in many research field areas such as bioinformatics, biotechnology and computational biology and have been used in many applications to solve variety of problems. Optimization as

stated by Banga [23], is an algorithm with an aim to make a design or system as efficient and functional as possible. Optimization algorithms is not a new method or concept as it is widely used to help researchers to analyze, enhance, encoded and optimize many of biological data.

Therefore, optimization algorithms can be taken into consideration to be used to support and assist deep learning method in order to achieve better finding result. By using several advantages that content in optimization algorithm such as provide great flexibility of algorithm and reduce the computational burden [24], it can be greatest help in training large layer of network and cover up drawbacks that occur in deep learning method. In this section, optimization algorithm called Differential Search Algorithm (DSA) will be discussed briefly.

A. Differential Search Algorithm (DSA)

Differential search algorithm (DSA) is a relatively new novel meta-heuristic algorithm that has been proposed by Civicioglu [25]. It is inspired by Brownian-like random walk movement which is used by an organism to migrate [26]. Basically how DSA work is similar like behavior of organisms to move away from habitat that has low food capacity to a new habitat with a more food capacity. They will stay in that new habitat for a time being until it finds another new habitat with more food capacity than in their current habitat. In DSA algorithm, search space is pictures as the food area or habitat while each point in the search space is pictures as organism migration. The main goal of DSA as stated in [26] is that to find the global optimal solution of the problem. Standard DSA pseudo code is as describe as Fig. 3 below.

```

(1) begin
(2) Set the generation counter  $G = 0$ ; and randomly initialize a population of
     $NP \times D$  individuals  $X_i$ . Initialize the parameter  $p1, p2$ 
(3) Evaluate the fitness for each individual in  $P$ .
(4) while stopping criteria is not satisfied do
(5)   scale = randg(2 * rand) * (rand - rand)
(6)   for  $i = 1$  to  $NP$  do
(7)     select randomly  $a \neq i$ 
(8)      $s_i = x_i + scale \times (x_a - x_i)$ 
(9)   end
(10)   $r = rand(NP, D)$ ;
(11)  If  $rand < p1$  then
(12)    If  $rand < p2$  then
(13)      for  $i = 1$  to  $NP$  do
(14)         $r(i,:) = r(i,:) < rand$ 
(15)      end
(16)    else
(17)      for  $i = 1$  to  $NP$  do
(18)         $r(i, randj(D)) = 0$ 
(19)      end
(20)    end
(21)  else
(22)    for  $i = 1$  to  $NP$  do
(23)       $d = randj(D, 1, [p2 \cdot rand])$ 
(24)      for  $j = 1$  to size( $d, 2$ ) do
(25)         $r(i, d(j)) = 0$ 
(26)      end
(27)    end
(28)  end
(29)   $r = r > 0$ ;
(30)   $s(r) = X(r)$ ;
(31)  for  $i = 1$  to  $NP$  do
(32)    Evaluate the offspring  $s_i$ 
(33)    If  $s_i$  is better than  $X_i$  then
(34)       $X_i = s_i$ 
(35)    end if
(36)  end for
(37)  Memorize the best solution achieved so far
(38) end while
(39) end

```

Fig. 3 DSA pseudo code [26].

There are four main steps that involved in DSA. First is the initialization step. Minimum and maximum bounds will be prescribed to initiate organism which utilizes $NP \times D$ -dimension parameter vector. After the initialization, Brownian-like random walk model can be describe by generate the stopover vectors between organisms. Third steps will be the search process of stopover site which can be calculated by an organism of the superorganism and lastly the selection step. Purpose of selection step is to choose the next population between stopover site and the organism population.

B. DSA for Proteomic Analysis Pipeline

As mentioned before, DSA has been widely used in bioinformatics and many other research fields. One of the applications of DSA can be seen in Xie *et al.* work [27]. They stated that electron transfer dissociation (ETD) is exceptionally helpful for peptide fragmentation in mass spectrometry. Unfortunately, ETD spectra usually receive low score in the identifications of 2+ ions. Therefore, in order to solve this problem, Xie *et al.* have been proposed a new method by combining both ion charges enhancing method with DSA algorithm. By using this new method, they observed that the complementary identification result shows great improvement in ETD identification.

C. Comparative study of optimization algorithms

Optimization plays an important role in solving variety of problems that occur in many applications and research fields. It was already gained wide attention in many research field areas such as biotechnology, bioinformatics, computational biology and many more. For the past few years, there were many other optimization algorithms that have been used by researchers in order to solve their research problems.

For this section, comparative study has been made between several other optimization algorithms such as ABC, GA, FA and also DSA. This comparison can be seen as shown in Table 1 below:

TABLE 1. COMPARISON TABLE FOR OPTIMIZATION ALGORITHMS

Algorithm(s)	Reference(s)	Advantage(s)	Disadvantage(s)
Ant Bee Colony (ABC)	[28]	• Good at global exploration	• Poor in the exploitation process
Genetic Algorithm (GA)	[28-29]	• Has ability for global search • Strong robustness	• Lower convergence rate • Easily trap at local optima
Firefly Algorithm (FA)	[30]	• Suitable to be used for high dimensional and non-linear problem	• Difficult to reach optimal solution within reasonable time

Algorithm(s)	Reference(s)	Advantage(s)	Disadvantage(s)
Differential Search Algorithm (DSA)	[10-11],[25]	<ul style="list-style-type: none"> • DSA has unique crossover and mutation operators. • More powerful compared to PSO, ABC, DE and GSA 	<ul style="list-style-type: none"> • Lack of strategy that may affect its local search ability

IV. IMPLEMENTATION OF DSA IN DEEP LEARNING METHOD

As mention in [31], there are three main reasons behind the fame and successfulness of deep learning method. First is that deep learning can drastically increase chip processing abilities, second is that it can significantly lowered cost of computing hardware and lastly is that deep learning is recent advanced computational technology in machine learning and information processing research. Not only that, advantages such as it can generate nature of model and it has an unsupervised pre-training step [31] are the key of successfulness of deep learning.

However, as deep learning offers powerful and efficient approach to be used to solve variety of problems, several limitations still occurs in deep learning. Main problem that always occur in deep learning is overfitting problem and it always stuck at local minima. Once these problems happen, it will result in lower performance and high computational time in deep learning.

Therefore, optimization algorithm such as differential search algorithm (DSA) can be taken into consideration to help cover up limitations that occur in deep learning method. There are several advantages that DSA offer such as good at finding optimal solution and also good at exploring search space and last but not least, the main highlight of DSA is that it is good at locating region of global minima. This advantage is very useful in help deep learning to solve their problem which is always trapped at local minima. As mentioned in [32], global optimization technique is required in order to avoid local minima problem. Once the problem with trap at local minima has been solved, it might increase back the performance of deep learning method and lower the computational time.

V. IMPLEMENTATION OF PROPOSED METHOD IN BIOINFORMATICS FIELD

Recently, bio-based chemical products such as xylitol have been in the spotlight due to their major contribution and advantages in both pharmaceutical and also food industries. Xylitol has been widely used as a sugar substitute in pharmaceutical industries to help diabetic patients as it will not increase the blood sugar level and insulin respond of the diabetic patients. While in food industries, xylitol has been widely used in sugar free chewing gum and it is believed that by eating sugar free chewing gum that content xylitol can prevent tooth decay problems. Due to these several advantages

that occur in xylitol, researchers start to show an interest on how to produce more xylitol within a short time.

In accordance to this, powerful approaches are needed in order to manipulate and simulate microbial hosts such as *E. Coli* to produce more xylitol. The proposed method in this paper which is the combination of DSA into DL can be used in this case. It can be used as a medium to predict which particular genes or pathways in *E. Coli* that may affect the production of xylitol. By doing so, analysis and simulation of *E. Coli* model can be done within short time period as computational method providing fast result if compared to wet experiment in the laboratory setting.

VI. CONCLUSION

High demand from bioinformatics field regarding the needed of advanced computational technologies and tools to help researchers to manipulate, analyze and extract knowledge encoded in biological data has motivated many others computational biologist and computer scientist to develop many software and tools that suitable with the current situation. Nowadays, deep learning has become one of the artificial intelligence methods that offer many advantages and has already making major advances by solving many problems that haunt artificial intelligence for many years. In this paper, review on deep learning, fundamental of deep learning and their involvement in bioinformatics field have been reviewed. However, despite the fame and advantages offered by deep learning, they still suffer from several limitations such as stuck at local minima, lower performance and high computational time [7-9].

Optimization algorithm needed in this situation to help and assist deep learning to get better finding result. Global optimization technique is required in order to avoid local minima problem [27]. Therefore, this paper has been focused on global optimization technique called DSA. Briefly review on DSA and their involvement in bioinformatics field have been made in this paper. With several advantages such as good at locating region of global minima, good at finding the optimal solution and also good at exploring search space, DSA can be taken into consideration as an algorithm that can be implemented in deep learning to cover limitations that occur in deep learning. With the implementation of DSA algorithm in deep learning method, it is hope that better finding results can be achieved.

For future work, it is hope that many researchers, computational biologist or computer scientist to proposed more hybrid methods between deep learning and optimization algorithm in order to get better finding results and data. As mentioned by Mohamad *et al.* [28,31], hybrid methods are highly recommended compared to filter methods to produce better results.

ACKNOWLEDGMENT

We would like to express our appreciation to Malaysia Ministry of Higher Education for supporting this project under Fundamental Research Grant Scheme (Project Vot No. 11H26). We also would like to thank to Research Management Center, Universiti Teknologi Malaysia for managing this project.

REFERENCES

- [1] Hapudeniya, M. (2010). Artificial neural networks in bioinformatics. Sri Lanka Journal of Bio-Medical Informatics, 1(2).
- [2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [3] Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 55, 263–274.
- [4] Leung, M. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. (2014). Deep learning of the tissue-regulated splicing code. *Bioinformatics*, 30, i121–i129.
- [5] Xiong, H. Y. et al. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218).
- [6] Helmstaedter, M. et al. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500, 168–174.
- [7] LeCun, Y., Huang, F. J., & Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on (Vol. 2, pp. II-97)*. IEEE.
- [8] Tirumala, S. S. (2014). Implementation of Evolutionary Algorithms for Deep Architectures. In *AIC*, 164-171.
- [9] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929-1958.
- [10] P. Civicioglu (2012). Transforming geocentric Cartesian coordinates to geodetic coordinates by using differential search algorithm, *Comput. Geosci.* 46, 229–247.
- [11] J. Liu, K. Teo, X. Wang, C. Wu, (2015). An exact penalty function-based differential search algorithm for constrained global optimization, *Soft Comput.* 1–9.
- [12] Deng, L.; Yu, D. (2014). Deep Learning: Methods and Applications. *Foundations and Trends in Signal Processing* 7, 3–4.
- [13] Bengio, Yoshua (2009). Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning* 2(1), 1–127.
- [14] Bengio, Y.; Courville, A.; Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8), 1798–1828.
- [15] Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. *Neural Networks* 61, 85–117.
- [16] Bengio, Y.; Courville, A.; Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8), 1798–1828.
- [17] Exploring Deep Learning & CNNs - RSIP Vision. (2015). Retrieved June 20, 2016, from <http://www.rsipvision.com/exploring-deep-learning/>
- [18] Erhan, D., Courville, A., Bengio, Y., & Vincent, P. (2010). Why Does Unsupervised Pre-training Help Deep Learning? 9, 201-208.
- [19] Hinton, G., & Salakhutdinov, R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313, 504-507.
- [20] Montana, D. J., & Davis, L. (1989). Training Feedforward Neural Networks Using Genetic Algorithms. In *IJCAI* 89, 762-767.
- [21] Park, Y., & Kellis, M. (2015). Deep learning for regulatory genomics. *Nat Biotechnol Nature Biotechnology*, 33(8), 825-826.
- [22] Alipanahi, B., Delong, A., Weirauch, M., & Frey, B. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol Nature Biotechnology*, 33(8), 831-838.
- [23] Banga, J. R. (2008). Optimization in computational systems biology. *BMC systems biology*, 2(1), 1.
- [24] Lamos-Sweeney, J. (2012). Deep learning using genetic algorithms. 3066-3072.
- [25] Civicioglu P. (2012). Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm, *Comput. Geosci.* 46, 229-247.
- [26] Liu, B. (2014). Composite Differential Search Algorithm. *Journal of Applied Mathematics*, 2014.
- [27] Xie, L. Q., Shen, C. P., Liu, M. B., Chen, Z. D., Du, R. Y., Yan, G. Q., ... & Yang, P. Y. (2012). Improved proteomic analysis pipeline for LC-ETD-MS/MS using charge enhancing methods. *Molecular BioSystems*, 8(10), 2692-2698.
- [28] Li, B., Li, Y., & Gong, L. (2014). Protein secondary structure optimization using an improved artificial bee colony algorithm based on AB off-lattice model. *Engineering Applications of Artificial Intelligence*, 27, 70-79.
- [29] Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics*, 22(24), 3096-3098.
- [30] Ali N., Othman A.Z., Husain M.N & Misran H.M. (2014). A REVIEW OF FIREFLY ALGORITHM. *Engineering and Applied Sciences*, 9(10), 1-5.
- [31] Deng, L. (2012). Three classes of deep learning architectures and their applications: a tutorial survey. *APSIPA transactions on signal and information processing*.
- [32] Dai, C., Chen, W., & Zhu, Y. (2010). Seeker optimization algorithm for digital IIR filter design. *IEEE transactions on industrial electronics*, 57(5), 1710-1718.
- [33] Mohamad, M. S., Omatu, S., Deris, S., Yoshioka, M., Abdullah, A., & Ibrahim, Z. (2013). An enhancement of binary particle swarm optimization for gene selection in classifying cancer classes. *Algorithms for Molecular Biology*, 8(1), 1.