# New Hybridization Algorithm of Differential Evolution and Particle Swarm Optimization for Efficient Feature Selection

**Koon Meng Ang, Mohd Rizon Bin Mohamed Juhari, Wei Hong Lim*, Sew Sun Tiang, Chun Kit Ang, Eryana Eiyda Hussin, Li Pan, Ting Hui Chong**

*Faculty of Engineering, Technology and Built Environment, UCSI University, 1, Jalan Puncak Menara Gading, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia.*

*E-mail: 1001436889@ucsiuniversity.edu.my, mohdrizon@ucsiuniversity.edu.my, limwh@ucsiuniversity.edu.my, tiangss@ucsiuniversity.edu.my, angck@ucsiuniversity.edu.my, eryanaeiyda@ucsiuniversity.edu.my, 1002060534@ucsiuniversity.edu.my, 1001747667@ucsiuniversity.edu.my*

*http://www.ucsiuniversity.edu.my/*

**Abstract**

Feature selection is a popular pre-processing technique applied to enhance the learning performances of machine learning models by removing irrelevant features without compromising their accuracies. The rapid growth of input features in big data era has increased the complexities of feature selection problems tremendously. Given their excellent global search ability, differential evolution (DE) and particle swarm optimization (PSO) are considered as the promising techniques used to solve feature selection problems. In this paper, a new hybrid algorithm is proposed to solve feature selection problems more effectively by leveraging the strengths of both DE and PSO. The proposed feature selection algorithm is reported to achieve an average accuracy of 89.03% when solving 7 datasets obtained from UCI Machine Learning Repository.

*Keywords*: Feature Selection; Particle Swarm Optimization (PSO); Differential Evolution (DE); Metaheuristic Search Algorithm; Hybridization.

## 1. Introduction

Feature selection[1] is a popular pre-processing technique used to address the "curse of dimensionality" issue by eliminating redundant features from large-scale datasets. Feature selection is widely used in real-world problems due to its ability to reduce storage space and computational time required for training the predictive models without sacrificing their performances[2,3]. Feature selection is formulated as a non-deterministic polynomial-time (NP) hard combinatorial problem that is not trivial to solve, especially when it involves the large input feature size of $D > 100$[4].

Nature-inspired algorithms have emerged as effective approach to solve the complex real-world optimization problems due to their promising global search ability. Differential evolution (DE)[5] and particle swarm optimization (PSO)[6] are the two most popular nature-inspired algorithms widely used to solve different types of optimization problems[7,8,9,10,11,12,13,14]. However, the capability of conventional DE and PSO to tackle large-scale feature selection problems remains unexplored. The presence of excessive irrelevant features in original datasets can introduce massive number of local optima in search space and increase the complexity of feature selection problem. For conventional DE and PSO, the random initialization scheme adopted do not fully consider any information around their search environments, therefore the quality of initial population obtained is questionable[15]. The "No Free Lunch Theorem"[16] is another factor that restrict the performances of conventional DE and PSO to tackle various optimization problems. More robust optimization algorithms are required to handle the feature selection problems with different complexity levels effectively.

In this article, a hybrid DE and PSO with chaotic-opposition-based initialization scheme (HDPCIS) is proposed to address the aforementioned challenges in performing feature selection. A chaotic-opposition-based initialization scheme (CIS) is first incorporated into

HDPCIS to generate an initial population with better solution quality. A hybridization scheme that aims to leverage the benefits of DE and PSO is also introduced to achieve better tradeoff in terms of search efficiency and diversity preservation. The performance of HDPCIS to solve different feature selection problems are assessed with 7 datasets of UCI Machine Learning Repository.

## 2. Related Works

### 2.1. *Conventional DE*

Given a set of randomly initialized DE solution with the population size of $N$, each $n$-th solution of $X_n = [X_{n,1}, \ldots, X_{n,d,\ldots,} X_{n,D}]$ represents a candidate solution of a given problem with total dimension size of $D$, where $d \in [1, D]$ and $n \in [1, N]$ refer to the dimension and solution indices, respectively. During the evolution process of DE, a mutation process is first performed using "DE/rand/1" strategy to generate a donor vector $U_n = [U_{n,1}, \ldots, U_{n,d,\ldots,} U_{n,D}]$ for each $n$-th target vector $X_n = [X_{n,1}, \ldots, X_{n,d,\ldots,} X_{n,D}]$ as follow:

$$U_n = X_a + F\left(X_b - X_c\right) \tag{1}$$

where $F$ is a scaling factor in range of 0 to 1; $X_a$, $X_b$ and $X_c$ are three randomly selected solutions from population with $n \neq a \neq b \neq c$.

An offspring vector $Y_n = [Y_{n,1}, \ldots, Y_{n,d,\ldots,} Y_{n,D}]$ is then produced for each $n$-th solution by performing crossover on the target vector $X_n = [X_{n,1}, \ldots, X_{n,d,\ldots,} X_{n,D}]$ and donor vector $U_n = [U_{n,1}, \ldots, U_{n,d,\ldots,} U_{n,D}]$. Define $C_r \in [0.5, 1]$ as the crossover probability, the $d$-th dimension of each $n$-th offspring vector can be computed as:

$$Y_{n,d} = \begin{cases} U_{n,d}, & \text{if } rand_d \leq Cr \\ X_{n,d}, & \text{otherwise} \end{cases} \tag{2}$$

For selection process, the fitness value of $Y_n$ is compared with that of $X_n$ in terms of their fitness values as a selection process. The latter solution is replaced if the former one has better fitness value.

### 2.2. *Conventional PSO*

Each $n$-th candidate solution or particle consists of two vectors define its current state, i.e., the velocity $V_n = [V_{n,1}, \ldots, V_{n,d,\ldots,} V_{n,D}]$ and position $X_n = [X_{n,1}, \ldots, X_{n,d,\ldots,} X_{n,D}]$, where $d \in [1, D]$ and $n \in [1, N]$. Each PSO particle can memorize the best solution found by itself and population that are denoted as $P_n^{best} = \left\lfloor P_{n,1}^{best}, \ldots, P_{n,d}^{best}, \ldots, P_{n,D}^{best} \right\rfloor$ and $G^{best} = \left\lfloor G_1^{best}, \ldots, G_d^{best}, \ldots, G_D^{best} \right\rfloor$, respectively. The velocity $V_n$ and position $X_n$ of each particle $n$ are updated as:

$$V_n^{new} = \omega V_n + c_1 r_1 \left(P_n^{best} - X_n\right) + c_2 r_2 \left(G^{best} - X_n\right) \tag{3}$$

$$X_n^{new} = X_n + V_n^{new} \tag{4}$$

where $\omega$ js inertia weight; $c_1$ and $c_2$ are acceleration coefficients; $r_1$ and $r_2$ are two random numbers obtained from uniform distribution in range of 0 to 1. The fitness value of $X_n$ is compared with those of $P_n^{best}$ and $G^{best}$. The latter two solutions are replaced if the former one has better fitness value.

### 2.3. *Feature Selection Problem*

Feature selection is considered as a bi-objective optimization problem, aiming to minimize the number of selected features and maximize the classification accuracy, simultaneously. In order to satisfy these objectives, a fitness function is formulated to measure the quality of each candidate solution as follow[17]:

$$f(\bullet) = \chi\varepsilon + \gamma \frac{|F_S|}{|F_T|} \tag{5}$$

where $\chi \in [0,1]$ and $\gamma = (1-\chi)$ refer to the parameters measuring the weightage of classification quality and subset length, respectively; $\varepsilon$ represents the classification error; $|F_S|$ and $|F_T|$ indicate the selected subset of features and the total number of features in original dataset, respectively.

## 3. The Proposed HDPCIS

At the beginning of search process, a chaotic-opposition-based initialization scheme (CIS)[8] is incorporated to replace random initialization scheme. A chaotic swarm $\Psi^{CS}$ and an opposite swarm $\Psi^{OS}$ are produced by CIS based on the modified sine map and opposition-based-learning strategy, respectively. $\Psi^{CS}$ and $\Psi^{OS}$ are then combined to form a merged population $\Psi^M$. After all the solution members of $\Psi^M$ are sorted from the worst to the best based on their fitness values, the first best $N$ members of $\Psi^M$ are selected to construct the initial population $\Psi^I = \left[X_1, \ldots, X_n, \ldots, X_N\right]$ of HDPCIS.

For the proposed hybridization scheme, DE and PSO are employed as the primary and secondary algorithms used to evolve candidate solutions, respectively. During the DE stage, a mutation scheme of Eq. (1) is performed to generate a donor vector $\partial_n$ for each solution $n$. The corresponding offspring vector $\wp_n$ is computed based on Eq. (2). A greedy selection scheme is applied to compare the fitness of $\wp_n$ with that of $X_n$. The latter solution is replaced by the former solution if the former solution has better fitness value. The greedy selection scheme is also used to update the best solution in population $G^{best}$.

If $\wp_n$ computed by DE has better fitness than original $X_n$, PSO is triggered as secondary optimizer to refine $X_n$. In PSO stage, the velocity component of solution $n$ is updated as follow:

$$V_n^{new} = \omega V_n + c\left(G^{best} - X_n\right) \tag{6}$$

where $c$ refers to acceleration coefficient. Notably, the velocity update equation of HDPCIS in Eq. (6) only considers social component. For HDPCIS, $X_n$ is essentially equivalent to self-cognitive component $P_n^{best}$ because it only updated when a better solution is found in optimization process. Given $V_n^{new}$, the new position $X_n^{new}$ is calculated with Eq. (4). Both of $X_n$ and $G^{best}$ are replaced by $X_n^{new}$, if the latter solution is more superior than the former solutions. Otherwise, $X_n^{new}$ with worse fitness will be discarded.

The overall framework of HDPCIS is summarized in Fig. 1. The optimization process is iterated until the termination criteria $\tau > \tau^{\max}$ is satisfied, where $\tau$ and $\tau^{\max}$ represent the fitness evaluation counter and the predefined maximum fitness evaluation number.

```
Algorithm: HCPCIS
Inputs: D, N, Ub, Lb Cr, F, ω, c, τ, τ^max
Initialize τ ← 0;
Produce Ψ^I = [X_1,...,X_n,...,X_N] using CIS;
τ ← τ + 2N;
while τ ≤ τ^max do
    for each solution n do //Execute DE.
        Produce ∂_n using Eq. (1);
        Produce ℘_n using Eq. (2);
        Evaluate fitness of ℘_n using Eq. (5);
        τ ← τ + 1;
        Update X_n and G^best with greedy selection;
        if f(X_n) < f(℘_n) then // Execute PSO
            Calculate V_n^new using Eq. (6);
            Calculate X_n^new using Eq. (4);
            Evaluate fitness of X_n^new using Eq. (5);
            τ ← τ + 1;
            Update X_n and G^best with greedy selection;
        end if
    end for
end while
Outputs: G^best
```

Fig. 1. Pseudocode of HDPCIS.

## 4. Performance Evaluations of HDPCIS

### 4.1. *Simulation settings*

The performance of HDPCIS to solve feature selection problem is evaluated using seven datasets obtained from the UCI Machine Learning Repository[18], i.e., (a) glass identification, (b) lymphography, (c) lung cancer, (d) multiple features, (e) statlog (heart), (f) ionosphere and (g) iris. The proposed HDPCIS is compared with four peer algorithms known as: chaotic-opposition-based hybridized DE with PSO (CO-HDEPSO)[8], chaotic-opposition-based differential evolution (CO-DE), conventional DE (DE)[5] and conventional PSO (PSO)[6], in terms of the mean accuracy $Acc^{mean}$ and average numbers of selected features $nF^{avg}$. The population size and maximum fitness evaluations numbers of all algorithms are set as $N$ = 10 and $\tau^{\max} = 1000$, respectively. All algorithms are simulated for 30 times to solve the selected datasets.

Table 1. Mean accuracy $Acc^{mean}$

| Datasets | HDPCIS | CO-HDEPSO | CO-DE | DE | PSO |
|---|---|---|---|---|---|
| (a) | **0.7952** | 0.7762 | 0.7619 | 0.7143 | 0.7286 |
| (b) | **0.5725** | 0.5724 | 0.5448 | 0.4621 | 0.5448 |
| (c) | **1.0000** | **1.0000** | **1.0000** | 0.9600 | 0.6400 |
| (d) | **0.9825** | 0.9730 | 0.9715 | 0.9705 | 0.9695 |
| (e) | **0.9074** | 0.8926 | 0.8667 | 0.7852 | 0.8556 |
| (f) | **0.9743** | 0.9714 | 0.9543 | 0.9457 | 0.9114 |
| (g) | **1.0000** | **1.0000** | **1.0000** | **1.0000** | 0.9667 |

Table 2. Average number of selected features $nF^{avg}$.

| Datasets | HDPCIS | CO-HDEPSO | CO-DE | DE | PSO |
|---|---|---|---|---|---|
| (a) | **3.8** | 4.8 | 4.8 | **3.8** | 5.4 |
| (b) | **4.2** | 7.0 | 4.8 | 6.0 | 8.4 |
| (c) | **12.4** | 14.8 | 12.8 | 14.2 | 21.2 |
| (d) | **273.4** | 308.2 | 295.2 | 285.6 | 308.6 |
| (e) | **3.2** | 5.0 | 4.4 | 4.4 | 4.8 |
| (f) | **9.6** | 10.4 | 10.8 | 11.0 | 14.4 |
| (g) | **1.0** | **1.0** | 2.0 | **1.0** | **1.0** |

### 4.2. *Comparisons between selected algorithms*

The $Acc^{mean}$ and $nF^{avg}$ values obtained by all algorithms in solving 7 selected image datasets are reported in Tables 1 and 2, respectively. The best and second-best results are indicated by boldface and underlined fonts, respectively. Table 1 reported that the HDPCIS has the best feature selection performance by producing the $Acc^{mean}$ values for seven datasets. This is followed by CO-HDEPSO, CO-DE, DE, and PSO that produce the best $Acc^{mean}$ values in 2, 2, 1, and 0 dataset, respectively. On the other hand, Table 2 reported that the proposed HDPCIS has the best performance in selecting optimal number of features by producing best $nF^{avg}$ for all seven datasets. This is followed by DE, CO-HDEPSO, PSO, and CO-DE that obtain the best $nF^{avg}$ values for 2, 1, 1, and 0 datasets, respectively.

## 5. Conclusions

A new hybridization algorithm of HDPCIS is introduced to solve feature selection problem effectively. A CIS module is first employed to initialize a population with better quality to reduce the possibility of premature convergence. A hybridization scheme is designed with DE as the primary algorithm and PSO as the secondary algorithm, to achieve better balancing of exploration and exploitation search behaviors. The simulation studies reported that the proposed HDPCIS can outperform its peer algorithms by solving feature selection problems with higher accuracy and lesser 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 Proj-FRGS/1/2019/TK04/UCSI/02/1 and Proj-FRGS/1/2020/TK0/UCSI/02/4.

## References

1. B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Transactions on Evolutionary Computation,* vol. 20, no. 4, pp. 606-626, 2015.
2. J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature extraction, construction and selection*: Springer, 1998, pp. 117-136.
3. F. Hafiz, A. Swain, N. Patel, and C. Naik, "A two-dimensional (2-D) learning framework for particle swarm based feature selection," *Pattern Recognition,* vol. 76, pp. 416-433, 2018.
4. H. Faris *et al.*, "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems," *Knowledge-Based Systems,* vol. 154, pp. 43-67, 2018.
5. K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.
6. J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995, vol. 4, pp. 1942-1948 vol.4.
7. O. T. Chong, W. H. Lim, N. A. M. Isa, K. M. Ang, S. S. Tiang, and C. K. Ang, "A Teaching-Learning-Based Optimization with Modified Learning Phases for Continuous Optimization," in *Science and Information Conference*, 2020, pp. 103-124: Springer.
8. Z. C. Choi *et al.*, "Hybridized Metaheuristic Search Algorithm with Modified Initialization Scheme for Global Optimization," *Advances in Robotics, Automation and Data Analytics: Selected Papers from ICITES 2020,* vol. 1350, p. 172, 2021.
9. W. S. Koh *et al.*, "Multi-objective Particle Swarm Optimization with Alternate Learning Strategies," Singapore, 2022, pp. 15-25: Springer Singapore.
10. K. M. Ang *et al.*, "Modified Particle Swarm Optimization with Unique Self-cognitive Learning for Global Optimization Problems," Singapore, 2022, pp. 263-274: Springer Singapore.
11. K. M. Ang, W. H. Lim, S. S. Tiang, C. K. Ang, E. Natarajan, and M. Ahamed Khan, "Optimal Training of Feedforward Neural Networks Using Teaching-Learning-Based Optimization with Modified Learning Phases," in *Proceedings of the 12th National Technical Seminar on Unmanned System Technology 2020*, 2022, pp. 867-887: Springer.
12. K. Ang, Z. Yeap, C. Chow, W. Cheng, and W. Lim, "Modified Particle Swarm Optimization with Chaotic Initialization Scheme for Unconstrained Optimization Problems," *MEKATRONIKA,* vol. 3, no. 1, pp. 35-43, 2021.
13. K. M. Ang *et al.*, "A Constrained Teaching-learning-based Optimization with Modified Learning Phases for Constrained Optimization," *Journal of Advanced Research in Dynamical and Control Systems,* vol. 12, no. 04, p. 15, 2020.
14. K. M. Ang, W. H. Lim, N. A. M. Isa, S. S. Tiang, and C. H. Wong, "A constrained multi-swarm particle swarm optimization without velocity for constrained optimization problems," *Expert Systems with Applications,* vol. 140, p. 112882, 2020.
15. X.-S. Yang, "Metaheuristic optimization: algorithm analysis and open problems," in *International Symposium on Experimental Algorithms*, 2011, pp. 21-32: Springer.
16. D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation,* vol. 1, no. 1, pp. 67-82, 1997.
17. E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing,* vol. 172, pp. 371-381, 2016.
18. M. Lichman. UCI Machine Learning Repository [Online]. Available: http://archive.ics.uci.edu/ml/index.php

## Authors Introduction

Mr. Koon Meng Ang

He received the B.Eng. degree in Mechatronic Engineering with Honours from UCSI University, Malaysia, in 2019. He is currently pursuing Ph.D. degree in UCSI University, Malaysia. His research interests are swarm intelligence, machine learning and deep learning.

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, patterm recognition and vision for mobile robot.

Dr. Wei Hong Lim

He is an Assistant 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.

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. Ang Chun Kit

He is an Associate Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Mechanical and Manufacturing Engineering from Universiti Putra Malaysia in 2014. His research interests are artificial intelligence, soft computing, robotics and mechatronics.

Ms. Eryana Eiyda Hussin

She received her Master's degree from the Faculty of Electronics and Computer System, Universiti Teknikal Malaysia Melaka. She is currently persuing her study under Doctor of Philisophy in Electrical and Electronic Engineering in Universiti Teknology Petronas, Malaysia. She is also a lecturer from the Department of Electrical and Electronics of UCSI University, Malaysia.

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.

Mr. Ting Hui Chong

He is currently pursuing Bachelor of Mechatronics Engineering with Honours as final year student in Faculty of Engineering, Technology and Built Environment, UCSI University, Malaysia. His research interests are swarm intelligence and feature selection.