# Chaotic African Vultures Optimization Algorithm for Feature Selection

**Wy-Liang Cheng[1], Li Pan[1], Mohd Rizon Bin Mohamed Juhari[1], Chin Hong Wong[2,3], Abhishek Sharma[4], Tiong Hoo Lim[5], Sew Sun Tiang[1,*], Wei Hong Lim[1,*]**

[1]*Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia*
[2]*Maynooth International Engineering College, Maynooth University, Maynooth, Co Kildare, Ireland*
[3]*Maynooth International Engineering College, Fuzhou University, Fujian, 350116, China*
[4]*Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun 248002, India*
[5]*Faculty of Engineering, Universiti Teknologi Brunei, Bandar Seri Begawan 1410, Brunei Darussalam*
*E-mail: 1001437466@ucsiuniversity.edu.my, 1002060534@ucsiuniversity.edu.my, mohdrizon@ucsiuniversity.edu.my, chinhong.wong@mu.ie, abhishek15491@gmail.com, lim.tiong.hoo@utb.edu.bn, tiangss@ucsiuniversity.edu.my limwh@ucsiuniversity.edu.my*

**Abstract**

Feature selection is a widely used technique to remove the undesirable, noisy and inaccurate information from raw input dataset while maintaining the accuracy and efficiency of classifier. Tremendous researches have explored the feasibility of metaheuristic search algorithms (MSAs) such as African Vultures Optimization Algorithm (AVOA) to solve feature selection problem. Similar with many original MSAs, the conventional initialization scheme of AVOA has undesirable drawbacks that can lead to entrapment of local optima, especially when dealing with complex dataset. In this paper, a new variant known as Chaotic African Vultures Optimization Algorithm (CAVOA) is proposed to solve feature selection problem with enhanced classification accuracy by incorporating the chaotic map concept into the initialization scheme. Twelve datasets obtained from UCI Machine Learning Repository are used to investigate the capability of CAVOA in feature selection and compared with four other peer algorithms. Simulation results show that CAVOA can produce the best classification accuracies and lowest feature numbers in most datasets.

*Keywords*: African Vultures Optimization Algorithm, feature selection, metaheuristic search algorithm

## 1. Introduction

Enormous growth of dataset is observed in the era of Big Data due to the generation of datasets from various domains of sources. Complexity of datasets has increased significantly due to the presence of unwanted and noisy data, leading to “curse of dimensionality” issue that can compromise the effectiveness of classifier. It remains a challenging issue to eliminate unwanted information from raw input datasets while maintaining the accuracy and computation time of classifier.

Feature selection[1] is a popular technique used to eliminate redundant information from raw datasets when solving real-world problems such as fault detection[2],[3] and automatic modulation recognition[4],[5] Existing feature selection can be classified as the filter and wrapper approaches. Filter approach is implemented to identify the potential useful feature subset based on predefined data or metric content. Meanwhile, wrapper approach is designed by incorporating a selected classifier to evaluate the quality of feature subset. Despite incurring longer computation time, the wrapper methods tend to produce better accuracy than those of filter approaches when solving feature selection problems.

Metaheuristic search algorithm (MSA) has emerged as a promising wrapper approach used to solve feature selection problems due to its good global search ability. Depending on the sources of inspirations, existing MSAs can be categorized into: (a) Darwin’s theory of evolution, (b) swarm intelligence, (c) physics-based and (d) human-based[6]. African Vultures Optimization Algorithm (AVOA) is a new MSA proposed in 2021[7] motivated by the behavior of vulture in hunting for food sources. Similar with other MSAs, the initial population of AVOA is randomly produced by discarding the information of search environment[8]. This drawback can restrict the ability of AVOA to solve tackle real-world problem such as feature selection due to premature convergence issue.

In this paper, a new variant of Chaotic African Vultures Optimization Algorithm (CAVOA) is designed as a wrapper-based method to solve the feature selection problems competitively. A chaotic initialization method is introduced to produce the initial population with better solution quality that can lead to the better optimization results in terms of feature subset selected. The capability of CAVOA to solve feature selection problems is evaluated with twelve datasets selected from the UCI Machine Learning Repository[9].

## 2. Related Works

### 2.1. *Inspiration of AVOA*

Vulture is a hunting bird that searches for the animal carcasses as a food source. Popular physical feature of a vulture is its bald head to avoid any infection when consuming the carcasses. According to some researches, the bare skin of vulture is used to maintain its body heat. Different species of vultures have unique characteristics in terms of flight pattern and strength to fend off another animal. When hunting for food, vultures fly around to search for other species of vultures with food supply. This might cause other vultures to move to the same place and fight with each other for food sources. The weaker vultures tend to move around the stronger vultures and steal their food by tiring the stronger vultures, resulting in some vultures becoming more aggressive.

### 2.2. *Feature Selection*

Feature selection can be formulated as a bi-objective optimization problem by considering the classification accuracy and number of selected features. The fitness function used to measure the quality of each AVOA solution $X$ when solving the feature selection problem is:

$$f(X) = q\gamma + n\frac{|F_s|}{|F_o|} \tag{1}$$

where $q \in [0, 1]$ and $n \in [1 - q]$ are parameters used to represent the weightage of classification error $\gamma$ and the length of selected feature subset $|F_s|$. Meanwhile, $|F_O|$ is the lengths of original input datasets. Smaller $f(X)$ value is more desirable for feature selection problem.

## 3. Proposed CAVOA

The search mechanisms of proposed CAVOA to solve feature selection problems can be described in four stages. Stages 1 is the modified initialization scheme with a chaotic map to produce initial population of CAVOA. Stage 2 determines the satiation rate of each CAVOA solution. Stages 3 and 4 focus on balancing exploration and exploitative states of proposed CAVOA.

In stage 1, the initial value of a chaotic variable is randomly produced as $\alpha_0 \in [0,1]$. A circle map is then used to update the value of chaotic variable at every $k$-th iteration, denoted as $\alpha_k$, where $k = 1, \dots, k_{max}$ and $k_{max}$ is the maximum iteration. The circle map used to update $\alpha_{k+1}$ in the next $(k + 1)$-th iteration is formulated as:

$$\alpha_{k+1} = mod\left(\alpha_k + c - \left(\frac{b}{2\pi}\right)\sin(2\pi\alpha_k), 1\right) \tag{2}$$

where $b = 0.5$ and $c = 0.2$ are two constant values. After obtaining the chaotic result $\alpha_{k(end)}$ at the final iteration of $k_{max}$, the initial position of each $i$-th CAVOA solution in the $d$-th dimension is obtained as:

$$X_{i,d} = X_d^{LB} + \alpha_{k(end)}(X_d^{UB} - X_d^{LB}) \tag{3}$$

where $X_{i,d}^{LB}$ and $X_{i,d}^{UB}$ are the lower and upper boundaries of $d$-th decision variable, respectively, and $d = 1, .., D$. The initial population of CAVOA can be generated using Eq. (3) and it is sorted in ascending order based on fitness values to obtain the first- and second-best results.

During Stage 2, the satiation rate used to mimic the status of vultures (i.e., hungry or fulfilled) are calculated to determine the exploration and exploitation state of CAVOA as shown below:

$$m = (2 \times r_1 + 1) \times \aleph \times \left(1 - \frac{\mu}{\mu_{max}}\right) + Q \tag{4}$$

where $r_1$ is a random number between 0 and 1; $\aleph$ is a random number between -1 and 1; $\mu$ and $\mu_{max}$ represents the current fitness evaluation number and maximum fitness evaluation number, respectively. $Q$ is a variable utilized to enhance the performance of CAVOA when dealing with complex optimization problems, i.e.,

$$Q = r_2 \times \left(sin^f\left(\frac{\pi\mu}{2\mu_{max}}\right) + cos\left(\frac{\pi\mu}{2\mu_{max}}\right) - 1\right) \tag{5}$$

where $r_2$ indicates a random number with value between -2 and 2; $f$ is a predefined parameter used to represents the optimization operation disrupts the exploration and exploitation state. At the end of Stage 2, a roulette wheel approach is used to randomly select one CAVOA solution from the best and second-best performing vultures. When the values of parameter $|m|$ in Stage 2 is bigger or equal to one (i.e., $|m| \geq 1$), the searching process of CAVOA will move forward to Stage 3 that aims to promote the exploration search of algorithm.

During the Stage 3 of CAVOA, all vultures are guided to perform searching in different random locations by using two possible search methods. In particular, the search method performed by every $i$-th CAVOA solution to update its position is randomly selected based on a probability denoted as $P_1$. If $P_1$ is less than generated random number, Eq. (6) is used to calculate the updated position of $i$-th CAVOA solution as $X_i^{new}$. Otherwise, Eq. (7) is used instead:

$$X_i^{new} = X_{row} - |(2 \times r_3) \times X_{row} - X_i| \times m \tag{6}$$

$$X_i^{new} = X_{row} - m + r_3 \times \left((X^{UB} - X^{LB}) \times r_4 + X^{LB}\right) \tag{7}$$

where $m$ is the vulture satiation rate obtained from Eq. (4); $X^{LB}$ and $X^{UB}$ represent the lower and upper boundary of decision variables; $X_{row}$ refers to the CAVOA solution randomly selected from the best or second-best vultures using the roulette wheel approach; $r_3$ and $r_4$ are two random numbers between 0 and 1.

Stage 4 of CAVOA is triggered when the $|m|$ value obtained in Stage 2 is smaller than one (i.e.,$|m| < 1$). Note that this stage is more exploitative and it has two subphases that can be triggered based on the value of $|m|$. The searching procedure of first subphase starts when $|m|$ is found larger than or equal to 0.5 (i.e., $|m| \geq 0.5$). Referring to Eq. (8), a search method is chosen based on the probability $P_2 \in [0,1]$ to calculate the new position of every $i$-th CAVOA solution $X_i^{new}$ as follow:

$$X_i^{new} = \begin{cases} (C + D)/2 & , P_2 > rand \\ X_{row} - |X_{row} - X_i| \times m \times L & , \text{otherwise} \end{cases} \tag{8}$$

where $C$ and $D$ represent the movements of the best vulture $X_1^{best}$ and second-best vulture $X_2^{best}$ when they are competing for the food sources, i.e.,

$$C = X_1^{best} - \frac{X_1^{best} \times X_i}{X_1^{best} \times (X_i)^2} \times m \tag{9}$$

$$D = X_2^{best} - \frac{X_2^{best} \times X_i}{X_2^{best} \times (X_i)^2} \times m \tag{10}$$

It is also notable that $L$ presented in Eq. (8) refers to the levy flight characteristic used to further enhance the search potential of CAVOA as follow:

$$L = \frac{0.01}{|Y|^{\frac{1}{\omega}}} \left( T \left( \frac{\delta(1+\omega) \times sin\left(\frac{\pi\omega}{2}\right)}{\delta(1+2\omega) \times \omega \times 2\left(\frac{\omega-1}{2}\right)} \right)^{\frac{1}{\omega}} \right) \tag{11}$$

where $T$ and $Y$ are two random numbers between 0 and 1; ω is predefined constant and its value is set as 1.5; δ(·) refers to the gamma function.

Meanwhile, the searching mechanism for the second subphase of Stage 4 is triggered when $|m|$ is lesser than 0.5 (i.e., $|m| < 0.5$). The search strategy method in the second subphase is also selected based on a predefined probability $P_3 \in [0,1]$. If $P_3$ is larger than the randomly generated value (*rand*), Eq. (12) is used to calculate the updated position $X_i^{new}$ of every $i$-th CAVOA solution. Otherwise, Eq. (13) is applied as shown below:

$$X_i^{new} = |(2 \times r_5) \times X_{row} - X_i| \times (m + r_6) - X_{row} - X_i \tag{12}$$

$$X_i^{new} = X_i - (S_1 - S_2) \tag{13}$$

where $r_5$ and $r_6$ are two random numbers with value between 0 and 1; $X_{row}$ represent a randomly chosen vulture using roulette wheel selected; $S_1$ and $S_2$ are calculated using Eqs. (14) and (15), respectively, where $r_7$ and $r_8$ are two random numbers between 0 and 1.

$$S_1 = X_i \times \left(\frac{r_7 \times X_i}{2\pi}\right) \times cos(X_i) \tag{14}$$

$$S_2 = X_i \times \left(\frac{r_8 \times X_i}{2\pi}\right) \times sin(X_i) \tag{15}$$

The overall flow for CAVOA is summarized in Fig 1. The search process continues until stopping condition is achieved, i.e., when the current fitness evaluation number exceeds the maximum fitness evaluation number ($\mu > \mu_{max}$).

**Algorithm 1:** CAVOA based Feature Selection

**Inputs:** $i,\ k,\ \mu,\ \mu_{max},\ X^{UB},\ X^{LB}$

Load parameters and selected dataset;
**for** each $i$-th vulture **do**
    **for** each $d$-th dimension **do**
        Generate chaotic variable using Eq. (2);
        Initialize the position using Eq. (3);
    **end for**
**end for**
Evaluate vulture's fitness using Eq. (1)
Select the first and second-best vultures as $X_1^{best}$ and $X_2^{best}$, respectively;
**while** $\mu \leq \mu_{max}$ **do**
    **for** each $i$-th vulture **do**
        Calculate $m$ using Eq. (4) and (5);
        Perform the roulette wheel selection on $X_1^{best}$ and $X_2^{best}$ to select $X_{row}$;
        **if** $|m| \geq 1$ **then**
            Calculate $X_i^{new}$ using Eq. (6) and (7);
        **else if** $|m| < 1$ **then**
            **if** $|m| \geq 0.5$ **then**
                Calculate $X_i^{new}$ using Eq. (8);
            **else if** $|m| < 0.5$ **then**
                Calculate $X_i{}^{new}$ using Eq. (12);
            **end if**
        **end if**
        Boundary check of $X_i^{new}$;
        Fitness evaluation of $X_i^{new}$ using Eq. (1);
        Update the $X_1^{best}$ and $X_2^{best}$;
        $\mu \leftarrow \mu + 1$;
    **end for**
**end while**

**Output:** $X_1{}^{best}$ (i.e., optimal feature subset)

Fig 1. Overall flow of CAVOA for feature selection.

## 4. Performance Evaluation on CAVOA

### 4.1. *Simulation Settings*

The capability of proposed CAVOA to solve various feature selection problems is investigated using twelve datasets obtained from UCI Machine Learning Repository[9], i.e., (a) Breast Cancer Wisconsin (Original), (b) Dermatology, (c) Lymphography, (d) Iris, (e) Arrhythmia, (f) Echocardiogram, (g) Waveform Database Generator (Version 1), (h) Zoo, (i) Indian Liver Patient Dataset (ILPD), (j) Semeion Handwritten Digit, (k) Letter Recognition and (l) Balance Scale. The performance of the proposed CAVOA to solve these selected datasets are also compared with another four peer feature selection algorithms developed using the state-of-art MSAs, known as the African Vultures Optimization Algorithm (AVOA)[7], Bezier Search Differential Evolution Algorithm (BeSD)[10], Generalized Normal Distribution Optimization (GNDO)[11] and Sperm Swarm Optimization (SSO)[12]. The performances of all compared feature selection algorithms in solving all selected datasets are evaluated using two metrics, i.e., the mean classification accuracy $Acc^{mean}$ and average number of selected features $N^{feature}$. For fair comparison, the population sizes and maximum fitness evaluation numbers of all algorithms are set as $N = 20$ and $\tau^{max} = 2000$, respectively. Each algorithm is used simulated for 30 independent runs to solve each dataset during performance comparative studies.

### 4.2. *Performance Comparisons*

An algorithm that can produce higher $Acc^{mean}$ and lower $N^{feature}$ values simultaneously is considered to be better when solving feature selection problem. The value with boldface represents the best result among the compared result while italic and underlined valued indicate the second-best result. Table 1 reports that the CAVOA has the best feature selection performance by solving 11 out 12 datasets with best $Acc^{mean}$ results. This is followed by DSSA and BeSD with two best $Acc^{mean}$, AVOA with one best $Acc^{mean}$, GNDO with no best $Acc^{mean}$. Table 2 result reports that the proposed CAVOA has the best performance in selecting the least number of features when solving four datasets, followed by AVOA, BeSD and SSO that solve lesser datasets with the least numbers of features.

Table 1. Comparison of $Acc_{mean}$.

| Dataset | AVOA | BeSD | GNDO | SSO | CAVOA |
|---|---|---|---|---|---|
| (a) | 0.987 | 0.988 | *0.991* | 0.963 | **1.000** |
| (b) | 0.990 | 0.981 | 0.985 | *0.993* | **1.000** |
| (c) | 0.590 | 0.544 | *0.617* | 0.535 | **0.655** |
| (d) | **1.000** | **1.000** | 0.967 | *0.994* | 0.967 |
| (e) | 0.710 | 0.687 | 0.673 | *0.711* | **0.759** |
| (f) | **1.000** | **1.000** | **1.000** | *0.995* | **1.000** |
| (g) | *0.831* | 0.815 | 0.830 | 0.813 | **0.838** |
| (h) | *0.927* | 0.888 | 0.890 | 0.773 | **0.963** |
| (i) | *0.760* | 0.726 | 0.724 | 0.739 | **0.793** |
| (j) | *0.932* | 0.886 | 0.928 | 0.882 | **0.949** |
| (k) | 0.953 | 0.946 | *0.954* | 0.923 | **0.955** |

| | | | | | |
|---|---|---|---|---|---|
| (l) | *0.840* | 0.808 | 0.712 | 0.822 | **0.888** |

Table 2. Comparison of $N_{feature}$.

| Dataset | AVOA | BeSD | GNDO | SSO | CAVOA |
|---|---|---|---|---|---|
| (a) | 4.07 | 5.20 | *3.83* | 4.97 | **3.30** |
| (b) | **11.43** | 16.57 | 13.97 | 16.87 | *11.87* |
| (c) | **3.37** | 9.40 | 7.23 | 6.23 | *3.90* |
| (d) | 2.00 | 2.40 | **1.00** | *1.97* | **1.00** |
| (e) | *38.13* | 134.73 | 126.97 | **17.90** | 47.87 |
| (f) | **1.00** | 4.37 | **1.00** | *3.90* | **1.00** |
| (g) | 16.97 | **9.67** | 14.23 | *10.47* | 16.90 |
| (h) | *5.23* | 7.63 | 6.80 | 6.57 | **3.77** |
| (i) | 2.90 | 4.63 | *2.87* | **1.73** | 3.37 |
| (j) | 144.70 | *129.10* | 129.37 | **127.83** | 185.03 |
| (k) | 12.17 | **7.87** | 12.30 | *10.53* | 12.20 |
| (l) | 4.00 | **2.17** | 4.00 | *3.83* | 4.00 |

## 5. Conclusions

A modified algorithm known as CAVOA is presented in this paper to solve feature selection problems more efficiently. In particular, the ergodicity and non-repetitive characteristics of circle chaotic map are leveraged during the initialization stage to produce the initial population of CAVOA with enhanced solution quality. This is followed by the search mechanisms used to achieve the better balancing of exploration and exploitation states of proposed. A total of 12 datasets are selected to evaluate the performance of CAVOA and its peer algorithm to solve feature selection problems. The simulation results show that CAVOA can outperform another four feature selected algorithms developed with state-of-art MSAs by solving more selected benchmark problems with the highest classification accuracy and the least number of selected features.

## Acknowledgements

This research is supported by Ministry of Higher Education Malaysia (MOHE) under the Fundamental Research Grant Scheme with the project codes of FRGS/1/2019/TK04/UCSI/02/1 and FRGS/1/2020/TK0/UCSI/02/4. This research is also supported by UCSI Research Excellence & Innovation Grant (REIG) with project code of REIG-FETBE-2022/038.

## References

1. Ang, K. M. et al. New Hybridization Algorithm of Differential Evolution and Particle Swarm Optimization for Efficient Feature Selection. (2022)
2. Alrifaey, M., Lim, W. H. & Ang, C. K. A Novel Deep Learning Framework Based RNN-SAE for Fault Detection of Electrical Gas Generator. IEEE Access 9, 21433-21442 (2021).
3. Alrifaey, M. et al. Hybrid Deep Learning Model for Fault Detection and Classification of Grid-Connected Photovoltaic System. IEEE Access 10, 13852-13869 (2022).
4. Jdid, B., Hassan, K., Dayoub, I., Lim, W. H. & Mokayef, M. Machine Learning Based Automatic Modulation Recognition for Wireless Communications: A Comprehensive Survey. IEEE Access 9, 57851-57873 (2021).
5. Jdid, B., Lim, W. H., Dayoub, I., Hassan, K. & Juhari, M. R. B. M. Robust Automatic Modulation Recognition Through Joint Contribution of Hand-Crafted and Contextual Features. IEEE Access 9, 104530-104546 (2021).
6. Ahmad, M. F., Isa, N. A. M., Lim, W. H. & Ang, K. M. Differential evolution: A recent review based on state-of-the-art works. Alexandria Engineering Journal 61, 3831-3872 (2022).
7. Abdollahzadeh, B., Gharehchopogh, F. S. & Mirjalili, S. African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. Computers & Industrial Engineering 158, 107408 (2021).
8. Ahmad, M. F., Isa, N. A. M., Lim, W. H. & Ang, K. M. Differential evolution with modified initialization scheme using chaotic oppositional based learning strategy. Alexandria Engineering Journal 61, 11835-11858 (2022).
9. Dua, D. & Graff, C. UCI machine learning repository. (2017).
10. Civicioglu, P. & Besdok, E. Bezier Search Differential Evolution Algorithm for numerical function optimization: A comparative study with CRMLSP, MVO, WA, SHADE and LSHADE. Expert Systems with Applications 165, 113875 (2021).
11. Zhang, Y., Jin, Z. & Mirjalili, S. Generalized normal distribution optimization and its applications in parameter extraction of photovoltaic models. Energy Conversion and Management 224, 113301 (2020).
12. Shehadeh, H., Ahmedy, I. & Idris, M. Sperm Swarm Optimization Algorithm for Optimizing Wireless Sensor Network Challenges. (2018).

## Authors Introduction

Mr. Wy – Liang Cheng

He received the B.Eng. degree in Mechatronic Engineering with Hnours from UCSI University, Malaysia in 2020. He is currently pursuing Master of Philosophy in Engineering in UCSI University, Malaysia. His research interests are swarm intelligence and feature selection.

Ms. Li Pan

She received her Master of Engineering in Computer Technology degree from Huazhong University of Science and Technology, China in 2008. She is currently a Doctoral research student in UCSI University, Malaysia.

Prof. Dr. Mohd Rizon Mohamad Juhari

He is a Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Engineering from Oita University, Japen in 2002. His research interests are face analysis, pattern recognition and vision for mobile robot.

Dr. Chin Hong Wong

He is a Lecturer in Maynooth International Engineering College at Fuzhou University in China. He received his PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2017. His research interests are Energy harvesting and control system.

Dr. Abhishek Sharma

He is a Research Assistant Professor at Graphic Era Deemed to be University in India. He received his PhD from University of Petroluem & Energy Studies in 2022. His research interests are artificial intellgence and power electronics.

Dr. Sew Sun Tiang

She is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. She received her PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2014. Her research interests are optimization and antenna design.

Dr. Wei Hong Lim

He is an Associate Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Computational Intelligence from Universiti Sains Malaysia in 2014. His research interests are optimization and artificial intelligence.