# Feature selection for sky image classification based on self adaptive ant colony system algorithm

**Montha Petwan[1], Ku Ruhana Ku-Mahamud[2,3]**
[1]Faculty of Science and Technology, Suratthani Rajabhat University, Surat Thani, Thailand
[2]School of Computing, Universiti Utara Malaysia, Kedah, Malaysia
[3]Department of Information and Communications Technology, College of Engineering, Shibaura Institute of Technology, Tokyo, Japan

## ABSTRACT

Statistical-based feature extraction has been typically used to purpose obtaining the important features from the sky image for cloud classification. These features come up with many kinds of noise, redundant and irrelevant features which can influence the classification accuracy and be time consuming. Thus, this paper proposed a new feature selection algorithm to distinguish significant features from the extracted features using an ant colony system (ACS). The informative features are extracted from the sky images using a Gaussian smoothness standard deviation, and then represented in a directed graph. In feature selection phase, the self-adaptive ACS (SAACS) algorithm has been improved by enhancing the exploration mechanism to select only the significant features. Support vector machine, kernel support vector machine, multilayer perceptron, random forest, k-nearest neighbor, and decision tree were used to evaluate the algorithms. Four datasets are used to test the proposed model: Kiel, Singapore whole-sky imaging categories, MGC Diagnostics Corporation, and greatest common divisor. The SAACS algorithm is compared with six bio-inspired benchmark feature selection algorithms. The SAACS algorithm achieved classification accuracy of 95.64% that is superior to all the benchmark feature selection algorithms. Additionally, the Friedman test and Mann-Whitney *U* test are employed to statistically evaluate the efficiency of the proposed algorithms.

*This is an open access article under the CC BY-SA license.*

*Corresponding Author:*

Montha Petwan
Faculty of Science and Technology, Suratthani Rajabhat University
Khun Taleay, Muang, Surat Thani, 84100, Thailand
Email: monta.pet@sru.ac.th

## 1. INTRODUCTION

The classifying of the cloud type from ground-based sky images is continually receiving attention. The different forms of cloud have an impact on both weather prediction and the exchange of energy between the atmosphere and the Earth's surface [1], [2]. The variations of cloud images which depend on various atmospheric circumstances are the primary distinction between cloud images and other images. A cloud does not always have a definite spatial distribution. Even clouds of the same genus can vary in size and shape. Additionally, sophisticated examples of curving shapes, crossing borders, and angles can be seen in the structure information and cloud distribution [3]–[5]. The various identification technology equipment to collect sky photographs include meteorological balloons, satellites-based, and ground-based [6], [7]. The meteorological balloon and satellite-based approach's cloud-system enable the direct observation on how clouds affect the earth's radiation at the top of the atmosphere. The purpose of a ground-based approach is to use the local area and observe cloud bottoms in order to get whole data of the

cloud. These ground-based sky images are readily available, relatively inexpensive, and of high spatial resolution [8].

Creating significant features that can be used to distinguish between different cloud types in ground-based sky image is a crucial issue in this field. Different algorithms have been used to extract many visual attributes from sky images, such as texture, color, and shape, which will be taken into account when determining the type of cloud. The work in [9] was the first to accomplish automated cloud classification and combines spectral and textural attributes using the grey level co-occurrence matrix. The k-nearest neighbor (KNN) classifier with random test sample successfully categorizes seven different sky patterns with an accuracy of 87.52%. In [4], a method for feature extraction that uses the average ranking of occurrence patterns of all rotation-invariant patterns offered in the local binary pattern (LBP) was described. A cloud image becomes more robust when the occurrence rates are in various changing patterns. Four patch-LBP and region LBP technique have also been proposed in [1]. Support vector machine (SVM) and linear discriminant analysis are used to classify the various cloud types represented in a histogram. However, insignificant extracted features occur during the selection of the most important feature set.

Feature selection technique is the preprocessing process that aims to minimize the dimensionality of the data which gets only the significant feature subset as small as conceivable that are supplied to the learning algorithm [10]. The majority of researchers have addressed the feature selection problem for image classification task by using the bio-inspired algorithms [11]. The grey wolf optimization (GWO) [12], ant colony optimization (ACO) [13], and bat algorithm (BA) [14] are examples of these techniques, which are used for prioritizing features and improving classification accuracy. The development of an intelligent transportation system with traffic sign detection and recognition system based on the ACO algorithm was proposed by Jayaprakash and KeziSelvaVijila [15]. This work has achieved significant progress when tested on public road sign database. Liu *et al.* [14] has presented a feature selection method to increase the final detection accuracy for image steganalysis. The relevant binary feature subset was retrieved from the whole feature set by using BA. A levy flight based and GWO was also proposed in [12] to address feature selection for image steganalysis.

A modified version of the ant lion optimizer (ALO) algorithm with a wavelet SVM classifier was proposed in [16] for overcoming the high correlation bands in hyperspectral image classification. This method performed better than previous algorithms because of its ability to leverage the global best ants in the local search. In [17], a feature subset of computed tomography images was generated using the ACO algorithm with a rough dependency measure. SVM and naïve Bayes classifier were trained on the selected subsets to predict lung cancer disease. These techniques experience early convergence and are drawn towards regional optimum areas. ACO-based feature selection has also been proposed in the high dimensional space of image classification problem [13]. Nevertheless, the ACO algorithm has low classification accuracy due to premature convergence and poor balance between exploitation and exploration mechanism.

The ant colony system (ACS) was first presented by Dorigo and Stützle [18]. The ACS improved the ACO algorithmic in three different ways [19], [20]. The state transition rule for determining the route to the next node is first replaced with an aggressive action choice between exploitation and exploration. Second, the procedure for updating the global pheromone rule will deposit pheromone on the routes of the best ant's tour. Finally, the ants ignored some pheromone trails along the previously traveled path to increase deliberation of other remaining paths [21]. The accuracy and effectiveness of ACS are considerably higher than those of other algorithms because there are not many parameters that have to be modified to aid in the identification of significant features. This ability dealt with the nonlinear global optimization problem and was initially developed for finding the shortest path of the travelling salesman problem [22]. Other combinatorial optimization problems that have been tackled using ACS, including vehicle routing [23], job scheduling [24], network communication [19], [25] and image processing [26]. However, the ACS's exploring mechanism is ineffective. In the search step, which is carried out using the trail-and-error method and is error prone while the fundamental of the local pheromone update parameter are fixed and maintained constant. Additionally, effective results might not always be reproducible. In this paper, the ACS exploration mechanism for selecting significant features is also proposed. The remainder of the paper is organized as follows: section 2 describes the method of the proposed algorithm and dataset collection while section 3 presents the experimental findings and discussion. Finally, section 4 gives the conclusion and recommendations for future works.

## 2. METHOD
### 2.1. Proposed algorithm

This section provides the self-adaptive ACS (SAACS) algorithm for feature selection. Figure 1 displays three main components i.e., i) the feature extraction and graph modeling, ii) the SAACS algorithm,

and iii) the classification process. The first component will produce the weighted directed graph data to SAACS. The SAACS will provide the significant feature subset to be used in the classification process. The SAACS algorithm uses the sigmoid function to adaptively control the amount of local pheromone update value, drawing inspiration from the activation function of neural networks [27]. The experiments are assembled using a different number of features in order to determine the influence of these different selections on accuracy results. First, the sky images were rescaled in fixed size 500×500 pixels and converted into greyscale. Next, we randomly select 70% of sky images for training and 30% of sky images for testing. The classifier is SVM, multilayer perceptron (MLP), kernel support vector machine (KSVM), KNN, decision tree (DT), and random forest (RF).
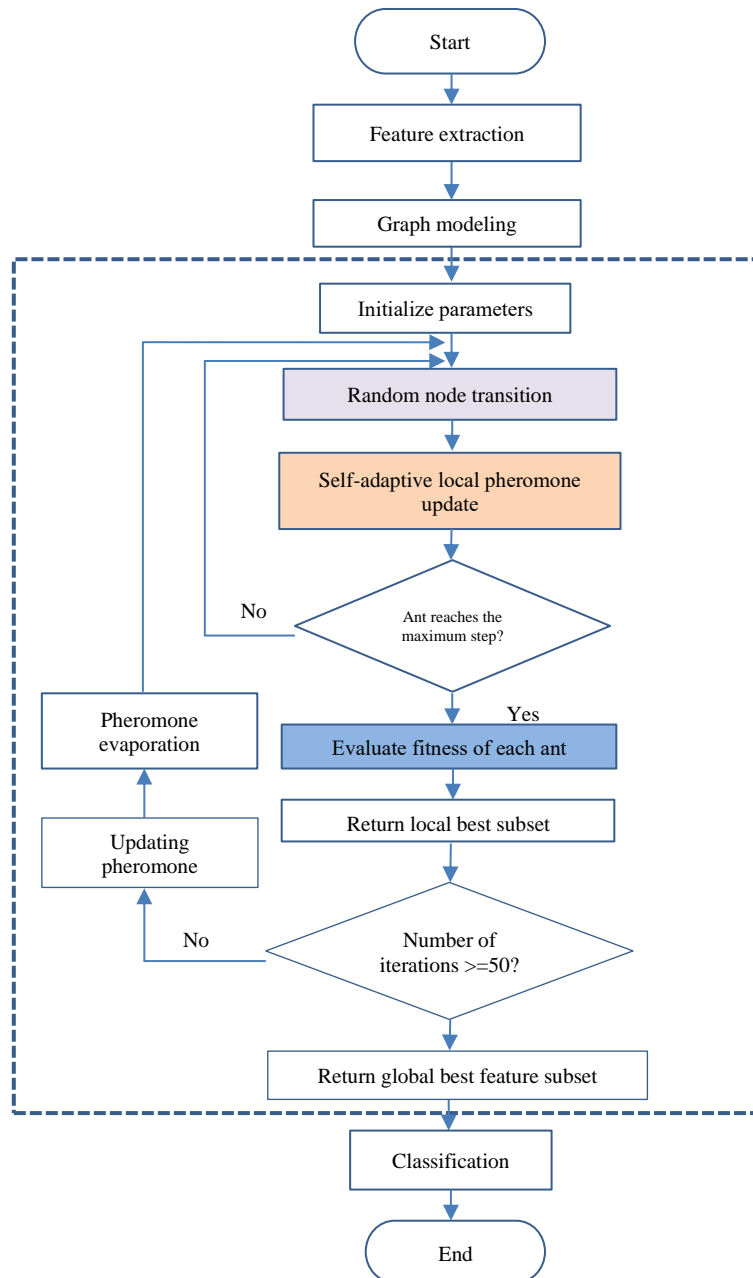


Figure 1. SAACS feature selection algorithm

The graph modeling part is intended to represent the sky image obtained from the feature extraction technique and thus provide the search space for selecting the significant features. The search space presented as a pair $(F, V)$, where $F$ is a feature and $V$ assigned as an intensity value $V(f_i) \in [0,255]$ to each feature

$f_i = (x_{f_i}, y_{f_i}) \in F$. $x_{f_i}$ and $y_{f_i}$ indicate the position $x$ and $y$ of feature $f_i$ respectively. The $n$ features are presented as $f_1, f_2, \ldots, f_n$ in the graph modeling $G = (v, e)$, where node $v_1$ is presented as feature $f_1$ and $e$ is the edge link to the nearest node. The directed graph is formed by connecting the two nodes $v_i$ and $v_j$ if the weight of the related features of two nodes is more than 1.

The SAACS for feature selection starts by initialization of the parameters as shown in Table 1. Nodes $f_i$ are related to pheromone values $\tau_i$. Each ant $k$ is distributed to a particular node in the first step, which can travel and consider any other node in the graph $s_k \leftarrow \{f_i\}$, where $s_k$ is the subset by ant $k$. Ants carry out a forward selection in which each ant $k$ grows its subset $s_k$ incrementally by adding new features. Each ant $k$ explores all features in the set $N - s_k$ and selects the following feature to include in $s_k$ based on the ACS-based feature selection algorithm.

Table 1. Parameter setting for SAACS algorithm

| Parameter | Description | Value |
|---|---|---|
| $m$ | Amount of ant | 50 |
| $I$ | Iterations | 50 |
| $\rho$ | Pheromone decay | 0.5 |
| $\tau_0$ | Initial amount of pheromone on the edge | 0.5 |
| $\alpha$ | Importance of the pheromone value | 0.12 |
| $\eta$ | Heuristic value | 0.5 |
| $\beta$ | Importance of heuristic value | 1.22 |
| $q_0$ | Exploration/exploitation parameter | 0.3 |

Ant transition in SAACS algorithm is where each ant will choose a feature subset at random from a total of $m$ features. In the first stage, each ant selects a node that corresponds to its direction in time $t$. The random transition utilizing the proposed in (1) and (2),

$$J = \begin{cases} argmax_{i \in N_i^m} \{\tau_{il} [\eta_{il}]^\beta\} & if\ q \le q_0 \\ Z_{ij}^m & otherwise \end{cases} \quad (1)$$

where parameter $q_0$ is specified as $0 \le q_0 \le 1$ and $q$ is a random variable with uniform distribution in the closed interval [0,1]. $Z_{ij}^m$ is established using:

$$Z_{ij}^m = \lambda [\tau_{ij}]^\alpha [\eta_{ij}]^\beta, j \in N_i^m \quad (2)$$

where $\alpha$ and $\beta$ are two parameters for balancing the weights between the pheromone and heuristic value, and $N_i^m$ refers to the neighbor nodes of node $i$ which ant $m$ has not yet visited.

The total number of features is given by the value of $N$. The amount of pheromone for each feature in the initial iteration was assigned a minimal random value. Heuristic value also influences a feature's productivity. The mean decrease in impurity, which offers a relative feature importance, is used to set the value of $\eta_{ij}$. The relative by ordering the relevance of the features is provided by this feature significance score [28].

The ACS-based selection process included the ant transition and local pheromone update rule as two crucial steps to improve the exploration mechanism. The transition probability of ACS is used to determine whether the current $\lambda$ is dominated or not. As a result, the ants determine which features are sufficiently significant to include in the feature set based on the value of a parameter $\lambda$. In other words, these important steps are guided by the appropriate $\lambda$ value for direct subset formation which can maximize dependency and minimize redundancy among the features.

In the global pheromone updating process, the quantity of pheromones of each chosen node and feature is updated once each iteration $i$ of the graph has been fully traversed by all the ants. The updating rule is carried out by utilizing the (3) [29] to update the quantity of pheromone for each feature.

$$\tau_{ij} \leftarrow f(x) = \begin{cases} (1 - \rho)\tau_{ij} + \rho\Delta\tau_{ij}^{bs} & \forall (i, j) \in T^{bs} \\ \tau_{ij} & otherwise \end{cases} \quad (3)$$

where $\rho$ is the evaporation factor, $\tau_{ij}$ corresponds to the current quantity of pheromone on the link $(i, j)$, $T^{bs}$ is the iteration's best tour so far, and $\Delta\tau_{ij}^{bs}$ is the reward provided to the best tour.

Each node that has been visited will undergo the local pheromone update procedure by the corresponding ant [29]. As a result, choosing an appropriate pheromone level is essential to navigate the search space and discovering the most comprehensive optimal solution. In the early stages of the evolution process, all individuals are motivated to thoroughly explore the entire search space. The ants are then encouraged towards convergence to the global optimum and establish the optimal solution in a later stage of the process. The local pheromone update is proposed to be adaptively controlled using an adaptive weighting strategy through the proposed (4). This is done by employing a sigmoid function based on feedback collection and reward mechanism for determining the significant features to be included in the final subset. Significant features are determined according to (5) [27],

$$\tau_{ij} = (1 - \rho)\tau_{ij} + (1 - \lambda(\tau_{ij}))\tau_0 \tag{4}$$

where $\lambda$ is weighted activation function and $\tau_0$ is the initial pheromone value,

$$\lambda(x) = \left(\frac{1}{1+e^{-x}}\right) \tag{5}$$

where $e$ is the natural logarithm and $x$ is the input to the function which is determined by (4).

The evaluation function is the core component of any feature selection method. The function evaluates the quality of engaging features based on their abilities to distinguish between various classes to determine the optimality of subsets. The wrapper technique is used in this study's evaluation process. This approach evaluates various combinations of features using a learning algorithm, and after a number of evaluation rounds, the best optimal features are shown.

The experiments are repeated 50 times and the average accuracy is used for comparison. This implies some sort of classification decision feedback mechanism and evaluation criteria to modify the searching of significant features. The average of all the fitness functions has been calculated. Following convergence, the best ant's relevant feature set was chosen to prune the feature dimension. The fitness function $f(s)$ of solution $s$ is defined as in (6) [28].

$$f(s) = argmax(accuracy(s)) \tag{6}$$

## 2.2. Dataset

The widely accepted sky image datasets i.e., Kiel, Singapore whole-sky imaging categories (SWIMCAT), MGC Diagnostics Corporation (MGCD), and greatest common divisor (GCD) were used in this study. Experiments are conducted on different benchmark classification problems available in the literature to demonstrate the performance of the proposed algorithm. The description of the datasets is shown in Table 2. Figure 2 displays the images and classes of all four datasets. The Kiel and SWIMCAT is a small dataset, but MGCD and GCD are very large datasets.

A calibrated ground-based WSI created by [30] in Singapore between January 2013 and May 2014 is presented in Figure 2(a). The automatic wide angled high-resolution sky imaging camera collected image patches for the whole-sky imaging categories database. This database included images from each of the 5 cloud categories. A total of 784 image patches were chosen. Each image patch size is 125×125 pixels.

Kiel dataset categorizes sky images into 7 classes with a resolution of 2,272×1,704 pixels as shown in Figure 2(b). The sky images were captured when a German researcher's team was studying a project named "Polarstern" during a transit of a research vessel. The sky images are different in illumination and intra-class variation. Different climates, seasons, and sunray angles impact upon the various images. This sky image dataset is programmed by capturing one sky image every 15 second [9]. Heinle *et al.* [9] selected approximately 1,500 sky images from a total of 75,000 based on independent time and with respect to predefined universal cloud classification system.

The MGCD dataset as shown in Figure 2(c) contains 8,000 cloud samples in JPEG format. This dataset was captured using a sky camera with a fisheye lens of size 1,024×1,024 pixels. The WMO's taxonomic classification guidelines and similarities in cloud appearance were used by the MGCD to categorize the sky conditions into 7 sky classes. "Mix cloud" are a category of cloud where at least two distinct cloud types are typically presented. Additionally, sky image with 10% or less cloud coverage is regarded as clear sky [31], [32].

The largest GCD dataset in Figure 2(d) which comprises 19,000 cloud images, was collected by camera sensors in nine Chinese regions. GCD has a lot of variation in sky condition because it was gathered over an extended period. GCD includes 7 different types of clouds according to the WMO's classification rules. The resolution of cloud images in GCD is 512×512 pixels, and they are recorded in JPEG format [8], [32].

Table 2. Datasets summary

| Dataset | Number of classes | Number of images | Training | Testing | Size |
|---|---|---|---|---|---|
| Kiel | 7 | 1,500 | 1,050 | 450 | Small |
| SWIMCAT | 5 | 784 | 548 | 235 | Small |
| MGCD | 7 | 8,000 | 5,600 | 2,400 | Very large |
| GCD | 7 | 19,000 | 13,300 | 5,700 | Very large |

| Clear sky | Pattern | Thick dark | Thick white | Veil |
|---|---|---|---|---|

(a)

| Cumulus | Cirrus | Altocumulus | Clear sky | Stratocumulus | Stratus | Cumulonimbus |
|---|---|---|---|---|---|---|

(b)

| Cumulus | Altocumulus and Cirrocumulus | Cirrus and Cirrostratus | Clear sky | Stratocumulus and Stratus and Altostratus | Cumulonimbus and Nimbostratus | Mixed cloud |
|---|---|---|---|---|---|---|

(c)

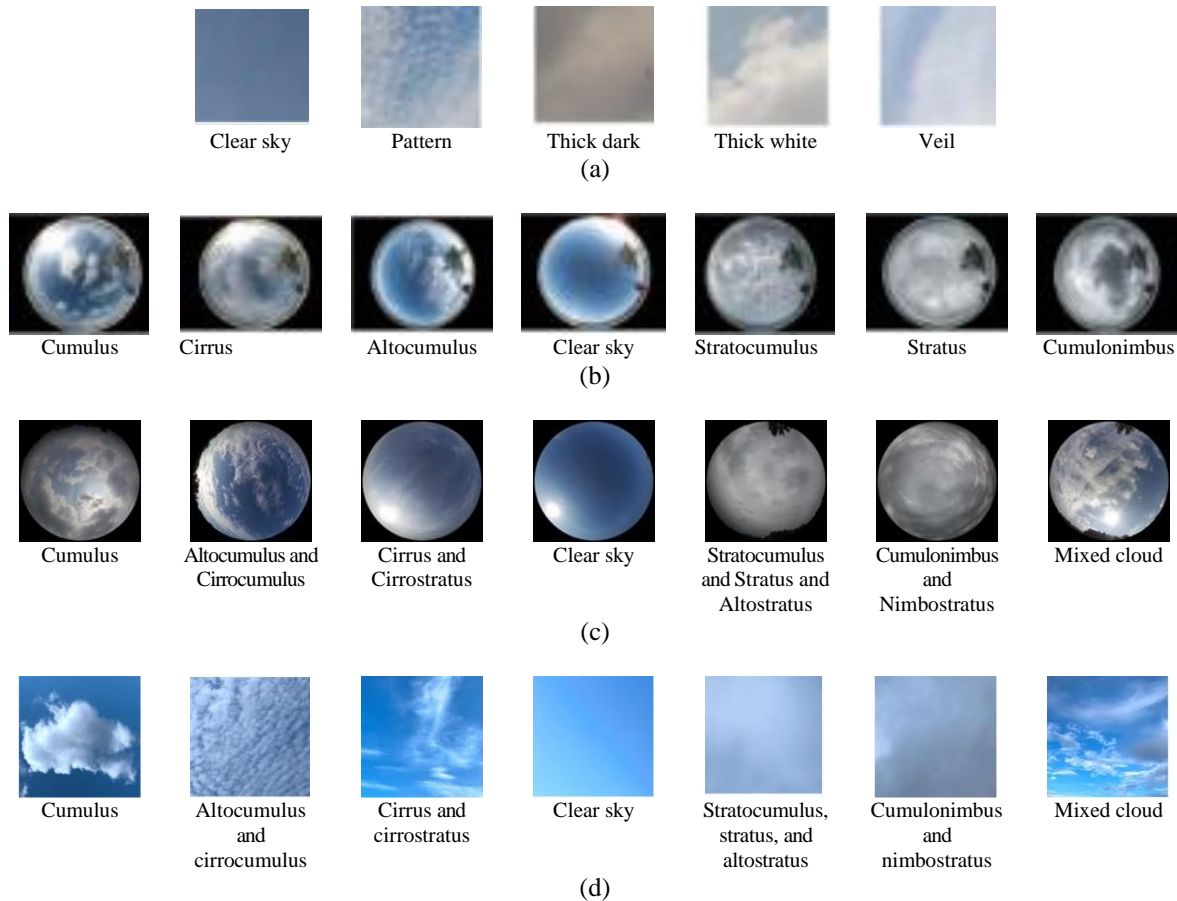| Cumulus | Altocumulus and cirrocumulus | Cirrus and cirrostratus | Clear sky | Stratocumulus, stratus, and altostratus | Cumulonimbus and nimbostratus | Mixed cloud |
|---|---|---|---|---|---|---|

(d)

Figure 2. Representation of four sky image datasets (a) SWIMCAT, (b) Kiel, (c) MGCD, and (d) GCD

## 3. RESULTS AND DISCUSSION

Experiments were carried out on the four datasets to evaluate the SAACS algorithm. The classifiers that were used in the classification process are SVM, KSVM, MLP, RF, KNN and DT. Six benchmark bio-inspired algorithms have been successfully used for performance comparison namely, road sign detection and recognition (RSDR) [15], levy flight-based GWO (LFGWO) [12], binary BA (BBA) [14], ACO [13], modified ALO (MALO) [16], and ACO with rough dependency measure (ACO_RDM) [17]. The classification accuracy, number of selected features, similarity score, precision, recall and f-measure metrics are used as the algorithm performances. Calculation of the accuracy is as given in (7) [33] which indicates of the correctly classified cloud type from the total number of samples in the dataset.

$$\%Accuracy = \left( \frac{number\ of\ correctly\ classify\ samples}{total\ number\ of\ samples} \right) \times 100 \tag{7}$$

Precision detects the rate of true positives among all positive values while recall is used to compute the capability of the positive case. A harmonic mean of recall and precision is measured using the f-measure. Precision, recall and f-measure are computed using (8), (9) and (10) respectively as:

$$Precision = \left( \frac{\sum True\ Positve}{\sum Predicted\ condition\ positive} \right) \times 100 \tag{8}$$

$$Recall = \left( \frac{\sum True\ Positve}{\sum Condition\ positive} \right) \times 100 \tag{9}$$

$$F\text{-}measure = 2 \times \frac{Positive\ predictive\ value \times True\ positive\ rate}{Positive\ predictive\ value + True\ positive\ rate} \tag{10}$$

Cosine similarity is used to determine the angle between two features' cosine value. This measurement provides details regarding the direction of two feature vectors without taking into account their magnitudes. The similarity is measured using (11) [17] as:

$$sim\left(f_i f_j\right) = \frac{\sum_{a=1}^{m}(f_{ia}f_{ja})}{\sqrt{(\sum_{a=1}^{m}f_{ia}^2)(\sum_{a=1}^{m}f_{ia}^2)}} \tag{11}$$

where $f_i$, $f_j$ are any two features in $m$ feature vectors. The performance of the algorithms was also compared using Friedman test and Mann-Whitney $U$ test. Friedman test is used to generate a ranking across multiple algorithms. This nonparametric statistical test is to ascertain the average categorization accuracy rank of the algorithms. The smallest value for the rank implies the best performance [34]. Mann-Whitney $U$ test is used to show a significant difference between two independent groups. The $p$ value of the Mann-Whitney $U$ test will reveal whether there is significant difference in the algorithm's performance [35].

The results of number of selected features and cosine similarity of the algorithms is shown in Table 3. In summary, SAACS has achieved best similarity value in Kiel, SWIMCAT and MGCD datasets while for the GCD dataset, the SAACS has obtained the smallest number of selected features. Table 4 tabulates the average classification accuracy for all four datasets which are combined to form a single dataset. The figure in parenthesis is the performance rank. The results demonstrate that SAACS reaches the best average accuracy compared to other algorithms. Through the use of the Gaussian smoothness standard deviation, the possibility of capturing the dominant information in sky images as well as applicable to the accurate classification is increased. This technique contributes the mean information of pixel distribution in each of the patch hierarchical properly in the sky image. Then, graph modeling technique is introduced to create the relationship between features for highlighting the high relevance features. Furthermore, the final feature subset is selected by formulating activation function in local pheromone update value and heuristic information to selecting the significant features in every step of ant.

Table 3. Number of selected features and cosine similarity of each algorithm for each dataset

| Algorithm | Kiel | | SWIMCAT | | MGCD | | GCD | |
|---|---|---|---|---|---|---|---|---|
| | No. of selected features | Similarity value | No. of selected features | Similarity value | No. of selected features | Similarity value | No. of selected features | Similarity value |
| RSDR | 1,898 | 0.81 | 1,936 | 0.88 | 1,890 | 0.94 | 1,912 | 0.92 |
| LFGWO | 526 | 0.92 | 636 | 0.97 | 632 | 0.87 | 654 | **0.94** |
| BBA | 1,806 | 0.91 | 1,850 | 0.97 | 1,865 | 0.9 | 1,902 | 0.92 |
| ACO | 1,871 | 0.92 | 1,864 | 0.96 | 1,840 | 0.87 | 1,738 | 0.92 |
| MALO | 1,834 | 0.93 | 1,845 | 0.96 | 1,734 | 0.89 | 1,749 | 0.88 |
| ACO_RDM | 1,643 | 0.84 | 1,482 | 0.92 | 1,368 | 0.91 | 1,412 | 0.85 |
| SAACS | **63** | **0.97** | **64** | **0.98** | **83** | **0.96** | **100** | 0.93 |

Table 4. Average classification accuracy for combined datasets

| Classifier | Algorithm | | | | | | |
|---|---|---|---|---|---|---|---|
| | RSDR (%) | LFGWO (%) | BBA (%) | ACO (%) | MALO (%) | ACO_RDM (%) | SAACS (%) |
| KSVM | 80.92 | 84.7 | 83.68 | 83.32 | 82.05 | 81.69 | **89.86** |
| | (7) | (2) | (3) | (4) | (5) | (6) | **(1)** |
| MLP | 80.07 | 86.15 | 84.76 | 84.14 | 82.85 | 83.84 | **88.59** |
| | (7) | (2) | (3) | (4) | (6) | (5) | **(1)** |
| RF | 82.96 | 88.24 | 87.7 | 86.32 | 87.15 | 87.81 | **95.64** |
| | (7) | (2) | (3) | (6) | (4) | (5) | **(1)** |
| SVM | 74.1 | 85.35 | 79.74 | 81.41 | 82.99 | 77.38 | **85.55** |
| | (7) | (2) | (5) | (4) | (3) | (6) | **(1)** |
| KNN | 79.72 | 84.02 | 83.41 | 77.34 | 81.82 | 79.32 | **85.28** |
| | (6) | (2) | (3) | (7) | (4) | (5) | **(1)** |
| DT | 76.2 | 82.77 | 80.89 | 79.18 | 81.67 | 78.48 | **86.09** |
| | (7) | (2) | (4) | (5) | (3) | (6) | **(1)** |
| Average accuracy | 78.91 | 85.21 | 83.36 | 81.95 | 83.62 | 81.51 | **89.17** |
| | (6) | (2) | (3) | (4) | (2) | (5) | **(1)** |

The benchmark image feature selection algorithms have two main drawbacks i.e., trapped into local optimum and immature convergence in later stage [12]–[17] due to incomplete search space exploration. In other words, those algorithms rely on the positive feedback principle to reinforce the best solution which converges prematurely before the best solution is founded. Thus, the concept of local pheromone update of ACS has proved to be effective to search the solution in wider feature space. ACS provides the local pheromone update parameter which can avoid stagnation and premature convergence by decreasing the pheromone value on previously used edges and makes them less attractive for other ants. The SAACS parameters were designed considering the input from the feedback of fitness function and are all automatically adaptable when the solution is generated. In this fashion, the path that leads to significant features may be chosen by the ants using the pheromone level and heuristic information. Tables 5 to 8 show the results of three performance metrics including average precision, recall and F-measure on each dataset. It can be concluded that the SAACS provides the best average performance for all datasets.

Table 5. Precision, recall, and f-measure for Kiel dataset

| Algorithm | Evaluation metric | | |
|---|---|---|---|
| | Precision | Recall | f-measure |
| RSDR | 0.8167 | 0.7997 | 0.8022 |
| LFGWO | 0.8728 | 0.8652 | 0.8661 |
| BBA | 0.8389 | 0.8214 | 0.8243 |
| ACO | 0.8252 | 0.8091 | 0.8109 |
| MALO | 0.8482 | 0.8354 | 0.8382 |
| ACO_RDM | 0.8251 | 0.8128 | 0.8331 |
| SAACS | **0.9047** | **0.8989** | **0.9003** |

Table 6. Precision, recall, and f-measure for SWIMCAT dataset

| Algorithm | Evaluation metric | | |
|---|---|---|---|
| | Precision | Recall | f-measure |
| RSDR | 0.7479 | 0.7316 | 0.7358 |
| LFGWO | 0.8607 | 0.8539 | 0.8537 |
| BBA | 0.8227 | 0.8082 | 0.8110 |
| ACO | 0.8089 | 0.7898 | 0.7932 |
| MALO | 0.8081 | 0.7965 | 0.7966 |
| ACO_RDM | 0.7894 | 0.7795 | 0.7774 |
| SAACS | **0.8894** | **0.8799** | **0.8806** |

Table 7. Precision, recall, and f-measure for MGCD dataset

| Algorithm | Evaluation metric | | |
|---|---|---|---|
| | Precision | Recall | f-measure |
| RSDR | 0.8205 | 0.8084 | 0.8087 |
| LFGWO | 0.8593 | 0.8494 | 0.8504 |
| BBA | 0.8646 | 0.8552 | 0.8555 |
| ACO | 0.8562 | 0.8446 | 0.8452 |
| MALO | 0.8552 | 0.8442 | 0.8449 |
| ACO_RDM | 0.8503 | 0.8361 | 0.8368 |
| SAACS | **0.8911** | **0.8862** | **0.8863** |

Table 8. Precision, recall, and f-measure for GCD dataset

| Algorithm | Evaluation metric | | |
|---|---|---|---|
| | Precision | Recall | f-measure |
| RSDR | 0.8363 | 0.8201 | 0.822 |
| LFGWO | 0.8503 | 0.8387 | 0.8402 |
| BBA | 0.8625 | 0.8483 | 0.85 |
| ACO | 0.8688 | 0.8498 | 0.8513 |
| MALO | 0.8605 | 0.8404 | 0.8442 |
| ACO_RDM | 0.8483 | 0.8285 | 0.8318 |
| SAACS | **0.884** | **0.8771** | **0.8761** |

Table 9 depicts the performance rank of 6 benchmark algorithms and the proposed algorithm in different classifiers using the Friedman test. This statistical test computes the rank for the algorithm in terms of the classification accuracy. The smallest value for the classification accuracy indicates the highest rank. SAACS has obtained the first rank for all the classifiers.

Table 9. Performance rank of all datasets for each classifier

| Classifier | Algorithm | | | | | | |
|---|---|---|---|---|---|---|---|
| | RSDR | LFGWO | BBA | ACO | MALO | ACO_RDM | SAACS |
| KSVM | 5 | 4.251 | 3.625 | 4.523 | 4.753 | 4.752 | **1.125** |
| MLP | 5.254 | 4.252 | 3.756 | 4.502 | 4.514 | 4.522 | **1.25** |
| RF | 5.754 | 3.756 | 4.511 | 4.783 | 3.756 | 4.251 | **1** |
| SVM | 6.252 | 3 | 4.25 | 3.5 | 3.114 | 5.75 | **2.25** |
| KNN | 5.444 | 2.5 | 2.753 | 6.517 | 3.521 | 5.515 | **1.752** |
| DT | 4.468 | 2.254 | 3.75 | 5.25 | 3.5 | 5.75 | **1** |

In Table 10, Mann-Whitney $U$ test shows the result of average accuracy of the first and second place algorithms (SAACS and LFGWO) are tested for any significance difference. The method used for this test is the nonparametric Mann-Whitney $U$ test with confidence interval of 95%. Any $p$ value which is less than 0.05 indicates a significant difference. The result of Mann-Whitney $U$ test indicates that SAACS is significantly better than the LFGWO benchmark algorithm in all classifiers.

Table 10. $p$ values from Mann-Whitney $U$ test

| Classifier | Algorithm | Average rank | $p$ value |
|---|---|---|---|
| KSVM | LFGWO | 4.251 | 0.0148 |
| | SAACS | 1.125 | |
| MLP | LFGWO | 4.252 | 0.0151 |
| | SAACS | 1.25 | |
| RF | LFGWO | 3.756 | 0.0145 |
| | SAACS | 1 | |
| SVM | LFGWO | 3 | 0.0151 |
| | SAACS | 2.25 | |
| KNN | LFGWO | 2.5 | 0.0303 |
| | SAACS | 1.752 | |
| DT | LFGWO | 2.254 | 0.0156 |
| | SAACS | 1 | |

## 4. CONCLUSION

This paper has proposed a new feature selection algorithm that improves the local pheromone update value and heuristic information of the original ACS in classifying the cloud type from ground-based sky images. The level of local pheromone update value and heuristic information are adaptively controlled by employing a sigmoid activation function based on feedback information and reward mechanism. Therefore, the most significant feature subset of the extracted features is generated. The benchmark comparison of six image feature selection algorithms on four sky image datasets has shown that the SAACS algorithm outperforms the benchmark algorithms. SAACS was able to leap out from the local optimum because of its ability to explore a wider feature space, which significantly increases classification accuracy with a small number of features. This paper emphasizes the exploration mechanism. However, as a recommendation for future work, the balance between the exploration and exploitation mechanisms has also to be improved. The SAACS can also be used to determine the most important features in other domains, with the purpose of selecting significant information from the image. Disaster management, medical diagnosis, industrial inspection, sports management, and content-based image retrieval are examples of these domains.

## REFERENCES

[1] S. Oikonomou, A. Kazantzidis, G. Economou, and S. Fotopoulos, "A local binary pattern classification approach for cloud types derived from all-sky imagers," *International Journal of Remote Sensing*, vol. 40, no. 7, pp. 2667–2682, Apr. 2019, doi: 10.1080/01431161.2018.1530807.

[2] Y. Cao and H. Yang, "Weather prediction using cloud's images," in *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, Jan. 2022, pp. 820–823, doi: 10.1109/BDICN55575.2022.00157.

[3] T. Kliangsuwan and A. Heednacram, "FFT features and hierarchical classification algorithms for cloud images," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 40–54, Nov. 2018, doi: 10.1016/j.engappai.2018.08.008.

[4] Y. Wang, C. Shi, C. Wang, and B. Xiao, "Ground-based cloud classification by learning stable local binary patterns," *Atmospheric Research*, vol. 207, pp. 74–89, Jul. 2018, doi: 10.1016/j.atmosres.2018.02.023.

[5] S. Liu, L. Duan, Z. Zhang, X. Cao, and T. S. Durrani, "Multimodal ground-based remote sensing cloud classification via learning heterogeneous deep features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7790–7800, Nov. 2020, doi: 10.1109/TGRS.2020.2984265.

[6] R. Tapakis and A. G. Charalambides, "Equipment and methodologies for cloud detection and classification: a review," *Solar Energy*, vol. 95, pp. 392–430, Sep. 2013, doi: 10.1016/j.solener.2012.11.015.

[7] C. Fang, K. Jia, P. Liu, and L. Zhang, "Research on cloud recognition technology based on transfer learning," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov. 2019, pp. 791–796, doi: 10.1109/APSIPAASC47483.2019.9023267.

[8] S. Liu, L. Duan, Z. Zhang, X. Cao, and T. S. Durrani, "Ground-based remote sensing cloud classification via context graph attention network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022, doi: 10.1109/TGRS.2021.3063255.

[9] A. Heinle, A. Macke, and A. Srivastav, "Automatic cloud classification of whole sky images," *Atmospheric Measurement Techniques*, vol. 3, no. 3, pp. 557–567, May 2010, doi: 10.5194/amt-3-557-2010.

[10] N. Naheed, M. Shaheen, S. Ali Khan, M. Alawairdhi, and M. A. Khan, "Importance of features selection, attributes selection, challenges and future directions for medical imaging data: a review," *Computer Modeling in Engineering and Sciences*, vol. 125, no. 1, pp. 315–344, 2020, doi: 10.32604/cmes.2020.011380.

[11] M. Petwan and K. R. Ku-Mahamud, "A review on bio-inspired optimization method for supervised feature selection," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022, doi: 10.14569/IJACSA.2022.0130516.

[12] Y. Pathak, K. V Arya, and S. Tiwari, "Feature selection for image steganalysis using levy flight-based grey wolf optimization," *Multimedia Tools and Applications*, vol. 78, no. 2, pp. 1473–1494, Jan. 2019, doi: 10.1007/s11042-018-6155-6.

[13] R. Sharma and A. Purohit, "Image feature selection using ant colony optimization," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 2669–2674, Nov. 2019, doi: 10.35940/ijitee.L3396.119119.

[14] F. Liu, X. Yan, and Y. Lu, "Feature selection for image steganalysis using binary bat algorithm," *IEEE Access*, vol. 8, pp. 4244–4249, 2020, doi: 10.1109/ACCESS.2019.2963084.

[15] A. Jayaprakash and C. KeziSelvaVijila, "Feature selection using ant colony optimization (ACO) and road sign detection and recognition (RSDR) system," *Cognitive Systems Research*, vol. 58, pp. 123–133, Dec. 2019, doi: 10.1016/j.cogsys.2019.04.002.

[16] M. Wang, C. Wu, L. Wang, D. Xiang, and X. Huang, "A feature selection approach for hyperspectral image based on modified ant lion optimizer," *Knowledge-Based Systems*, vol. 168, pp. 39–48, Mar. 2019, doi: 10.1016/j.knosys.2018.12.031.

[17] J. D. Sweetlin, H. K. Nehemiah, and A. Kannan, "Computer aided diagnosis of pulmonary hamartoma from CT scan images using ant colony optimization based feature selection," *Alexandria Engineering Journal*, vol. 57, no. 3, pp. 1557–1567, Sep. 2018, doi: 10.1016/j.aej.2017.04.014.

[18] M. Dorigo and T. Stützle, "Ant colony optimization: overview and recent advances," in *International Series in Operations Research and mathsemicolon Management Science*, 2010, pp. 227–263.

[19] H. J. A. Nasir, K. R. Ku-Mahamud, and E. Kamioka, "Parameter adaptation for ant colony system in wireless sensor network," *Journal of Information and Communication Technology*, vol. 18, no. 2, pp. 167–182, Apr. 2019, doi: 10.32890/jict2019.18.2.3.

[20] K. R. Ku-Mahamud and M. M. Alobaedy, "New heuristic function in ant colony system algorithm for optimization," *Proceedings of MAMECTIS'13: The 15th international Conference on Mathematical Methods, Computational Techniques and Intelligent Systems*, pp. 13–18, 2013.

[21] N. Sakthipriya and T. Kalaipriyan, "Variants of ant colony optimization-a state of an art," *Indian Journal of Science and Technology*, vol. 8, no. 31, Nov. 2015, doi: 10.17485/ijst/2015/v8i31/87296.

[22] S. Yasear and K. Ku-Mahamud, "Fine-tuning the ant colony system algorithm through Harris's hawk optimizer for travelling salesman problem," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 4, pp. 136–145, Aug. 2021, doi: 10.22266/ijies2021.0831.13.

[23] O. S. S. Júnior and J. E. Leal, "A multiple ant colony system with random variable neighborhood descent for the vehicle routing problem with time windows," *International Journal of Logistics Systems and Management*, vol. 40, no. 1, 2021, doi: 10.1504/IJLSM.2021.117687.

[24] Y. Wan, T.-Y. Zuo, L. Chen, W.-C. Tang, and J. Chen, "Efficiency-oriented production scheduling scheme: an ant colony system method," *IEEE Access*, vol. 8, pp. 19286–19296, 2020, doi: 10.1109/ACCESS.2020.2968378.

[25] D. Moussaoui, M. Hadjila, S. M. H. Irid, and S. Souiki, "Clustered chain founded on ant colony optimization energy efficient routing scheme for under-water wireless sensor networks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5197–5205, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5197-5205.

[26] K. Benhamza and H. Seridi, "Improvement on image edge detection using a novel variant of the ant colony system," *Journal of Circuits, Systems and Computers*, vol. 28, no. 5, May 2019, doi: 10.1142/S0218126619500804.

[27] B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks," in *2018 Chinese Control And Decision Conference (CCDC)*, Jun. 2018, pp. 1836–1841, doi: 10.1109/CCDC.2018.8407425.

[28] B. H. Menze *et al.*, "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC Bioinformatics*, vol. 10, no. 1, Dec. 2009, doi: 10.1186/1471-2105-10-213.

[29] R. Bello, A. Puris, A. Nowe, Y. Martínez, and M. M. García, "Two step ant colony system to solve the feature selection problem," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2006, pp. 588–596.

[30] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "WAHRSIS: A low-cost high-resolution whole sky imager with near-infrared capabilities," in *SPIE Proceedings*, May 2014, doi: 10.1117/12.2052982.

[31] S. Liu, M. Li, Z. Zhang, X. Cao, and T. S. Durrani, "Ground-based cloud classification using task-based graph convolutional network," *Geophysical Research Letters*, vol. 47, no. 5, Mar. 2020, doi: 10.1029/2020GL087338.

[32] S. Liu, M. Li, Z. Zhang, B. Xiao, and T. S. Durrani, "Multi-evidence and multi-modal fusion network for ground-based cloud recognition," *Remote Sensing*, vol. 12, no. 3, Feb. 2020, doi: 10.3390/rs12030464.

[33] G. Saleem, M. Akhtar, N. Ahmed, and W. S. Qureshi, "Automated analysis of visual leaf shape features for plant classification," *Computers and Electronics in Agriculture*, vol. 157, pp. 270–280, Feb. 2019, doi: 10.1016/j.compag.2018.12.038.

[34] J. Demvsar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine learning research*, vol. 7, pp. 1–30, 2006.

[35] P.-H. Dinh, "An improved medical image synthesis approach based on marine predators algorithm and maximum Gabor energy," *Neural Computing and Applications*, vol. 34, no. 6, pp. 4367–4385, Mar. 2022, doi: 10.1007/s00521-021-06577-4.

## BIOGRAPHIES OF AUTHORS

**Montha Petwan** received the B.B.A degree in Computer Business from Sripatum University, Thailand, 2001. Master's degree is in Management Information Technology from Prince of Songkhla University in 2006. She is currently working toward the Ph.D. degree in computer science at the Universiti Utara Malaysia, Malaysia. She is lecturer in Suratthani Rajabhat University, Thailand. Her research interests include machine learning, artificial intelligence, and computer vision. She can be contacted at monta.pet@sru.ac.th.

**Ku Ruhana Ku-Mahamud** received the B.Sc. (Hons) in Mathematical Sciences, University of Bradford, England. Her PostGrad Diploma is in Education, Universiti Kebangsaan Malaysia. Her M.Sc. is in Computing, University of Bradford, England, and Ph.D. in Computer Science from Universiti Pertanian Malaysia. She is currently a professor in the Department of Computer Science, School of Computing, Universiti Utara Malaysia. Her research interests include computer systems performance modeling, ant colony optimization and intelligent agent. She can be contacted ruhana@uum.edu.my.