

Intelligent Framework With Controlled Behavior for Gene Regulatory Network Reconstruction

Bhavana Bansal, Jaypee Institute of Information Technology, Noida, India*

Aparajita Nanda, Jaypee Institute of Information Technology, Noida, India

Anita Sahoo, Jaypee Institute of Information Technology, Noida, India

ABSTRACT

Gene regulatory networks (GRNs) are the pioneering methodology for finding new gene interactions getting insights of the biological processes using time series gene expression data. It remains a challenge to study the temporal nature of gene expression data that mimic complex non-linear dynamics of the network. In this paper, an intelligent framework of recurrent neural network (RNN) and swarm intelligence (SI)-based particle swarm optimization (PSO) with controlled behaviour has been proposed for the reconstruction of GRN from time-series gene expression data. A novel PSO algorithm enhanced by human cognition influenced by the ideology of Bhagavad Gita is employed for improved learning of RNN. RNN guided by the proposed algorithm simulates the nonlinear and dynamic gene interactions to a greater extent. The proposed method shows superior performance over traditional SI algorithms in searching biologically plausible candidate networks. The strength of the method is verified by analyzing the small artificial network and real data of *Escherichia coli* with improved accuracy.

KEYWORDS

Gene Regulatory Networks, Human Cognition-Based PSO, Recurrent Neural Networks, Time-Series Gene Expression Data

1. INTRODUCTION

Recent years have seen the advent of DNA microarray technology and information retrieval has been proved essential for the reconstruction of gene regulatory networks (GRNs) from temporal gene expression data. GRNs proves important for understanding many unknown biological functionalities and processes. It gives insights of the activities of genes and provide knowledge about transcriptional regulations among them (Aalto et al., 2020). GRNs is a virtual network of genes and their mutual influences, where node of the network is a gene and edges are the influence from the regulator to the target gene which either activates or suppress target gene's ability of protein formation (Morgan et al., 2019). GRNs have been successfully applied in diagnostics and contributes in identification

DOI: 10.4018/IJIRR.2022010104

*Corresponding Author

This article published as an Open Access Article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

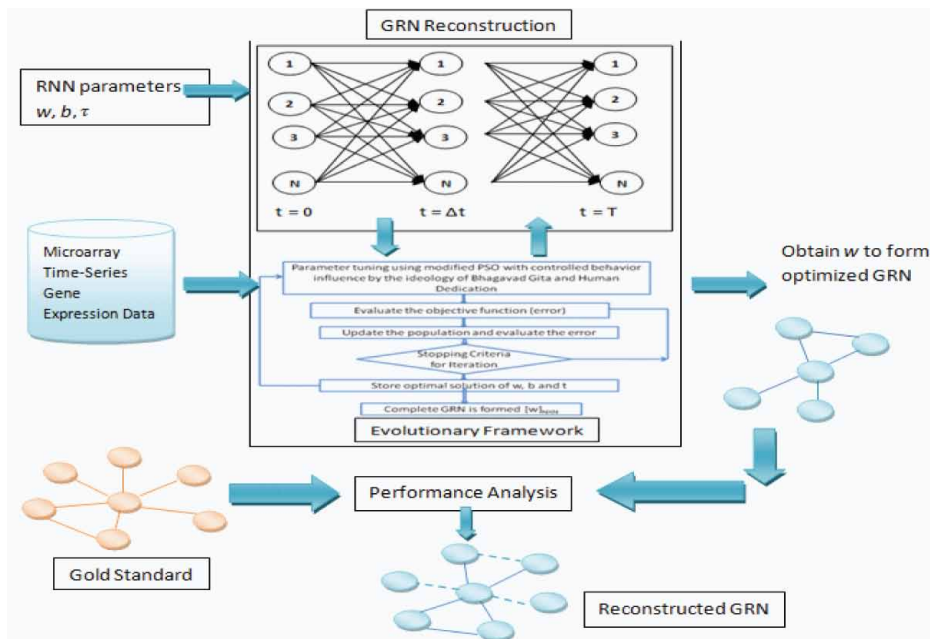
of essential genes (Xie et al., 2020). The well known issue encountered in the analysis of temporal data for GRN reconstruction problem is the curse of dimensionality (Altman & Krzywinski, 2018).

In context of the computational models used for GRNs reconstruction from time series data, researchers has adopted several methods (Delgado & Gómez-Vela, 2019; Razaghi-Moghadam & Nikoloski, 2020) such as Boolean networks, Bayesian networks (BNs), dynamic Bayesian networks (DBNs) and linear additive genetic model. Boolean networks model (Barman & Kwon, 2018) considers only two states for each gene: active and inactive. This model does not take into consideration the intermediary effects on the genes which cause information loss. Bayesian networks (BNs) model (Sanchez-Castillo et al., 2018) are graph based models forming a genetic network as a directed acyclic graphs. This model effectively handles noise, missing values and the random nature of gene expression data, however it does not take into account the dynamical nature of GRNs and the temporal aspect of the data. The limitations of BNs were overcome by dynamic Bayesian networks (DBNs) (Adabor & Acquah-Mensah, 2019). The linear additive genetic model (Luque-Baena et al., 2014) may identify linear regulatory relationships but does not consider the non-linear behaviour of GRNs.

Motivation: Considering the limitations of these models researchers adopted recurrent neural network (RNN) for the problem of GRNs reconstruction. RNN model clearly manifested the temporal nature of gene expression data and non-linear dynamics among gene regulations which is essential for GRN reconstruction. This model has an ability to consider the feedforward and feedback loops of the genetic regulation network (Biswas & Acharyya, 2016, 2018). Time-series data is the input to the RNN model. The data contains expression levels ($x_i(t)$) of genes at consecutive time points. The gene expression level $x_i(t+1)$ of a gene (i) of the current time point ($t+1$) of an RNN.

Layer is simulated from the expression levels $x_j(t)$ of genes (j) (regulator genes) at previous time point t accompanied by the set of genetic network parameters. In terms of GRNs topology very few regulatory genes j influence a target gene i which concludes that genetic network connectivity is

Figure 1. The overview of the work flow



sparse. In sense of the modelling of GRNs with RNN model, the set of parameters which requires training are weights w_{ij} , bias term b_i and the time constant τ_i associated with each gene.

Training of RNN model parameter is a difficult task. Most widely used machine learning algorithm which effectively learns RNN model parameters was back propagation through time (BPTT) algorithm (Gudise & Venayagamoorthy, 2003; Lillicrap & Santoro, 2019). BPTT algorithm encounters difficulties in handling local minima. Further, existing literature has also reported various swarm intelligence (SI) algorithms for RNN parameter training for the problem of GRN reconstruction. Some of such algorithms are genetic algorithm (GA) (Kordmahalleh et al., 2017), particle swarm optimization (PSO) (Jana et al., 2019) and ant colony optimization (ACO) (Kentzoglanakis & Poole, 2012). PSO algorithm has been proved ground breaking for unrolling RNN when multiple local optima exists (R. Xu et al., 2007), but suffers from the problem of slow global optimization and premature convergence. There is always a need to improve the learning of RNN parameters to provide improved credibility to obtain biologically plausible candidate GRN structure hence the motivation of the paper is to improve the learning mechanism of RNN model parameters. Various PSO variants have been proposed which certainly improved the learning strategies leading to better search efficiency (Lynn & Suganthan, 2015; G. Xu et al., 2019). Some human cognition based variants of PSO with controlled behavior was reported in the literature such as self-regulating PSO (SRPSO) (Tanweer et al., 2015), human bagavad Gita PSO (HBGPSO) and human dedication PSO (HDPSO) (Gajawada et al., 2019). These algorithms significantly improves the learning strategy of PSO leading to improved exploration and exploitation ability. Thus, accommodating the intelligent framework using human cognitive PSO for the training of RNN model parameters associated with GRN may provide an intelligent GRN reconstruction framework.

Contribution: This paper contributes by proposing an intelligent framework of modified PSO with controlled behavior for training the RNN model parameters for the problem of GRN reconstruction using time series gene expression data. The overview of the work flow is explained in fig. 1. The proposed algorithm is employed for improving learning strategies by incorporating human cognition of self-learning from his own performance. The learning strategy is further enhanced by the ideology of Bhagavad Gita and human dedication. The proposed method adopts performance analyser mechanism which intakes the performance of all the particles and each particle is provided with different learning strategy based on its performance. This practice effectively guides the searching of the particle thus achieving faster convergence. The quality of the biologically plausible network architectures are evaluated by predicting minimum error and as a result the proposed intelligent framework of GRN reconstruction involving improved human cognition based learning strategies for training RNN model parameters produced better accuracy of reconstructed GRN with respect to the gold standard GRN when compared with other state-of-art human cognition based SI algorithms in terms of performance metrics for both small artificial network and E.Coli real world dataset.

The rest of the paper has been organized as follows. Section 2 gives brief overview of literature survey. Section 3 describes of the problem of GRN using RNN model. Section 4 gives the explanation of the proposed framework. Section 5 explains the experimental outcomes and finally section 6 gives the conclusion highlighting the future scopes for the research.

2. LITERATURE SURVEY

GRN has become a pioneering methodology for the modeling of typical biological activities. In the recent years various computational models have been developed for the reconstruction of GRNs from gene expression data based on mathematical formulations (Barbuti et al., 2020). These models are

Boolean network, Bayesian network (BNs), dynamic Bayesian network (DBNs) and linear additive genetic model. Boolean network model has a binary nature where each gene state can be regarded as either turned on or off (Manica et al., 2020). These models do not consider the intermediate effects on the gene state which certainly cause the loss of information in data discretization. The Boolean network has a deterministic approach thus cannot deal with missing expression data and performs task in huge computational time. Researchers also proposed Bayesian network model (BNs) (Grzegorzczuk et al., 2019; Li et al., 2016). It is represented in the form of acyclic graphs where vertices of the graph represent genes and edges represent the conditional dependency relationship in between the genes. BNs can adequately eradicate noise and stochastic form of gene data but however failed to use the temporal gene expression data and does not consider dynamical nature of gene regulatory networks. This model also requires huge computational time and fails to determine self-regulation of the gene which is a needful requirement GRN modeling. DBNs overcome the drawbacks of BNs (Yu et al., 2017). This model takes into account the temporal nature of gene expression data and effectively incorporates the non-linear dynamics of gene regulations. The linear additive genetic model performs the weighted sum of the expression level of all genes at previous time point to calculate the expression level of a gene at the current time point, but this model does not consider the non-linear behaviour of GRNs (Luque-Baena et al., 2014).

Recurrent neural network is however the successful method for the modeling of GRNs, since it employs the non-linear and dynamic interactions of gene regulations (Liu & Liu, 2020; Raza & Alam, 2016). RNN permits continuous variables and produces the expression level of a gene (i) of RNN at $t + 1$ time point by the influence of the regulatory gene (j) at the previous time points t using w_{ij} weight parameter. The gene (i) then produces the expression level $x'_i(t + 1)$ as predicted output. To minimize the error between original and predicted expression level of gene (i) there is a need to train the parameters of RNN efficiently.

Researchers have adopted BPTT algorithm and SI algorithms to train RNN parameters (Khan et al., 2020). BPTT algorithm unfolds each layer of the network through time finding the derivatives with respect to the weights and rolling back these derivatives using gradient descent with the objective to minimize the error. However, it faces difficulty in dealing with local optima. SI methods proves efficient for training RNN model parameters for GRN reconstruction some such algorithms are genetic algorithms (GA), particle swarm optimization (PSO) and ant colony optimization (ACO). Adopting improved learning mechanism for training RNN parameters has improves the quality of functional and structural parameters of GRN obtained. Khan et al., (2016) proposed hybrid SI framework of bat algorithm (BA) inspired PSO for GRN reconstruction problem and obtained better results when compared to contemporary literature because of improved learning strategies adopted for RNN model training. ACO-PSO framework was used by Kentzoglanakis & Poole, (2012) wherein ACO is employed for GRN structure optimization and PSO is used for RNN model parameter optimization for the corresponding structure obtained. A hybrid differential evolution (DE) and PSO was proposed R. Xu et al., (2007) for RNN model parameter learning proves to perform better for GRN reconstruction problem. Some human cognition based PSO variants exists in the contemporary literature such as SRPSO, HBGPSO and HDPSO (Gajawada et al., 2019; Tanweer et al., 2015). These algorithms are inspired by human thought process in contributing problem solving skills and demonstrated improved convergence characteristics. Hence, in this paper, we proposed a human cognition based variants of PSO with controlled behavior which possess human like behavior to improve the performance of RNN parameter learning.

3. PROBLEM

Mathematically, GRN is represented as a graph $G = [V, E]$ where V vertices represent genes and E edges represents the regulatory interactions among them. The problem of GRN reconstruction aims

to obtain the genetic network parameters. Basically, there are three parameters associated with GRN problem, these are connection weight $W = [w_{ij}]_{N \times N}$ (represented in the form of matrix), bias expression parameter $B = [b_i]_{1 \times N}$ and time constant $T = [\tau_i]_{1 \times N}$, where N is the total number of genes. The GRN is computationally denoted as an adjacency matrix $G = [g_{ij}]_{N \times N}$, and on the basis of presence and absence of an edge the values of g_{ij} is 0 or 1 (Kentzoglanakis & Poole, 2012). GRN reconstruction is a two-fold process which involves the structure optimization finding appropriate biologically plausible network architecture that explains the connectivity between the genes and second is parameter estimation for the given model structure. Gene expression dataset is fed as an input to GRN reconstruction problem. Data is represented in the form of matrix $Data = [x_i(t)]_{N \times T}$, where $x_i(t)$ is the gene expression of $gene(i)$ at time (t), N is the total number of genes and T is total time points.

A neural network based RNN formalism is an efficient method for GRN reconstruction as it allows feed forward and feedback loops which is necessary of GRN reconstruction. RNN model captures the non-linear dynamics of gene interactions and reproduces the nature of temporal gene expression data Vohradsky, (2001). The gene expression level x_i of target gene (i) in time point $(t + \Delta t)$ represents the regulatory effect and is given by (1):

$$x_i(t + \Delta t) = \frac{\Delta t}{\tau_i \left(1 + \exp \left(- \left(\sum_{j=1}^N w_{ij} x_j(t) + b_i \right) \right) \right)} - \left(1 - \frac{\Delta t}{\tau_i} \right) x_i(t) \quad (1)$$

where x_j is the gene expression level of regulators, w_{ij} , b_i and τ_i are the parameters of GRN associated with RNN model which requires training. The expression level of gene (i) obtained in the range [0, 1].

The parameters of the model are predicted so as to mimic the nature of the time series gene expression data. In the sense, the model-driven GRN reconstruction problem becomes an optimization problem where model parameters are trained with the objective to minimize the mean square error (MSE) in between the original and predicted gene expression value of each time point of the time-series gene expression data which is given by (2):

$$MSE = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(x_i(t) - \hat{x}_i(t) \right)^2 \quad (2)$$

where $x_i(t)$ is the original gene expression level and $\hat{x}_i(t)$ is the predicted gene expression level of i_{th} gene at t time point.

In context of structure optimization, GRN connectivity is usually sparse i.e. only few regulators influence a target gene. For the 8-gene network we have considered maximum 4 genes influences per target gene. This property limits the search space from 2^N to N_4C , where N is the total number of genes in the GRN (Khan et al., 2016). The intense searching is done on reduced search space using the combination of all possible regulators this obtains more likable candidate architectures. Regulators restriction process also improves model quality. For parameter estimation, the global problem has $N \times (N + 2)$ parameters which becomes very large with large values of N . Thus researchers suggested

the problem decomposition strategy where parameters associated with each target gene i.e. $(N + 2)$ is estimated this reduces computational overhead (Noman et al., 2013). In this work, the proposed novel SI algorithm is given the objective function of parameter estimation for each target gene (i) to minimize the estimated error ($error_i$) is given by (3):

$$error_i = \frac{1}{T} \sum_{t=1}^T (x_i(t) - \hat{x}_i(t)) \quad (3)$$

where $x_i(t)$ and $\hat{x}_i(t)$ is the observed and predicted gene expression level at time point (t) . Subsequently, the $(error_i)$ is used to obtain the total predicted mean square error ($Total_MSE$) for all N genes given by (4):

$$Total_MSE = \frac{1}{N} \sum_{i=1}^N error_i \quad (4)$$

4. PROPOSED FRAMEWORK

In this paper, an intelligent framework of modified PSO with controlled behaviour for training the RNN model parameters for the problem of GRN reconstruction using time series gene expression data has been proposed. The problem takes time-series gene expression data as input. The data is then passed to the proposed intelligent GRN reconstruction framework. In specific to GRN modelled with RNN, w, b and τ parameters are tuned with the objective to minimize the error between the actual and predicted gene expression time-series. The obtained model is analysed against the gold standard network so that true biological understanding can be fetched from it. The overview of the work process is described in fig. 1. In the proposed approach, the performance of RNN model is improved by training its parameters with modified PSO algorithm with controlled behaviour. This algorithm is inspired by the human cognition enhanced by the ideology of Bhagavad Gita and human dedication. The algorithm adopts a performance analyser and adopts an intelligent move by dividing the population of the particles into subpopulations. The performance analyser on the basis of the performance of the particle divides the population into the best particle, ideal/dedicated particle, non-ideal particle when got success and non-ideal particle when got failure and adaptively adjust their learning strategies.

The learning strategy for the best particle is incorporated with the feature of human cognition i.e. when a person knows his current performance is best he will adopt best learning strategy for the desired outcome, this is done by accelerating its weight inertia without taking consent of his own previous personal experience and social information which induces higher exploration. Best particle's velocity vel_i and position pos_i is updated using (5), (6) and (7) respectively:

$$w_I = w_I + dw \quad (5)$$

$$vel_i(t+1) = w_I * vel_i(t) \quad (6)$$

$$pos_i(t+1) = pos_i(t) + vel_i(t+1) \quad (7)$$

where:

$$dw = \frac{\max_w - \min_w}{\text{maximum_iteration}}$$

w_I ranges linearly from maximum weight inertia $\max_w = 0.9$ to minimum weight inertia $\min_w = 0.4$.

The second improvement in learning strategy of other remaining particles is done by incorporating the ideology of Bhagavad Gita and dedication dividing them into ideal/dedicated, non-ideal particle which got success and non-ideal particle which got failure characteristics, considering the lesson of Bhagavad Gita that the human is not always classified as best or worst, the ideal human also exist who always moves with some dedication whether they have achieved success or failure. A particle is ideal/dedicated in a population is decided with the random probability considering $\text{rand} < \text{IdealCandidateProbability}$ is true, where rand is within the range [0, 1] and $\text{IdealCandidateProbability} = 0.5$. Dedication characteristic is induced by making the particle move with the factor of 0.9 in the search space. The random nature and increased movement effects in particle's learning strategy provides the intelligent exploitation of the search space. The inertia weight, velocity and position of these particles are obtained using (8), (9) and (10) respectively:

$$w_I = w_I - dw \quad (8)$$

$$\text{vel}_i(t+1) = w_I * \text{vel}_i(t) + c1 * \text{rand1} * (pbest - pos_i(t)) \quad (9)$$

$$pos_i(t+1) = pos_i(t) + 0.9 * \text{vel}_i(t+1) \quad (10)$$

The non-ideal person can move ahead with some dedication only when he gets success. The random probability decides non-ideal particles achieves success when $\text{rand} < \text{SuccessProbability}$ is true, where $\text{SuccessProbability} = 0.5$ and rand is within the range [0, 1]. The inertia weight for these particles is updated using (5) because if these particles are provided with higher acceleration they can perform better leading to higher exploration of the search space. Velocity vel_i and position pos_i of these particles are updated using (11) and (12), because these particles achieved some success they will show some dedication and will move with the factor of 0.9 in the search space. Further, non-ideal person when fails will have no dedication so the velocity vel_i and position pos_i of these particles are updated using (13) and (14):

$$\text{vel}_i(t+1) = w_I * \text{vel}_i(t) + c1 * \text{rand1} * (pbest - pos_i(t)) + c2 * \text{rand2} * (gbest - pos_i(t)) \quad (11)$$

$$pos_i(t+1) = pos_i(t) + 0.9 * \text{vel}_i(t+1) \quad (12)$$

$$\text{vel}_i(t+1) = \text{vel}_i(t) \quad (13)$$

$$pos_i(t+1) = pos_i(t) + 0.1 * vel_i(t+1) \quad (14)$$

The improved learning strategies provide better convergence and balanced exploration and exploitation which helps to obtain better biologically plausible network architecture. The position vector of the proposed algorithm encodes the RNN model parameters associated with each sub problem and training is performed. The error for each candidate solution is evaluated by using (3). The proposed framework of RNN learning using novel SI methodology is explained in Algorithm 1 and compared with other state-of-art human cognition based SI algorithms in terms of MSE and accuracy of prediction of true interactions i.e. true positives (TP).

Algorithm 1. Steps in RNN learning using proposed SI method

```

1. for each gene  $g = 1: N$  do // Genes Loop
2.   Given the population of the particles  $ps$ , the set of
     particles is represented  $X = (x_1, x_2, \dots, x_{ps})$ 
     Initialize randomly the position and velocity vector of the
      $ps$  particles of  $D$  dimension.
3.   The structure of the position vector for each particle
      $iPos(i) = [w_{i1}, w_{i2}, \dots, w_{iN}, b_i, \tau_i]$ , where  $i$  is the number of genes.
4.   The weight, bias and time constant should be in assigned
     boundaries,  $[wmin, wmax], [bmin, bmax]$  and  $[\tau min, \tau max]$  and velocity
      $[-Vmax, Vmax]$ 
5.   for each particle  $i$  do
6.     Using the RNN model formalism predict the gene
     expression of gene  $g$  at time point,  $\hat{x}_i(t)$  from the original
     gene expression of genes at previous time point  $(t-1)$ ,
      $\hat{x}_i(t-1)$  as in (1) then evaluate the error for each
     particle  $i$  using (3).
7.     Find the personal best ( $pbest$ ) for each particle and
     global best ( $gbest$ ).
8.   end for
// Evolution of Particles

9.   for iter = 1:MaxIterations do
10.    Find the  $gbest$  particle
11.    for the gbest particle do
12.      Evaluate the inertia weight using (5) and then
      update the velocity and position of the best
      particle using (6) and (7) respectively.
13.    end for
14.    for other remaining particles do
15.      Randomly generate a number and check whether a
      particle is ideal/dedicated
16.      if ( $rand < IdealCandidateProbability$ ), do

```

```

17.         Evaluate the inertia weight using (8) and
           then update the velocity and position of the
           particle using (9) and (10) respectively.
18.         else // Particle is Non-ideal
19.             if (rand < SuccessProbability), do //
           Non-ideal particle achieving success
20.                 Evaluate the inertia weight using (5)
                   and then update the velocity and
                   position of the particle using (11) and
                   (12) repectively.
21.             else // Non-ideal particle got failure
22.                 Update the velocity and position of the
                   particle using (13) and (14) repectively.
23.             endif
24.         endif
25.     end for
26.     Update the personal best and global best.
27. end for MaxIterations
28. Store gbest for each gene at the end of MaxIterations.
29. end for Genes loop

```

Collect and combine the obtained *gbest* result i.e. $N(N+2)$. Extract the first N elements from each row to form an NN adjacency matrix i.e. the reconstructed GRN. The proposed framework of RNN learning using novel SI algorithm for the reverse engineering problem of GRNs reconstruction is stochastic in nature; i.e. considering the time-series data each reconstructed network has differences in their topology when run for $L=10$ independent experiments. These L networks are ensembled and the adjacency matrix of size NN which is considered as the best solution of each independent experiment is noted. Subsequently, the scoring methodology $Score_{ij}$ is used to build a final reconstructed GRN $\widehat{G} = [\widehat{g}_{ij}]_{NN}$ whose each edge $\widehat{g}_{ij}$ is either 0 or 1, is obtained by (15):

$$Score_{ij} = \frac{1}{L} \sum_{L=1}^L g_{ij} \quad (15)$$

where $g_{ij} \in G$ reconstructed GRN at independent experiment. An edge in final predicted GRN will be included or not is monitored by the inclusion threshold value $\theta \in [0, 1]$ is obtained by (16):

$$\widehat{g}_{ij} = \begin{cases} 1 & \text{if } Score_{ij} > \theta \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

5. EXPERIMENTATION AND DISCUSSIONS

In this paper, the proposed framework for the problem of the reconstruction of gene regulatory networks was first tested on the small artificial network consisting of 4 genes and then the methodology has been applied on 4 different temporal datasets of SOS DNA damage repair network of Escherichia coli (in vivo) consisting of 8 major genes. The final reconstructed GRN is obtained with the binary values in the matrix $\widehat{g}_{ij}$, is validated for its biological significance. Network validation provides

insights for many meaningful regulations among genes. Section 5.1 explains the evaluation criteria on which the performance of the reconstructed GRN has been validated when compared with true Gold Standard. The results are shown in the section 5.2. The simulations are done on Intel Core i7 CPU with 3.4 GHz and 8 GB RAM system, the program is implemented in MATLAB 2015a with Windows 10 64-bit operating system environment.

5.1 Evaluation Criteria

The performance of the proposed method is validated when comparison is made between the final reconstructed GRN and the Gold Standard GRN using some statistical properties. Edge e_{ij} in final GRN is categorized as true positive (TP), true negative (TN), false positive (FP), false negative (FN). TP: when an edge is predicted correctly, TN: when an edge is not predicted in final network and also not present in gold standard network, FP: when an edge in final network is predicted but not present in gold standard network and FN: when an edge in final network is not predicted but is present in gold standard network.

When gold standard network is known then final reconstructed network is compared with it using these statistical properties for performance analysis. True Positive Rate (TPR) (sensitivity/recall) which is the fraction of the count of edges predicted correctly (TP) to the count of total number of edges actually have to be predicted (TP + FN), another is false positive rate (FPR) (specificity) which is the fraction of the count of incorrect predicted edges (FP) to the count of total number of edges that should not be predicted (FP + TN), last metric is positive predictive value (PPV) (precision) which is the fraction of count of edges predicted correctly (TP) to the count of total number of edges predicted (TP + FP).

5.2 Results and Analysis

The performance of the proposed approach is tested on the synthetic dataset in which gene expression data is generated and then used as input for GRN reconstruction, it is elaborated in sub section 5.2.1. Further sub section 5.2.2 shows the performance of the approach on the 4 different experiments of E. Coli SOS DNA damage repair real-world dataset.

5.2.1 Small Artificial Network

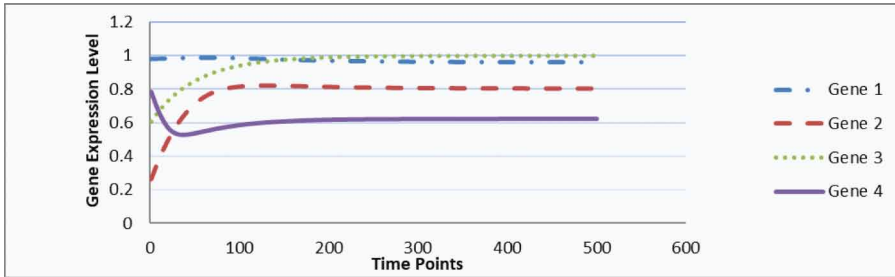
The proposed methodology reconstructs the 4-genes and 8 regulatory interactions small artificial network given by (R. Xu et al., 2007). The purpose of reconstructing small artificial network is for primary justification of the proposed method for the problem of GRNs reconstruction. The training dataset has been generated using (1) considering the parameters of RNN model given in the Table 1.

Using (1) total 500 time points are generated assuming time interval $\Delta t = 0.1$, the expression values of the genes get saturated very fast. Figure 2 shows the dynamics of the training data generated. Out of 500 time points 50 time points are evenly sampled because in real-world such huge time-points

Table 1. RNN parameters used to generate training data of small artificial network (R. Xu et al., 2007)

w_{ij}				b_i	τ_i
20	-20	0	0	0	10
15	-10	0	0	-5	5
0	-8	12	0	0	5
0	0	8	-12	0	5

Figure 2. Dynamics of the training dataset for small artificial network consisting of 4-genes



are not possible. Now considering the training data generated with 50 time points as input for the GRN reconstruction, $L=10$ independent trails of the proposed RNN learning with the novel SI algorithm framework has been performed, taking swarm population of ${}_m^4C$ where $m = 1, 2, 3$ and 4 and performing the learning for 50,000 iterations.

Discussion: Considering the parameters proposed SI algorithm the weight inertia decreases linearly with $max_w = 0.9$, $min_w = 0.4$, $c1 = c2 = 2$. The initial value of weight (bias) and time constant parameters of RNN model are $[-1, 1]$ and $[1, 10]$ respectively that encodes the position vector of the proposed SI algorithm and maximum velocity is $V_{max} = 2$. Table 2 gives the count of TP, FP, TN and FN of the reconstructed GRN formed by the proposed framework of RNN learning using novel SI algorithm and results are compared with the framework of RNN learning with PSO and RNN learning framework with the other human cognition based SI algorithms Self-Regulating PSO (SRPSO) (Tanweer et al., 2015), Human Bhagavad Gita PSO (HBPSO) and Human Dedication PSO (HDPSO) (Gajawada et al., 2019). It is observed that with inclusion threshold $\theta = 0.8$, novel SI algorithm obtained 7 TPs and no FP, using single time-series. The average mean square error (MSE) obtained by novel SI algorithm is approximately in magnitude of 10^{-2} and 10^{-3} . Figure 3 shows that proposed algorithm performed better when comparison is made for TPR with other state-of-art algorithms. FPR and PPV are 0 and 1 respectively for all the algorithms.

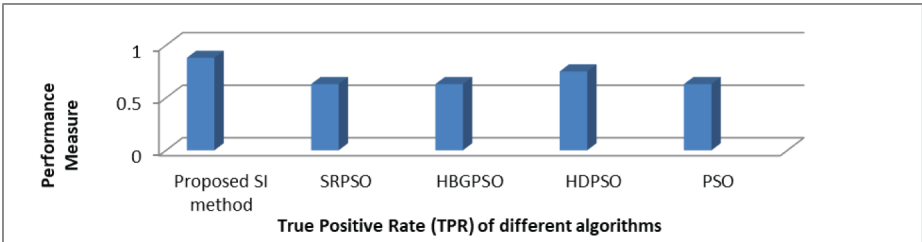
5.2.2 E. Coli SOS DNA Damage Repair Real-World Network

The proposed methodology for reverse engineering of GRNs from time-series data has been applied on the (in vivo) real-world network of E. Coli SOS repair networks. The dataset appropriately shows the dynamical nature of E. Coli SOS repair system; this network contains the proteins which participate in the repair mechanism. The actual SOS network consists of 40 genes for the repair mechanism of

Table 2. Performance Analysis of the proposed algorithm and other state-of-art algorithms

Algorithms	TP	FP	TN	FN
Proposed SI method	7	0	8	1
SRPSO	5	0	8	3
HBGPSO	5	0	8	3
HDPSO	6	0	8	2
PSO	5	0	8	3

Figure 3. Performance measure of TPR between the proposed SI algorithm and other state-of-art algorithms



DNA (Michel, 2005). LexA and recA deeply contribute in the changes in expression level of other genes. Ronen et al. (2002) elaborated about the temporal behavior of 8 major genes of the SOS repair mechanism capturing the dynamical nature of response system showing 9 interactions. These eight major genes are lexA, recA, uvrD, uvrA, uvrY, umuD, ruvA, and polB. Total 4 experiments are conducted using two separate UV light intensities on the E.Coli, on 1st and 2nd experiment UV light of 20 Jm-2 was used and on 3rd and 4th experiment UV light of 5 Jm-2 was used and temporal datasets are formulated, each dataset contains 50 time points (columns in the training data) sampled evenly at the interval of 6 minutes for 8 major genes (rows in the training data). The dataset is obtained from the link (<http://wws.weizmann.ac.il/mcb/UriAlon/sites/mcb.UriAlon/files/uploads/DownloadableData/sosdata.zip>).

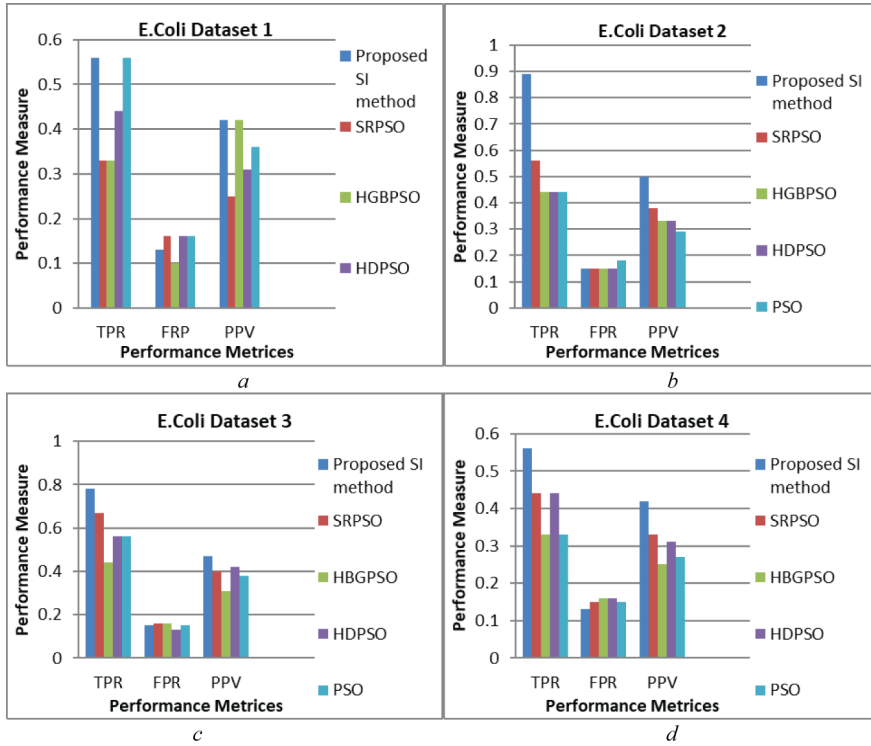
The experimental setup used for the inference of SOS repair system from the given temporal data, $L = 10$ independent trials of the proposed framework has been performed taking the population size of swarm as 8C_m where m be the plausible combinations of regulations per target gene ($m = 1, 2, 3$ and 4). Each experiment is trained for 5000 generations. In all 4 datasets of SOS repair system the gene expression value at first time point is zero so in experimentation it has been removed leaving with total 49 time points in the dataset. The dataset then needs to be normalized within the range of [0, 1].

Discussion: Considering the initial random value of weight (bias) and time constant of the RNN model parameters encoded in the position vector of the proposed SI algorithm are [-5, 5] and [1, 10] respectively and maximum velocity taken is $v_{max} = 5$. Table 3 shows the statistical properties and MSE obtained for the reconstructed GRN with the threshold value of $\theta = 0.9$ using the framework of RNN learning with proposed SI algorithm. Figure 4 shows the comparison between the performance (TPR, FPR and PPV) of the proposed algorithm with other state-of-art algorithms. The 2nd experiment achieves highest prediction with 16 edges in network topology. Kentzoglanakis and Poole (2012) lacks to obtain true positive in 4th experiment but our proposed method identifies true positive in each experiment. It has been observed that results show significant improvement. It is because of the improved learning

Table 3. Statistical Properties of the proposed SI algorithm with E.Coli dataset

Dataset	TP	TN	FP	FN	TPR	FPR	PPV	MSE
1	5	48	7	4	0.56	0.13	0.42	0.0036
2	8	47	8	1	0.89	0.15	0.5	0.0074
3	7	47	8	2	0.78	0.15	0.47	0.0072
4	5	48	7	4	0.56	0.13	0.42	0.0062

Figure 4. Performance measure for E.Coli datasets (a)Dataset 1 (b) Dataset 2 (c) Dataset 3 (d) Dataset 4



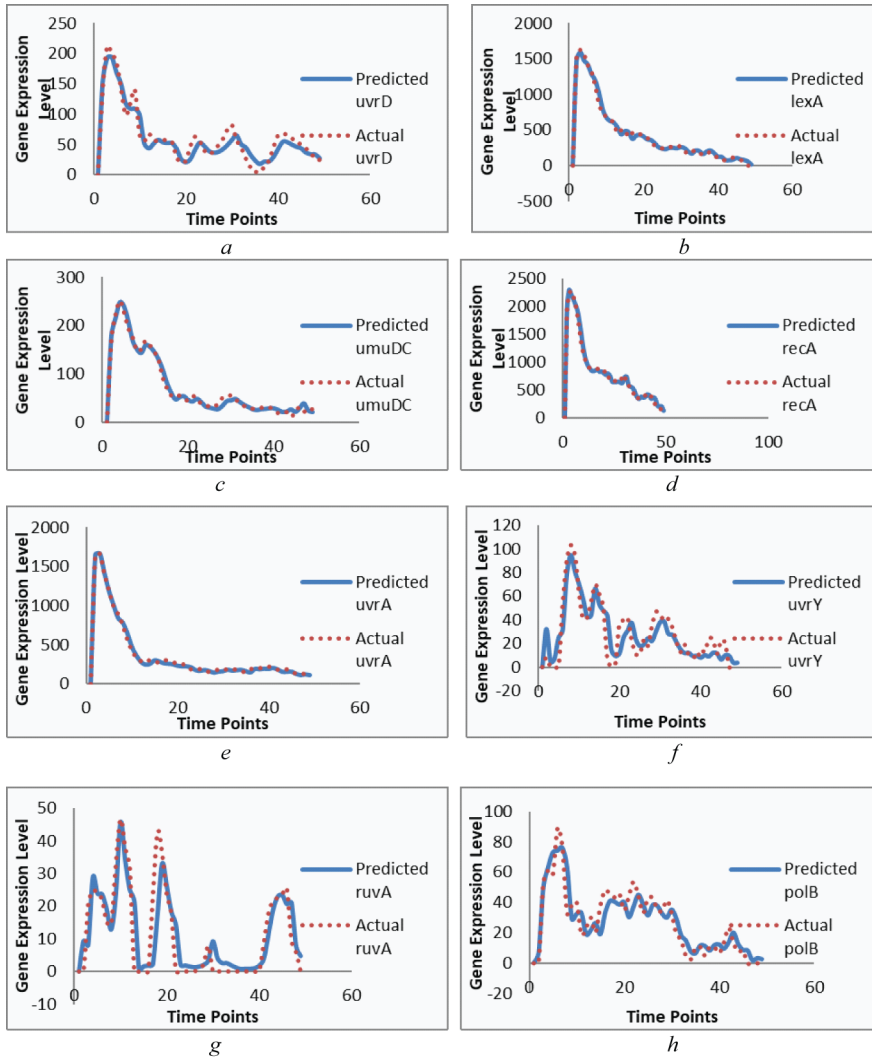
strategy adopted in proposed SI algorithm for both the best particle providing it with high acceleration leading to enhanced exploration along with it other particles are incorporated with the ideology of Bhagavad Gita empowering its velocity and position update mechanism and some randomness which help escape the local optimum, improving its exploration and exploitation ability providing high precision solution.

Figure 5 shows the actual and predicted gene expression level obtained by RNN learning with proposed algorithm from the 2nd experiment of E.Coli dataset consisting 8 major genes efficiently mimicking the dynamics of the original time-series data.

6. CONCLUSION

The problem of the reconstruction of gene regulatory networks (GRNs) from high throughput time-series gene expression data using computational intelligence is a well-posed challenge in bioinformatics. RNN model is an efficient modeling approach for GRN reconstruction. In this paper, an intelligent framework for GRN reconstruction problem has been proposed which employs the modified PSO based on the human cognition for the training of RNN parameters (GRN parameters). The proposed swarm intelligence based algorithm along with SRPSO, HGBPSO and HDPSO human cognition based SI algorithms are fitted to the GRN model and each algorithm is executed separately on small artificial network dataset (4 genes and 8 interactions) and four different experiments of E.Coli real-world network dataset (8 genes and 9 interactions). Experiments show that proposed SI

Figure 5. Actual and Predicted dynamics for genes of E.Coli Dataset 2 (a) Gene *uvrD* (b) Gene *lexA* (c) Gene *umuDC* (d) Gene *recA* (e) Gene *uvrA* (f) Gene *uvrY* (g) Gene *ruvA* (h) Gene *polB*



methodology outperforms all other methods for both artificial and real-world network datasets. In the experiment, the proposed framework obtains highest true positive count (7) for artificial network dataset. In the 2nd experiment of the E.Coli dataset proposed framework achieves the highest sensitivity (0.89). The GRN reconstruction problem requires highly-scalable modeling approaches to handle large-scale datasets. Large datasets causes curse of dimensionality that makes suffer the performance of modeling methods. Further, complete biological insight will be obtained by the integration of multi-omics data obtained from multiple-sources, and it can be the major focus of GRN research. In the conclusion, GRN modeling is a powerful tool in system biology which allows the discovery of complex relationships among biological entities.

REFERENCES

- Aalto, A., Viitasari, L., Ilmonen, P., Mombaerts, L., & Gonçalves, J. (2020). Gene regulatory network inference from sparsely sampled noisy data. *Nature Communications*, 11(3493), 3493. Advance online publication. doi:10.1038/s41467-020-17217-1 PMID:32661225
- Adabor, E. S., & Acquah-Mensah, G. K. (2019). Restricted-derestricted dynamic Bayesian Network inference of transcriptional regulatory relationships among genes in cancer. *Computational Biology and Chemistry*, 79, 155–164. doi:10.1016/j.compbiolchem.2019.02.006 PMID:30822674
- Altman, N., & Krzywinski, M. (2018). The curse(s) of dimensionality. In *Nature Methods* (pp. 15, 399–400). doi:10.1038/s41592-018-0019-x
- Barbuti, R., Gori, R., Milazzo, P., & Nasti, L. (2020). A survey of gene regulatory networks modelling methods: From differential equations, to Boolean and qualitative bioinspired models. *Journal of Membrane Computing*, 2(3), 207–226. doi:10.1007/s41965-020-00046-y
- Barman, S., & Kwon, Y. K. (2018). A Boolean network inference from time-series gene expression data using a genetic algorithm. *Bioinformatics (Oxford, England)*, 34(17), i927–i933. Advance online publication. doi:10.1093/bioinformatics/bty584 PMID:30423074
- Biswas, S., & Acharyya, S. (2016). Neural model of gene regulatory network: A survey on supportive meta-heuristics. *Theory in Biosciences*, 135(1), 1–19. doi:10.1007/s12064-016-0224-z PMID:27048512
- Biswas, S., & Acharyya, S. (2018). A Bi-Objective RNN Model to Reconstruct Gene Regulatory Network: A Modified Multi-Objective Simulated Annealing Approach. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(6), 2053–2059. doi:10.1109/TCBB.2017.2771360 PMID:29990170
- Delgado, F. M., & Gómez-Vela, F. (2019). Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artificial Intelligence in Medicine*. doi:10.1016/j.artmed.2018.10.006
- Gajawada, S., & Mustafa, H. (2019) Ten Artificial Human Optimization Algorithms. *Transactions on Machine Learning and Artificial Intelligence*. 10.14738/tmlai.73.6631
- Grzegorzczak, M., Aderhold, A., & Husmeier, D. (2019). Overview and Evaluation of Recent Methods for Statistical Inference of Gene Regulatory Networks from Time Series Data. In *Methods in Molecular Biology* (pp. 49–94). doi:10.1007/978-1-4939-8882-2_3
- Gudise, V. G., & Venayagamoorthy, G. K. (2003). Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks. *2003 IEEE Swarm Intelligence Symposium, SIS 2003 - Proceedings*, 110–117. doi:10.1109/SIS.2003.1202255
- Jana, B., Mitra, S., & Acharyya, S. (2019). Repository and Mutation based Particle Swarm Optimization (RMPSO): A new PSO variant applied to reconstruction of Gene Regulatory Network. *Applied Soft Computing*, 74, 330–355. doi:10.1016/j.asoc.2018.09.027
- Kentzoglanakis, K., & Poole, M. (2012). A swarm intelligence framework for reconstructing gene networks: Searching for biologically plausible architectures. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(2), 358–371. doi:10.1109/TCBB.2011.87 PMID:21576756
- Khan, A., Mandal, S., Pal, R. K., & Saha, G. (2016). Construction of Gene Regulatory Networks Using Recurrent Neural Networks and Swarm Intelligence. *Scientifica*, 2016, 1–14. Advance online publication. doi:10.1155/2016/1060843 PMID:27298752
- Khan, A., Saha, G., & Pal, R. K. (2020). Modified Half-System Based Method for Reverse Engineering of Gene Regulatory Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(4), 1303–1316. doi:10.1109/TCBB.2019.2892450 PMID:30640623
- Kordmahalleh, M. M., Sefidmazgi, M. G., Harrison, S. H., & Homaifar, A. (2017). Identifying time-delayed gene regulatory networks via an evolvable hierarchical recurrent neural network. *BioData Mining*, 10(1), 29. Advance online publication. doi:10.1186/s13040-017-0146-4 PMID:28785315

- Li, Y., Chen, H., Zheng, J., & Ngom, A. (2016). The Max-Min High-Order Dynamic Bayesian Network for Learning Gene Regulatory Networks with Time-Delayed Regulations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(4), 792–803. doi:10.1109/TCBB.2015.2474409 PMID:26336144
- Lillicrap, T. P., & Santoro, A. (2019). Backpropagation through time and the brain. In *Current Opinion in Neurobiology* (pp. 55, 82–89). doi:10.1016/j.conb.2019.01.011
- Liu, L., & Liu, J. (2020). Reconstructing gene regulatory networks via memetic algorithm and LASSO based on recurrent neural networks. *Soft Computing*, 24(6), 4205–4221. doi:10.1007/s00500-019-04185-y
- Luque-Baena, R. M., Urda, D., Subirats, J. L., Franco, L., & Jerez, J. M. (2014). Application of genetic algorithms and constructive neural networks for the analysis of microarray cancer data. *Theoretical Biology and Medical Modelling*, 11(1). 10.1186/1742-4682-11-S1-S7
- Lynn, N., & Suganthan, P. N. (2015). Heterogeneous comprehensive learning particle swarm optimization with enhanced exploration and exploitation. *Swarm and Evolutionary Computation*, 24, 11–24. doi:10.1016/j.swevo.2015.05.002
- Manica, M., Polig, R., Purandare, M., Mathis, R., Hagleitner, C., & Martinez, M. R. (2020). FPGA Accelerated Analysis of Boolean Gene Regulatory Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6), 2141–2147. doi:10.1109/TCBB.2019.2936836 PMID:31494553
- Melkman, A. A., Cheng, X., Ching, W. K., & Akutsu, T. (2018). Identifying a Probabilistic Boolean Threshold Network from Samples. *IEEE Transactions on Neural Networks and Learning Systems*, 29(4), 869–881. doi:10.1109/TNNLS.2017.2648039 PMID:28129190
- Michel, B. (2005). After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biology*. doi:10.1371/journal.pbio.0030255
- Morgan, D., Studham, M., Tjärnberg, A., Weishaupt, H., Swartling, F. J., Nordling, T. E. M., & Sonnhämmer, E. L. L. (2019). Perturbation-based gene regulatory network inference to unravel oncogenic mechanisms. In *bioRxiv*. 10.1101/735514
- Noman, N., Palafox, L., & Iba, H. (2013). *Reconstruction of Gene Regulatory Networks from Gene Expression Data Using Decoupled Recurrent Neural Network Model*. 10.1007/978-4-431-54394-7_8
- Raza, K., & Alam, M. (2016). Recurrent neural network based hybrid model for reconstructing gene regulatory network. *Computational Biology and Chemistry*, 64, 322–334. doi:10.1016/j.compbiolchem.2016.08.002 PMID:27570069
- Razaghi-Moghadam, Z., & Nikoloski, Z. (2020). Supervised learning of gene-regulatory networks based on graph distance profiles of transcriptomics data. *NPJ Systems Biology and Applications*, 6(21), 21. Advance online publication. doi:10.1038/s41540-020-0140-1 PMID:32606380
- Ronen, M., Rosenberg, R., Shraiman, B. I., & Alon, U. (2002). Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.152046799
- Sanchez-Castillo, M., Blanco, D., Tienda-Luna, I. M., Carrion, M. C., & Huang, Y. (2018). A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics (Oxford, England)*, 34(6), 964–970. doi:10.1093/bioinformatics/btx605 PMID:29028984
- Tanweer, M. R., Suresh, S., & Sundararajan, N. (2015). Self regulating particle swarm optimization algorithm. *Information Sciences*, 294(September), 182–202. doi:10.1016/j.ins.2014.09.053
- Vohradsky, J. (2001). Neural Model of the Genetic Network. *The Journal of Biological Chemistry*, 276(39), 36168–36173. doi:10.1074/jbc.M104391200 PMID:11395518
- Xie, J., Zhao, C., Sun, J., Li, J., Yang, F., Wang, J., & Nie, Q. (2020). Prediction of Essential Genes in Comparison States Using Machine Learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1–1. doi:10.1109/TCBB.2020.3027392 PMID:32991286
- Xing, L., Guo, M., Liu, X., Wang, C., Wang, L., & Zhang, Y. (2017). An improved Bayesian network method for reconstructing gene regulatory network based on candidate auto selection. *BMC Genomics*, 18(S9), 844. Advance online publication. doi:10.1186/s12864-017-4228-y PMID:29219084

Xu, G., Cui, Q., Shi, X., Ge, H., Zhan, Z. H., Lee, H. P., Liang, Y., Tai, R., & Wu, C. (2019). Particle swarm optimization based on dimensional learning strategy. *Swarm and Evolutionary Computation*, 45, 33–51. doi:10.1016/j.swevo.2018.12.009

Xu, R., Wunsch, D. C., & Frank, R. L. (2007). Inference of genetic regulatory networks with recurrent neural network models using particle swarm optimization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(4), 681–692. Advance online publication. doi:10.1109/TCBB.2007.1057 PMID:17975278

Yu, B., Xu, J. M., Li, S., Chen, C., Chen, R. X., Wang, L., Zhang, Y., & Wang, M. H. (2017). Inference of time-delayed gene regulatory networks based on dynamic Bayesian network hybrid learning method. *Oncotarget*, 8(46), 80373–80392. doi:10.18632/oncotarget.21268 PMID:29113310

Bhavana Bansal is a PhD candidate in the department of CS & IT, Jaypee Institute of Information Technology, Noida. Her research interest includes evolutionary computation, computational biology, machine learning.

Aparajita Nanda received the Ph.D. degree in Computer Science and Engineering from NIT Rourkela, India. She is currently an Assistant Professor with the CS&IT Department, Jaypee Institute of Information Technology, Noida. She has authored or co-authored in research articles in journals, conferences, and book chapters. She is a member of the ACM and IAENG. She has worked on a R&D projects funded by SERB. Area of work: Computer vision, Data Mining, Machine learning and Visual surveillance.

Anita Sahoo has done her Ph.D. in the area of medical image analysis. Her research focuses on utilizing computational intelligence to develop different meta-heuristic guided modules to produce a comprehensive framework for solving identification/ recognition problems. She has published many research papers in proceedings of international conference and journals of repute. She has more than 20 years of teaching experience. Presently, she is working as an assistant Professor (Sr. Grade) in the department of CSE&IT at JIIT, Sec-62, NOIDA, India. Her career goal has been to imbibe virtues of being a good educator and researcher and deliver in best possible way to help students and the society.