



Hybrid Feature Selection with Parallel Multi-Class Support Vector Machine for Land Use Classification

Bharani Basapathy Rudra^{1*} Gururaj Murtugudde²

¹*Department of Information Science & Engineering, Cambridge Institute of Technology, Bangalore, India*

²*Department of Computer Science & Engineering, Sapthagiri College of Engineering, Bangalore, India*

* Corresponding author's Email: bharanisuri@gmail.com

Abstract: Land use classification in remote sensing is required in various applications like natural resource management, urban mapping and agriculture etc. Existing methods in the Land use classification which has the limitation of overfitting problem due to the improper feature selection in the method. In this research, the hybrid feature selection methods with Parallel Multi-Class Support Vector Machine (MSVM) is proposed to improve the land use classification performance. The UC Merced and AID datasets were applied to validate the performance of the hybrid feature selection method with the parallel MSVM method. The input images were applied in Histogram Equalization to enhance the image quality which removes the artifacts in the preprocessing stage. The Speeded Up Robust Feature (SURF), Local Ternary Pattern (LTP), Discrete Wavelet Transform (DWT) were applied for feature extraction. The extracted features are applied to hybrid feature selection of Particle Swarm Optimization (PSO) and Grey Wolf Optimization (GWO) method to select the relevant features. The hybrid feature selection method has the advantages of good convergence with higher efficiency in search analysis. The PSO model provided good search exploration to find better solution and GWO method has good convergence of local and global solution. The hybrid method has effective exploration and exploitation for the feature selection. The proposed hybrid features with the MSVM method have 99.15 % accuracy and the existing SVM has 94 % accuracy in land use classification.

Keywords: Grey wolf optimization (GWO), Land use classification, Parallel multi-class support vector machine (MSVM), Particle swarm optimization (PSO), Speeded up robust feature (SURF).

1. Introduction

Land cover is the pattern analysis of the human activities and ecological resources in various Earth's surface regions and this type of data supporting land management, and many environmental sciences at global, local and regional scales. Considering Land cover information is important in environmental sustainability research and global change, land cover datasets were created at various scales [1-3]. The classification of land cover like airport, residential, a wetland is challenging due to the presence of complex heterogeneous land cover [4]. This Land use classification provides a series of semantic classes that assist in the land cover information in remote sensing images. Land use classification is required in many applications like natural resource management,

precision agriculture, target detection, and urban mapping. Recently, various researches have been carried out in land use classification for feature representation and classification tasks for land use classification [5]. Remote sensing classification consists of two aspects, a feature extractor that provides discriminative feature vectors from the transform of spatial, spectral, and temporal data, and a classifier that labels each data based on feature representation [6-8].

Land use classification is based on low-level visual features to represent the region of interest in the images and low-level features are either global or local features. The global features are extracted from the images like shape, texture, color (spectral) features. Local features are extracted from the image patches based on the point of interest [9, 10]. Standard classification methods like Random Forest

(RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) have been successfully applied in various image classifications [11]. SVM method has a low sensitivity to high dimensions and provides better performance in pattern recognition applications [12]. Existing methods have the limitation of overfitting problem in the Land use classification which affects the performance [13]. In this research, the hybrid feature selection with the Parallel MSVM method is applied to improve the classification performance of Land use in remote sensing. The PSO method improves the exploration of the search and GWO method increases the exploitation of the search. The selected features are applied in parallel MSVM to improve the classification performance. The AID and UC Merced datasets were used to validate the performance of the model.

The organization of the paper is given as follows: a literature survey of the existing land use classification is given in Section 2. The explanation of the proposed method is given in Section 3 and the simulation setup is given in Section 4. The conclusion of this paper is given in Section 5.

2. Literature survey

Remote Sensing classification is required in many applications like environment monitoring, urban planning, classification of land use/land cover, etc. Various methods were applied in remote sensing classification to improve the efficiency of the classification. Recently, some of the researches in the Remote sensing classification was reviewed in this section.

Wang [14] proposed Attention Recurrent Convolutional Network (ARCNet) method for remote sensing classification. High-level features are extracted from the key location or region of the images where less information is extracted from other locations. High-level semantic features were extracted using ARCNet and several simplex vectors are applied on spatial features for the reduction of learning parameters. ARCNet adaptively selects the series of Attention regions and generates the powerful prediction based on sequential learning. The ARCNet method is tested with the UC Merced Land Use dataset in remote sensing classification. The ARCNet has the limitation of lower performance in imbalance data and has an overfitting problem.

Mahdianpari [15] applied various deep learning methods like InceptionResNetV2, ResNet50, Xception, VGG19, VGG16, InceptionV3, and DenseNet121 in wetland mapping of Canada. The deep learning methods performance is tested based

on the UC Merced Land Use dataset. The performance of CNN is compared with standard classifiers like Random Forest and Support Vector Machine (SVM). Five spectral bands fully-trained convnets have higher performance than other methods. The InceptionResNet V2 has a higher performance in wetland classification than other methods in the analysis. The developed method has a lower performance in multi-class classification in large datasets.

Zhang [16] proposed Capsule Network (CapsNet) that applies the neurons group as vector or a capsule in Traditional Neural Network. The feature spatial information with the properties is encoded in the neural network to improve the classification performance. CNN is applied for the initial feature map extractor where the fully trained CNN is applied for the feature extraction. The CapsNet uses the selected features to provide the final classification results. The CNN-CapsNet method is tested on the AID and UC Merced Land used dataset. The CNN-CapsNet method has the limitation of overfitting problem in the classification that affects the performance.

Zhang [17] proposed Object-based CNN (OCNN) for the classification of remote sensing images. OCNN method depends on the segmented object as it performs on functional units instead of pixel-wise convolution processes. CNN analyses and label object based on the partition of between-object variation and within an object. Two CNN structures with various model structures and window sizes are developed to predict linear-shaped objects and general objects. OCNN methods with small and large window sizes perform higher when compared to other CNN methods. The OCNN model has a lower performance in detecting the smaller object in the remote sensing images.

Chaib [18] proposed Pretrained Visual Geometry Group Network (VGG-Net) to extract information features from the remote sensing images. The VGG-Net fully connected layer is applied as a separated feature discriminator and Discrimination Correlation Analysis (DCA) is applied as a feature fusion method. The VGG-Net model was tested on the UC Merced dataset to evaluate the performance in remote sensing classification. Raw deep features based on the feature fusion method provide higher performance in the classification. The developed method has the limitation of overfitting problem that affects the performance of the method.

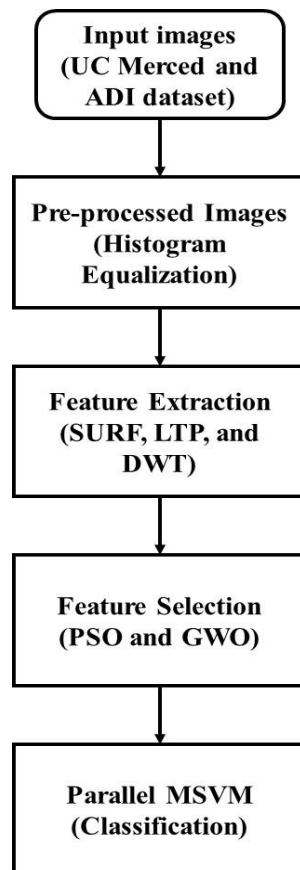


Figure. 1 The hybrid feature selection with Parallel MSVM model block diagram

3. Proposed method

The hybrid feature selection with the Parallel MSVM method is proposed to improve the performance of Land use classification in Remote sensing. In Pre-processing, Histogram equalization is applied to enhance the image quality and remove the artifacts in images. The SURF, LTP, and DWT features were extracted from the pre-processed image. The hybrid of the PSO and GWO methods is applied to select the relevant features for the classifier. The selected features were applied to MSVM and the classification is performed parallel based on the Parallel Computation Toolbox of MATLAB. The overall diagram of the proposed method is shown in Fig. 1.

3.1 Histogram equalization – pre-processing

The histogram Equalization method is applied to enhance the quality of the images and remove unwanted artifacts of the input images. Sometimes, the input images don't provide clear information about the images and this degrades the quality of the results of the model. The contrast enhancement method helps to maintain the image with natural aspects and smoothness in the image. Histogram

Equalization [19] is applied to improve the visualization by preserving the details.

According to Eq. (1), the Histogram Equalization maps the intensity level of input i to its output level X_i .

$$X_i = (N - 1) \times C_{X(i)} \quad (1)$$

Where the image total intensity levels are represented as N and a cumulative histogram is represented as $C_{X(i)}$. This method is suitable to continue the intensity levels of images and equalizes the histogram.

3.2 Feature extraction

The SURF, LTP and DWT feature extraction methods were applied in this method for the feature representation.

3.2.1. SURF method

The SURF method provides the effective feature information [20] and the SURF method process is described as follows.

Step 1: The Hessian matrix is given in Eq. (2).

$$H(L(x, y), \sigma) = \begin{bmatrix} \frac{\partial^2 L}{\partial x^2} & \frac{\partial^2 L}{\partial x \partial y} \\ \frac{\partial^2 L}{\partial x \partial y} & \frac{\partial^2 L}{\partial y^2} \end{bmatrix} \quad (2)$$

Where second-order derivative of Gaussian convolution is given as $L(x, y)$ with input signal $f(x, y)$ and the Gaussian filter variance is denoted as $g(x, y, \sigma)$ is σ^2 .

The second order of Gaussian derivatives are transformed to box filters and second-order of Gaussian derivatives approximation is denoted as D_{xx}, D_{yy}, D_{xy} , to simplify the Hessian matrix determinants computation. The Hessian matrix determinants approximation is given in Eq. (3).

$$\det(H_{approx}) = D_{xx}D_{yy} - (\omega D_{xy})^2 \quad (3)$$

Where, the weighted coefficient is denoted as ω .

Step 2: Scaled varied box filter is applied to construct the image pyramid. The starting filter size is set as 9×9 along with the filter size L in SURF is calculated using Eq. (4).

$$L = 3 \times (2^{octave} \times interval + 1) \quad (4)$$

Where Octave value and interval start from 1.

The 2D Gaussian kernel filtering is approximated using size L integral images, which is given in Eq. (5).

$$\sigma = \sigma_0 \times \frac{L}{9} \quad (5)$$

Step 3: In the image pyramid, interest points were located using Non-Maximal Suppression is applied in $3 \times 3 \times 3$ neighborhood. Every pixel is compared with 26 remaining pixels which consist of two nine pixels above and below the scale layer which has eight pixels' scale layer. If the pixel value is more than neighbor values, then the pixel is considered an interesting point.

Step 4: The SURF descriptor is generated and interest points dominant orientation is obtained based on the SURF descriptor. The Haar wavelet transform is applied to calculate the SURF descriptor.

3.2.2. Local ternary pattern (LTP)

The LTP consists of 3-valued codes as $(-1, 0, 1)$ and the texture operator is robust to noise. The LTP mathematical expression is given in Eq. (6).

$$LTP_{P,R} = \sum_{p=0}^{P-1} 2^p s(i_p - i_c), s(x) = \begin{cases} 1, & x \geq t, \\ 0, & -t < x < t, \\ -1, & x < -t, \end{cases} \quad (6)$$

Where user threshold is denoted as t , and P , i_c , i_p , R , are defined. The lower and upper pattern are coded and constructed after the threshold step. The upper and lower pattern code concatenation is the LTP operator.

3.2.3. Discrete wavelet transform

DWT provides the time resolution and frequency information in the input images based on the ability of localization. DWT revealed local characteristics of the input image and provides feature degradation [21]. DWT achieves improved information of low and high frequency for short and long time windows. The DWT is widely used to analyze non-stationary signals. The DWT mathematical formula is provided in Eq. (7).

$$DWT(j, k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x(N) \mu\left(\frac{N-2^jK}{2^j}\right) dN \quad (7)$$

Where μ denotes the wavelet function, $x(N)$ denotes the actual wavelet, and scaling shift parameters is denoted as 2^j and 2^jK .

3.3 Feature selection

The PSO and GWO are applied for the feature selection process and the hybrid method has the advantages of good convergence and better search efficiency.

3.3.1. Particle swarm optimization

The PSO consists of some parameters and definitions in the optimization process and starts with an initial random population, named particles [22]. Each particle consists of solutions for the main problem and process in n -dimensional space. Each particle consists of velocity $V_i^k = (V_{i1}^k, V_{i2}^k, \dots, V_{in}^k)$ and position $x_i^k = (x_{i1}^k, x_{i2}^k, \dots, x_{in}^k)$ in the variable space. The velocity and position of Each particle are updated using Eq. (8) and (9), respectively. If a particle has the best solution, then this is carried out to the next process. The best position is represented as $pbest$ and the best position of every particle is represented as $gbest$.

$$v_i^{k+1} = w_i v_i^k + c_1 r_1 (pbest_i - x_i^k) + c_2 r_2 (gbest_i - x_i^k), \quad (8)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}, \quad (9)$$

The weight function is given in Eq. (10)

$$w_i = w_{max} - \frac{w_{max} - w_{min}}{k_{max}} \times k, \quad (10)$$

Where r_1 and r_2 are random values in the range of $[0, 1]$, particle motion inspect acceleration coefficients are denoted as c_1 and c_2 .

3.3.2. Grey wolf optimization

The hunting behavior of the grey wolf pack is mimic for the optimization method as GWO [23]. GWO consists of various processes like prey, track, encircle and attack. The grey wolf is the hierarchy structure of four levels like alpha α , beta β , omega ω , and delta δ . The alpha α is responsible for decision-making in attacks and beta β helps the alpha in decision making. The beta executes the alpha decision to the lower hierarchy and provides feedback to the alphas. The omega plays the role of scapegoat and the delta are other wolves such as sentinels and scouts. The delta has provided the information to alpha and beta, dominate the omega.

The grey wolves hunting process is simulated where first three best solutions of GWO is saved as $(x_\alpha, x_\beta, x_\delta)$. Every wolf position is updated by a

group of evolving equations as given in Eqs. (11) to (14).

$$y_1 = x_\alpha^j - a_1 \times |c_1 \times x_\alpha^j - x_i^j(t)| \quad (11)$$

$$y_2 = x_\beta^j - a_2 \times |c_2 \times x_\beta^j - x_i^j(t)| \quad (12)$$

$$y_3 = x_\delta^j - a_3 \times |c_3 \times x_\delta^j - x_i^j(t)| \quad (13)$$

$$x_i^j(t+1) = \frac{y_1 + y_2 + y_3}{3} \quad (14)$$

Where $x_i^j(t)$ is the value of i^{th} solution in the j^{th} dimension at the t^{th} iteration, c_1 , c_2 and c_3 random numbers uniformly distributed in a range of $[0, 2]$; a_1 , a_2 and a_3 are the random numbers uniformly distributed in $[-2 \times (1 - t/G), 2 \times (1 - t/G)]$; $i = 1, 2, \dots, m$. Each solution population in size m ; $j = 1, 2, \dots, n$, indexes each dimension of the n -dimensional problem to be solved; $t = 1, 2, \dots, G$, is the number of iterations.

3.4 Parallel multi-class support vector machine

Binary classifiers f_1, f_2, \dots, f_N is constructed for $1 \dots N$ classes, each trained to be different from one class to the others [24]. The multi-class category is obtained based on the maximal output before applying the sgn function.

Where $\text{argmax } g^k(x)$

Where $g^k(x) = \sum_{i=1}^n y_i \alpha_i^k k(x, x_i) + b^k$

Where $k = 1, \dots, N$

Where hyperplane distance to the point x of a signed real value is denoted as $g^k(x)$ which is referred to as the confidence value. The higher value increases the confidence where x belongs to the positive class. The highest confidence value is assigned with x .

The input data is denoted as $X = \{x_1, x_2, \dots, x_m\} \in R^d$, the hypersphere radius is denoted as r , and the center is denoted as $c \in R^d$. The minimum hypersphere which encloses the optimization problem is given in Eq. (15).

Minimize r^2

Subject to $|\Phi(x_j) - c| \leq r^2, j = 1, \dots, m$

$$L(c, r, \alpha) = r^2 + \sum_{j=1}^m \alpha_j \{|\Phi(x_j) - c|^2 - r^2\} \quad (15)$$

Derive $\frac{\partial L(c, r, \alpha)}{\partial c} = 2 \sum_{j=1}^m \alpha_j (\Phi(x_j) - c) = 0$

Eq. (16) is obtained.

$$\sum_{j=1}^m \alpha_j = 1 \text{ and } \sum_{j=1}^m \alpha_j \Phi(x_j) \quad (16)$$

Hence, Eq. (15) becomes, Eq. (17).

$$L(c, \gamma, \alpha) = \sum_{j=1}^m \alpha_j k(x_j, x_j) - \sum_{i,j=1}^m \alpha_i \alpha_j k(x_i, x_j) \quad (17)$$

The optimization problem is solved based on a dual form of α , as given in Eq. (18).

Maximizing,

$$W(\alpha) = \sum_{i=1}^m \alpha_i k(x_i, x_i) - \sum_{i,j=1}^m \alpha_i \alpha_j k(x_i, x_j) \quad (18)$$

Subject to $\sum_{i=1}^m \alpha_i = 1$ and $\alpha_i \geq 0, i = 1 \text{ to } m$.

Lagrange multiplier possibilities of non-zero if the inequality constraints are solution equality.

Optimal solutions complementarity conditions for $\alpha, (c, \gamma)$ are given in Eq. (19).

$$\alpha_i \{|\Phi(x_i) - c|^2 - r^2\}, i = 1, \dots, m \quad (19)$$

Training samples x_i lie on the surface of the optimal hypersphere related to $\alpha_i > 0$.

Eq. (20) provides the decision function solution.

$$f(x) = \text{sgn}(r^2 - |\Phi(x) - c|^2) \quad (20)$$

Eqs. (21) and (22) is provided.

$$f(x) = \text{sgn}(r^2 - \{\Phi(x) \cdot \Phi(x) - 2 \sum_{i=1}^m \alpha_i \Phi(x) \cdot \Phi(x_i) + \sum_{i,j=1}^m \alpha_i \alpha_j (\Phi(x_i) \cdot \Phi(x_j))\}) \quad (21)$$

$$f(x) = \text{sgn}(r^2 - \{k(x, x) - 2 \sum_{i=1}^m \alpha_i k(x, x_i) + \sum_{i,j=1}^m \alpha_i \alpha_j k(x_i, x_j)\}) \quad (22)$$

The method aims to obtain minimum enclosing hypersphere consists of satisfy all training samples.

4. Simulation setup

The implementation details like datasets, metrics and system requirements were discussed in this section.

Datasets: The UC Merced dataset and AID dataset were used in this method to analyze the classification performance. The UC Merced dataset consists of 2100 images with 21 land-use scene classes in the US region. The dataset images are in the size of 256×256 pixels with 0.3 m pixel spatial resolution in RGB color space. Each class consists of 100 images where some classes have high overlap. The AID dataset is collected from Google Earth



Figure. 2 The sample images of UC Merced dataset: (a) Aeroplane, (b) Baseball diamond field, and (c) Beach

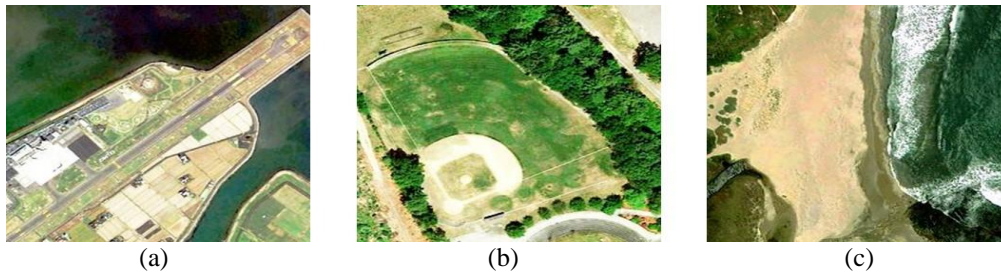


Figure. 3 The sample images of AID dataset: (a) Airport, (b) Baseball field, and (c) Beach

images which consist of 10,000 images. The images are in the size of 600×600 pixels and 30 classes are present in the dataset. The images are collected from various countries and increases the diversity of the images. The sample UC Merced dataset images of the Aeroplane, Baseball diamond field, and the beach is shown in Fig. 2. The sample AID images of the Airport, Baseball field, and the beach is shown in Fig. 3.

Metrics: The performance metrics like Accuracy, Sensitivity, Specificity, MCC, and F-Score are measured from the proposed method and the formula is shown in Eqs. (23) to (26), respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (23)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (24)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (25)$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \times 100 \quad (26)$$

Where TP represents the True Positive, TN represents the True Negative, FP represents the False Positive, and FN represents the False Negative.

System Requirement: The proposed hybrid feature selection with the MSVM method is tested in the system that consists of an Intel i7 processor with 16 GB of RAM and 6 GB of Graphics card. The MATLAB 2018a tool is used to implement the proposed method.

5. Results and discussion

In this research, the hybrid feature selection of PSO and GWO with parallel MSVM is proposed to improve the efficiency of Land use classification. The UC Merced and AID datasets used to test the performance of the proposed hybrid feature selection with parallel MSVM method. The histogram equalization method is applied to enhance the image quality and remove the artifacts of the images. The SURF, LTP, and DWT features were extracted from the input images and applied for feature selection. This section provides a detailed description of the results of the hybrid feature selection with parallel MSVM in land use classification.

The performance analysis of the hybrid feature selection with the parallel MSVM method is tested and shown in Table 1. The standard classifiers like Random Forest (RF), K-Nearest Neighbor (K-NN), linear SVM were compared with proposed hybrid feature selection with parallel MSVM method. The

Table 1. The hybrid feature selection with MSVM method Performance analysis on UC Merced dataset

	Accuracy	Sensitivity	Specificity	MCC	F-score
RF	92.39	93.52	34.15	73.73	75.70
KNN	78.00	57.06	26.42	72.53	35.55
SVM(Linear)	94.10	95.41	71.56	65.73	81.45
Proposed MSVM	99.15	96.59	94.39	91.18	93.76

Table 2. Feature Selection performance analysis in UC Merced dataset

	Accuracy	Sensitivity	Specificity	MCC	F-score
Without FS	94.83	95.03	90.58	86.45	93.40
PSO	95.60	92.75	91.59	86.87	91.96
GWO	96.39	91.89	91.52	88.11	89.58
PSO + GWO	99.15	96.59	94.39	91.18	93.76

Table 3. Performance analysis of proposed method on AID dataset

	Accuracy	Sensitivity	Specificity	MCC	F-score
RF	87.98	90.16	31.25	72.08	73.25
KNN	77.69	56.18	25.93	77.66	34.07
SVM(Linear)	93.41	92.12	71.17	65.35	77.33
Proposed MSVM	96.45	93.18	93.96	89.46	92.50

Table 4. Feature selection method performance analysis on AID dataset

	Accuracy	Sensitivity	Specificity	MCC	F-score
Without FS	90.62	91.82	89.33	83.18	89.59
PSO	93.38	90.58	88.54	81.90	88.05
GWO	94.31	92.72	90.34	83.43	88.02
PSO + GWO	96.45	93.18	93.96	89.46	92.50

hybrid feature selection with MSVM has higher performance than the existing method. The proposed hybrid feature selection method has the advantage of good convergence and searches efficiency. The proposed parallel MSVM method has the advantage of effective analysis of the non-linear relationship of the features and processes the data parallel to provide efficient classification. The existing RF method has the limitation of lower efficiency with more number of trees and overfitting problem in less number of trees. The SVM has a lower performance which imbalanced data classification and lower efficiency in the analysis of the non-linear relationship in the data. The proposed parallel MSVM method has an accuracy of 99.15 % and existing RF has 92.39 % accuracy, and SVM has 94.10 % accuracy. The proposed hybrid feature selection method of PSO and GWO method is compared with individual feature selection in the parallel MSVM method is shown in Table 2. The table shows that the hybrid feature selection method of PSO and GWO has higher efficiency due to its capacity to have good convergence and search efficiency. The PSO method has the limitation of lower convergence and GWO has lower search performance. The hybrid method applies GWO for convergence and PSO provides effective search performance. The hybrid feature selection method has an accuracy of 99.15 %, the existing PSO has 95.6 % and the existing GWO method has 96.39 % accuracy.

The proposed hybrid feature selection method of PSO and GWO with parallel MSVM method is compared with the existing method in the AID dataset, as shown in Table 3. The proposed hybrid method

with the MSVM method with parallel MSVM method has the higher performance due to the higher efficiency in the feature selection and parallel classification. Similarly, UC Merced dataset, the RF has a limitation of overfitting problem. SVM has a data imbalance problem and KNN has lower performance in feature relation analysis. The proposed hybrid method has an accuracy of 96.45 %, the existing RF has 87.98 % accuracy, KNN has 77.69 % accuracy and Linear SVM has 93.41 % accuracy.

The proposed hybrid feature selection method of PSO and GWO with parallel MSVM classifier is compared with the individual feature selection method, as shown in Table 4. The table shows that hybrid feature selection has a higher performance than the individual methods. The PSO method has limitations of lower convergence and GWO has lower efficiency in the search analysis. The hybrid feature selection method has an accuracy of 96.45 %, the PSO method has 93.38 % accuracy and the GWO method has 94.31 % accuracy.

The proposed hybrid feature selection method with parallel MSVM on two datasets such as UC Merced and AID datasets, as shown in Fig. 4. The proposed hybrid feature selection with the MSVM method has higher performance in both datasets due to its efficiency in feature selection and parallel classification. The hybrid method has the advantages of good convergence and improved search performance. The parallel MSVM method can analyze the non-linear relationship among the features.

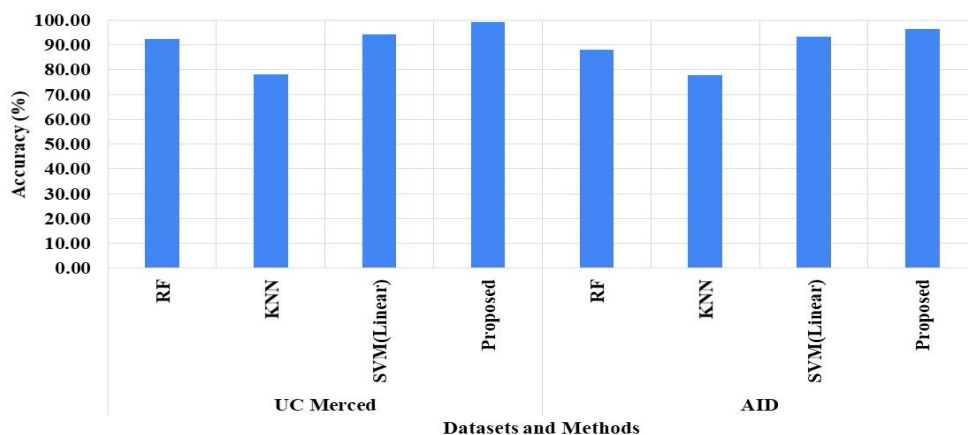


Figure. 4 The proposed hybrid method with Parallel MSVM performance on two datasets

Table 5. Hybrid feature selection with MSVM method Comparative analysis

Method	Dataset	Accuracy (%)
ARCNet [14]	UC Merced	99.12
Inception-v3-CapsNet [16]	UC Merced	99.05
Inception-v3-CapsNet [16]	AID	96.32
VGG-Net [18]	UC Merced	97.42
VGG-Net [18]	AID	91.86
Proposed	UC Merced	99.15
Proposed	AID	96.45

5.1 Comparative analysis

The proposed hybrid feature selection with the Parallel MSVM method is compared with existing methods in UC Merced and AID datasets to analyze the efficiency. The hybrid feature selection with the Parallel MSVM method is compared with existing deep learning methods.

The proposed hybrid feature selection with Parallel MSVM method is compared with existing deep learning methods of ARCNet [14], Inception-v3-CapsNet [16], and VGG-Net [18], as shown in Table 5. The proposed hybrid feature selection with Parallel MSVM method has the advantages of better convergence and high search efficiency to select the relevant features for the classification. The existing deep learning methods have the limitation of overfitting problem that degrade model efficiency. The proposed method has accuracy of 99.15 % in the UC Merced dataset and existing ARCNet [14] has 99.12 %, Inception-v3-CapsNet [16] has 99.05 % and VGG-Net [18] has 97.42 % accuracy.

The hybrid feature selection with the Parallel MSVM method is compared with existing deep learning methods, as shown in Fig. 5. The proposed method has a higher accuracy which compares to the existing deep learning method due to the advantages of better convergence in the feature selection process. The existing deep learning methods have the limitation of overfitting problems in the classification.

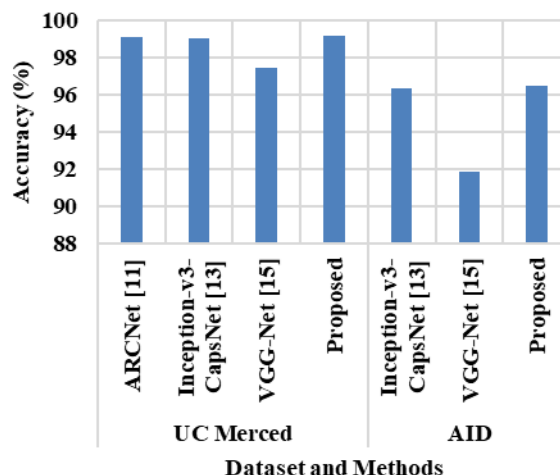


Figure. 5 The hybrid feature selection with Parallel MSVM method Comparative analysis on two datasets

6. Conclusion

Reliable land use classification is required in many applications like Urban planning, Natural Management, and Agriculture, etc. An existing method in the land use classification method has the limitation of overfitting problems due to improper feature selection. In this research, the hybrid feature selection method with the Parallel MSVM method is proposed to improve the performance of the land use classification. The UC Merced and AID datasets were used to test the hybrid feature selection with Parallel MSVM method performance. The input images were applied to Histogram Equalization to enhance the

image quality and remove the artifacts. The SURF, LTP and DWT methods were applied in pre-processed images to extract the features and applied to the hybrid feature selection method. The hybrid feature selection method of PSO and GWO selects the relevant features and applies them to the Parallel MSVM method. The hybrid feature selection method has the advantages of good convergence and higher search efficiency. The hybrid feature selection method has higher performance when compared to the individual PSO and GWO methods. This result shows that the proposed hybrid feature selection with the Parallel MSVM method has 99.15 % accuracy and the existing RF method 94.1 % accuracy. The future direction of the method involves applying the deep learning method to improve the performance of Land use classification.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper background work, conceptualization, methodology have been done by 3rd Author. Dataset collection, implementation, result analysis and comparison, preparing and editing draft, visualization have been done by first author. The supervision, review of work and project administration, have been done by second author.

References

- [1] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high-resolution remote sensing imagery using a fully convolutional network", *Remote Sensing*, Vol. 9, No. 5, pp. 498, 2017.
- [2] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data", *IEEE Geoscience and Remote Sensing Letters*, Vol. 14, No. 5, pp. 778-782, 2017.
- [3] B. R. Bharani and G. Murtugudde, "A Survey on Predictive Analytics and Parallel Algorithms for Knowledge Extraction from Data Received through Various Satellites", *International Research Journal of Engineering and Technology*, Vol. 6, No. 11, pp. 998-1002, 2019.
- [4] M. Rezaee, M. Mahdianpari, Y. Zhang, and B. Salehi, "Deep convolutional neural network for complex wetland classification using optical remote sensing imagery", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 11, No. 9, pp. 3030-3039, 2018.
- [5] W. Li, H. Liu, Y. Wang, Z. Li, Y. Jia, and G. Gui, "Deep learning-based classification methods for remote sensing images in urban built-up areas", *IEEE Access*, Vol. 7, pp. 36274-36284, 2019.
- [6] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, "3D convolutional neural networks for crop classification with multi-temporal remote sensing images", *Remote Sensing*, Vol. 10, No. 1, pp. 75, 2018.
- [7] G. Murtugudde, "Review on Statistical Relational Approach to Mining Bio Medical Databases", *AIRO International Publication*, Vol. 1, 2013.
- [8] G. Murtugudde, "Anomaly Detection and Local Outlier Factor for Credit Card Fraud Detection", *JETIR – Journal of Emerging Technologies and Innovative Research*, Vol. 7, No. 2, 2020.
- [9] R. P. D. Lima and K. Marfurt, "Convolutional neural network for remote-sensing scene classification: Transfer learning analysis", *Remote Sensing*, Vol. 12, No. 1, p. 86, 2020.
- [10] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 5, pp. 4340-54, 2020.
- [11] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification", *Neural Networks*, Vol. 95, pp. 19-28, 2017.
- [12] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu, "Deep fusion of remote sensing data for accurate classification", *IEEE Geoscience and Remote Sensing Letters*, Vol. 14, No. 8, pp. 1253-1257, 2017.
- [13] R. P. D. Lima and K. Marfurt, "Convolutional neural network for remote-sensing scene classification: Transfer learning analysis", *Remote Sensing*, Vol. 12, No. 1, p. 86, 2020.
- [14] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 2, pp. 1155-1167, 2018.
- [15] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery", *Remote Sensing*, Vol. 10, No. 7, p. 1119, 2018.
- [16] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-

- CapsNet”, *Remote Sensing*, Vol. 11, No. 5, p. 494, 2019.
- [17] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson, “An object-based convolutional neural network (OCNN) for urban land use classification”, *Remote Sensing of Environment*, Vol. 216, pp. 57-70, 2018.
- [18] S. Chaib, H. Liu, Y. Gu, and H. Yao, “Deep feature fusion for VHR remote sensing scene classification”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 8, pp. 4775-4784, 2017.
- [19] F. Nazir, M. M. Riaz, A. Ghafoor, and F. Arif, “Flood detection/monitoring using adjustable histogram equalization technique”, *The Scientific World Journal*, 2014.
- [20] R. Wang, Y. Shi, and W. Cao, “GA-SURF: A new speeded-up robust feature extraction algorithm for multispectral images based on geometric algebra”, *Pattern Recognition Letters*, Vol. 127, pp. 11-17, 2019.
- [21] Y. Liu, L. Guan, C. Hou, H. Han, Z. Liu, Y. Sun, and M. Zheng, “Wind power short-term prediction based on LSTM and discrete wavelet transform”, *Applied Sciences*, Vol. 9, No. 6, p. 1108, 2019.
- [22] H. Yapıcı and N. Çetinkaya, “An improved particle swarm optimization algorithm using eagle strategy for power loss minimization”, *Mathematical Problems in Engineering*, 2017.
- [23] M. Dehghani, H. R. Madvar, F. Hooshyaripor, A. Mosavi, S. Shamsirband, E. K. Zavadskas, and K. W. Chau, “Prediction of hydropower generation using grey wolf optimization adaptive neuro-fuzzy inference system”, *Energies*, Vol. 12, No. 2, p. 289, 2019.
- [24] I. S. Thaseen and C. A. Kumar, “Intrusion detection model using fusion of chi-square feature selection and multi class SVM”, *Journal of King Saud University-Computer and Information Sciences*, Vol. 29, No. 4, pp. 462-472, 2017.