Precise Feature Selection in Predictive Genetic Models Using Grey Wolf Optimization Algorithm

Mohamed Abbas
College of Engineering
Delta University
Egypt

***Abstract*—Genetic models are overarching ideas on how genes work in individuals to affect phenotypes. The process of feature selection is a vital process to be merged in the genetic model. Genetic models specifically relate genotype to phenotype. In lots of instances, the found facts are steady with those sorts of fashions and a few analyses of genetic records are in large part descriptive.The feature selection process in genetic models faces some difficulties arising from the correlation structures present in DNA structure data. An optimization algorithm is neededto solve these difficulties The main objective of this paper is to introduce an accurate feature selection process in genetic models. This aim is verified using an accurate and green optimization algorithm entitled grey Wolf Optimization algorithm GWO. The results depict that GWO solve the problem of the correlation structures in DNA sequence data more precise and efficiently with average efficiency equals 1.19E+02.**

Keywords: Genetic model, feature selection, correlation structures, proximal gradient, Grey wolf optimization.

1 INTRODUCTION

Genetic model is a theatrical performance of an mRNA transcript of a gene. This method should contains information about lineament of the transcript such as coding DNA - intron limit. Genetic models can explain how number of factors affect phenotype.Also, it explains if the gene have one or more gene model. Hence there is a tilt of feature of speech , and it is important to select a fixed identification number of those features to include in the model. Previous methods divided the methods of feature selection into two general methods: filter method and Wrapper method [1,2]. Filter out type methods pick variables regardless of the model. They're based handiest on trendy features just like the correlation with the variable to are expecting.

Filter techniques suppress the least exciting variable quantity. The other variables will be part of a classification or a regression model used to classify or to predict information [I]. However, wrapper strategies examine subset of variables which lets in, not like filter approaches, to locate the viable interactions among variables [2].

Xiao, J., Xiao, Y., Huang, A. et al. in [3] propose a feature-selection based dynamic carry-over ensemble model that intention to introduce transfer learning hypothesis for utilizing the customer data in both the aim and related source domains. Fig.1 depicts the feature selection based dynamic transfer ensemble model. Where target information set and the source domain are the initial stages of this method. After selection of N feature subsets ; the process arrived to its end. Then the finial classification is finished [3].

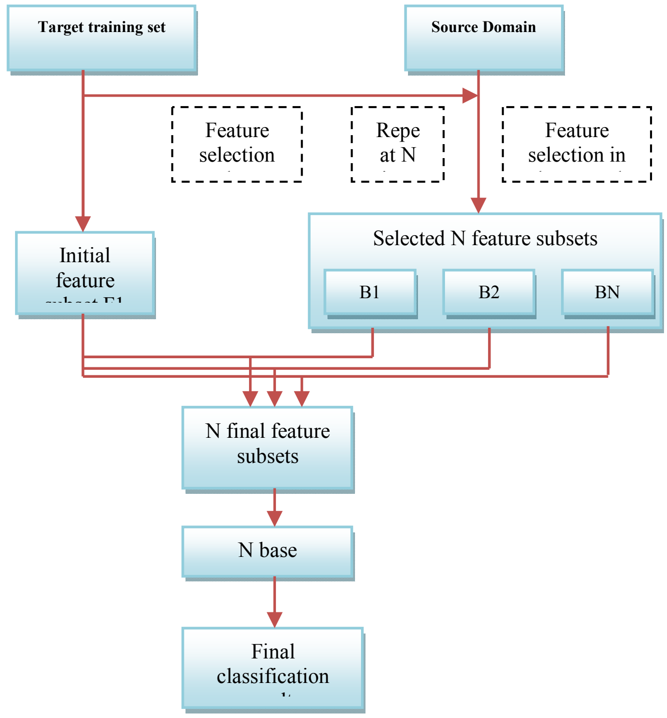

Fig 1 Feature-selection-based dynamic transfer ensemble model[3]

AytuğOnan et al in [4] presents an ensemble glide slope for characteristic excerption, which aggregates the several someone feature film of speech listing obtained by the different feature option method s so that a more robust and efficient feature sub stage set can be obtained. In orderliness to aggregate the individual feature lists, a genetic algorithm ic rule has been utilized. AlexejGossmann , Shaolong Cao in [5] used LASSO, least absolute shrinkage and pick operator, to address the feature selection problem.

They set up an algorithm for a proximal gradient technique to remedy its optimization trouble. The main objective of this newspaper is to introduce an accurate feature selection in the genetic model. The proposed function choice model depending on a modern optimization set of rules entitled grey Wolf Optimization, GWO. This algorithm approves its capability in solving optimization problem more accurate, precise and efficiently. The rest of this paper is prepared as comply with: section II is an in depth verbal description about the trouble system of this research. Section III introduces a brief description about GWO algorithm. Section IV presents the proposed feature selection approach using GWO algorithm. At the end, section V analyzes the results of the proposed approach comparing with old ones.

III PROBLEM FORMULATION

AlexejGossmann, Shaolong Cao in [5] consider the following linear model as an representation of a genetic model: $D = \mu A + \phi B + C$ where the matrix $\mu, \phi \in R^n$ constitutes the matrix; $D \in R^n$.This model contains the alien coefficient which we want to estimate, given the prior knowledge that b is sparse, this problem is known as model extract or lineament option. Apart from the identification of relevant forecaster variables, once the non-zero coefficients of b are estimated, they can also be used to make predictions from new data. For example, if A,B contain the genotype information of the matter and contains their phenotype values, then the non-zero coefficients in b represent the genetic modelsassociatedwith the phenotype. To reference the feature extract problem consider the following conceptualization :

Let the vector $z = \left(z_1^T, z_2^T, z_3^T, \ldots. z_m^T\right)$ be defined by

$z_i = \sqrt{l_i b_i}$ $\forall i \in \{1,2,\ldots\ldots,m\}$ Let D be the diagonal matrix with entries $\sqrt{l_i}$

such that z= Db. The minimization problem of LASSO is equivalent to the problem

$$\min_{b \in R^P} v1(z) + v2(z)$$

$$v1(z) = \frac{1}{2}\left\|y - XD^{-1}z\right\|_2^2$$

$$v2(z) = \sum_{i=1}^{m} \lambda i \left\|z(i)\right\|_2$$

and $\|z_1\|_2 \geq \|z_2\|_2 \geq \|z_3\|_2 \geq \ldots\ldots\ldots \geq \|z_m\|_2$ are the Euclidean norms of the blocks of the vector c in non-increasing order.

IV GREY WOLF OPTIMIZATION GWO ALGORITHM

This phase introduces the gray wolf optimization GWO ; the technique that is used inside the function selection method. Grey wolf optimization is a group method evolved by Mirjalili et al., in [6] which mimics the management hierarchy of wolves are widely known for his or her group searching. Grey wolf belongs to a family and usually opt to stay in a pack.To start the algorithm , first the wolves should be generated randomly based on size of the pack. Mathematically, these wolves can be expressed as,

$$Wolves = \begin{bmatrix} WO_1^1 & WO_2^1 & WO_d^1 \\ WO_1^2 & WO_2^2 & WO_d^2 \\ & & \\ WO_1^{GS} & WO_2^{GS} & WO_d^{GS} \end{bmatrix}$$

Where, Gij is the initial value of the j th pack of the i th wolves. Then, the algorithm estimate the fitness value of each hunt agent using the following:

$$\vec{D} = \left| \vec{n}\, rand. \vec{WO}_P(t) - \vec{WO}(t) \right|$$

$$\vec{WO}(t+1) = \vec{WO}_P(t) - \vec{p}\, rand. \vec{D}$$

$$\vec{n} = 2\, \vec{a}\, .rand1 - \vec{n}$$

$$\vec{p} = 2rand2$$

The values , a , are linearly decreased from 2 to 0 over the course of iterations. Where the magnitudes of vectors n, p are calculated randomly.

V PROPOSED FEATURE SELECTION APPROACH USING GWO ALGORITHM

This section presents the proposed method in details using GWO as depicted in algorithm (A).The proposed technique have a strict social dominant hierarchy. The leader is more often than not accountable for selection making. The orders of the dominant wolf have to be followed.

Algorithm(A)ProposedFeature Selection Approach Using GWO Algorithm

Fix $x \in (0, MP \min(1, \frac{1}{x})), b^{(0)} \in R^P, c^{(0)} = Db^{(0)}$

MP: multiplication parameter of the linear model
For k=0,1,2,……do
Generate initial search agents Gi (i=1, 2,…., n)
Initialize the vectors n,p
Estimate the fitness value of each hunt agent
repeat : determine the best hunt agent
    Renew the location of the current hunt agent.
    End for
    Estimate the fitness value of all hunt agents

Fix $\gamma_k \in \left[\varepsilon, \frac{2}{\varepsilon} - \varepsilon\right]$

    Update the value of WO
    Update the vectors n,p
Iter=Iter+1
until Iter>= maximum number of iterations
GWO = G

$z^{(k+1)} \longleftarrow GWO$

End For

The subordinate wolves which help the chief in selection making.The subordinate is an advisor to the leader and discipliner for the pack. The minorgrade grey wolf has to present all other dominant wolves.The hunting techniques and the social hierarchy of wolves as search agents (Vs), design variable size (Vd), vectors n ,p and maximum number of iteration.

The proposed algorithm is started by fixing the minimization problem of the linear model. That is required to select their features. The selection process will be repeated regarding to the number of the features. The selection process will be repeated according to the estimated number of features. Then GWO starts to work to verify its optimum value iteration number with a precise value of the minimization problem. First, it starts by generating an initial search agents. Then, estimate the fitness value of the best hunt agent. Renew the location of the current hunt agent. Then estimate the fitness value of all hunt agents. Hence update the value of the previous position and in order update the different values of the vectors, each step calculates the maximum number of iterations. On the cease, it calculates the final price of the linear equation that represents the genetic version.

VI ANALYSIS OF THE PROPOSED APPROACH AND RESULTS

This section introduces an analysis of the proposed approach. The analysis is based on a comparison between two approaches; Feature selection using proximal gradient method and feature selection using GWO. The comparison is based on the variations of the multiplication parameters. Case I: at a multiplication parameter equals 0.5 for the linear model and by using the proximal gradient method. The minimum value of the linear equation is 9.93e-07 after an iteration number 123 as depicted in Fig 2. In case of using GWO the minimum value of the linear model is 8.3436e-06 at 200 number of iterations as depicted in Fig 3.

Case II: at a multiplication parameter equals 1 for the linear model and by using the proximal gradient method. The minimum value of the linear equation is 9.95e-07after an iteration number 98as depicted in Fig 4. In case of using GWO the minimum value of the linear model is 4.1534e-06 at 200 number of iterations as depicted in Fig 5.

Case III: at a multiplication parameter equals 1.5 for the linear model and by using the proximal gradient method. The minimum value of the linear equation is 9.40e-07 after an iteration number 85 as depicted in Fig 6. In case of using GWO the minimum value of the linear model is 1.2346e-05at 200 number of iterations as depicted in Fig 7.

Case IV: at a multiplication parameter equals 2 for the linear model and by using the proximal gradient method. The minimum value of the linear equation is 9.66e-07after an iteration number 102 as depicted in Fig 8. In case of using GWO the minimum value of the linear model is 2.519e-05 at 200 number of iterations as depicted in Fig 9.

Case V: at a multiplication parameter equals 2.5 for the linear model and by using the proximal gradient method. The minimum value of the linear equation is 9.72e-07after an iteration number 94 as depicted in Fig 10.

In case of using GWO the minimum value of the linear model is 1.0079e-05 at 200 number of iterations as depicted in Fig 11.Case VI: at a multiplication parameter equals 3 for the linear model and by using the proximal gradient method.

The minimum value of the linear equation is 9.19e-07 after an iteration number 88 as depicted in Fig 12. In case of using GWO the minimum value of the linear model is 2.1928e-05 at 200 number of iterations as depicted in Fig 13.

This means GWO achieved the minimum value of the linear model due to its capability attains a maximum number of iterations equals 200. Fig 14 depicts a comparison between LASSO and GWO at different values of multiplication parameter MP. The figure shows that GWO is more efficient than LASSO by an average 1.19E+02.

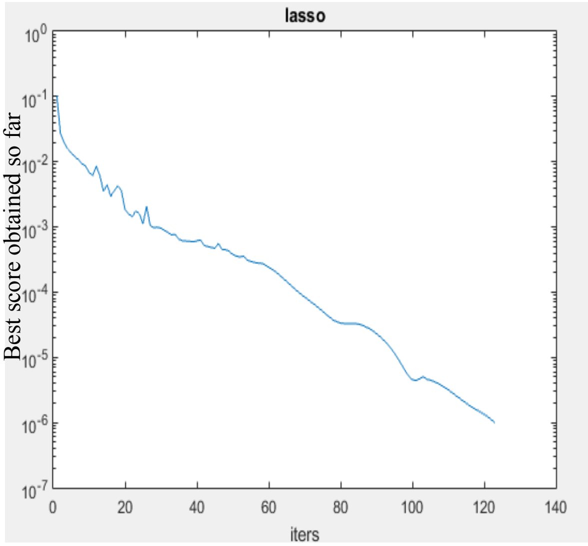

Fig 2 MP equals 0.5 by using the proximal gradient

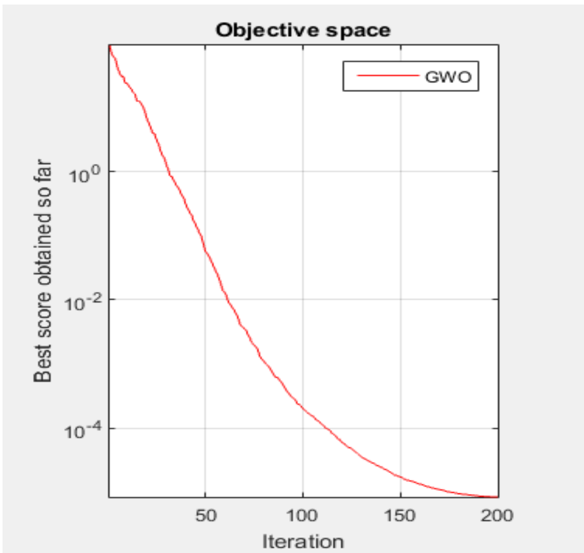

Fig 3MP equals 0.5 by using GWO

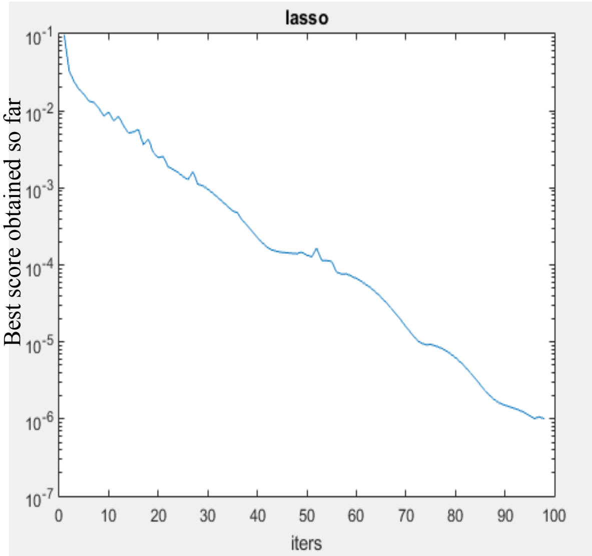

Fig 4MP equals 1 by using the proximal gradient

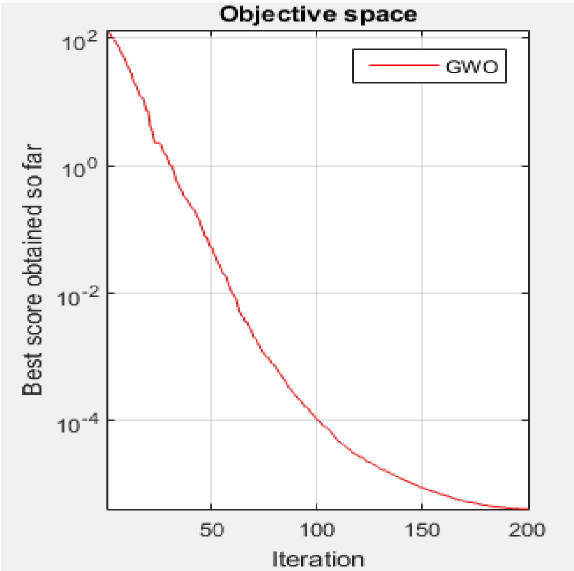

Fig 5MP equals 1 by using GWO

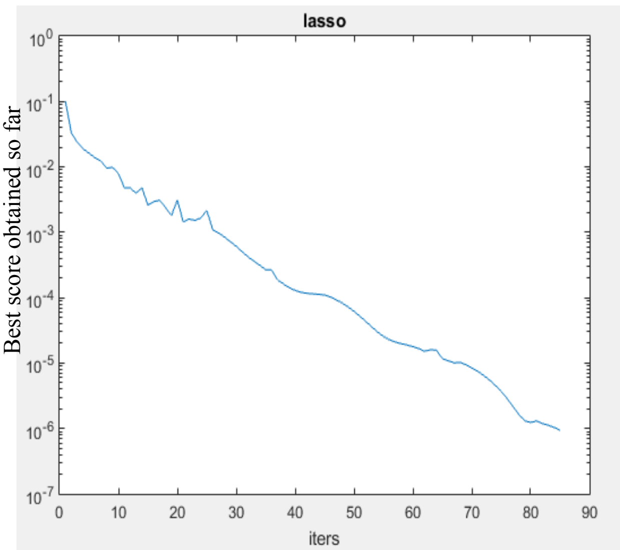

Fig 6MP equals 1.5 by using the proximal gradient

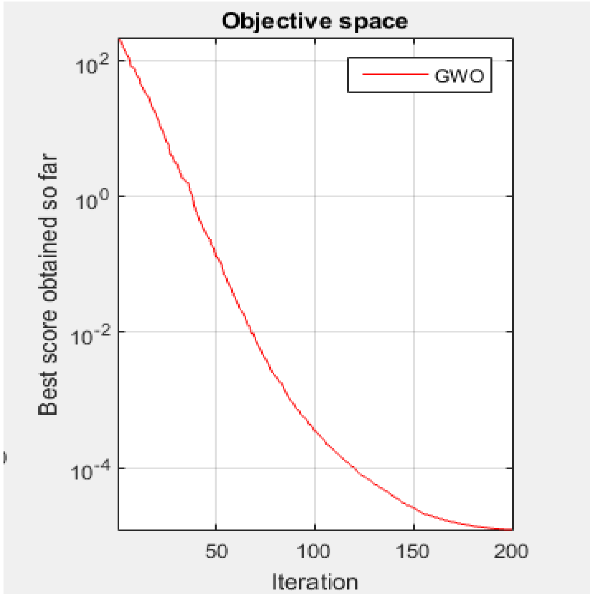

Fig 7MP equals 1.5 by using GWO

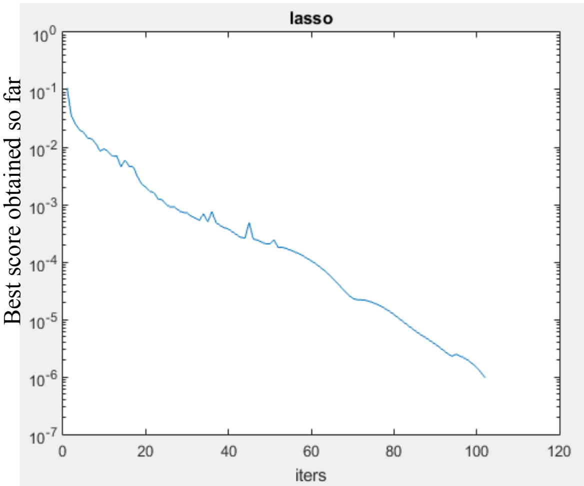

Fig 8MP equals 2 by using the proximal gradient

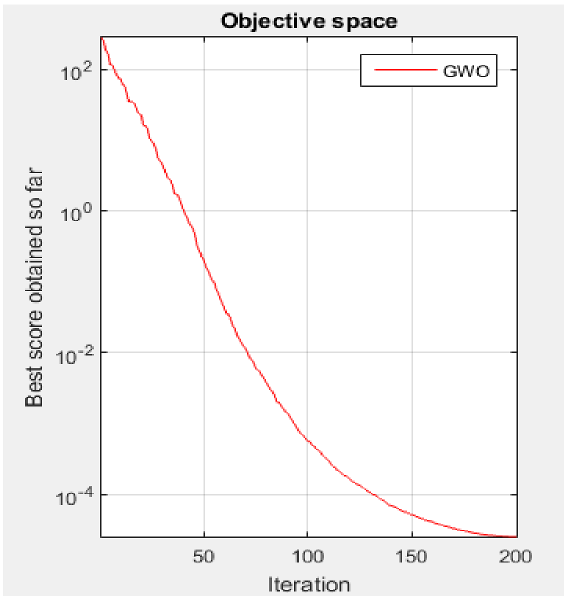

Fig 9MP equals 2 by using GWO

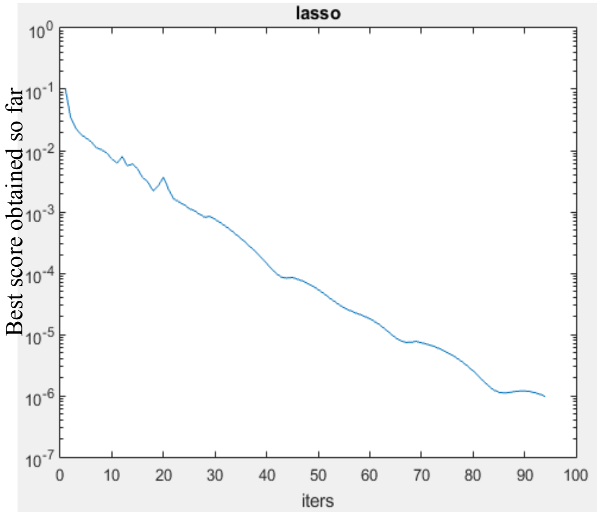

Fig 10MP equals 2.5 by using the proximal gradient

Fig 11MP equals 2.5 by GWO

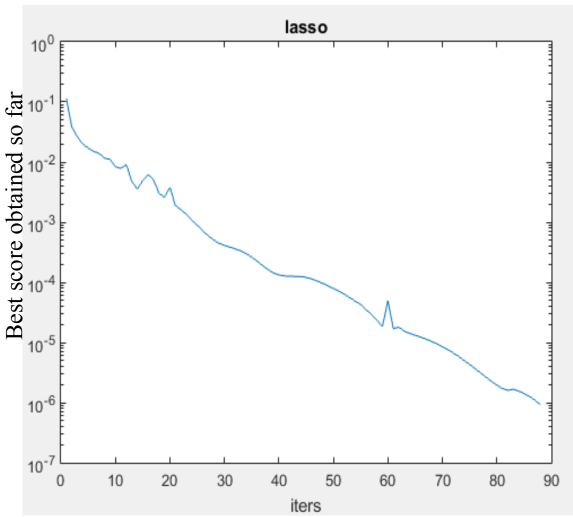

Fig 12MP equals 3 by using the proximal gradient

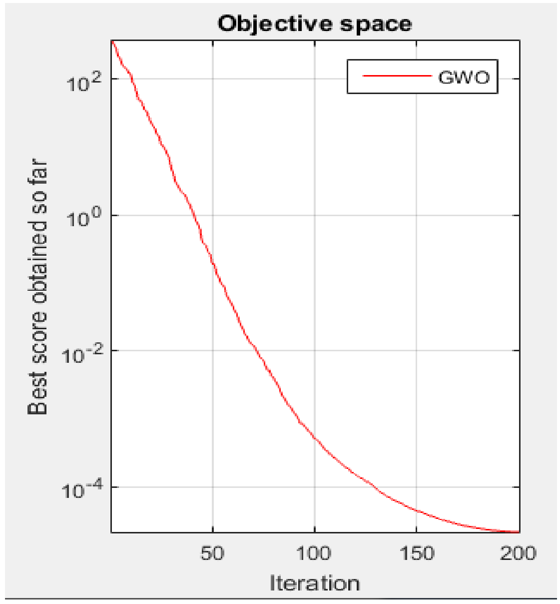

Fig 13MP equals 3 by GWO

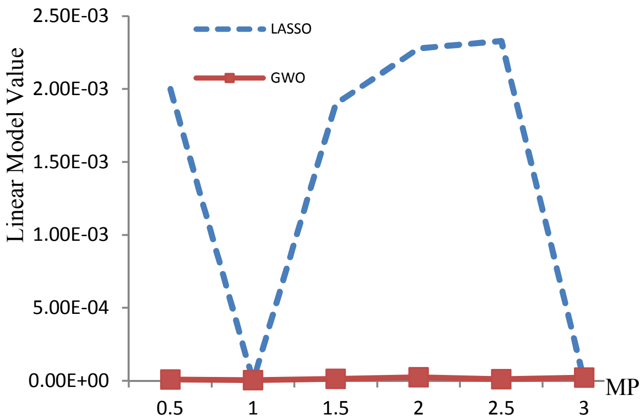

Fig 14 Comparison between the proximal gradient GWO at different values of MP

VI Conclusion

Genetic models depict how traits affect phenotypes. The technique of selection of the feature is a basic procedure that must be linked in the genetic model to study the effect of these qualities. Previous research has used the LASSO method to improve the method of feature selection. An improvement account is introduced to address these challenges. The main objective of this research is to provide a precise identification of the components in the genetic models using the calculation of accurate rationalization and efficiency entitled Gray Wolf Optimization GWO. The results indicate that GWO has additional performance efficiency compared to LASSO by up to E + 02.

References

[1] Julie Hamon. "Optimisation combinatoire pour la

sélection de variables en régression en grande dimension : Application en génétique animale. 2013, " Applications [stat.AP]. Université des Sciences et Technologie de Lille - Lille I,Français.
[2] T. M. Phuong, Z. Lin et R. B. Altman. "Choosing SNPs using feature selection." 2005, Proceedings / IEEE Computational Systems Bioinformatics Conference, CSB., pages 301-309,. PMID 16447987
[3] Jin Xiao,Yi Xiaoet al."Feature-selection-based dynamic transfer ensemble model for customer churn prediction", April 2014, Knowledge and Information Systems 43(1):29-51, DOI: 10.1007/s10115-013-0722-y.
[4] Aytuğ Onan and Serdar Korukoğlu,"A feature selection model based on genetic rank aggregation for text sentiment classification" First Published November 5, 2015, Journal of Information Science, Vol 43, Issue 1, pp. 25 - 38, https://doi.org/10.1177/0165551515613226
[5] Alexej Gossmann , Shaolong Cao,etal."A Sparse Regression Method for Group-Wise Feature Selection with False Discovery Rate Control", 2018 Jul-Aug, IEEE/ACM Trans Comput Biol Bioinform.;15(4):1066-1078.doi: 10.1109/TCBB.2017.2780106. Epub 2017 Dec 6.
[6]Seyed ali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, "Grey Wolf Optimizer", March 2014, Advances in Engineering Software,Volume 69,Pages 46-61.