Binary Enhanced Gray Wolf Optimization Approach For Feature Selection

MINOR PROJECT 2

Submitted by:

Suhani Kohli (9916103215)
Kashish Chugh (9916103164)
Manika Kaushik (9916103099)

Under the supervision of
Mr. Avinash Pandey

Department of CSE
Jaypee Institute of Information Technology University, Noida

MAY 2019

ACKNOWLEDGEMENT

We would like to place on record our deep sense of gratitude to Mr. Avinash Pandey, Assistant Professor at, Jaypee Institute of Information Technology, India for his generous guidance, help and useful suggestions.

We express our sincere gratitude to the faculty, Dept. of Computer Science, India, for his/her stimulating guidance, continuous encouragement and supervision throughout the course of present work.

We also wish to extend our classmates for their insightful comments and constructive suggestions to improve the quality of this project work.

Signature(s) of Students

Manika Kaushik (9916103099)

Suhani Kohli (9916103215)

Kashish Chugh (9916103164)

Signature of Mentor

Mr. Avinash Pandey

ABSTRACT

Nowadays, as more and more exploration in the data extraction techniques, data is represented using hundreds of features with very high dimension. Some of these features are redundant in nature while some does not contribute to the effective solution. Feature selection is a process used to select a subset of features from the extracted features or target data. These selected features can efficiently describe the input data while reducing effects from noise or irrelevant variables. For electing optimal features from a feature vector sized N, there will be 2^N possible solutions which is a huge space of features. Since, evaluating 2^N subsets becomes a np-hard problem, traditional methods of feature selection do not perform well. A better way to find suboptimal subsets is by employing search algorithms which find a subset heuristically. The Swarm Intelligence subclass of heuristic strategies impersonates the conduct of biological and physical frameworks in the nature and it has been proposed as solid techniques for global improvements. One such bio-inspired optimization technique is the Grey Wolf Algorithm which stimulates the leadership hierarchy and hunting mechanism of grey wolves in nature and is used to select optimal feature subset for classification purposes.

Therefore, a variant of GWO is introduced using Lévy flight . The proposed method uses Lévy flight for enhancing the exploration rate of grey wolf optimization method to overcome the problem of being trapped in a local minima/maxima. The performance of variant is tested on 14 benchmark functions. Then this variant is used to create an enhanced binary metaheuristic feature selection method bGWOLF is introduced which is based on grey wolf optimization and Lévy flight. The performance of the proposed binary feature selection method has been tested on the 5 feature selection benchmark datasets taken from UCI repository and compared with, binary particle swarm optimization, binary grey wolf optimization, Genetic algorithm. Statistical analysis and Experimental results validate the efficacy of proposed method.

LIST OF FIGURES

Figure Title **Page**

2.1 Social hierarchy of grey wolves……………………………………………..3

2.2 2D and 3D position vectors and their possible next positions…………………4

2.3 Position updating in GWO……………………………………………….....4

2.5 Lévy Flight……………………………………………………………..6

4.1 Work Flow………………………………………………………….......8

5.2 bGWO-LF on UCI datasets……………………………………………….12

6.1 Box plots for all the considered methods and proposed method GWO-LF……15

7.1 Gantt Chart……………………………………………………………..18

LIST OF TABLES

Table Title **Page**

4.1 Accuracy Of different classifiers with and

without Feature Selection ……………………...5

6.1 Benchmark Functions……………………………………………..……........14

6.2 Parameter settings for experiments………………………..……………………14

6.3 Mean Fitness and standard deviation of benchmark functions

on different algorithms…………………………………………………….……..15

6.4 Dataset Description…………………………………………………………….16

6.5 Average accuracy by different classifiers on datasets with and

without Feature Selection……………………………………………...…...17

ABBREVIATIONS

GWO	GREY WOLF OPTIMIZATION
FS	FEATURE SELECTION
bGWO	BINARY GREY WOLF OPTIMIZATION

NOMENCLATURE

GWO-LF	Variant of Grey Wolf Optimizer Using Lévy Flight
bGWO-LF	Enhanced binary version of Grey Wolf Optimizer

Contents

Acknowledgement………………………………..*i*

Abstract……………..……………………………………………………………………………...... *ii*

List of Figures…………………………………………………………………..…………………....*iii*

List of Tables……………………………………………………………………………………..... *iii*

List of Abbreviations……………………………………………………………………………....*iii*

List of Nomenclature………………………………………………………………………………....*iii*

CHAPTER 1- INTRODUCTION...1

1.1 GENERAL...1

1.2 OBJECTIVE...1

1.3 MOTIVATION...1

CHAPTER 2- BACKGROUND STUDY...2

2.1 GREY WOLF OPTIMIZER...3

2.2 LÉVY FLIGHT……………………………………………………………………..6.

CHAPTER 3 -REQUIREMENT ANALYSIS...7

3.1 HARDWARE REQUIREMENTS...7

3.2 SOFTWARE REQUIREMENTS...7

3.2.1 SOFTWARE USED...7

3.2.2 LIBRARIES USED...7

3.3 FUNCTIONAL REQUIREMENTS...7

3.4 NON-FUNCTIONAL REQUIREMENTS...7

CHAPTER 4-DETAILED DESIGN...8

4.1 PHASE 1...9

4.2 PHASE 2...9

CHAPTER 5- IMPLEMENTATION...11

CHAPTER 6- TESTING...14

CHAPTER 7- CONCLUSION AND FUTURE SCOPE...18

7.1 CONCLUSION...18

7.2 FUTURE SCOPE...18

CHAPTER 8- REFERENCES………………………………………………………………19

CHAPTER 1- INTRODUCTION

1.1 GENERAL

Feature selection is a process for identifying the important features and removing irrelevant (redundant) ones from the dataset. The feature selection objectives are data dimensionality reduction, improving prediction performance, and good data understanding for different machine learning applications. Moreover, the relevant (interdependence) features have an influence on the output and contain important information that will be obscure if any of them is excluded. Previously, an exhaustive search for the optimal set of features (attributes) in a high dimensional space may be unpractical. Various heuristic techniques mimic the behavior of biological and physical systems in the nature and it has been proposed as strong methods for global optimizations, one of them is grey wolf optimization. The binary version of this optimization algorithm can be used to select relevant features from the gamut of features.

1.3 OBJECTIVE

The objective of our project is to create an enhanced binary version of the grey wolf optimization to find optimal regions of the complex search space. The aim of our project is to:-

- To generate a new variant of GWO
- Use the new variant of GWO to introduce an enhanced binary version of GWO to further improve results.
- To use enhanced binary version of GWO for feature selection.

1.4 MOTIVATION

The major motivating factor for the creation of this project is the restriction behind classical techniques in solving the problems, for a feature vector sized N the different feature reduction would be $\mathbf{2}^{\boldsymbol{N}}$ which is a huge space of features. Since, evaluating $\mathbf{2}^{\boldsymbol{N}}$ subsets become np-hard problem, suboptimal subsets are found by employing search algorithms which find a subset heuristically. It becomes computationally intensive while searching so the evolutionary computation (EC) algorithms are the alternative for solving these limitations and searching for the optimum solution of the problems. Evolutionary computation (EC) algorithms are inspired from nature, social behaviour, and biological behaviour of (animals, birds, fish, bat, firefly, wolves, etc.) in a group. Many researchers have proposed different computational methods, in order to mimic the behaviour of these species to seek for their food (optimal solution). But according to no free lunch theorem, a general-purpose, universal optimization strategy is impossible. Therefore, a single optimization technique cannot have a generalized solution [9].

CHAPTER 2- BACKGROUND STUDY

In research paper [1], the GWO algorithm has been discovered where they have discussed the approach of this algorithm to handle optimization problems and have talked in much detail about leadership hierarchy and hunting mechanism of grey wolves in nature. The paper also presents a real application of the proposed method in the field of optical engineering. The results of the classical engineering design problems and real application prove that the proposed algorithm is applicable to challenging problems with unknown search spaces.

In [2] ,The main objective of this paper is applying a comparison between the Wolf Pack Search (WPS) as a newly introduced intelligent algorithm with several other known algorithms including Particle Swarm Optimization (PSO), Shuffled Frog Leaping (SFL), Binary and Continues Genetic algorithms. All algorithms are applied on two benchmark cost functions. The aim is to identify the best algorithm in terms of more speed and accuracy in finding the solution, where speed is measured in terms of function evaluations. The simulation results show that the SFL algorithm with less function evaluations becomes first if the simulation time is important, while if accuracy is the significant issue, WPS and PSO would have a better performance.

In research paper [4], a novel binary version of the grey wolf optimization (GWO) is proposed and used to select optimal feature subset for classification purposes. The binary version introduced here is performed using two different approaches. In the first approach, individual steps toward the first three best solutions are binarized and then stochastic crossover is performed among the three basic moves to find the updated binary grey wolf position. In the second approach, sigmoidal function is used to squash the continuous updated position, then stochastically threshold these values to find the updated binary grey wolf position. The two approaches for binary grey wolf optimization (bGWO) are hired in the feature selection domain for finding feature subset maximizing the classification accuracy while minimizing the number of selected features. The proposed binary versions were compared to two of the common optimizers used in this domain namely particle swarm optimizer and genetic algorithms

In research paper [5], they have discussed Wrapper methods to check the applicability of feature selection techniques. It uses the predictor as a black box and the predictor performance as the objective function to evaluate the variable subset.

Levy Flight that resembles food searching path of many animals like albatross, bumblebees and deer was added to nature-inspire algorithms to ensure improvement of the algorithms. In research paper [10], Xin-She Yang and Suash Deb used Levy Flight distribution to create new cuckoo in Cuckoo Search.

In [11], PSO is combined with Levy flight in this study to get rid of local minima and improve global search capability. The performance and accuracy of the proposed method called as Levy flight Particle Swarm Optimization (LFPSO) are examined on well-known unimodal and multimodal benchmark functions.

5.1 GRAY WOLF OPTIMIZER

Grey wolves are considered as apex predators, meaning that they are at the top of the food chain. Grey wolves mostly prefer to live in a pack. The group size is 5–12 on average. Of particular interest is that they have a very strict social dominant hierarchy as shown in Fig. 1. According to [1] grey wolf pack consists of the following:

1. The alpha wolf is also called the dominant wolf since his/her orders should be followed by the pack. The alphas decisions are dictated to the pack ´

2. The betas are subordinate wolves that help the alpha in decision making or other activities. The beta can be either male or female, and he/she is probably the best candidate to be the successor alpha.

3. The deltas have to submit alphas and betas, but they dominate the omega. Scouts, sentinels, elders, hunters, and caretakers belong to this category.

In addition to the social hierarchy of wolves, group hunting is another interesting social behaviour of grey wolves. According to Muro et al. [5] the main phases of grey wolf hunting are: Tracking the prey, pursuing, attack the prey.

This hunting technique and the social hierarchy of grey wolves are mathematically modelled in order to design GWO and perform optimization**.**

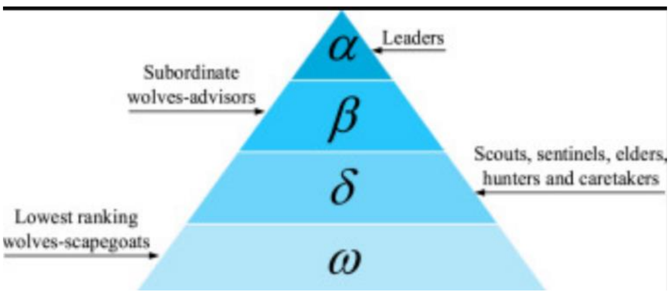


Fig 2.1: Social hierarchy of grey wolves

MATHEMATICAL REPRESENTATION OF GWO

In the mathematical model for the GWO the fittest solution is called the alpha (α). The second and third best solutions are named beta (β) and delta (δ) respectively. The rest of the candidate solutions are assumed to be omega (ω). The hunting is guided by α, β, δ, and ω follow these three candidates. In order for the pack to hunt a prey they first encircle it.

In order to mathematically model encircling behaviour the following equations (1), (2), (3), and (4) are used.

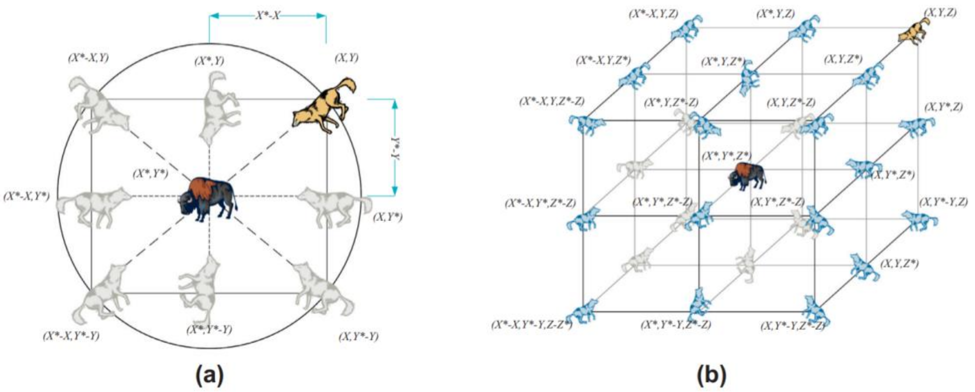


Fig 2.2: 2D and 3D position vectors and their possible next positions.

$$\vec{\mathbf{X}}(\mathbf{t}+\mathbf{1})=\vec{\mathbf{X}}_{\mathbf{p}}(\mathbf{t})+\vec{\mathbf{A}}.\vec{\mathbf{D}}, \quad (1)$$

Where $\vec{\boldsymbol{D}}$ is as defined in equation (2) t is the iteration number, $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{C}}$ are coefficient vectors, $\vec{\boldsymbol{X}}$p is the prey position, and $\vec{\boldsymbol{X}}$ is the gray wolf position.

$$\vec{\boldsymbol{D}}=|\vec{\boldsymbol{C}}.\vec{\boldsymbol{X}}_{\boldsymbol{p}}(\boldsymbol{t})-\vec{\boldsymbol{X}}(\boldsymbol{t})| \quad (2)$$

The $\vec{\boldsymbol{A}}$, $\vec{\boldsymbol{C}}$ vectors are calculated as in equations (3) and (4).

$$\vec{\boldsymbol{A}}=2\text{a}.\overrightarrow{\boldsymbol{r1}_{\mathbf{1}}}-\text{a} \quad (3)$$

$$\vec{\boldsymbol{C}}=2\vec{\boldsymbol{r}}_{\mathbf{2}} \quad (4)$$

Where ‘a’ is linearly decreased from 2 to 0 over the course of iterations, and r1, r2 are random vectors in [0, 1].So the updating for the wolves positions is as in equation (5).

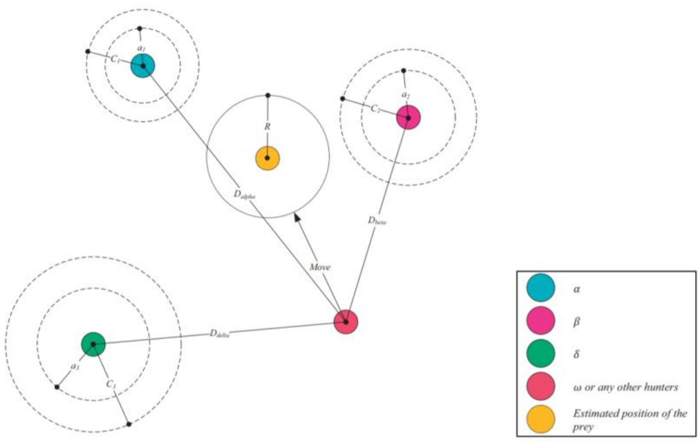


Fig 2.3: Position updating in GWO

$$\vec{\boldsymbol{X}}(\text{t}+1)=(\vec{\boldsymbol{X}}_1+\vec{\boldsymbol{X}}_2+\vec{\boldsymbol{X}}_3)/3 \quad (5)$$

Where $\vec{\boldsymbol{X}}$1, $\vec{\boldsymbol{X}}$2, $\vec{\boldsymbol{X}}$3 are defined as in equations (6), (7), and (8) respectively

$$\vec{\boldsymbol{X}}_1=|\vec{\boldsymbol{X}}_\alpha-\vec{\boldsymbol{A}}_1.\vec{\boldsymbol{D}}_\alpha| \quad (6)$$

$\vec{X}_2 = | \vec{X}_\beta - \vec{A}_2 . \vec{D}_\beta |$ (7)

$\vec{X}_3 = | \vec{X}_\delta - \vec{A}_3 . \vec{D}_\delta |$ |(8)

where $\vec{X}_\alpha$, $\vec{X}_\beta$, $\vec{X}_\delta$ are the first three best solutions in the swarm at a given iteration t, $\vec{A}_1$, $\vec{A}_2$, $\vec{A}_3$ are defined as in equation (3), and $\vec{D}_\alpha$, $\vec{D}_\beta$, $\vec{D}\gamma$ are defined using equations (9), (10), and (11) respectively.

$\vec{D}_\alpha = |\vec{C}_1 . \vec{X}_\alpha - \vec{X}|$ (9)

$\vec{D}_\beta = |\vec{C}_2 . \vec{X}_\beta - \vec{X}|$ (10)

$\vec{D}_\delta = | \vec{C}_3 . \vec{X}_\delta - \vec{X}|$ (11)

Where $\vec{C}_1$, $\vec{C}_2$, and $\vec{C}_3$ are defined as in equation (4). A final remark about the gray wolf optimizer (GWO) is the updating of the parameter 'a' that controls the tradeoff between exploration and exploitation. The parameter 'a' is linearly updated in each iteration to range from 2 to 0 according to the equation (12).

$$a = 2 - t\frac{2}{MaxIter} \quad (12)$$

where 't' is the iteration number and MaxIter is the total number of iteration allowed for the optimization. Algorithm 1 outlines the continuous grey wolf optimization (CGWO) algorithm.

To see how GWO is theoretically able to solve optimization problems, some points may be noted:

- The proposed social hierarchy assists GWO to save the best solutions obtained so far over the course of iteration.
- The proposed encircling mechanism defines a circle-shaped neighbourhood around the solutions which can be extended to higher dimensions as a hyper-sphere.
- The random parameters A and C assist candidate solutions to have hyper-spheres with different random radii.
- The proposed hunting method allows candidate solutions to locate the probable position of the prey.
- Exploration and exploitation are guaranteed by the adaptive values of 'a' and A.
- The adaptive values of parameters 'a' and A allow GWO to smoothly transition between exploration and exploitation.
- With decreasing A, half of the iterations are devoted to exploration ($|A| >= 1$) and the other half are dedicated to exploitation ($|A| < 1$)

- Set the initial population of grey wolf population X i(i = 1, 2, 3, , n)
- Initialize a, A, and R
- Figure out the fitness of each wolf
  Xα is the best wolf
  Xβ is the next best wolf
  Xδis the next to next best wolf
- t = 1
- while t <= MaxIter do
- for each wolf
- Change the position of the current best wolf using hunting location update equation
- end for
- Update a, A and R
- Compute the fitness of all the individual
- Update Xα, Xβ, Xδ
- t=t+1
- end while

Fig 2.4: Algorithm of GWO

2.2 LÉVY FLIGHT

Lèvy flights, also referred to as Lèvy motion, named for French mathematician Paul Lèvy, is a class of non-Gaussian random processes, the step-lengths of whose random walk have a probability distribution that is heavy-tailed. This distribution is a simple power law formula given in equation

$$\textbf{Lèvy}(\beta) \sim |t|^{-1-\beta} \sim \frac{u}{|v|^{\frac{1}{\beta}}} \quad , \quad 0 < \beta \leq 2 \tag{13}$$

It is observed that after numerous steps, the distance from the origin of the random walk tends to a stable distribution. Various studies have shown that the flight demeanour of many insects and animals exhibits the typical features of Lèvy flights [28]. By and large, the rummaging way of a creature is viably an arbitrary stroll since the next move is based on both the current location/state and the transition probability to the following location. The selected direction implicitly relies on a probability, which can be modelled mathematically.

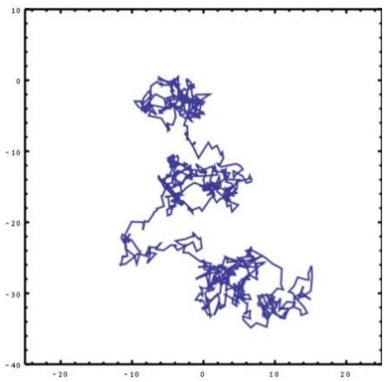


Fig 2.5 : Lévy Flight

CHAPTER 3 -REQUIREMENT ANALYSIS

3.1 Hardware Requirements:

- Ram- 4 GB
- Processor- Intel core i-3 (2GHz)
- Windows XP/7/8/8.1/10/linux/mac

3.2 Software Requirements:

3.2.1 Software Used

- Anaconda(Jupyter notebook, Spyder)
- NLTK(Natural Language Toolkit)(preprocessing)
- Linguistic Inquiry and Word Count (LIWC 2015)

3.2.2 Libraries Used

- PANDAS
- NUMPY
- MATPLOTLIB
- SKLEARN

3.3 Functional Requirements

- Improve performance of GWO
- Enhance bGWO
- Use bEGWO for better feature selection

3.4 Non-Functional Requirements

- Reduce computational time
- Improve fitness of search agents

CHAPTER 4-DETAILED DESIGN

The success of any meta-heuristic method relies upon the balance between exploration and exploitation. The GWO method generally suffers from the risk of confining into local optima owing to the lack of exploration in the wolves for some cases [1]. In this work, a new variant of GWO is proposed by introducing some randomization using Levy Flight. This new variant is validated by benchmarking it against 14*** functions and comparing its mean fitness against various algorithms. Then this model is described to build an enhanced version of binary grey wolf optimization (bGWO-LF) for the feature selection task that would maximize accuracy along with minimizing the considered features. Further, this optimizer is used to select the best features from the numerous features extracted from UCI repository datasets in such a way that would maximize accuracy along with minimizing the considered features.

The performance of the proposed binary feature selection method has been tested on the **** feature selection benchmark datasets taken from UCI repository and compared with binary particle swarm optimization, binary grey wolf optimization, Genetic Algorithm. Statistical analysis and Experimental results validate the efficacy of proposed method.

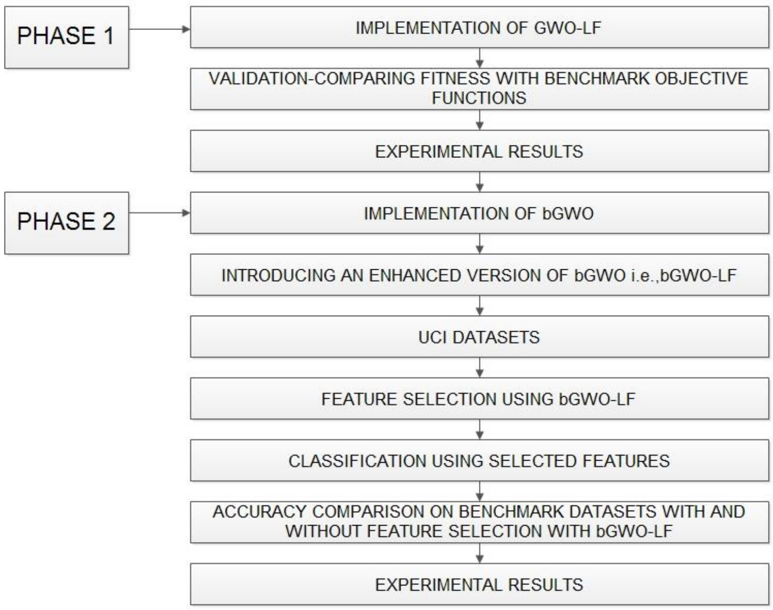


Fig 4.1: Workflow

Figure 4.1 shows the workflow of the project and describes it. In which phase 1 will result in an optimized version of the standard GWO by overcoming its limitations and in the other phase an enhanced binary version of GWO will be generated and then further used for feature selection on some dataset and these selected features are fed to the classifier to check the applicability of the constructed model.

The project is divided into two phases as shown in Figure 4.1

4.1 PHASE 1-

1. IMPLEMENTATION OF GWO-LF:

The algorithm consists of two parts- exploration (searching for prey) and exploitation (attacking prey). In the standard GWO we see the problem of local minima arises due to accelerated exploitation because of which premature exploitation takes place. Hence, the solution does not provide the global minimum solution. To overcome this problem, a new variant of GWO is be implemented in this project which introduces some randomization and enhance exploration using Lévy Flight.

2. VALIDATION COMPARING FITNESS WITH BENCHMARK OBJECTIVE FUNCTIONS:

GWO-LF is tested on 14 benchmark functions by evaluating their mean fitness obtained using each benchmark function.

3. EXPERIMENTAL RESULTS:

The comparative analysis of GWO-LF performance with nature inspired algorithms provide us the information how well the variant is performing compared to other algorithms under standard conditions.

4.2 PHASE 2-

1. IMPLEMENTATION OF bGWO:

A novel binary grey wolf optimization (bGWO) is proposed for the feature selection task. The wolves updating equation is a function of three position vectors namely $x\alpha$, $x\beta$, $x\delta$ which attracts each wolf towards the first three best solutions. In the bGWO, the pool of solutions is in binary form at any given time; all solutions are on the corner of a hypercube. To update the positions of a given wolf according to the CGWO principle, while keeping the binary restrictions following any of the two approaches given [3].

2. INTRODUCING AN ENHANCED VERSION OF bGWO-LF:

The variant of GWO, i.e., GWO-LF is converted to bGWO-LF and thus we have an enhanced version of binary GWO.

3. UCI DATASETS:

We have considered 5 standard datasets from UCI repository to test our model.

4. FEATURE SELECTION USING bGWO-LF:

Feature selection also called as attribute selection or variable subset selection is a process of selecting appropriate features with respect to target data. Feature selection is important since it:

1. Removes redundant data
2. Select attributes that are significant.
3. Reduces chances of over fitting.
4. Reduce training time.

5. CLASSIFICATION USING SELECTED FEATURES:

Classification models are used as the objective function for calculating the accuracy and other performance measures with and without optimal set of features on various datasets which guides the way for comparison. A model or a classifier is constructed to predict the categorical labels. We will implement certain base-line classification models on our dataset.

6. ACCURACY COMPARISON ON BENCHMARK DATASETS:

bGWO-LF is tested on 5 feature selection benchmark datasets where it selects best features out of the considered dataset and is fed to a classifier and then accuracy is calculated with and without considering the best features.

7. EXPERIMENTAL RESULTS:

The comparative analysis of bGWO-LF performance with nature inspired algorithms provide us the information how well the variant is performing, how it has enhanced the feature selection results compared to other algorithms under standard conditions.

CHAPTER 5- IMPLEMENTATION

5.1 Hybrid GWO-LF method

The algorithm consists of two parts- exploration (searching for prey) and exploitation (attacking prey). In the standard GWO we see the problem of local minima arising due to accelerated exploitation because of which premature exploitation takes place. Hence, the solution does not provide the global minimum solution. To overcome this, Lévy flight algorithm has been used alongside to create the balance between intensification and diversification. A variant of GWO, i.e., GWO-LF, is implemented in this project which will introduce some randomization on first best agent. This will enhance exploration.

Furthermore, the proposed GWO-LF method is described for the feature selection problem. The robustness of grey wolf optimization and Lévy flight are used to select the most adequate features from the high dimensional data set. In standard GWO method, wolves ceaselessly change their areas to any point in the space utilizing chasing position update condition. The positions of continuous GWO method are further refined using proposed GWO-LF method. Notwithstanding, in feature selection problem, the search space is discrete (0 and 1), while the GWO-LF technique has been proposed to tackle continuous optimization issue. Along these lines, keeping this limitation binary form of GWO-LF has been created.

$$\vec{\boldsymbol{X}}(\boldsymbol{t}+\mathbf{1}) = \vec{\boldsymbol{X}}_{\boldsymbol{p}}(\boldsymbol{t}) + \boldsymbol{\alpha} \otimes \mathbf{Levy}\ (\boldsymbol{\beta}) \tag{14}$$

Where, α > 0 is the step size which should be related to the scales of the problem of interests. In most cases, we can use α = 1. The product ⊗ means entry-wise walk while multiplications. Lévy flights essentially provide a random walk while their random steps are drawn from a Lévy Distribution for large steps.

$$\boldsymbol{\alpha} \otimes \mathbf{L\grave{e}vy}\ (\boldsymbol{\beta}) = \mathbf{0.01} * \mathbf{u} * |\boldsymbol{v}|^{\mathbf{1/\beta}} \left(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{t}} - \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{alpha}}\right) \ ; \quad \mathbf{0} < \boldsymbol{\beta} \leq \mathbf{2} \tag{15}$$

Where, u and v are drawn from normal distribution. That is,

$$\boldsymbol{u} \sim \boldsymbol{N}\left(\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}\right), \boldsymbol{v} \sim \boldsymbol{N}\left(\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{v}}^{\mathbf{2}}\right) \tag{16}$$

$$\boldsymbol{\sigma}_{\mathbf{u}} = \left\{\frac{\boldsymbol{\Gamma}(\mathbf{1}+\boldsymbol{\beta}) \sin\frac{\boldsymbol{\pi\beta}}{\mathbf{2}}}{\boldsymbol{\Gamma}\left[\frac{(\mathbf{1}+\boldsymbol{\beta})}{\mathbf{2}}\right]\boldsymbol{\beta}\mathbf{2}^{\frac{\boldsymbol{\beta}-\mathbf{1}}{\mathbf{2}}}}\right\}, \boldsymbol{\sigma}_{\boldsymbol{v}} = \mathbf{1} \tag{17}$$

Here Γ is the standard Gamma function.

The proposed binary methodology has been discussed below.

5.2 Enhanced binary GWO (bGWO-LF)

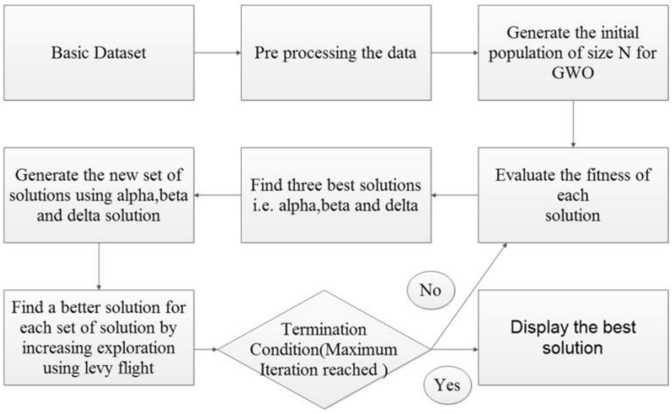


Fig 5.1 bGWO-LF on UCI dataset

In GWO, positions are updated to any point in space. But, whenever it comes to problems like feature selection, which has a binary outcome i.e. either a feature will be selected(1) or not selected(0), continuous GWO does not perform well. Therefore, a binary version of GWO was created **** [cite binary emary]. The wolves updating equation is a function of three position vectors namely Yα, Yβ, Yδ which attracts each wolf towards the first three best solutions. In the bGWO, the pool of solutions is in binary form at any given time. All solutions are confined to the corners of a hypercube. In this paper, a binary enhanced grey wolf optimization (bGWOLF) has been developed which aides in selecting optimum features. In the hybrid GWO-LF model, Lévy flight method has been used to find the global best solution. This process improves the exploration capability of GWO method.

$$\vec{X}(t+1) = \frac{\vec{X}_{rand1}+\vec{X}_{rand2}+\vec{X}_{rand3}}{3} \quad (18)$$

The best solution obtained using GWOLF have continuous values. Therefore, sigmoid function as given in eq. 18 is used to convert continuous values to discrete values (0 and 1).

$$X(t+1) = \begin{cases} 1 \text{ if sigmoid } (\mathrm{di}) \geq \text{ rand} \\ 0 \text{ otherwise} \end{cases} \quad (19)$$

Where, rand is random number obtained from uniform distribution and $\vec{X}$(t+ 1) is the updated binary position vector at last iteration t. The sigmoid(s) function is defined in the following eq. (19).

$$Sigmoid(s) = \frac{1}{1+e^{-10(x-0.5)}} \quad (20)$$

- Generate the initial population of grey wolves Xi(i = 1, 2, 3, n)
- Initialize a, A, and C
- Evaluate the fitness of each individual using benchmark functions
  - Xα is the best wolf
  - Xβ is the next best wolf
  - Xδ is the next to next best wolf
- t = 1
- Initialize the initial population of grey wolves Xi (i = 1, 2, 3, n) randomly between 0 and 1 .
- Calculate the fitness of these ‘n’ solutions by using fitness function
- while t <= MaxIter do
  - for each wolf
  - Update the location of the current agent using hunting position update formula of GWO-LF
  - Convert continuous positions to binary using equation 18, 19, 20.
  - end for
- Update a, A and C
- Compute the fitness of all the individual
- Update Xα, Xβ, Xδ
- t=t+1
- end while

Fig 5.2: Enhanced binary GWO algorithm

CHAPTER 6- TESTING

The GWO-LF algorithm is benchmarked on 14 benchmark functions. Despite the simplicity, we have chosen these test functions to be able to compare our results to those of the current meta-heuristics. These benchmark functions are listed in Table 6.1

Table 6.1: Benchmark Functions

S.No	Function Name	Function Equation	Type
1	Sphere	F(x)= $\sum_{i=1}^{d} x_i^2$	Uni-modal
2	Schwefe13	F(x)= $\sum_{i=1}^{d} \lvert x_i \rvert + \prod_{i=1}^{d} \lvert x_i \rvert$	Uni-modal
3	Schwefel4	F(x)=max $\{\lvert x_i \rvert, 1 \leq i \leq n\}$	Uni-modal
4	Rosenbrock	F(x)= $\sum_{i=1}^{d-1}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	Uni-modal
5	Quartic	F(x)= $\sum_{i=1}^{d} i x_i^4$	Multi-modal
6	New Schwefel	F(x)=$\sum_{i=1}^{n} -x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$	Multi-modal
7	Rastrigin	F(x)= 10d+$\sum_{i=1}^{d}(x_i^2 - 10\cos(2\pi x_i))$	Multi-modal
8	Ackley	F(x)=-20$exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$	Multi-modal
9	Griewank	F(x)=$\frac{1}{4000} + \sum_{i=1}^{d} x_i^2 - \prod_{i=1}^{d}\cos(\frac{x_i}{\sqrt{i}}) + 1$	Multi-modal
10	Penalty1	F(x)= $10\sin^2(\pi y_1) + \sum_{i=1}^{d}(y_i - 1)^2[1 + 10\sin^2 \pi y_{i+1}] + (y_d - 1)^2 + \sum_{i=1}^{d} u_i$	Multi-modal
11	Penalty 2	F(x)=$0.1\{\sin^2 3\pi x_1 + \sum_{i=1}^{n}(x_i - 1)^2[1 + \sin^2(3\pi x_i + 1)] + (x_n - 1)^2[1 + \sin^2 2\pi x_n]\} + \sum_{i=1}^{n} u(x_i, 5,100,4)$	Multi-modal
12	Alpine	F(x)=$\sum_{i=1}^{d}\lvert x_i \sin(x_i) + 0.1 x_i \rvert$	Multi-modal
13	Brown	F(x)=$\sum_{i=1}^{d-1}(x_i^2)^{(x_{i+1}^2+1)} + (x_{i+1}^2)^{(x_i^2+1)}$	Uni-modal
14	Powell's Second Singular	F(x)=$\sum_{i=1}^{d-1}(x_{i-1} + 10x_i)^2 + 5(x_{i+1} - x_{i+2})^2 + (x_i - 2x_{i+1})^4 + 10(x_{i-1} - x_{i+2})^4$	Uni-modal

The GWO algorithm was run 30 times on each benchmark function. For verifying the results, the GWO algorithm is compared to PSO as an SI-based technique and GSA as a physics-based algorithm. In addition, the GWO algorithm is compared with three EAs: DE and standard GWO by comparing average fitness and standard deviation.

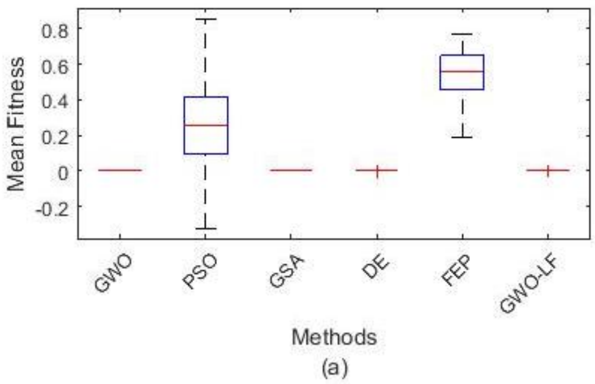
Table 6.2: Parameter Settings for experiments

S.no	Parameter	CS	PSO	GWO	WoaSA	BA	GWOLF
1.	No. of search agents	50	50	50	50	50	50
2.	Probability	0.25	-	-	-	-	-
3.	Step scaling factor (α)	0.01	-	-	-	-	-
4.	Number of iterations	1000	1000	1000	1000	1000	1000
5.	Search domain	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]	[0,1]
6.	Number of repetitive runs	30	30	30	30	30	30
7.	Cognitive constant	-	2	-	-	-	-
8.	Social constant	-	2	-	-	-	-
9.	Inertia weight	-	0.8	-	-	-	-
10.	Crossover rate	-	-	-	-	-	-
11.	Mutation rate	-	-	-	-	-	-
12.	α-parameter	-	0.99	-	-	-	0.1
13.	β-parameter		0.01	-	-	-	0.5

Table 6.3: Mean Fitness and standard deviation of benchmark functions on different algorithms

	GWO		PSO		GSA		DE		GWO-LF	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
F1	5.71E-70	1.79E-21	0.000136	0.000202	2.53E-16	9.67E-17	8.2E-14	5.9E-14	2.15E-70	2.16E-21
F2	6.38E-40	1.18E-12	0.042144	0.045421	0.055655	0.194074	1.5E-09	9.9E-10	9.04E-40	3.92E-13
F3	2.41E-17	2.23E-05	1.086481	0.317039	7.35487	1.741452	0	0	1.54E-17	27.45E-05
F4	26.5725793	0.52217	96.71832	60.11559	67.54309	62.22534	0	0	26.3551	0.539162
F5	0.0005531	0.002932	0.122854	0.044957	0.089441	0.04339	0.00463	0.0012	0.00058	0.002803
F6	-1261.2999	137.045	-4841.29	-1152.81	-2821.07	493.0375	-11080.1	574.7	-1267.03	130.2414
F7	0.3162099	5.32837	46.70423	11.62938	25.96841	7.470068	69.2	38.8	0.15143	4.997720
F8	20.833474	0.08443	0.276015	0.50901	0.062087	0.23628	9.7E-08	4.2E-08	20.28628	0.065616
F9	0.0011054	0.01407	0.009215	0.007724	27.70154	5.040343	0	0	0.000742	0.0204229
F10	0.0740678	0.161815	0.006917	0.026301	1.799617	0.95114	7.9E-15	8E-15	0.08857	0.1483137
F11	0.28169102	0.302158	0.006675	0.008907	8.899084	7.126241	5.1E-14	4.8E-14	0.2440262	0.2631101
F12	2.15E-16	0.9823	3.63	4.75E-01	3.75	1.06E-03	6.82	2.11	1.66E-62	3.1140824
F13	0.0272819	4.16E-22	3.85E-09	7.29E-05	2.11E-04	3.75E-13	1.30E-03	2.42E-09	9.12E-26	0.0887091
F14	1.92E-35	5.01E-09	1.12E-01	3.87E-01	1.03E-01	3.87E-01	1.78E-01	2.25E-01	9.35E-17	2.25E-09

According to the results of Table 6.3, GWO-LF is able to provide very competitive results. This algorithm outperforms all others in 10 benchmark functions including which are uni-modal functions.



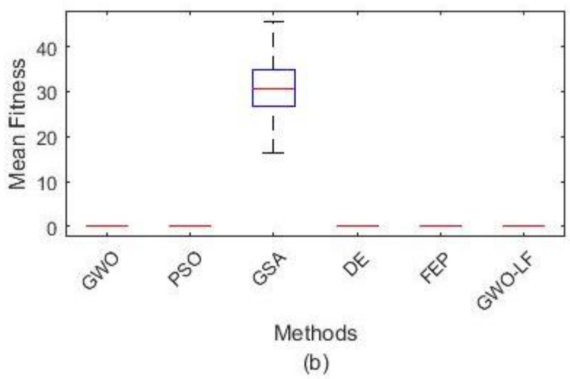


Fig 6.1: Box plots for all the considered methods and proposed method GWO-LF

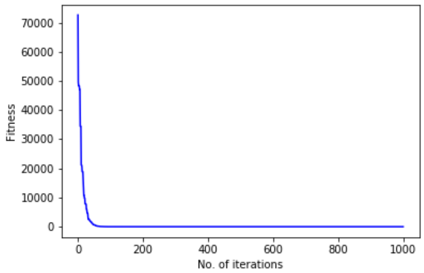


Fig 6.2: Convergence of GWO on F1

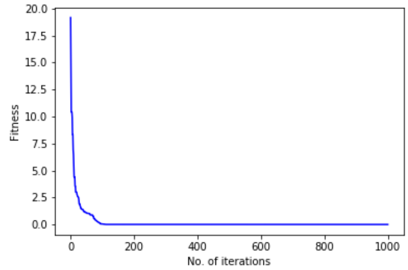
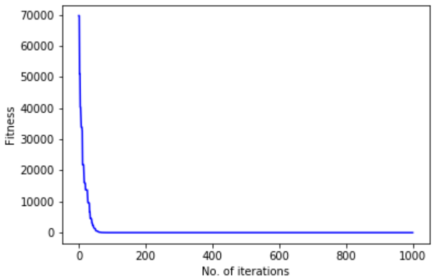


Fig 6.3: Convergence of GWO on F9



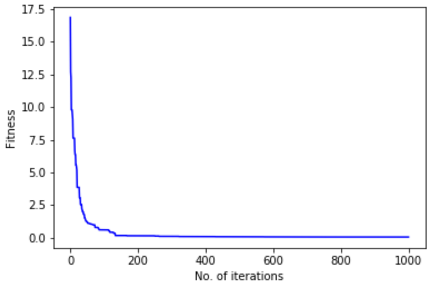


Fig6.4: Convergence of GWO-LF on F1

Fig6.5: Convergence of GWO on F9

To select the optimal features, a novel binary grey wolf optimization with levy flight method, bGWO-LF has been introduced**.** The performance of the proposed bGWO-LF method has been tested on the 5 feature selection benchmark datasets taken from UCI repository. All the considered datasets have different number of attributes and instances. Table 6.4 shows the brief description of all the considered datasets.

Table 6.4: Dataset Description

S.No	Dataset	Number of Features	Number of Instances
1.	WineEW	13	178
2.	LymphographyEW	18	148
3.	Zoo	16	101
4.	IonosphereEW &	34	351
5.	SonarEW	60	208

The proposed feature selection method minimizes the classification error and maximizes the accuracy. SVM and KNN classification techniques have been used to appraise the efficiency of proposed bGWO-LF method. Without FS refers to accuracy of classifiers considering all features of datasets while bGWOlf1 refers to binary version using sigmoid and bGWOlf2 refers to binary version using a threshold value(0.5).

Table 6.5 Average accuracy by different classifiers on datasets with and without ferature selection

Datasets	SVM			KNN		
	without FS	bGWOlf1	bGWOlf2	Without FS	bGWOlf1	bGWOlf2
Zoo	0.56	0.68	0.56	0.8	0.92	0.92
Wine	0.733333	0.888889	0.733333	0.733333	0.888889	0.73333
Lymphography	0.648649	0.648649	0.72973	0.756757	0.783784	0.864865
Sonar	0.519230	0.519230	0.519230	0.7115384	0.711538	0.75
Ionosphere	0.636364	0.636364	0.636364	0.909091	0.909091	0.920455

Since, meta heuristic methods are randomized in nature therefore each method has been executed 30 times and their mean value (average) has been taken for fair comparison. The performance of proposed bGWO-LF feature selection method has been measured in terms of mean accuracy. The mean accuracy of bGWO-LF and other methods have been given in and Table 6.5.

As seen in table 6.5, the enhanced version of bGWO- GWOlf1 and bGWOlf2 selects minimal features and provides almost same accuracy, sometimes even better mean accuracy in case of SVM and KNN as compared to when classifier is fed with all the features of the dataset. This shows the efficiency of the model to reduce redundant and irrelevant features.

CHAPTER 7- CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

GWO performs well but due to imbalance between exploration and exploitation it converges to local optimum solution. To solve this problem; a new variant using levy flight for introducing randomization is created which will enhances the exploration rate. This new variant is tested on 14 benchmark functions, outperforming GWO in 10 functions. Furthermore, this variant describes an enhanced version of the binary grey wolf algorithm. To test the performance of the proposed binary feature selection method 5 feature selection benchmark datasets have been used. From the statistical analysis and Experimental results efficacy of proposed feature selection methods has been observed.

7.2 FUTURE SCOPE

To further improve Grey Wolf Algorithm to overcome the problem of local minima and come up with a better discrete binary version of it. This binary version will be used to again extract features and test the classifier prediction accuracy by comparing the one achieved while taking these features and the one achieved earlier in the previous binary version. In the future, better feature extraction techniques can be applied to create more novel model of GWO. A parallel version of GWO can also be created in future.

However, the proposed feature selection methods outperform the existing methods; improvement in all the performance parameter is still required. Therefore, future work will incorporate to investigate the possibilities of accuracy enhancement by introducing different variants of metaheuristic methods.

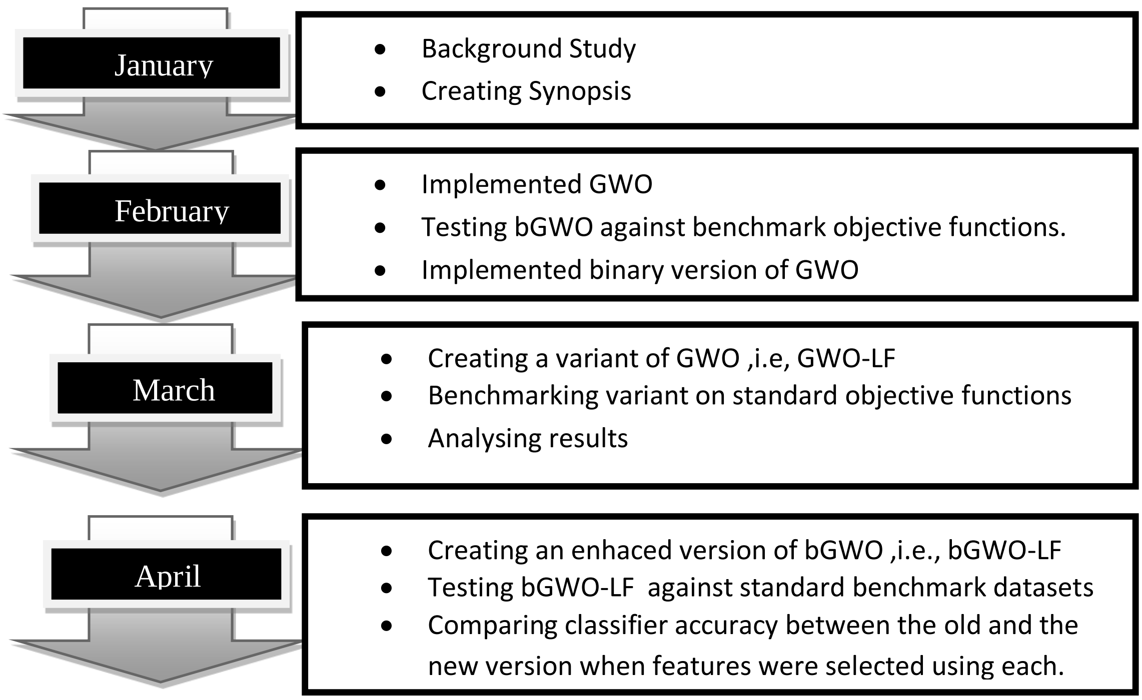


Fig 7.1: Gantt Chart

REFERENCES

[1] S. Mirjalili, S. M. Mirjalili, A. Lewis, ”Grey Wolf Optimizer”, Advances in Engineering Software, Vol. 69, pp. 46-61 , 2014.

[2] S. Shoghian, M. Kouzehgar, ”A Comparison among Wolf Pack Search and Four other Optimization Algorithms”, World Academy of Science, Engineering and Technology, Vol. 6, 2012.

[3] E. Emary, Hossam M. Zawbaa, Aboul Ella Hassanien,” Binary Gray Wolf Optimization Approaches for Feature Selection”, Neurocomputing, July 30, 2015

[4]Girish Chandrashekhar, Ferat Sahin,” A survey on feature selection methods”, Computers & Electrical Engineering, Vol. 40, Issue 1, January 2014.

[5] Muro C, Escobedo R, Spector L, Coppinger R. Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations. Behav Process; 88:192–7, 2011.

[6] J. Kennedy, R. C. Eberhart, ”A discrete binary version of the particle swarm algorithm”, IEEE International Conference on Systems, Man, and Cybernetics, Vol. 5, pp. 4104-4108, 1997.

[7] B. Chizi, L. Rokach, O. Maimon, ”A Survey of Feature Selection Techniques”, Encyclopedia of Data Warehousing and Mining, Second Edi-tion, IGI Global, pp. 1888-1895, 2009.

[8] X. S. Yang,”Engineering optimizations via nature-inspired virtual bee algorithms”, In: Lecture notes in computer science, Springer (GmbH), pp. 317323, 2005.

[9] F. Wilcoxon,”Individual Comparisons by Ranking Methods”, Biometrics Bulletin, Vol. 1, No. 6., pp. 80-83, 1945.

[10] X.-S. Yang, S. Deb, “Multiobjective cuckoo search for design optimization”,
Journal Computers & Operations Research,Volume 40, Issue 6, pp.1616–1624,June,2013.

[11] Haklı, H., & Uğuz, H.,” A novel particle swarm optimization algorithm with Levy flight”, Applied Soft Computing, Volume 23, pp.333–345, October 2014.

[12] Wolpert DH, Macready WG,” No free lunch theorems for optimization”, Evolut Comput, IEEE Trans , Volume 1, Issue 1, pp. 67–82, April 1997

Web links-

http://www.alimirjalili.com/

https://www.sciencedirect.com/

https://www.analyticsvidhya.com/