

Binary Multi-Verse Optimization (BMVO) Approaches for Feature Selection

Rahul Hans^{1,2}, Harjot Kaur¹ *

¹ Department of Computer Science and Engineering, Guru Nanak Dev University, Regional Campus, Gurdaspur, Punjab (India)

² Department of Computer Science and Engineering, DAV University, Jalandhar, Punjab (India)

Received 19 October 2018 | Accepted 11 May 2019 | Published 24 July 2019



ABSTRACT

Multi-Verse Optimization (MVO) is one of the newest meta-heuristic optimization algorithms which imitates the theory of Multi-Verse in Physics and resembles the interaction among the various universes. In problem domains like feature selection, the solutions are often constrained to the binary values viz. 0 and 1. With regard to this, in this paper, binary versions of MVO algorithm have been proposed with two prime aims: firstly, to remove redundant and irrelevant features from the dataset and secondly, to achieve better classification accuracy. The proposed binary versions use the concept of transformation functions for the mapping of a continuous version of the MVO algorithm to its binary versions. For carrying out the experiments, 21 diverse datasets have been used to compare the Binary MVO (BMVO) with some binary versions of existing meta-heuristic algorithms. It has been observed that the proposed BMVO approaches have outperformed in terms of a number of features selected and the accuracy of the classification process.

KEYWORDS

Machine Learning,
Feature Selection,
K-Nearest Neighbors,
Binary Multi-Verse
Optimization.

DOI: 10.9781/ijimai.2019.07.004

I. INTRODUCTION AND RATIONALE

IN the current era, due to swift advancement in the field of Computer Science, an enormous amount of information is generated and is included in the datasets in the form of a large number of features. In Machine learning, a feature is defined as the individual assessable quality of the process being observed [1]. However, the datasets may consist of a huge number of redundant features and the features which may not be relevant to the target concept [2]. These irrelevant features require more storage and lead to increase the computational cost. Thus, in order to reduce the computational constraint and storage requirement, as well as to improve the classification accuracy, feature selection has become a vital task for complex or large datasets [1][3].

In literature, there is an enormous amount of machine learning applications where feature selection has been applied. Some of these applications involve medical diagnosis [4], facial expression recognition [5], diagnose of bronchitis [6], gene selection and cancer classification [7], image steganalysis [8], big data classification [9], obstructive sleep apnea diagnosis [10], sentiment classification [11], Mobile Agent Platform Protection [70], Irony Detection [71], categorize text documents [72], classification of Plant Diseases [73], Breast Masses Detection [74].

A. Motivation

The fundamental objectives of the feature selection are to

reduce the dimensionality of the data and to improve the prediction performance [17], or in other words to employ fewer features to represent data without fading its discriminative capability. Pondering over, feature selection also aids in avoiding over fitting, resisting noise and improving classification accuracy [12]. This is the reason why finding the optimal subset of features to solve a task has become a necessary practice in machine learning. However, the complexity of task makes it impractical to find the promising solutions in a good enough time. Furthermore, searching for an optimal feature subset from a high dimensional feature space can be considered as an NP-complete problem [15].

In recent times, a class of algorithms called nature-inspired algorithms has drawn to the attention of the researchers to find the solution of optimization problems, nature-inspired algorithms have been developed by drawing inspiration from nature. To get to the bottom of the problem of feature selection, the scope of this research is to investigate the use of one of the most recent nature-inspired optimization algorithms viz. the Multi-Verse Optimization algorithm, to discover the best feature subset combination in diverse datasets (belonging to different domains) leading to maximum classification accuracy and a minimum number of selected features in the feature set. In view of the stochastic nature of the metaheuristic algorithms, metaheuristic algorithms discover the optimal solution by evaluating the sub solutions by invoking a fitness function. The sub solutions are formed by randomly selecting some of the features and rejecting the other features. For that reason, the conversion of the continuous version of the algorithm to its corresponding binary version becomes a significant task. Furthermore, this research aims to explore the use of two different categories of transfer functions viz. S-Shaped and V-Shaped transfer functions, for position updating step of the optimization algorithm for the conversion of the continuous version of the algorithm to its corresponding binary versions.

* Corresponding author.

E-mail addresses: rahulhans@gmail.com (R. Hans),
harjotkaursohal@rediffmail.com (H. Kaur).

B. Feature Selection and Optimization

The process of feature selection can be categorized into two different approaches: (i) wrapper-based and (ii) filter-based, that weighs up the eminence of the selected features. The former approach uses a machine learning method to search through the space of potential solutions, whereas, the latter approach goes over the feature space based on the data-dependent norm, unlike former which is based on the classification-dependent norm [13]. Furthermore, out of two, the former algorithms obtain better results; because they consider the association among the learning algorithm and the training data. However, they are computationally costly [3] and slower than the latter algorithms because the learning algorithm executes over and over again for every chosen feature subset [14].

Clearly, the Search process of finding an optimal feature subset from the original set is an exigent task as the complication of the problem of concern makes it impracticable to search every possible solution, so the prime aim is to find the best feasible solution in an acceptable time, which can be defined as an optimization problem [15].

C. Meta-heuristic Algorithm and MVO

In recent years, a class of algorithms known as metaheuristics has been considered more reliable when solving various optimization problems [16]. Metaheuristic algorithms use trial and error approaches to produce acceptable solutions to complex problems in a reasonable time [15]. Basically, these algorithms require balance in two following important components: Exploration of the search space (diversification) and exploitation of the best solutions found during exploration (intensification). In other words, the diversification produces the dissimilar solutions so as to explore the search space on a global scale, while intensification concentrates on the search in the local region by making use of information that a current good solution is found in this region. Meta-heuristic algorithms can also be divided into two classes: population-based algorithms (for instance, swarm intelligence, evolutionary algorithm) which are exploration-oriented and single-solution based algorithms (for instance, local search, simulated annealing) which are exploitation-oriented [16]. Last decade has revealed numerous efforts from researchers (as shown in Table I), who has reaped the benefits of metaheuristic algorithms to get to the bottom of the feature selection problem. Multi-Verse Optimization (MVO) is a newly introduced metaheuristic algorithm, which is inspired by the Multi-Verse theory in Physics. In this paper, the binary versions of MVO have been proposed and applied to solve one of the important problems of machine learning namely feature selection. The proposed binary versions of MVO use the concept of transformation functions for mapping of continuous version of MVO to the binary version of MVO algorithm.

D. Organization of the Paper

The rest of this paper is structured as follows: the brief analysis of previous related research works is presented in Section II. Section III throws light upon the concept of Multi-Verse Optimization (MVO) algorithm, whereas Section IV puts forward the binary versions of the MVO algorithm (BMVO) by using the concept of transformation functions. Furthermore, in Section V, various materials and methods used for the proposed study are discussed. Section VI briefly presents the results and discussions. Finally, in Section VII, conclusions and future works are presented.

II. THE CURRENT STATE-OF-THE-ART IN METAHEURISTIC ALGORITHMS

In the past many years, flourishing work has been accomplished on selecting an optimal number of features that are likely to preserve higher

classification accuracy rate, by means of metaheuristic algorithms. In this regard, this section aims to analyze the use of various metaheuristic algorithms viz. Ant colony optimization(ACO), Particle Swarm Optimization(PSO), Firefly Optimization Algorithm(FOA), Artificial Bee Colony algorithm(ABC), Genetic Algorithm(GA), Grey Wolf Optimization(GWO), Binary Bat Algorithm (BA), Simulated Annealing (SA), Ant lion Optimization (ALO), Cuckoo Search (CS), Sine Cosine Optimization Algorithm(SCA), Differential Evolution(DE), Whale Optimization algorithm(WOA) and their variants. The summary of various metaheuristic algorithms and their variants explored in recent years for feature selection has been given in Table I.

In [14] Manizheh Ghaemi, et al. have proposed the use of Forest Optimization Algorithm (FOA) for feature subset selection which was tested on 11 different datasets and compared with GA, PSO, and ACO. Uros Mlakar, et al. in [5], have used the notion of Multi-Objective DE for feature selection in Facial Expression Recognition. Similarly, Indu Jain et. al in [7], have proposed an improved-Binary Particle Swarm Optimization with Correlation-based Feature Selection (CFS) which uses the scheme of increasing inertia weight and controls the searching capability of the iBPSO algorithm. The performance of iBPSO has been evaluated using 11 benchmark microarray datasets of different cancer types and the results have been compared with methods like SVM, Random Forest, Fast Correlation-Based Filter, BPSO, PSO-DT.

In [17] Chuang et al. proposed an improved binary PSO for feature selection for gene expression data classification problems. By returning the gbest, improved binary PSO can avoid getting trapped in a local optimum, and better classification result can be attained with a lesser amount of chosen genes. Results also demonstrate that improved binary PSO efficiently simplifies feature selection and reduces the number of features desired. The algorithm was tested on 11 gene expression data test problems and the proposed method has the highest classification accuracy in 9 of the 11 gene expression data test problems.

In [18] Yu-Peng Chen et al. have proposed the two novel variants of Bacterial foraging optimization (BFO) algorithms viz. Adaptive Chemotaxis Bacterial Foraging Optimization Algorithm (ACBFO) and Improved Swarming and Elimination-Dispersal Bacterial Foraging Optimization Algorithm (ISED BFO). The proposed algorithms have been tested on 10 different benchmarks datasets and performance has been compared with PSO, GA, SA, ALO, BBA, CS. Also, J. Dhalia Sweetlin et al. in [6] have proposed the combination of ACO with cosine similarity and Support Vector Machine (SVM). The outcome of the proposed algorithms is compared with PSO and hybrid PSO algorithms. GA is used by Babatunde Oluleye et al. in [19] in combination with the classifiers of weka (MLP, RF, J48, NB, RC) and weka feature selection algorithms (Correlation Feature Selection Subset Evaluator, Information gain). For evaluating the performance two different datasets are used viz. features extracted from Flavia dataset and ionosphere dataset.

Pedram Ghamisi et.al in [20] have proposed the hybridization of GA and PSO. The accuracy of SVM classifier obtained on validation samples is used as a fitness value and it is evaluated on the Indian Pines hyperspectral data set. Gang Wang et al., in [21] have explored the use of ACO by adaptively adjusting its parameters (such as pheromone evaporation rate, number of ants and exploration probability factor) for feature selection. The results are evaluated on 10 different datasets and performance is compared with GA, PSO, ACO, fuzzy adaptive ant system. In [22] Fadzil Ahmad, et al. have used GA for simultaneous feature selection and parameter optimization of an artificial neural network (ANN) and for feature selection the performance of resilient back-propagation (GAANN_RP) with and without feature selection is compared on breast cancer dataset. H. Hannah Inbarani et al. in [4] applied PSO integrating with a rough set theory for feature subset selection and compared the performance on 4 different datasets.

TABLE I. BRIEF SUMMARY OF METAHEURISTIC ALGORITHMS AND THEIR VARIANTS EXPLORED IN RECENT YEARS FOR FEATURE SELECTION

Authors	Year	Algorithm	Strategy Used	No. of Datasets	Algorithms for Comparison
Li-Yeh Chuang, et al. [30]	2009	Particle Swarm Optimization	Use of Chaotic maps with PSO	10	BPSO
Hema Banati,Monika Bajaj [31]	2011	Fire fly Algorithm	Fire fly with rough set theory	4	RSAR*,AntRSAR,GenRSAR, PSORSAR,BeeRSAR,FA_RSAR
Li-Fei Chen et al. [10]	2011	Particle swarm optimization	Integrating Particle swarm optimization with 1-Nearest Neighbour scheme	8	BPN LR SVM C4.5 GA + 1-NN PSO + 1-NN
Yuanning Liu, et al. [32]	2011	Particle Swarm Optimization	Employed sub-swarms and a multi-swarm scheduler which monitors each sub-swarm	10	PSO+SVM and GA+SVM
R. Y. M. Nakamura, et al. [33]	2012	Binary Bat Algorithm	Used Sigmoid function to restrict the new bat's position to only binary values	5	PSO, FFA, HS, GSA
XiaoHong Han, et al. [34]	2014	Gravitational search algorithm	Use of Piecewise linear chaotic map for increasing diversity, SQP* for increasing local exploitation	15	BPSO, GA
E. Emary, et al. [35]	2015	Firefly Algorithm	Attraction is randomized with α being the randomization parameter	18	GA, PSO
Hossam M. Zawbaa, et al. [13]	2016	Moth-flame optimization	-	18	GA, PSO
E. Emary, et al. [36]	2016	Gray Wolf Optimization	Use of Sigmoid function and Crossover technique	18	GA, PSO
Lin Shang, et al. [11]	2016	Particle swarm optimization	Used Fitness proportionate selection binary particle swarm optimization	2	BPSO F-BPSO FS-BPSO
Hong Wang, Ben Niu [37]	2016	Bacterial algorithm	It uses modified population updating strategies and control mechanisms.	10	BFO, BFO-LDC,FO-NDC, BCO, BAFS
Majdi M. Mafarja and Seyedali Mirjalili [16]	2017	Whale Optimization Algorithm	Whale Optimization Algorithm with simulated annealing	18	ALO, GA, PSO
Gehad Ismail Sayed, et al. [3]	2017	Crow Search	Use of Chaotic Maps	20	PSO, ABC, CSO, FPA, MFO, GWO, etc
Majdi Mafarja, et al. [38]	2017	Ant loin Optimization	Use of s-shaped and v-shaped functions	18	GSA, PSO
Ahmed A. Ewees, et al. [39]	2017	Multiverse Optimization	Use of Chaotic Maps	5	MVO, PSO, ABC
P. Shunmugapriya, S. Kanmani [40]	2017	Ant Colony optimization and Bee colony optimization	Ebees perform exploitation on the solutions generated by ACO	13	ABC-DE Hybrid, AC-ABC Hybrid
Mohammed Aladeemy, et al. [41]	2017	Self-Adaptive Cohort Intelligence	For the sampling interval and mutation rate, It employed tournament-based mutation and self-adaptive scheme	10	SVM-CI, SVM-ABC, SVM, GA, SVM-DE, and SVM-PSO
E. Tamimi [61]	2017	ACO GA ICA PSO	Optimization of SVM	2	Random forest
Sushama Nagpal, et al. [42]	2017	Gravitational search algorithm	GSA and k-nearest neighbors	3	GA, PSO, GSA
M. K Sohrabi et al. [59]	2017	Non-dominated Sorting GA-II and Multi-Objective PSO	Non-dominated Sorting GA-II and Multi-Objective PSO and ANN	1	Classical methods
Hossam Faris, et al. [43]	2017	Multi verse Optimization	MVO for selecting optimal features and optimizing the parameters of SVM	10	GA PSO BAT FF
Ibrahim Aljarah, et al. [44]	2017	Grasshopper optimization algorithm	Optimizing SVM based on the Grasshopper Optimization Algorithm (a hybrid approach)	12	GOA Grid search MVO GA PSO GSA
Majdi Mafarjaa, Seyedali Mirjalili [45]	2017	Whale optimization	Tournament and Roulette Wheel selection mechanisms instead of random operator	20	ALO GA PSO
M. Suganthi ,V. Karunakaran [46]	2017	Cuttlefish optimization algorithm	In this PCA* is used for feature extraction with Cuttlefish optimization	4	Algorithm with Instance selection, Algorithm with Feature extraction
D. Zouache et al. [60]	2018	FF and PSO	Quantum computation/rough set theory	11	PSO-Rough Set, FSA-Rough Set
Hongbin Dong, et al. [47]	2018	Genetic Algorithm	Neighborhood policy to granulate the sample space and improved neighborhood rough set	11	FCBF, BIRS, MBEGA, GA, mRMR, IBGAFG, INRSG
E.Emary , et al. [67]	2018	Ant lion optimization	Levy flights with Ant lion Optimization	21	GA PSO ALO LALO

RSAR* - Rough Set Based Attribute Reduction, SQP*- Sequential quadratic programming, PCA*- Principal component analysis

Similarly, Yumin Chen et al. in [23] have explored the use of ACO with rough sets and evaluated the performance on 9 different datasets.

S. Tiwari et al. in [62] used local searching algorithms for spawning relevant and non-redundant features; subsequently, a global optimization algorithm has been used to remove the restrictions of global optimization algorithms. The time and accuracy were improved by using a feature set obtained from sequential backward selection and mutual information maximization algorithm which is fed to a global optimization technique (GA, PSO etc). To test the proposed approach publicly available Sonar, Wdbc and German datasets were used.

A. Ekbal in [63] proposed an algorithm based on the notion of multiobjective optimization viz. multiobjective GA, along with the lines of NSGA-II, for performing parameter optimization and feature selection. The experiments were performed on four different classifiers viz. random field, support vector machine, memory based learner and maximum entropy. The proposed algorithms are evaluated for solving the problems of named entity recognition (used in text processing applications). Comparisons with existing researches demonstrated the usefulness of the proposed approach.

Shih-Wei Lin et al. in [64] proposed a particle swarm optimization based approach to obtain the suitable parameter settings for the back-propagation network and to select the relevant subset of features which resulted in improved accuracy. For experimental evaluation of the proposed algorithm 23 different datasets were considered from the UCI repository. When the performance was compared with some existing works, the results confirmed that the proposed approach improved the classification accuracy in the majority of test problems.

J. Vijaya in [65] proposed an algorithm for the telecom churn prediction that makes use of particle swarm optimization and proposes three different versions of PSO for churn prediction that is, PSO having feature selection as its pre-processing mechanism, PSO with simulated annealing and lastly PSO with a blend of both feature selection and simulated annealing.

R. P. S. Manikandan et al. in [66] proposed fish swarm optimization for feature selection in big data. To perform the experiments, Product review dataset that is obtained from Amazon along with synthetic data is used. The dataset contained 235,000 positive and 147,000 negative reviews. Results depicted that the technique proposed attains improved performance than that of the other techniques.

III. MULTIVERSE OPTIMIZATION ALGORITHM (MVO)

Multi-verse is another latest hypothesis among researchers in physics [68]. MVO algorithm is based on the ideas of white holes, black holes, and wormholes which are mathematically mapped to construct the MVO [15]. The word Multi-Verse implies “opposite of universe”, which means there exists one more universe besides the one in which we are living [13].

A. Notion

It is believed that more than one big bang happenings have taken place and each big bang has led to the dawn of a new universe. This theory also states that numerous universes act together and might even have a collision with each other. Furthermore, it puts forward a fact that there might be different physical laws in each of the universes.

Significantly, there are following three concepts of the multiverse theory which has inspired the MVO the most:

- White holes- these can be accounted for as a big bang which may be the source of the origin of the universe [24].
- Black holes- these are assumed as to pull towards them every object with their gravitational force [25].

- Wormholes- these can be assumed as time/space travel channels, where objects are able to travel instantaneously across the universe [26].

It has been assumed that every universe expands through the space caused by its inflation rate [27][28]. Moreover, it has been also argued that multiple universes interact via these three different holes in order to target a stable situation [27][29].

B. Continuous Version of Multi-Verse Optimization Algorithm(MVO)

As discussed in sub-section III.A, in MVO algorithm black holes receive the objects that are transmitted by white holes. So as the number of iterations add to, the fitness (inflation rates) of all the universes will become better. Improvement in the exploration and exploitation phase and to prevent getting rapt in local optima, preservation of the diversity of universes is obligatory which is made possible by wormholes [39]. To begin with MVO algorithm, random universes are generated, and in each iteration, objects are transferred through white and black holes from the universe with higher fitness to the universe with lower fitness by a random transmission via wormholes in the direction of the best universe. This process occurs until the end criterion is not satisfied.

In the MVO algorithm, the following set of laws is implemented on the universes:

1. More the inflation rate more is the likelihood of having a white hole and less is the likelihood of encompassing black holes.
2. High inflation rated universes have a tendency to propel objects through white holes while low inflation rated universes have a tendency to accept objects through black holes.
3. Despite the inflation rate, the objects in the universes face random movement towards the best universe.

To map the algorithm mathematically the white and black holes are represented as the population of universes U . It is supposed that each candidate solution is analogous to a universe and each variable in the solution is assumed as an object β_i^j (j^{th} parameter of the i^{th} universe) in that universe, the dimension of the problem is represented by d and the number of universes is represented by N . The different universes are arranged on the basis of their fitness values (inflation rates). Probing further, the object β_i^j is swapped by using the concept of Roulette Wheel mechanism which selects universe U_k as in equation (1), where $r1=[0,1]$ is a random number, U_i is the i^{th} universe, and $NF(U_i)$ is the normalized fitness value(inflation rate) of the U_i . This complete process of MVO is depicted in Fig. 1.

The smaller value of fitness specifies the more likelihood of transferring objects through white and black holes. By assuming that each U_i transports the objects of U_i in the course of space arbitrarily using wormholes, the exploitation is carried out and the assortment of universes is retained. The objects of the universes are altered by the wormholes in a stochastic manner, where wormhole existence probability (WEP) is defined in equation (3). These wormholes are also used to bring up to date the universe’s objects and perk up the inflation rate by altering the objects of the universe which has the finest inflation rate as the subsequent shown in equation (2) [39], where β_j shows the j^{th} parameter of best universe formed so far, lb_j and ub_j indicate the lower bound and the upper bound respectively corresponding to the j^{th} variable $r2,r3,r4$ are the random numbers between 0 and 1. The coefficient, Travelling Distance Rate (TDR) is employed to choose the distance letting a wormhole move the object towards the best universe, as shown in equation (4),

$$\beta_i^j = \begin{cases} \beta_k^j & , \text{ if } r1 < NF(U_i) \\ \beta_i^j & , \text{ if } r1 > NF(U_i) \end{cases} \quad (1)$$

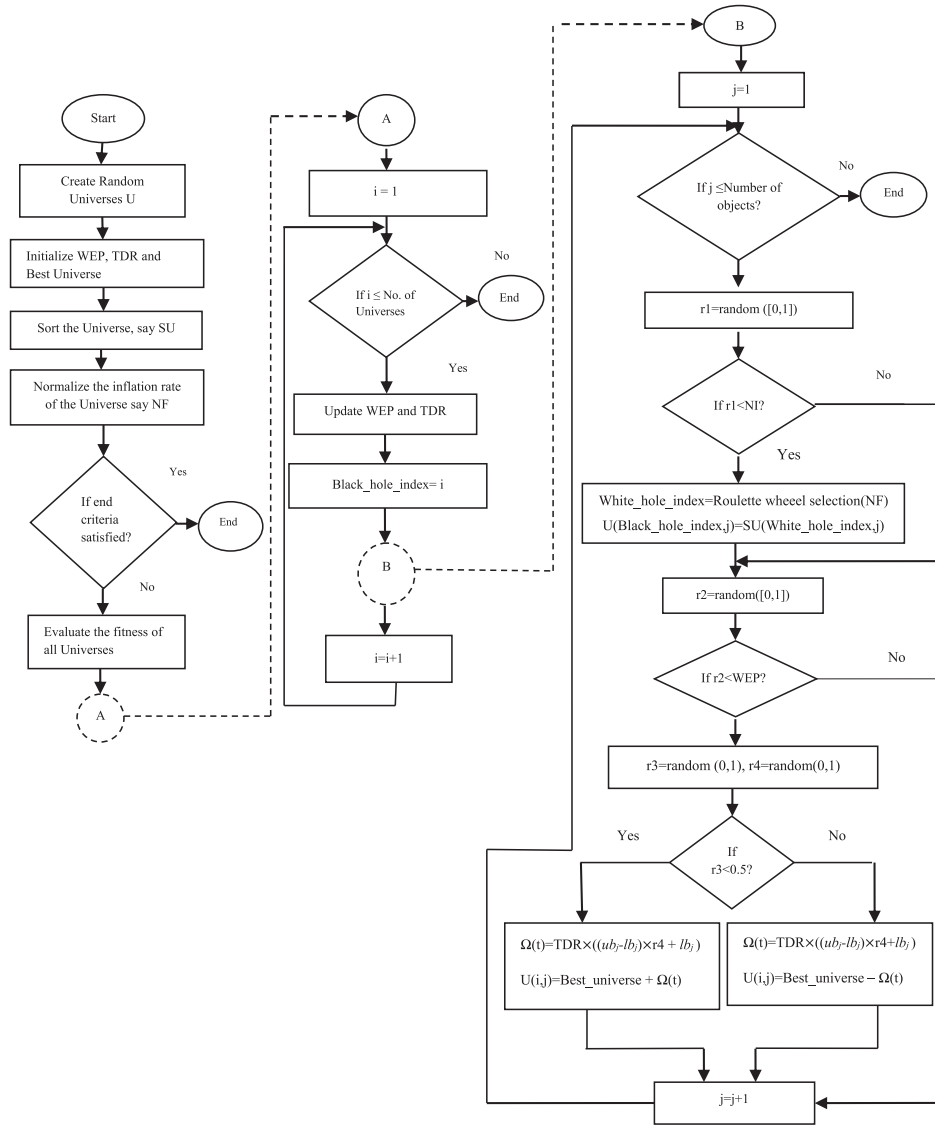


Fig. 1. General Flowchart representation of Multi-verse Optimization algorithm.

$$\beta_i^j = \begin{cases} \beta_j + TDR \times ((ub_j - lb_j) \times r4 + lb(j)) \geq 1 & r3 < 0.5 \\ \beta_j - TDR \times ((ub_j - lb_j) \times r4 + lb(j)) \geq 1 & r3 \geq 0.5 \\ \beta_i^j & \end{cases} \quad \begin{matrix} r2 < WEP \\ r2 > WEP \end{matrix} \quad (2)$$

$$WEP = MIN_{WEP} + l \times \frac{Max_{WEP} - MIN_{WEP}}{L} \quad (3)$$

$$TDR = 1 - \frac{l^{1/p}}{L^{1/p}} \quad (4)$$

Where l and L defines the current iteration and a total number of iterations respectively. The default value of p is 6. In equation 3, the default minimum (MIN_{WEP}) and maximum (MAX_{WEP}) values of WEP are 0.2 and 1 respectively.

IV. BINARY VERSIONS OF THE MULTI-VERSE OPTIMIZATION ALGORITHM (BMVO)

One of the obligatory requirements to find the feature subset containing a minimum number of features and achieving high

classification accuracy from those selected features is to transform the continuous version of MVO to binary version. In this transformation, the location of search agents, as well as the position of the search agents, is represented in the form of strings of 0s and 1s. A transformation function defines the likelihood of altering a position vector's elements from 0 to 1 and vice versa [48]. This section examines the functionality and performance of two different V-shaped and two different S-shaped transformation functions for developing different binary versions of MVO [49][57].

A. Proposed Approach

In the problems with search space consisting of continuous values can be transformed into binary problems by transforming their variables to binary [69]. Irrespective of the binary problem category, the search space of binary problem consists of binary values which may have its own structure. In general, the search space of a binary problem can be represented as a hypercube. So, by flipping some of the bits of the candidate solution (universe/agent) the solution of a binary optimization algorithm may reallocate to closer and farther points of the hypercube [48]. Therefore, to design a binary MVO, this research customizes the position updating procedure.

In binary search space, while dealing with 0s and 1s, equation (2) of MVO cannot perform the position updating directly. Therefore, another way has been proposed to use the equation. (2) for changing agent's location from 1 to 0 or from 0 to 1. In BMVO, the agent's position is transformed with the probability of position updating value (β_i^j) of the objects of the universe, depending upon random number r3. In this regard, four different transformation functions have been explored in this work to map equation (2) values to probability values for updating the agent's position.

1. Binary Mapping Using S-Shaped Family of Transfer Functions

a) BMVO-I uses Sigmoidal Transformation

$$\beta_i^j = \begin{cases} 0, & \text{trf1}(\beta_i^j) > \theta_1 \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

$$\text{trf1}(\beta_i^j) = \text{sigmoid}(\beta_i^j) = \frac{1}{1 + e^{-\beta_i^j}} \quad (6)$$

b) BMVO-II uses another version of sigmoidal Transformation

$$\beta_i^j = \begin{cases} 0, & \text{trf2}(\beta_i^j) > \theta_2 \\ 1, & \text{otherwise} \end{cases} \quad (7)$$

$$\text{trf2}(\beta_i^j) = M\text{-sigmoid}(\beta_i^j) = \frac{1}{1 + e^{-(\beta_i^j)^2}} \quad (8)$$

2. Binary Mapping Using V-Shaped Family of Transfer Functions

a) BMVO-III uses the tangent hyperbolic transfer function

$$\beta_i^j = \begin{cases} 0, & \text{trf3}(\beta_i^j) > \theta_3 \\ 1, & \text{otherwise} \end{cases} \quad (9)$$

$$\text{trf3}(\beta_i^j) = \tanh(\beta_i^j) = \left| \frac{(e^{-(\beta_i^j)} - 1)}{(e^{-(\beta_i^j)} + 1)} \right| \quad (10)$$

b) BMVO-III uses the inverse tangent transfer function

$$\beta_i^j = \begin{cases} 0, & \text{atan}(\beta_i^j) > \theta_4 \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

$$\text{trf4}(\beta_i^j) = \text{atanh}(\beta_i^j) = \frac{2}{\pi} \left| \arctan\left(\frac{\pi}{2}x\right) \right| \quad (12)$$

In equations, (5)-(12), *trf1*, *trf2*, *trf3*, and *trf4* can be defined as the transformation function applied on β_i^j and $\theta_1, \theta_2, \theta_3, \theta_4$ are the random numbers distributed over 0 and 1. The pseudo code of the proposed binary versions of MVO is shown in Algorithm 1 in which the position updating phase is modified using four different transfer functions.

V. MATERIALS AND METHODS

Generally, the problem of feature subset selection is assumed to be as a multi-objective optimization problem in which the occurrence of tradeoffs among two or more conflicting objectives is considered for making an optimal choice [50]. As discussed in the previous sections, the problem of feature subset selection consists of two most important objectives viz. to maximize the classification performance and to minimize the number of features selected or size of selected feature subset. The process feature selection is considered successful or

Algorithm 1. Pseudo code of Binary version of Multiverse Optimization algorithm (BMVO)

1. Creation of U, where U is random universe.
2. Initialization of TDR, Best_universe and WEP.
3. SU=Sorted universes; NF=Normalized inflation rate.
4. **while** end criterion is not met
5. Evaluate all the universes for the fitness or inflation rate
6. **for** each universe indexed by i
7. Update WEP and TDR ; Black_hole_index = i;
8. **for** each object indexed by j
 - r1=random([0,1]);
 9. **if** r1<NF(Ui)
 - White_hole_index= RouletteWheelSelection(-NF);
 - U(Black_hole_index,j)=SU(White_hole_index,j);
 10. **end if**
 - r2=random([0,1]);
 - if** r2<Wormhole_existance_probability
 - r3= random([0,1]);
 - r4= random([0,1]);
 11. **if** r3<0.5
 - $\Omega(t)=\text{trf}(\text{Best_universe}(j) + \text{Travelling_distance_rate} * ((\text{ub}(j) - \text{lb}(j)) * r4 + \text{lb}(j)));$
 - If** $\Omega(t) > \text{random}([0,1])$
 - $\Omega(t)=0;$
 - else**
 - $\Omega(t)=1;$
 - End if**
 - $U(i, j)=\Omega(t)$
 12. **else**
 - $\Omega(t)=\text{trf}(\text{Best_universe}(j) - \text{Travelling_distance_rate} * ((\text{ub}(j) - \text{lb}(j)) * r4 + \text{lb}(j)));$
 - If** $\Omega(t) > \text{random}([0,1])$
 - $\Omega(t)=0;$
 - else**
 - $\Omega(t)=1;$
 - End if**
 - $U(i,j)=\Omega(t)$
 13. **end if**
 14. **end if**
 15. **end for**
 - 16.**end for**
 - 17.**end while**

solution is considered as the best solution if the dimensionality of the data is reduced and the accuracy of the algorithm improves or remains the same [51].

For an optimization algorithm, to weigh up the discriminative potential of each subset of features a fitness function must be defined. The fitness function for which every candidate solution of the

population is evaluated is based on k-NN classifier [52] as shown in equation (13).

$$\text{Fitness} = W_1 \left| \frac{\text{FS}}{N} \right| + W_2 * \mu(D) \quad (13)$$

Where $\mu(D)$ represents the error rate in the classification process calculated from the k-NN classifier. Furthermore, $|\text{FS}|$ is the total number of features selected by an optimization algorithm in the subset and $|N|$ is the original size (Number of features present) of feature set, W_1 and W_2 are two constants equivalent to the significance of classification eminence and feature subset size, where $W_1 \in [0, 1]$ and $W_2 = (1 - W_1)$ adopted from [16][53].

A. k-Nearest Neighbors

In this study, the simplest algorithm known as k-nearest neighbor's (k-NN) serves as an assessor of the candidate solutions in the population [70]. This algorithm is a supervised machine learning algorithm and also a non-parametric method used for classification as well as regression [52], which is based on finding k-nearest neighbors by means of smallest distance between training examples and the query instance [36].

B. Parameter Settings and Dataset Description

To summarize the system configuration for performing the experiments a CPU with a processor configuration of Core 2 Duo 2.00 GHz, 3.00 GB RAM and 64-bit Operating system is selected. The parameter setting for the proposed BMVO approaches is shown in Table II. Also, the algorithms used for comparison viz. GWO, SCA, WOA, and MVO are implemented for 10 different runs with 100 iterations in each run and are initialized to 20 agents. To carry out the experiments, 21 datasets were obtained from various online sources [54][55] which are summarized in Table III.

TABLE II. VARIOUS PARAMETER SETTINGS FOR THE EXPERIMENTS

Parameter	Value
Runs	10
Number of Iterations	100
MIN_{WEP}	0.2
MAX_{WEP}	1
P(default value)	6
No. of Universes	20
Lb(Lower Bound)	0
Ub(Upper Bound)	1
Dimensions	Total features in the dataset
W_2 (fitness function)	0.99
W_1 (fitness function)	0.01

All the chosen datasets have dissimilar numbers of attributes plus instances representing a range of categories including medical domain also. The instances are arbitrarily divided into three different sets that are training, validation, and testing sets in a cross-validation manner for all data sets. The method for feature selection is based on k-NN classifier and works on trial and error logic. To limit the experiments, the value of k is set to 5 on all the datasets for its best output.

C. Evaluation Metrics

To quantify the performance of proposed algorithms and to compare with other existing metaheuristic algorithms, the following criteria are used which are summarized in Table IV along with their expected values:

a) **Average Classification Accuracy**- It can be described as the

TABLE III. DATASETS DESCRIPTION WITH NO. OF INSTANCES AND ATTRIBUTES

No.	Dataset	Instances	Attributes
D ₁	Zoo	101	16
D ₂	Statlog	1000	20
D ₃	Lung cancer	32	56
D ₄	Exactly	1000	13
D ₅	Exactly2	1000	13
D ₆	M-of-N	1000	13
D ₇	Heart	294	13
D ₈	Vote	300	16
D ₉	Spect Heart	267	22
D ₁₀	Australian	690	14
D ₁₁	Ionosphere	351	34
D ₁₂	Water treatment	521	38
D ₁₃	Wine	178	13
D ₁₄	Indian Liver	583	10
D ₁₅	Tic-Tac-Toe	958	9
D ₁₆	Wavform	1000	21
D ₁₇	Dermatology	366	34
D ₁₈	Glass Identification	214	9
D ₁₉	Breast cancer	699	9
D ₂₀	Sonar	208	60
D ₂₁	Vowel	990	13

proportion of samples taken for testing correctly classified by the algorithm. It generally evaluates the ability of a classifier in classifying the dataset in N runs, where Acc_t is the accuracy obtained in t^{th} run.

- b) **Average fitness** - This metric gives the average of the various fitness values achieved by a probabilistic algorithm in N runs, where Fit_t is the fitness obtained in t^{th} run.
- c) **Worst fitness** - This metric gives a maximum of the N fitness values gained by the algorithm in N runs.
- d) **Best fitness** - This measure gives a minimum of the N fitness values achieved by the algorithm in N runs.
- e) **Standard Deviation** - This metric is defined as the divergence of the finest achieved solutions found after running a stochastic optimizer for N runs [36].
- f) **Average Number of Features Selected** - This criterion is defined as the number of features selected averaged on all the runs.
- g) **F-Measure** - F-measure is another important criterion to measure the performance of the classification algorithm also known as F-score; it is an assessment of classifier's accuracy, which integrates both the precision as well as the recall as a harmonic mean.
- h) **Average Time** - This criterion measures the time taken by an algorithm averaged on all the runs.
- i) **Non-Parametric Testing** - Wilcoxon's signed rank test is one of the non-parametric tests that aims to detect significant differences between the two sample means. The test returns a parameter called p-value which verifies the level of significance of two algorithms [56].

TABLE IV. EVALUATION METRICS

Evaluation criteria	Formula	Expected Value
Average Classification Accuracy	$Avg_Acc = \frac{1}{N} \sum_{t=1}^N Acc_t$	Maximum
Average fitness	$Avg_fit = \frac{1}{N} \sum_{t=1}^N Fit_t$	Minimum
Worst fitness	$Worst_fit = \max(Fit_1 : Fit_N)$	Maximum
Best fitness	$Best_fit = \min(Fit_1 : Fit_N)$	Minimum
Standard Deviation	$St_dev = \sqrt{\frac{\sum_{t=1}^N (Fit_t - \overline{Fit})^2}{N-1}}$	Minimum
Average size of Feature Subset	$Avg_FS = \frac{1}{N} \sum_{k=1}^N SizeFS_t$	Minimum
F-Measure	$F_Measure = \frac{2 \cdot Precision \cdot Recall}{(Precision + Recall)}$	Maximum
Average Time	$Avg_Time = \frac{1}{N} \sum_{t=1}^N time_t$	Minimum

VI. RESULTS AND DISCUSSIONS

This section reports all the results obtained by implementing different variants of Binary Multi-verse optimization (BMVO). Also, the proposed binary MVO approaches are compared with other binary versions of metaheuristic algorithms viz. GWO, WOA, SCA, PSO, ALO, and bMVO (without transformation functions) based on the criteria shown in Table IV. It can be depicted from Table V that when algorithms are compared on the basis of average classification accuracy BMVO-I and BMVO-II outperforms the other versions of BMVO as well as another state of the art algorithms considered for comparison. Moving ahead, another important metric to assess the performance of the proposed variants of BMVO is the average fitness value obtained from the objective function on various datasets. Clearly, it can be depicted from Table VI that all the versions of BMVO outperform the binary versions of existing meta-heuristic algorithms viz. GWO, WOA, SCA, PSO, ALO bMVO (native version, without transformation function) on almost all the datasets but BMVO-I and BMVO-II perform better than other two versions i.e. BMVO-III and BMVO-IV in terms of average fitness values obtained.

Similarly Table VII and Table VIII compares the performance of proposed versions with the other state of the art metaheuristic algorithms on the basis of their worst (Maximum in this case) and best (Minimum in this case) fitness values, obtained in various runs on various datasets. On comparison it can be found that all the four versions outperform the other metaheuristic algorithms in terms of worst fitness. However, the performance of BMVO-I and BMVO-II is better than BMVO-III and BMVO-IV and GWO, SCA, WOA and bMVO in terms of worst fitness obtained, consequently in terms of best fitness value obtained, all the versions of BMVO obtain Best results on most of the datasets taken for comparison as shown in Table VIII.

Moving ahead Table IX shows the performance comparison of all the algorithms on the basis of standard deviation achieved. Clearly, it can be seen that proposed versions of BMVO perform better than

GWO, WOA, SCA, PSO, ALO, bMVO on almost 14 different datasets in terms of standard deviation out of remaining datasets MVO performs better on three different datasets and GWO, WOA, SCA performs better in terms of a standard deviation on 5 different datasets.

Moving further another important measure to assess the performance of the proposed versions of BMVO is the average number of features which are selected for feature subset by various algorithms. Clearly it can be deduced from Table X that proposed variants outperforms GWO, WOA, SCA, PSO, ALO, bMVO in terms of average numbers of selected features, also it can be noticed that out of four different variants of BMVO, the proposed BMVO-I and BMVO-II outperforms the other two variants of BMVO in terms of average number of selected features. Also, the graph shown in Fig. 2 presents a number of total features selected on all datasets by all algorithms. It can be clearly depicted from the graph that the proposed versions of BMVO outperform all the existing metaheuristic algorithms when the comparison was made on the basis of an average number of features selected on all dataset. Moving ahead, another important criterion to gauge the performance of proposed variants is F-Measure, which makes use of precision and recall parameter. It can be depicted from the graph in Fig. 3 that proposed versions of BMVO outperform the other algorithms when the comparison was made on the basis of average F-measure values obtained on all the datasets. The results of F-measure values can be seen in Table XV in the appendix section, that indicate that the proposed algorithms perform much better on the maximum number of datasets.

Similarly, when the comparison is made on the basis of average time taken by an algorithm per run in as shown in Table XIV in the appendix section, the proposed variants BMVO-II and BMVO-III show much better performance than other algorithms. Also it can be clearly seen from the graph shown in Fig. 4 that all the binary variants (apart from BMVO-I) of MVO outperform another state of the art metaheuristic algorithms but BMVO-II and BMVO-III takes lesser time when cumulative sum of average time per run is taken on all the datasets than all the other variants of BMVO as well as the metaheuristic algorithms considered for comparison. The overall performance of BMVO-I is better than the remaining algorithms in terms of an average number of features selected on nine datasets and average classification accuracy on eleven datasets. The use of transformation function leads to an increase in the execution time of the algorithm to some extent.

It can be clearly depicted from Table XI which indicates the p-values obtained, in most of the cases the results are significant at $p \leq 0.05$. To summarize the results it can be clearly seen that when comparison is made on the basis of average classification accuracy and average fitness value BMVO-I outperforms all the other versions of BMVO: $BMVO-I > BMVO-II > BMVO-III > BMVO-IV$, on the other hand when comparison was made on the basis of standard deviation BMVO-III outperforms all the other versions of BMVO: $BMVO-II > BMVO-III > BMVO-I > BMVO-IV$. However when the comparison was made on the basis of the average number of selected features and f-measure. BMVO-I outperforms all the other variants of BMVO: $BMVO-I > BMVO-II > BMVO-III > BMVO-IV$ and finally when the comparison is made on the basis of average time taken by an algorithm per run BMVO-III performs better than other algorithms: $BMVO-III > BMVO-II > BMVO-IV > BMVO-I$.

VII. CONCLUSIONS AND FUTURE WORK

In recent times, diverse metaheuristic algorithms have been developed to find the solution for feature subset selection problem in different applications. In this research work, four different wrapper based binary variants of Multi-Verses Optimization algorithm is proposed by converting the continuous version of Multi-verse

TABLE V. COMPARATIVE ANALYSIS OF AVERAGE CLASSIFICATION ACCURACY OBTAINED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	0.919	0.910	0.905	0.943	0.884	0.939	0.937	0.961	0.953	0.935
D ₂	0.717	0.718	0.716	0.710	0.713	0.721	0.743	0.740	0.730	0.727
D ₃	0.863	0.888	0.881	0.894	0.925	0.888	0.956	0.900	0.938	0.875
D ₄	0.703	0.726	0.718	0.723	0.704	0.738	0.879	0.974	0.791	0.808
D ₅	0.754	0.748	0.748	0.755	0.760	0.752	0.768	0.763	0.765	0.765
D ₆	0.880	0.911	0.831	0.900	0.881	0.912	0.970	0.977	0.950	0.937
D ₇	0.805	0.821	0.816	0.817	0.816	0.834	0.854	0.846	0.829	0.831
D ₈	0.943	0.943	0.939	0.931	0.946	0.954	0.962	0.957	0.954	0.959
D ₉	0.719	0.726	0.737	0.729	0.728	0.754	0.774	0.757	0.757	0.754
D ₁₀	0.799	0.777	0.790	0.810	0.790	0.820	0.870	0.860	0.849	0.844
D ₁₁	0.866	0.867	0.870	0.854	0.869	0.861	0.899	0.888	0.897	0.863
D ₁₂	0.815	0.821	0.815	0.807	0.821	0.828	0.855	0.855	0.841	0.843
D ₁₃	0.974	0.980	0.978	0.979	0.983	0.984	0.994	0.991	0.990	0.991
D ₁₄	0.703	0.717	0.710	0.704	0.705	0.717	0.723	0.734	0.729	0.728
D ₁₅	0.792	0.794	0.756	0.808	0.798	0.797	0.790	0.798	0.801	0.799
D ₁₆	0.823	0.832	0.801	0.830	0.825	0.844	0.833	0.836	0.830	0.830
D ₁₇	0.962	0.975	0.948	0.974	0.967	0.985	0.972	0.983	0.977	0.981
D ₁₈	0.664	0.674	0.682	0.680	0.674	0.686	0.687	0.699	0.700	0.698
D ₁₉	0.971	0.969	0.965	0.975	0.973	0.973	0.977	0.976	0.978	0.975
D ₂₀	0.783	0.767	0.772	0.785	0.776	0.794	0.840	0.834	0.837	0.825
D ₂₁	0.820	0.836	0.790	0.823	0.812	0.839	0.842	0.848	0.852	0.835

TABLE VI. COMPARATIVE ANALYSIS OF AVERAGE FITNESS VALUE OBTAINED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	0.162	0.165	0.144	0.101	0.218	0.140	0.101	0.063	0.101	0.103
D ₂	0.305	0.308	0.297	0.303	0.302	0.299	0.284	0.279	0.289	0.293
D ₃	0.251	0.192	0.252	0.190	0.129	0.192	0.128	0.250	0.251	0.253
D ₄	0.328	0.307	0.335	0.305	0.333	0.296	0.271	0.091	0.287	0.288
D ₅	0.269	0.267	0.271	0.264	0.273	0.258	0.255	0.252	0.255	0.252
D ₆	0.157	0.134	0.140	0.131	0.153	0.137	0.124	0.077	0.113	0.103
D ₇	0.242	0.225	0.238	0.220	0.219	0.200	0.192	0.197	0.195	0.225
D ₈	0.085	0.099	0.081	0.112	0.076	0.078	0.068	0.065	0.066	0.058
D ₉	0.329	0.308	0.307	0.329	0.316	0.296	0.262	0.276	0.271	0.278
D ₁₀	0.273	0.277	0.265	0.273	0.264	0.265	0.146	0.160	0.205	0.181
D ₁₁	0.173	0.175	0.169	0.199	0.155	0.186	0.138	0.137	0.151	0.184
D ₁₂	0.239	0.241	0.213	0.227	0.202	0.203	0.175	0.171	0.191	0.210
D ₁₃	0.053	0.040	0.031	0.041	0.050	0.030	0.017	0.037	0.018	0.027
D ₁₄	0.316	0.300	0.309	0.328	0.324	0.308	0.294	0.290	0.290	0.292
D ₁₅	0.229	0.244	0.230	0.217	0.227	0.231	0.236	0.224	0.226	0.230
D ₁₆	0.206	0.193	0.191	0.189	0.196	0.175	0.182	0.186	0.191	0.183
D ₁₇	0.070	0.047	0.057	0.050	0.053	0.030	0.043	0.027	0.043	0.033
D ₁₈	0.403	0.385	0.377	0.368	0.394	0.376	0.368	0.357	0.357	0.354
D ₁₉	0.053	0.052	0.043	0.036	0.042	0.043	0.034	0.037	0.037	0.038
D ₂₀	0.299	0.290	0.252	0.274	0.252	0.309	0.195	0.213	0.224	0.252
D ₂₁	0.229	0.209	0.228	0.200	0.233	0.220	0.203	0.191	0.175	0.200

TABLE VII. COMPARATIVE ANALYSIS OF WORST FITNESS VALUE OBTAINED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	0.043	0.044	0.046	0.028	0.046	0.004	0.003	0.024	0.005	0.025
D ₂	0.258	0.273	0.263	0.281	0.265	0.258	0.234	0.248	0.262	0.244
D ₃	0.005	0.067	0.007	0.066	0.003	0.067	0.003	0.003	0.003	0.005
D ₄	0.277	0.237	0.071	0.240	0.284	0.202	0.011	0.005	0.071	0.005
D ₅	0.225	0.246	0.221	0.221	0.221	0.241	0.205	0.217	0.209	0.219
D ₆	0.087	0.005	0.048	0.068	0.073	0.054	0.005	0.005	0.005	0.005
D ₇	0.150	0.141	0.132	0.153	0.158	0.131	0.125	0.117	0.153	0.137
D ₈	0.050	0.044	0.046	0.055	0.043	0.040	0.016	0.035	0.030	0.033
D ₉	0.242	0.250	0.227	0.228	0.234	0.231	0.189	0.205	0.211	0.204
D ₁₀	0.142	0.150	0.141	0.158	0.165	0.138	0.107	0.126	0.123	0.131
D ₁₁	0.106	0.101	0.082	0.118	0.121	0.090	0.081	0.083	0.073	0.101
D ₁₂	0.147	0.133	0.144	0.129	0.153	0.129	0.119	0.113	0.130	0.131
D ₁₃	0.006	0.017	0.006	0.008	0.007	0.017	0.003	0.003	0.005	0.004
D ₁₄	0.285	0.257	0.273	0.265	0.272	0.259	0.252	0.230	0.253	0.258
D ₁₅	0.196	0.190	0.182	0.177	0.190	0.190	0.188	0.183	0.187	0.184
D ₁₆	0.157	0.160	0.154	0.165	0.162	0.147	0.157	0.158	0.159	0.160
D ₁₇	0.021	0.014	0.013	0.015	0.026	0.017	0.020	0.015	0.020	0.015
D ₁₈	0.272	0.290	0.264	0.278	0.265	0.256	0.273	0.246	0.272	0.255
D ₁₉	0.025	0.024	0.028	0.024	0.024	0.020	0.023	0.020	0.020	0.021
D ₂₀	0.147	0.205	0.185	0.169	0.205	0.158	0.128	0.139	0.137	0.109
D ₂₁	0.129	0.144	0.157	0.169	0.163	0.130	0.136	0.138	0.134	0.151

TABLE VIII. COMPARATIVE ANALYSIS OF BEST FITNESS VALUE OBTAINED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	0.043	0.044	0.046	0.028	0.046	0.004	0.003	0.024	0.005	0.025
D ₂	0.258	0.273	0.263	0.281	0.265	0.258	0.234	0.248	0.262	0.244
D ₃	0.005	0.067	0.007	0.066	0.003	0.067	0.003	0.003	0.003	0.005
D ₄	0.277	0.237	0.071	0.240	0.284	0.202	0.011	0.005	0.071	0.005
D ₅	0.225	0.246	0.221	0.221	0.221	0.241	0.205	0.217	0.209	0.219
D ₆	0.087	0.005	0.048	0.068	0.073	0.054	0.005	0.005	0.005	0.005
D ₇	0.150	0.141	0.132	0.153	0.158	0.131	0.125	0.117	0.153	0.137
D ₈	0.050	0.044	0.046	0.055	0.043	0.040	0.016	0.035	0.030	0.033
D ₉	0.242	0.250	0.227	0.228	0.234	0.231	0.189	0.205	0.211	0.204
D ₁₀	0.142	0.150	0.141	0.158	0.165	0.138	0.107	0.126	0.123	0.131
D ₁₁	0.106	0.101	0.082	0.118	0.121	0.090	0.081	0.083	0.073	0.101
D ₁₂	0.147	0.133	0.144	0.129	0.153	0.129	0.119	0.113	0.130	0.131
D ₁₃	0.006	0.017	0.006	0.008	0.007	0.017	0.003	0.003	0.005	0.004
D ₁₄	0.285	0.257	0.273	0.265	0.272	0.259	0.252	0.230	0.253	0.258
D ₁₅	0.196	0.190	0.182	0.177	0.190	0.190	0.188	0.183	0.187	0.184
D ₁₆	0.157	0.160	0.154	0.165	0.162	0.147	0.157	0.158	0.159	0.160
D ₁₇	0.021	0.014	0.013	0.015	0.026	0.017	0.020	0.015	0.020	0.015
D ₁₈	0.272	0.290	0.264	0.278	0.265	0.256	0.273	0.246	0.272	0.255
D ₁₉	0.025	0.024	0.028	0.024	0.024	0.020	0.023	0.020	0.020	0.021
D ₂₀	0.147	0.205	0.185	0.169	0.205	0.158	0.128	0.139	0.137	0.109
D ₂₁	0.129	0.144	0.157	0.169	0.163	0.130	0.136	0.138	0.134	0.151

TABLE IX. COMPARATIVE ANALYSIS OF STANDARD DEVIATION VALUE

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	0.037	0.043	0.030	0.022	0.049	0.041	0.036	0.016	0.030	0.030
D ₂	0.012	0.010	0.011	0.007	0.011	0.013	0.014	0.010	0.009	0.016
D ₃	0.076	0.057	0.063	0.041	0.049	0.039	0.059	0.093	0.077	0.072
D ₄	0.015	0.027	0.095	0.020	0.018	0.030	0.093	0.033	0.088	0.099
D ₅	0.014	0.007	0.013	0.012	0.015	0.006	0.017	0.013	0.015	0.010
D ₆	0.024	0.036	0.026	0.019	0.025	0.025	0.040	0.022	0.046	0.033
D ₇	0.033	0.026	0.029	0.022	0.022	0.022	0.023	0.026	0.013	0.029
D ₈	0.011	0.016	0.009	0.017	0.011	0.011	0.015	0.009	0.013	0.009
D ₉	0.029	0.020	0.024	0.028	0.026	0.017	0.019	0.020	0.022	0.020
D ₁₀	0.049	0.037	0.042	0.034	0.032	0.036	0.012	0.011	0.024	0.015
D ₁₁	0.021	0.027	0.029	0.026	0.010	0.024	0.020	0.016	0.025	0.022
D ₁₂	0.028	0.034	0.026	0.027	0.014	0.022	0.019	0.016	0.020	0.025
D ₁₃	0.015	0.009	0.008	0.011	0.013	0.006	0.006	0.010	0.004	0.009
D ₁₄	0.011	0.013	0.010	0.018	0.015	0.015	0.015	0.019	0.010	0.012
D ₁₅	0.008	0.019	0.014	0.012	0.012	0.016	0.013	0.014	0.012	0.015
D ₁₆	0.016	0.013	0.012	0.009	0.012	0.010	0.008	0.009	0.010	0.007
D ₁₇	0.014	0.010	0.014	0.011	0.009	0.005	0.008	0.004	0.008	0.006
D ₁₈	0.047	0.034	0.037	0.032	0.041	0.036	0.025	0.036	0.032	0.029
D ₁₉	0.009	0.007	0.005	0.004	0.006	0.008	0.003	0.006	0.006	0.006
D ₂₀	0.050	0.024	0.020	0.031	0.014	0.045	0.024	0.020	0.025	0.040
D ₂₁	0.029	0.018	0.021	0.011	0.020	0.024	0.020	0.017	0.013	0.016

TABLE X. COMPARATIVE ANALYSIS OF NO. OF FEATURES SELECTED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	10.3	11.3	11.3	10.5	9.4	11.1	6.7	7.8	8.3	8.3
D ₂	11.3	11.7	10	12	9.8	12.8	6.8	7	9.3	8.5
D ₃	24	32.4	27.2	27	24.8	33.4	18.6	17	23	26.6
D ₄	10	10.3	8.8	9.9	10	9.6	7.1	6.3	7.9	7.3
D ₅	6.7	7.8	5.1	6	8.9	5.6	3.5	2.7	3.6	4.9
D ₆	10.9	10.1	10.1	10.2	11.8	9.7	6.8	6.6	7.6	8
D ₇	5.2	5.6	6.4	5.8	4.2	7.7	4.1	3.8	4.9	4.2
D ₈	8.7	7.6	9	8.1	8.6	9.6	4.5	6.6	7.2	8
D ₉	11.1	14.2	12.8	12.4	11.8	15.9	8.9	8.5	11.2	8.8
D ₁₀	6	6.8	6.9	6.2	6.1	6.5	3.7	4.2	4.7	5.2
D ₁₁	13.9	20.2	19	18.9	16.4	21	9	9.7	16.8	13.4
D ₁₂	19	22.6	21.4	20.7	17.2	25.9	14.4	15.2	18.4	17.8
D ₁₃	7.4	8.3	9.3	9.1	8.2	8.2	5.8	5.5	6.8	6.5
D ₁₄	3.8	4.9	5	4.3	5.1	5.1	2.3	3.1	2.6	4.5
D ₁₅	9	8.8	8.7	8.4	9	8	5.8	5.9	7.1	7.2
D ₁₆	19.5	18.8	17.5	19	19.4	17.8	13.7	13.3	13.9	13.6
D ₁₇	26.6	26.1	26.2	26.4	27.9	26.8	14.5	16.1	20.2	18.7
D ₁₈	4.9	6.3	5.7	6.3	5.5	5.8	4.3	4.7	4.3	4.5
D ₁₉	6	6.7	6.2	5.5	4.8	5.8	4.8	5	4.6	4.8
D ₂₀	32.5	28.1	32.7	31	28.9	42.9	22	23.9	30	26.9
D ₂₁	9.4	9.5	8.7	9.5	8.7	9.6	7.7	7.7	8.2	8.3

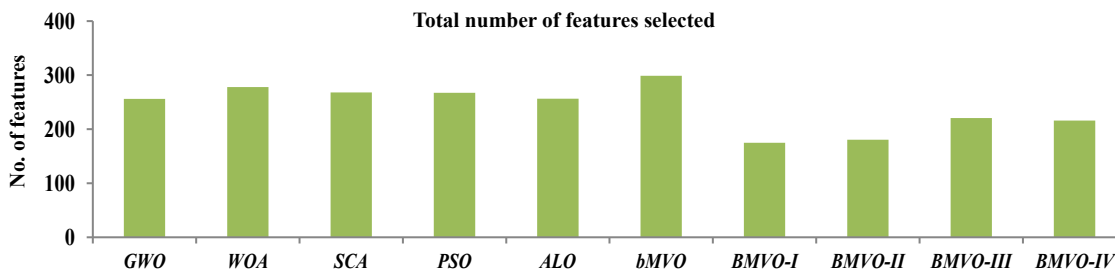


Fig 2. Graphical representation of No. of Features selected Averaged on all datasets by all algorithms.

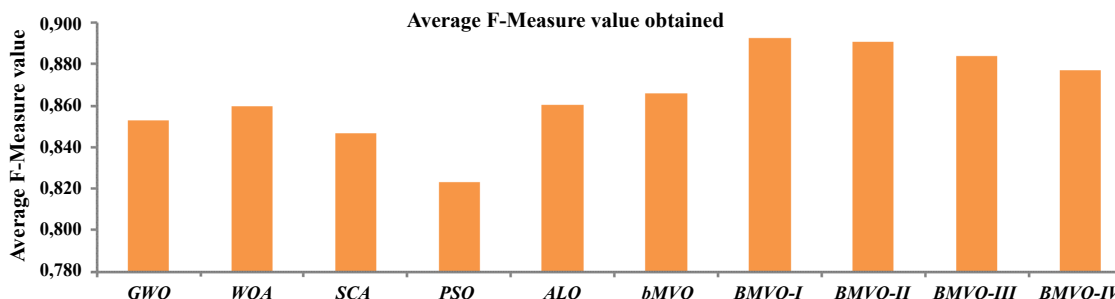


Fig 3. Graphical representation of average of cumulative sum of F-Measure values (average F-measure values on all the datasets).

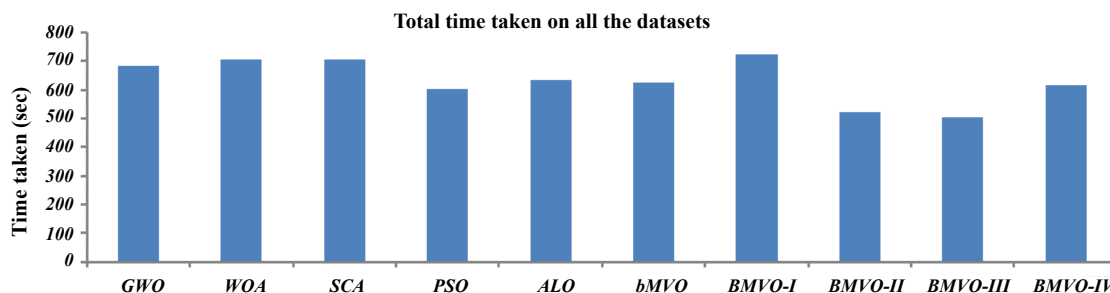


Fig 4. Graphical representation of cumulative sum of average time taken per run by an algorithm on all the datasets.

TABLE XI. WILCOXON SIGNED RANK TEST FOR THE AVERAGE FITNESS OBTAINED BY THE DIFFERENT ALGORITHMS

	GWO	WOA	SCA	PSO	ALO	bMVO	BMVO-I	BMVO-II	BMVO-III	BMVO-IV
bMVO	0.00014	0.00034	0.01352	0.00124	0.00614	-	0.0009	0.00012	0.00058	0.01016
BMVO-I	0	0.0001	0.00014	0.00038	0.0001	0.0009	-	0.8493	0.04236	0.00374
BMVO-II	0	0	8.00E-05	0.00014	0.00032	0.00012	0.8493	-	0.04338	0.00016
BMVO-III	0	8.00E-05	8.00E-05	0.00014	0.00001	0.00058	0.04236	0.04338	-	0.0466
BMVO-IV	6.00E-05	0.00058	0.00116	0.00158	.00116	0.01016	0.00374	0.00016	0.0466	-

optimization algorithm into its binary variants and are applied on one of the important problems of machine learning, i.e. feature selection. These binary variants make use of different types of transformation functions from the class of V-shaped as well as S-shaped transformation functions, which defines the likelihood of altering a position vector's elements from 0 to 1 and vice versa. Their performance is compared against the six most recent metaheuristic algorithms used for feature selection viz. Grey wolf Optimization (GWO), Whale Optimization Algorithm (WOA), Sine Cosine Optimization algorithm (SCA), Particle Swarm Optimization (PSO) Ant lion Optimization (ALO), and Multi-verse Optimization Algorithm (MVO) and their performance are analyzed on the basis of various evaluation criteria. To cope with the stochastic nature of the metaheuristic algorithms the results from

all the algorithms were averaged on ten different runs. The obtained evaluation results clearly indicate that the proposed variants of BMVO outperform the existing metaheuristic to find the solution of feature subset selection problem. Furthermore, it can be seen that the performance of all the variants of BMVO is better than algorithms taken for comparison but the performance of BMVO versions based on s-shaped(sigmoid and modified sigmoid) transformation function is better than the performance of variants based on V-shaped(tangent hyperbolic and inverse tangent) transformation function.

Furthermore, the MVO makes use of selection mechanism which is run for every variable in every universe over the iterations also the use of sorting algorithm in every iteration, along with these two basic operations the binary MVO makes use of transfer functions for the

conversion of continuous version of MVO to binary version which may lead to increase in the time complexity of the algorithm as compared to other binary algorithms considered for the comparison which can be gauged from the graph shown in Fig. 4.

In the future, the proposed binary version of the Multi-verse optimization algorithm can be used to explore for the solutions of other binary optimization problems, which may lead to an improvement in the results. Moreover, this work can be enhanced further by developing other different variants of MVO based on other transformation functions or combining the MVO algorithms with local search approaches like simulated annealing and tabu search for feature selection problem. Lastly, the hybrid versions of MVO can be developed by hybridizing MVO with some other metaheuristic algorithms.

Similarly Table XII and Table XIII shows the detailed comparative analysis of maximum and minimum classification accuracies obtained by different versions of BMVO with GWO, WOA, SCA and original MVO (without incorporating transformation functions) for feature selection in all the 10 runs, this measure gives an idea of variation of outcome in all the runs.

1. **Maximum Accuracy-** This criterion gives the highest percentage of test samples correctly classified among all the runs, where the function max calculates the maximum accuracy from all runs.
2. **Minimum Accuracy-** This measure gives the lowest percentage of test samples correctly classified among all N the runs, where the function min calculates the minimum accuracy from all runs.

TABLE XII. COMPARATIVE ANALYSIS OF MAXIMUM CLASSIFICATION ACCURACY OBTAINED

MAX	GWO	WOA	SCA	PSO	ALO	MVO	BMVO-I	BMVO-II	BMVO-III	BMVO-IV
D ₁	0.961	0.961	0.941	0.980	0.961	1.000	1.000	0.980	1.000	0.980
D ₂	0.748	0.732	0.730	0.724	0.736	0.744	0.768	0.752	0.740	0.758
D ₃	1.000	0.938	0.938	0.938	1.000	0.938	1.000	1.000	1.000	1.000
D ₄	0.730	0.768	0.894	0.764	0.722	0.802	0.994	1.000	0.934	1.000
D ₅	0.782	0.762	0.778	0.780	0.780	0.762	0.794	0.782	0.790	0.784
D ₆	0.918	1.000	0.864	0.938	0.932	0.952	1.000	1.000	1.000	1.000
D ₇	0.850	0.864	0.871	0.850	0.844	0.871	0.878	0.884	0.850	0.864
D ₈	0.953	0.960	0.947	0.953	0.960	0.967	0.987	0.967	0.973	0.973
D ₉	0.761	0.754	0.776	0.776	0.769	0.776	0.813	0.799	0.791	0.799
D ₁₀	0.861	0.852	0.861	0.843	0.838	0.864	0.896	0.875	0.878	0.872
D ₁₁	0.898	0.903	0.920	0.886	0.881	0.915	0.920	0.921	0.932	0.903
D ₁₂	0.858	0.874	0.858	0.874	0.851	0.877	0.885	0.889	0.874	0.874
D ₁₃	1.000	0.989	0.989	1.000	1.000	0.989	1.000	1.000	1.000	1.000
D ₁₄	0.716	0.747	0.723	0.736	0.729	0.740	0.747	0.771	0.747	0.747
D ₁₅	0.812	0.816	0.785	0.831	0.818	0.818	0.820	0.820	0.820	0.825
D ₁₆	0.852	0.848	0.832	0.842	0.846	0.860	0.848	0.848	0.846	0.844
D ₁₇	0.984	0.995	0.967	0.995	0.984	0.989	0.984	0.989	0.984	0.989
D ₁₈	0.729	0.710	0.738	0.729	0.738	0.748	0.729	0.757	0.729	0.748
D ₁₉	0.980	0.980	0.974	0.983	0.983	0.986	0.983	0.986	0.986	0.983
D ₂₀	0.856	0.798	0.817	0.837	0.798	0.846	0.875	0.865	0.865	0.894
D ₂₁	0.875	0.861	0.820	0.836	0.840	0.877	0.869	0.867	0.871	0.855

TABLE XIII. COMPARATIVE ANALYSIS OF MINIMUM CLASSIFICATION ACCURACY OBTAINED

MIN	GWO	WOA	SCA	PSO	ALO	MVO	BMVO-I	BMVO-II	BMVO-III	BMVO-IV
D ₁	0.843	0.843	0.863	0.902	0.784	0.863	0.902	0.941	0.902	0.902
D ₂	0.696	0.698	0.704	0.698	0.702	0.706	0.716	0.722	0.712	0.708
D ₃	0.750	0.813	0.750	0.813	0.875	0.813	0.875	0.750	0.750	0.750
D ₄	0.672	0.698	0.664	0.700	0.674	0.710	0.732	0.914	0.716	0.712
D ₅	0.734	0.734	0.730	0.740	0.730	0.742	0.746	0.750	0.746	0.752
D ₆	0.852	0.874	0.788	0.876	0.856	0.870	0.882	0.928	0.894	0.904
D ₇	0.762	0.776	0.762	0.782	0.782	0.803	0.810	0.803	0.810	0.776
D ₈	0.920	0.907	0.920	0.893	0.927	0.927	0.933	0.940	0.940	0.947
D ₉	0.672	0.694	0.694	0.672	0.687	0.709	0.739	0.724	0.731	0.724
D ₁₀	0.730	0.725	0.725	0.730	0.739	0.739	0.855	0.841	0.797	0.820
D ₁₁	0.830	0.830	0.835	0.807	0.847	0.818	0.864	0.864	0.852	0.818
D ₁₂	0.766	0.762	0.782	0.774	0.801	0.801	0.828	0.831	0.812	0.793
D ₁₃	0.955	0.966	0.966	0.966	0.955	0.978	0.989	0.966	0.989	0.978
D ₁₄	0.685	0.702	0.695	0.671	0.678	0.695	0.705	0.712	0.712	0.709
D ₁₅	0.779	0.764	0.733	0.789	0.781	0.777	0.766	0.781	0.777	0.777
D ₁₆	0.802	0.814	0.776	0.818	0.812	0.832	0.822	0.818	0.814	0.822
D ₁₇	0.934	0.962	0.923	0.956	0.956	0.978	0.962	0.978	0.962	0.973
D ₁₈	0.598	0.617	0.626	0.636	0.607	0.626	0.636	0.645	0.645	0.645
D ₁₉	0.954	0.954	0.957	0.969	0.963	0.963	0.971	0.969	0.969	0.966
D ₂₀	0.702	0.712	0.750	0.731	0.750	0.692	0.808	0.788	0.779	0.750
D ₂₁	0.776	0.798	0.756	0.806	0.774	0.786	0.802	0.812	0.830	0.804

TABLE XIV. COMPARATIVE ANALYSIS OF AVERAGE TIME TAKEN PER RUN (100 ITERATIONS) OBTAINED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	20.11	27.63	18.04	17.04	27.17	23.13	47.24	23.21	19.30	26.47
D ₂	45.68	39.64	46.78	36.45	54.48	38.06	47.83	26.02	42.25	71.19
D ₃	24.16	31.42	21.01	16.55	25.53	17.27	32.60	22.43	17.91	36.06
D ₄	29.88	34.47	33.96	31.28	39.47	31.57	40.00	28.17	26.43	30.86
D ₅	29.90	59.59	35.21	34.12	33.56	30.77	38.56	27.64	30.38	32.15
D ₆	30.32	51.98	34.59	30.48	35.64	31.53	34.02	34.25	27.88	27.95
D ₇	22.81	32.75	17.41	17.31	18.93	18.16	20.35	25.76	16.94	16.83
D ₈	20.63	35.64	18.59	17.18	20.09	17.42	20.57	21.20	17.21	18.90
D ₉	31.01	38.13	23.55	16.96	24.50	17.27	20.49	24.13	16.78	17.78
D ₁₀	32.08	41.62	30.41	20.58	29.24	23.74	22.85	27.60	20.19	20.37
D ₁₁	38.44	22.81	30.25	30.93	26.58	19.95	33.35	19.00	18.80	19.81
D ₁₂	35.91	34.98	31.25	35.05	31.04	29.72	20.05	22.29	22.64	25.52
D ₁₃	39.84	23.05	23.35	16.75	19.24	20.28	24.24	19.51	18.13	23.56
D ₁₄	39.96	20.45	24.37	27.65	21.19	20.05	22.97	22.75	21.13	20.30
D ₁₅	47.50	31.22	64.10	37.42	42.66	39.45	33.84	23.74	23.58	24.15
D ₁₆	46.19	36.67	72.19	47.06	45.80	39.59	55.11	28.17	32.25	32.41
D ₁₇	25.59	22.91	35.41	33.33	26.77	22.96	55.86	22.29	27.67	28.99
D ₁₈	22.38	36.31	34.33	29.79	20.54	20.59	39.84	20.34	22.76	41.20
D ₁₉	28.39	25.25	38.92	28.24	25.84	22.09	37.21	18.65	20.01	32.16
D ₂₀	22.42	18.64	31.43	16.87	25.71	17.69	24.97	18.48	30.94	25.07
D ₂₁	51.97	41.71	42.70	61.94	42.17	122.26	51.23	45.43	31.66	42.77

TABLE XV. COMPARATIVE ANALYSIS OF F-MEASURE VALUES OBTAINED

	<i>GWO</i>	<i>WOA</i>	<i>SCA</i>	<i>PSO</i>	<i>ALO</i>	<i>bMVO</i>	<i>BMVO-I</i>	<i>BMVO-II</i>	<i>BMVO-III</i>	<i>BMVO-IV</i>
D ₁	0.980	0.994	0.959	0.963	0.965	0.978	0.973	0.993	0.993	0.985
D ₂	0.814	0.812	0.811	0.795	0.807	0.812	0.828	0.825	0.819	0.817
D ₃	0.705	0.746	0.721	0.552	0.821	0.747	0.924	0.796	0.861	0.765
D ₄	0.800	0.813	0.813	0.795	0.806	0.820	0.913	0.981	0.854	0.867
D ₅	0.856	0.851	0.853	0.837	0.857	0.855	0.865	0.865	0.865	0.864
D ₆	0.837	0.876	0.763	0.836	0.836	0.882	0.958	0.969	0.933	0.914
D ₇	0.849	0.864	0.859	0.809	0.860	0.874	0.888	0.880	0.870	0.873
D ₈	0.929	0.927	0.926	0.892	0.935	0.942	0.953	0.946	0.945	0.949
D ₉	0.764	0.774	0.785	0.748	0.777	0.793	0.812	0.802	0.802	0.799
D ₁₀	0.822	0.802	0.815	0.731	0.816	0.839	0.879	0.870	0.859	0.857
D ₁₁	0.904	0.904	0.907	0.883	0.906	0.901	0.926	0.919	0.927	0.900
D ₁₂	0.891	0.897	0.892	0.859	0.895	0.900	0.915	0.912	0.906	0.907
D ₁₃	0.980	0.986	0.980	0.979	0.989	0.986	0.995	0.995	0.991	0.991
D ₁₄	0.806	0.816	0.809	0.786	0.805	0.814	0.825	0.829	0.825	0.820
D ₁₅	0.856	0.857	0.827	0.861	0.862	0.859	0.848	0.854	0.860	0.858
D ₁₆	0.780	0.789	0.745	0.779	0.780	0.809	0.787	0.791	0.781	0.784
D ₁₇	0.996	0.999	0.990	0.953	0.995	0.999	0.993	1.000	0.997	0.998
D ₁₈	0.698	0.698	0.725	0.678	0.721	0.713	0.735	0.740	0.755	0.751
D ₁₉	0.978	0.976	0.973	0.976	0.979	0.979	0.982	0.982	0.983	0.981
D ₂₀	0.743	0.728	0.734	0.678	0.739	0.753	0.819	0.808	0.809	0.794
D ₂₁	0.918	0.945	0.892	0.903	0.912	0.933	0.929	0.942	0.927	0.949

REFERENCES

- [1] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.
- [2] Hancer, E., Xue, B., Karaboga, D., & Zhang, M. (2015). A binary ABC algorithm based on advanced similarity scheme for feature selection. *Applied Soft Computing*, 36, 334-348.
- [3] Sayed, G. I., Hassanien, A. E., & Azar, A. T. (2017). Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications*, 1-18.
- [4] Inbarani, H. H., Azar, A. T., & Jothi, G. (2014). Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. *Computer methods and programs in biomedicine*, 113(1), 175-185.
- [5] Mlakar, U., Fister, I., Brest, J., & Potočnik, B. (2017). Multi-objective differential evolution for feature selection in facial expression recognition systems. *Expert Systems with Applications*, 89, 129-137.
- [6] Sweetlin, J. D., Nehemiah, H. K., & Kannan, A. (2017). Feature selection using ant colony optimization with tandem-run recruitment to diagnose bronchitis from CT scan images. *Computer methods and programs in biomedicine*, 145, 115-125.
- [7] Jain, I., Jain, V. K., & Jain, R. (2018). Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification. *Applied Soft Computing*, 62, 203-215.
- [8] Adeli, A., & Broumandnia, A. (2018). Image steganalysis using improved particle swarm optimization based feature selection. *Applied Intelligence*, 48(6), 1609-1622.
- [9] Lin, K. C., Zhang, K. Y., Huang, Y. H., Hung, J. C., & Yen, N. (2016). Feature selection based on an improved cat swarm optimization algorithm for big data classification. *The Journal of Supercomputing*, 72(8), 3210-3221.
- [10] Chen, L. F., Su, C. T., Chen, K. H., & Wang, P. C. (2012). Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis. *Neural Computing and Applications*, 21(8), 2087-2096.
- [11] Shang, L., Zhou, Z., & Liu, X. (2016). Particle swarm optimization-based feature selection in sentiment classification. *Soft Computing*, 20(10), 3821-3834.
- [12] ElAlami, M. E. (2011). A novel image retrieval model based on the most relevant features. *Knowledge-Based Systems*, 24(1), 23-32.
- [13] Zawbaa, H. M., Emary, E., Parv, B., & Sharawi, M. (2016, July). Feature selection approach based on moth-flame optimization algorithm. In *Evolutionary Computation (CEC), 2016 IEEE Congress on* (pp. 4612-4617). IEEE.
- [14] Ghaemi, M., & Feizi-Derakhshi, M. R. (2016). Feature selection using forest optimization algorithm. *Pattern Recognition*, 60, 121-129.
- [15] Yang, X. S., Deb, S., & Fong, S. (2014). Metaheuristic algorithms: optimal balance of intensification and diversification. *Applied Mathematics & Information Sciences*, 8(3), 977.
- [16] Mafarja, M. M., & Mirjalili, S. (2017). Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. *Neurocomputing*, 260, 302-312.
- [17] Chuang, L. Y., Chang, H. W., Tu, C. J., & Yang, C. H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry*, 32(1), 29-38.
- [18] Chen, Y. P., Li, Y., Wang, G., Zheng, Y. F., Xu, Q., Fan, J. H., & Cui, X. T. (2017). A novel bacterial foraging optimization algorithm for feature selection. *Expert Systems with Applications*, 83, 1-17.
- [19] Babatunde, O. H., Armstrong, L., Leng, J., & Diepeveen, D. (2014). A genetic algorithm-based feature selection. *International Journal of Electronics Communication and Computer Engineering*, 5(4), 899-905.
- [20] Ghamisi, P., & Benediktsson, J. A. (2015). Feature selection based on hybridization of genetic algorithm and particle swarm optimization. *IEEE Geoscience and Remote Sensing Letters*, 12(2), 309-313.
- [21] Wang, G., Chu, H. E., Zhang, Y., Chen, H., Hu, W., Li, Y., & Peng, X. (2015). Multiple parameter control for ant colony optimization applied to feature selection problem. *Neural Computing and Applications*, 26(7), 1693-1708.
- [22] Ahmad, F., Isa, N. A. M., Hussain, Z., Osman, M. K., & Sulaiman, S. N. (2015). A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. *Pattern Analysis and Applications*, 18(4), 861-870.
- [23] Chen, Y., Miao, D., & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letters*, 31(3), 226-233.
- [24] Eardley DM (1974) Death of white holes in the early Universe. *Phys Rev Lett* 33:442.
- [25] Davies PC (1978) Thermodynamics of black holes. *Rep Prog Phys* 41:1313.
- [26] Morris MS, Thorne KS (1988) Wormholes in spacetime and their use for interstellar travel: a tool for teaching general relativity. *Am J Phys* 56:395-412.
- [27] Mirjalili, S., Mirjalili, S. M., & Hatamlou, A. (2016). Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Computing and Applications*, 27(2), 495-513.
- [28] Guth AH (2007) Eternal inflation and its implications. *J Phys A Math Theor* 40:6811.
- [29] Steinhardt PJ, Turok N (2005) The cyclic model simplified. *New Astron Rev* 49:43-57.
- [30] Chuang, L. Y., Yang, C. H., & Li, J. C. (2011). Chaotic maps based on binary particle swarm optimization for feature selection. *Applied Soft Computing*, 11(1), 239-248.
- [31] Banati, H., & Bajaj, M. (2011). Fire fly based feature selection approach. *IJCSI International Journal of Computer Science Issues*, 8(4), 473-480.
- [32] Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An improved particle swarm optimization for feature selection. *Journal of Bionic Engineering*, 8(2), 191-200.
- [33] Nakamura, R. Y., Pereira, L. A., Costa, K. A., Rodrigues, D., Papa, J. P., & Yang, X. S. (2012, August). BBA: a binary bat algorithm for feature selection. In 2012 25th SIBGRAPI conference on graphics, Patterns and Images (pp. 291-297). IEEE.
- [34] Han, X., Chang, X., Quan, L., Xiong, X., Li, J., Zhang, Z., & Liu, Y. (2014). Feature subset selection by gravitational search algorithm optimization. *Information Sciences*, 281, 128-146.
- [35] Emary, E., Zawbaa, H. M., Ghany, K. K. A., Hassanien, A. E., & Parv, B. (2015, September). Firefly optimization algorithm for feature selection. In *Proceedings of the 7th Balkan Conference on Informatics Conference* (p. 26). ACM.
- [36] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371-381.
- [37] Wang, H., & Niu, B. (2017). A novel bacterial algorithm with randomness control for feature selection in classification. *Neurocomputing*, 228, 176-186.
- [38] Mafarja, M., Eleyan, D., Abdullah, S., & Mirjalili, S. (2017, July). S-shaped vs. V-shaped transfer functions for ant lion optimization algorithm in feature selection problem. In *Proceedings of the International Conference on Future Networks and Distributed Systems* (p. 14). ACM.
- [39] Ewees, A. A., El Aziz, M. A., & Hassanien, A. E. (2019). Chaotic multi-verse optimizer-based feature selection. *Neural Computing and Applications*, 31(4), 991-1006.
- [40] Shunmugapriya, P., & Kanmani, S. (2017). A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm and Evolutionary Computation*, 36, 27-36.
- [41] Aladeemy, M., Tutun, S., & Khasawneh, M. T. (2017). A new hybrid approach for feature selection and support vector machine model selection based on self-adaptive cohort intelligence. *Expert Systems with Applications*, 88, 118-131.
- [42] Nagpal, S., Arora, S., & Dey, S. (2017). Feature Selection using Gravitational Search Algorithm for Biomedical Data. *Procedia Computer Science*, 115, 258-265.
- [43] Faris, H., Hassonah, M. A., Ala'M, A. Z., Mirjalili, S., & Aljarah, I. (2018). A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. *Neural Computing and Applications*, 30(8), 2355-2369.
- [44] Aljarah, I., Ala'M, A. Z., Faris, H., Hassonah, M. A., Mirjalili, S., & Saadeh, H. (2018). Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm. *Cognitive Computation*, 10(3), 478-495.
- [45] Mafarja, M., & Mirjalili, S. (2018). Whale optimization approaches for wrapper feature selection. *Applied Soft Computing*, 62, 441-453.
- [46] Suganthi, M., & Karunakaran, V. (2018). Instance selection and feature

extraction using cuttlefish optimization algorithm and principal component analysis using decision tree. *Cluster Computing*, 1-13.

- [47] Dong, H., Li, T., Ding, R., & Sun, J. (2018). A Novel Hybrid Genetic Algorithm with Granular Information for Feature Selection and Optimization. *Applied Soft Computing*, 65, 33-46.
- [48] Mirjalili, S., & Lewis, A. (2013). S-shaped versus V-shaped transfer functions for binary particle swarm optimization. *Swarm and Evolutionary Computation*, 9, 1-14.
- [49] Saremi, S., Mirjalili, S., & Lewis, A. (2015). How important is a transfer function in discrete heuristic algorithms. *Neural Computing and Applications*, 26(3), 625-640.
- [50] Xue, B., Zhang, M., & Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE transactions on cybernetics*, 43(6), 1656-1671.
- [51] Hall, M. A. (1999). Correlation-based feature selection for machine learning.
- [52] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
- [53] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary ant lion approaches for feature selection. *Neurocomputing*, 213, 54-65.
- [54] Asuncion, A., & Newman, D. (2007). UCI machine learning repository.
- [55] Jensen, R., & Shen, Q. (2008). *Computational intelligence and feature selection: rough and fuzzy approaches* (Vol. 8). John Wiley & Sons.
- [56] Derrac, J., García, S., Molina, D., & Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1), 3-18
- [57] Reddy, K. S., Panwar, L. K., Panigrahi, B. K., & Kumar, R. (2018). A New Binary Variant of Sine-Cosine Algorithm: Development and Application to Solve Profit-Based Unit Commitment Problem. *Arabian Journal for Science and Engineering*, 43(8), 4041-4056.
- [58] Faris, H., Hassonah, M. A., Ala'M, A. Z., Mirjalili, S., & Aljarah, I. (2018). A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. *Neural Computing and Applications*, 30(8), 2355-2369.
- [59] Sohrabi, M. K., & Tajik, A. (2017). Multi-objective feature selection for warfarin dose prediction. *Computational biology and chemistry*, 69, 126-133.
- [60] Zouache, D., & Abdelaziz, F. B. (2018). A cooperative swarm intelligence algorithm based on quantum-inspired and rough sets for feature selection. *Computers & Industrial Engineering*, 115, 26-36.
- [61] Tamimi, E., Ebadi, H., & Kiani, A. (2017). Evaluation of different metaheuristic optimization algorithms in feature selection and parameter determination in SVM classification. *Arabian Journal of Geosciences*, 10(22), 478.
- [62] Tiwari, S., Singh, B., & Kaur, M. (2017). An approach for feature selection using local searching and global optimization techniques. *Neural Computing and Applications*, 28(10), 2915-2930.
- [63] Ekbal, A., & Saha, S. (2016). Simultaneous feature and parameter selection using multiobjective optimization: application to named entity recognition. *International Journal of Machine Learning and Cybernetics*, 7(4), 597-611.
- [64] Lin, S. W., Chen, S. C., Wu, W. J., & Chen, C. H. (2009). Parameter determination and feature selection for back-propagation network by particle swarm optimization. *Knowledge and Information Systems*, 21(2), 249-266.
- [65] Vijaya, J., & Sivasankar, E. (2017). An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing. *Cluster Computing*, 1-12.
- [66] Manikandan, R. P. S., & Kalpana, A. M. (2017). Feature selection using fish swarm optimization in big data. *Cluster Computing*, 1-13.
- [67] Emary, E., & Zawbaa, H. M. (2019). Feature selection via Lévy Antlion optimization. *Pattern Analysis and Applications*, 22(3), 857-876.
- [68] Tegmark, M. (2004). Barrow, JD Davies, PC Harper, CL, Jr eds. *Science and Ultimate Reality Cambridge University Press Cambridge*.
- [69] Rashedi, E., Nezamabadi-Pour, H., & Saryzadi, S. (2010). BGSA: binary gravitational search algorithm. *Natural Computing*, 9(3), 727-745.
- [70] Bagga, P., Hans, R., & Sharma, V. (2017). N-grams Based Supervised Machine Learning Model for Mobile Agent Platform Protection against Unknown Malicious Mobile Agents. *International Journal of Interactive*

Multimedia and Artificial Intelligence, 4(6), 33-39.

- [71] H. M. Keerthi Kumar, B. S. Harish. Automatic Irony Detection using Feature Fusion and Ensemble Classifier, *International Journal of Interactive Multimedia and Artificial Intelligence*, (2019).
- [72] Revanasiddappa, M. B., & Harish, B. S. (2018). A new feature selection method based on intuitionistic fuzzy entropy to categorize text documents. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(3), 106-117.
- [73] Pujari, D., Yakkundimath, R., & Byadgi, A. S. (2016). SVM and ANN based classification of plant diseases using feature reduction technique. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(7), 6-14.
- [74] Belkhdja, L., & Hamdadou, D. (2019). IMCAD: Computer Aided System for Breast Masses Detection based on Immune Recognition. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(5), 97-108.



Rahul Hans

Rahul Hans has received his B.Tech degree in Computer Science and Engineering from Punjab Technical University, Jalandhar and M.Tech. degree in Computer Science and Engineering from Guru Nanak Dev University, Amritsar. He is currently pursuing Ph.D. from Guru Nanak Dev University, Amritsar. His predominant research areas include Mobile Agent Systems, Machine Learning and Metaheuristics. He has published research papers in various reputed international conferences and journals.



Harjot Kaur

Harjot Kaur is presently working as Assistant Professor in the Department of Computer Science and Engineering at Guru Nanak Dev University, Regional Campus Gurdaspur, Punjab, India. She has completed her Ph.D. degree from the Department of Computer Science and Engineering, Guru Nanak Dev University, Amritsar, Punjab, India. Her predominant area of research is Distributed Artificial Intelligence. She has published research papers in various reputed International Journals. Also she has attended various national and international Conferences and presented many research papers.