INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Image Analysis and Classification Using HRSVM-CNN for Land-Cover Classification by Using Remote Sensing Images

G Vinuja[1], N Bharatha Devi[2]*, G Aloy Anuja Mary[3]

1 Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India
2 Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India
3 Professor, Department of Electronics and Communication Engineering, VelTech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India

OPEN ACCESS

Received: 15-09-2023
Accepted: 20-02-2024
Published: 14-03-2024

Citation: Vinuja G, Devi NB, Mary GAA (2024) Image Analysis and Classification Using HRSVM-CNN for Land-Cover Classification by Using Remote Sensing Images. Indian Journal of Science and Technology 17(12): 1129-1142. https://doi.org/10.17485/IJST/v17i12.2343

* **Corresponding author**.

bharathi.actech@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: To effectively analyze diverse satellite images and derive valuable insights, it's crucial to employ efficient methods for classification and image processing. However, due to imperfections in image formats and sensor data, satellite imagery often contains flaws and inaccuracies, requiring extra steps to enhance its quality. The proposed solution involves two key techniques: segmenting the input image using a Hybrid HRSVM-CNN and classifying the resulting high-resolution remote sensing image using a Convolutional Neural Network. This combined approach addresses the challenges posed by image inconsistencies and aims to improve the accuracy and efficiency of current methods for satellite image analysis. **Methods:** In this research, a high-resolution Support Vector Machine-Convolutional Neural Network (Hybrid HRSVM-CNN) and texture characteristics are used to create an automated land identification method for satellite Remote sensing (RS) images. This approach's main focus is segmentation using the Bendlet Transform and Improved Chan-Vese, and it also does classification using a Hybrid HRSVM-CNN based on feature extraction and gray-level co-occurrence matrix algorithm. **Findings:** The proposed classification method's accuracy was evaluated against several other classification algorithms, including Semi-Supervised Graph Based Method (SSG), Conditional Random Fields (CRF), k-Nearest Neighbor (KNN), and Bi-layer Graph-based Learning (BLGL), Support Vector Machine (SVM), and Artificial Neural Networks (ANN). When compared with existing methods, the findings of the proposed method display excellent accuracy of 98.83%. **Novelty:** To start, an adaptive median filter is used to pre-process the satellite Remote sensing images, removing unwanted noises and other impacts. Following pre-processing, the image is segmented using the Bendlet Transform

and Improved Chan-Vese algorithms. Gray-level co-occurrence matrix is utilized to extract texture information, and the Hybrid HRSVM-CNN is then used to categorize the various types of land. These applications frequently have a variety of issues that affect the categorization accuracy. A few significant factors, like location, irregularity, form, and diameter, reduce the process's overall accuracy. This research focuses on presenting a unique Land-Cover classification model in order to address such problems. The UC Merced Land Use dataset was considered in this research.

Keywords: Bendlet Transform and Improved ChanVese Segmentation; Hybrid HRSVMCNN; Land detection system; Satellite remote sensing images; Adaptive median filter; Gray level cooccurrence matrix

1 Introduction

Remote sensing is a technique that measures an object's or surface's qualities from a distance. Remote Sensing image sensors can gather electromagnetic spectrum bands such as visible, infrared, thermal, and microwave wavelengths. Each section of the spectrum has unique properties and provides unique information about the Earth's surface. It is used when the research region is large, and humans are unable to conduct the examination directly. Agriculture, forestry, geology, hydrology, atmospheric monitoring, and land cover / land use are examples of such fields. Remote sensing is mostly used to explore dam, bridge, and pipeline building sites in order to find construction materials and offer comprehensive geographic information. Images acquired by satellites and drones are used to examine the Earth's surface in remote sensing image analysis.

Remote sensing is the process of collecting data from the Earth's surface using passive or active means without making physical contact with the region being sensed. Active remote sensing occurs when a signal is transmitted by a satellite or aircraft and its reflection by an object is detected by a sensor. Passive remote sensing occurs when a sensor detects the reflection of sunlight, and the sensors collect radiation emitted or reflected by the item or nearby surroundings. Film photography, infrared, and charge-coupled devices are the most popular passive sensors. These sensors detect the amount of light reflected from the sun. Traditionally, satellite-based sensors record data in 4 to 6 different electromagnetic spectrum areas, including visible, infrared, and heat infrared wavelength bands, and are known as multispectral sensors. Hyperspectral imaging (imaging spectroscopy) has recently developed as a potent passive remote sensing method.

Machine learning (ML) is a branch of computer science that includes both supervised and unsupervised learning techniques[1–3]. It addresses both regression and classification issues[4]. In machine learning, a thorough dataset is created that includes all of the system parameters. ML is beneficial in situations where theoretical knowledge alone is insufficient to anticipate some facts[5,6]. It has a wide range of applications, including land use and cover problems[7], disaster management, and climate change[8]. Artificial intelligence (AI) is subdivided into machine learning (ML)[9]. It uses statistical approaches to automatically extract characteristics from data[10,11]. Initially, the categorization of remote sensing images was thought to be "shallow structures." There are several ways for doing remote sensing classification, such as decision trees, SVM, artificial neural network, bag of visual words, and many more[12–14].

For classification problems, there are numerous ML approaches accessible, such as K-means clustering and PCA, and for regression tasks, there are techniques such as SVM, decision trees, ANN, ensemble methods, random forest, and so on[15]. Existing CNN

approaches for remote sensing image classification can be used, but they require a lot of CPU power and a large, labelled dataset to work well. CNN model training takes a considerable time, but GPUs help us tackle this problem. Remote images obtained from satellite photos are extremely valuable, however there are certain challenges with image clarity when weather conditions are not clear, which affects the feature selection component of the ML process and hence performance. Because of advancements in deep learning techniques and parallel computing, these distant images may be readily identified by initializing weights in training layers such that scene prediction in later deep learning layers can be more accurate. There are several deep learning models, including AlexNet, GoogLeNet, VGG, and ResNet.

The rest of the paper is structured as follows. The prior research on image processing module to improve land-cover classification algorithms is discussed in Section II. In Section III, we investigate the proposed methodology of noise reduction algorithm-based adaptive median filter, discussion of Bendlet Transform and Improved Chan-Vese Segmentation, Hybrid HRSVM-CNN. Section V describes the simulation environment, the data that was gathered, and the outcomes. We wrap up our work with future directions in Section VI.

2 Materials and Methods

2.1 Input data set

The UC Merced Land Use dataset was considered in this research. As with the UC Merced Land Use dataset, there are 21 unique scene groups with 100 images per class and size of 256 x 256 pixels. Figure 1 depicts the proposed flow diagram.

Pre-processing

Pre-processing is the first stage of image processing in an image processing system. This stage is used to modify the input image by removing undesired elements such as noise reduction during the image capture process [4]. An adaptive median filter (AMF), which efficiently filters out noisy information while maintaining fine features in an image, was considered in this research. In order to enhance the edge portion of an image, we also applied a noise-protected edge detection approach. The system's detection probability performance is enhanced by supporting these steps. Typically, the filtering process is used to separate the items of relevance from the background and other objects [3].

2.2.1 Algorithm for the adaptive median filter

The technique below is used to eliminate noise from the corrupted image. Depending on whether the value of the window centre indicates noise, an adaptive median filter may alter the size of the filter window. The median value is used to replace the pixel if the noise point is the filter window centre pixel. If the centre pixel of the filter window has no noise, its current value is unaffected. Impulse noise with higher noise intensity can be handled with an adaptive median filter, where x_{min} is the filter window's minimal grey value, x_{med} is its median grey value $x_{i,j}$ is its minimum grey value (i, j), and S_{max} is its maximum value.

Step 1: Calculate $z_1 = x_{med} - x_{min}$ and $z_2 = x_{med} - x_{max}$

Step 2: If $z_1 > 0$ and $z_2 < 0$, go to step 3

Otherwise, we extent the window $S_{i,j}$

If $S_{i,j} < S_{max}$ repeat step 1 and step 2

if not, output $x_{i,j}$ directly.

Step 3: Calculate $k_1 = x_{i,j}\, x_{i,j} - x_{min}$ and $k_2 = x_{i,j} - x_{max}$

Step 4: If $k_1 > 0$ and $k_2 < 0$, output $x_{i,j}$ directly.

Otherwise, we regard x_{med} as the output value.

This algorithm's strength is its ability to smooth non-impulse noise while filtering impulse noise. Additionally, it may simultaneously maintain image details and reduce image distortion.

2.3 Segmentation

To simplify further processing or analysis of the image and to minimize the complexity of the image, image segmentation is a technique that divides a digital image into a number of subgroups known as Image segments [14].

2.3.1 Bendlet Transform and Improved Chan-Vese Model

In this research, propose an automated Chan-Vese (CV) model with enhanced Bendlet transform for segmenting land cover images. Bendlet transform is used to characterize the images, map them to the feature space, and ultimately retrieve the feature

set since the Bendlet system is based on the sparse approximation concept. The border of the land cover area is more precisely defined by the Bendlet transform, which can aid in exploring the mapping interaction between land cover lesions. In this paper used Bendlet to convert cross-sectional images to the frequency domain, and we looked at the transformation coefficients at each scale. The coefficient response was quite significant when the bending components and the image's curvature coincided, and Bendlet was able to extract contour information from medical images. As registration components, we added bending elements to Bendlet to characterize medical images at various sizes and orientations. The location with a high coefficient response was identified as the registration point when the bending element was compatible with the curvature of the image. Then, to perform image registration and correction, stable locations at various scales were chosen, and feature vectors were created. By comparing the brightness, structure, and contrast of normal and land cover images, we may determine how similar they are to one another and identify the area that contains abnormal one.

Here is a luminance comparison:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{1}$$

The contrast-based comparison is:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2}$$

Comparison of the structure can be obtained through Equation (3):

$$S(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3}$$

The three parts come together to form a single expression that is weighted with the exponents, α, β *and* γ.

$$SSIM(x,y) = [l(x,y)]^{\alpha} . [c(x,y)]^{\beta} . [S(x,y)]^{\gamma} \tag{4}$$

Where, μ_x and μ_y are the mean of pixels in the image blocks x and y, σ_x and σ_y are the standard deviations of the pixels in the image blocks x and y, σ_{xy} is the covariance between x and y, C_1, C_2, C_3 are constants. The formulae:

$$\sigma_{xy} = \frac{1}{H*W-1}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j) - \mu_x)(Y(i,j) - \mu_y) \tag{5}$$

$$\mu_x = \frac{1}{H*W}\sum_{i=1}^{H}\sum_{j=1}^{W}X(i,j) \tag{6}$$

$$\sigma_x = \left(\frac{1}{H*W-1}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j) - \mu_x)^2\right)^{\frac{1}{2}} \tag{7}$$

$\{x_1, x_2, x_3, \ldots x_n\}$ show that the normal image is divided into n image blocks. The values $\{y_1, y_2, yx_3, \ldots y_n\}$ show that the land cover image is divided into n blocks. We can calculate SSIM values of every two image blocks using Equation (3), $SSIM = \{s_1, s_2, s_3, \ldots s_n\}$.

The image block containing the land cover is the block with the least SSIM value. Segmenting within the unit region can increase segmentation accuracy and decrease interference from other organs. To divide the image into overlapping blocks, we first establish the overlap percentage and sliding window. The structural resemblance of the matching patches is then determined by taking into account three factors: brightness, structure, and contrast. The unit with land cover is the block with the least value. If the step size is chosen too small during the window sliding operation, accuracy of the resulting land cover unit is enhanced but time complexity rises. The inaccuracy increases and the intricate characteristics are lost when the step size is set too large. We increased the stride size to 50 in order to achieve a balance between computational effectiveness and the final performance.

The Chan-Vese (CV) model was proposed by Chan and Vese using the image's regional feature information. The specific energy function is:

$$F(\varnothing, c_1, c_2) = \mu_{\Omega} H(\varnothing(x,y)\,dxdy + \lambda_{1\Omega}|I(x,y) - c_1|^2 H(\varnothing(x,y)\,dxdy + \lambda_{2\Omega}|I(x,y) - c_2|^2 (1 - H(\varnothing(x,y))\,dxdy \tag{8}$$

Where, c_1 and c_2 are gray mean of the target area

$$c_1 = \frac{\Omega I(x,y) H(\varnothing(x,y)\, dxdy}{\Omega H(\varnothing(x,y)\, dxdy} \tag{9}$$

$$c_2 = \frac{\Omega I(x,y)\,[1 - H(\varnothing(x,y)]dxdy}{\Omega[1 - H(\varnothing(x,y)]dxdy} \tag{10}$$

The background area is, $H(\varnothing) = \begin{Bmatrix} 1, & \varnothing \geq 0 \\ 0, & \varnothing \leq 0 \end{Bmatrix}$

The average grey values c_1 and c_2 of the foreground and background are first determined according to the initialized level set, which is necessary for the CV model to segment the image. The evolution equation is then used to update each point on the level set. The value of the matching level set for this point increases if the current point's grey value is close to the foreground's grey average value; if not, it drops. The borders between land cover and background images are hazy because the grayscale difference between them is so minor. When using the CV model for land cover directly on cross-sectional images, it is challenging to obtain appropriate segmentation. We can segment the land cover through the improved CV model, so that the land cover can be accurately detected and segmented.

2.4 Feature Extraction

The importance of feature extraction is for improved performance. One of the most important elements in determining the effectiveness of a classification system is the feature extraction technique.

2.4.1 Gray Level Co-occurrence Matrix (GLCM)

- **Step 1:** Count the pixel pairs which has first pixel value as 'i' and the other pixel of the pixel pair displaced from the first pixel by 'd' and which has a value of 'j'.
- **Step 2**: The matrix Pd[i,j] receives the count value computed for the pixel pair, and they are assigns to i^{th} row and j^{th} column of the matrix.
- **Step 3**: Pd[i,j] isn't symmetric because the number of pixels pairs having gray levels [i,j] require not necessarily equal the number of pixel pairs having gray levels [j,i].
- **Step 4:** The elements of Pd [i,j] could be normalized. The normalization is made by dividing each entry by the total number of pixel pairs.
- **Step 5:** Normalized GLCM N [i,j], defined by,

$$P(i,j) = \frac{Pd(i,j)}{\sum_{i,=0}^{G-1} Pd(i,j)} \tag{11}$$

Let,

- p(i,j) be the $(i,j)^{th}$entry in the given gray level co-occurrence matrix.
- N is the number of distinct gray level in the image.
- P is the number of points in the image.
- i,j are the pixel values.

• **Energy:** It is the sum of squared GLCM components. This estimation is likewise called as Uniformity or Angular second moment. It quantifies pixel match duplication called as textural consistency or uniformity. A component which estimates the general likelihood of having unmistakable gray scale designs in the image, characterized as,

$$Energy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p^2(i,j) \tag{12}$$

• **Contrast:** Contrast is a measure of intensity or gray-level variations between the seed pixel and its neighbor. It quantifies the adjacent varieties in the gray level co- occurrence matrix of the image. It is characterized as

$$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 p(i,j) \tag{13}$$

• **Correlation:** Correlation feature demonstrates the linear dependency of gray level values in the co-occurrence matrix. It measures the joint probability occurrence f the predefined pixel combines in the GLCM network. It introduces how a reference pixel is identified with its neighbor, 0 is uncorrelated and 1 is perfectly correlated. characterized as,

$$Correlation = \sum_{i=0}^{N_{p-1}} \sum_{j=0}^{N_{p-1}} \frac{(i-\mu_i)(j-\mu_j)\,p(i.j)}{\sigma_i \sigma_j} \tag{14}$$

• **Homogeneity:** It is the proportion of closeness of GLCM appropriation components to the GLCM inclining components. This strategy is likewise called as Inverse Difference Moment. It has the most extreme esteem when all components in the picture are same. The range of the homogeneity ranges from 0 to 1. It is 1 for a diagonal GLCM. It is characterized as,

$$Homogeneity \;=\; \sum_{i=0}^{N_{p-1}} \sum_{j=0}^{N_{p-1}} \frac{p(i,j)}{1+(i-j|} \tag{15}$$

• **Dissimilarity:** It is a statistic proportion of the equality with which two gatherings are circulated crosswise over part geographic zones that make up a bigger zone.

$$Dissimilarity = \sum_{i-0}^{N-1} \sum_{j=0}^{N-1} \; p(i,j)\,|i-j| \tag{16}$$

• **Sum of Average:** It is the total of a rundown of numbers partitioned by the quantity of numbers in the rundown. In science and insights, this would be known as the arithmetic mean.

$$Sum\,of\,Average = \sum_{j=2}^{2N_2} \sum_{i=2}^{2N_2} ijp(i,j) \tag{17}$$

• **Entropy:** Entropy is a notoriously difficult term to understand, and the idea originates from thermodynamics. It shows the proportion of information of the image that is required for the image compression. It quantifies the loss of image data.

$$Entropy = -\sum_i \sum_j p(i,j)\,log\,(p(i,j)) \tag{18}$$

2.5 High-Resolution Support Vector Machine-Convolutional Neural Network (Hybrid HRSVM-CNN)

In this step, the developed features are sent into the Hybrid RSVM-CNN (HRSVM-CNN) classifier. It classifies the image into ten categories: airplane, beach, desert, commercial area, forest, lake, overpass, river, tennis court, and wetland. CNN is a well-known deep architecture in deep learning. CNN makes use of multi-layered visualizations. Because of its deep architecture, CNN may automatically achieve the representation trait as of the [raw] data via non-linear transformations and approximation non-linear functions. CNN has demonstrated its remarkable ability to demand information and useful features. However, the default classifier, known as the softmax classifier, lacks common generalization competence, but the RSVM performs rather well. As a result, conducting such study on those 'two' approaches are both required and meaningful. As a result, this research proposes an HRSVM-CNN classification. Figure 1 depicts the proposed HRSVM-CNN architecture.

The proposed hybridized model, HRSVM-CNN, which combines CNN and RSVM theories, is used for OBC in RS images. As a function extractor The CNN component is used, while the RSVM serves as the classifier. The first stage in this process is to structure hybrid architecture. The RSVM classifier replaces CNN's top-most layer when structuring. In this $O_j(j = 0,1,2,\ldots\ldots,10)$ denotes the dimension of the number jth layer. Presuming that the convolution filter's sizes and pooling size are (L_i, L_i) and $(G_i, G_i)\,(i = 1,2,3)$, each layer's dimension is assayed spotlighted on the subsequent functions.

$$O_{j1} = O_{j1-1} - L_{i+1},\; L_i = 2,\; 4,\; 6,8,10;\; i = 1,\; 2,\; 3\ldots\ldots..10. \tag{19}$$

$$O_{j2} = \frac{O_{j2-1}}{G_i};\; L_i = 1,3,\; 5,\; 7,9;\; i = 1,\; 2,\; 3\ldots.10. \tag{20}$$

$$O_{10} = n_k \times O_9 \times O_9 \tag{21}$$

where, n_k signifies the kernel number of the final pooling or convolutional layer, O_{10} indicates the selected output dimension. Algorithm flowchart of the HRSVMCNN is depicts in Figure 2.

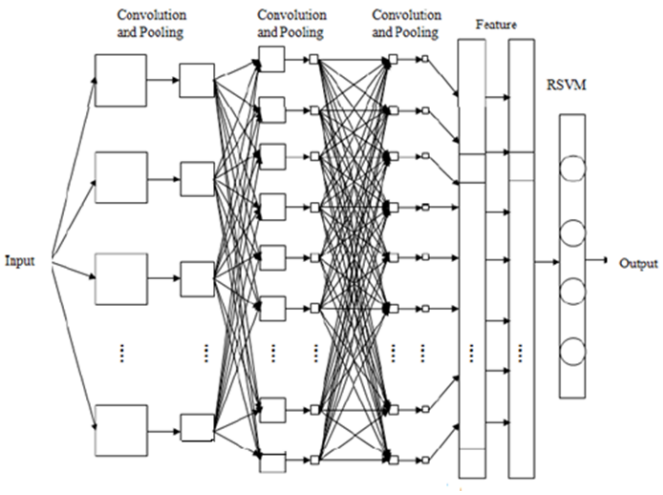

Fig 1. Architecture diagram of the proposed HRSVM-CNN

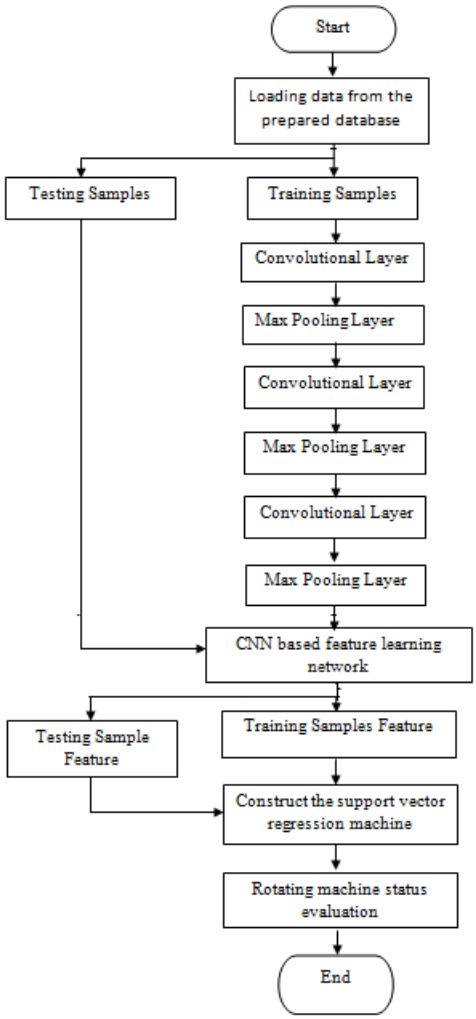

Fig 2. Algorithm flowchart of the HRSVM-CNN

The segmented image-related extracted features are sent as input to the classifier's preliminary layer. The progression of the feature maps is indicated by,

$$q_j^l = \eta\left(\sum_{i\in M_j} F_j^l . q_i^{l-1} + b_j^l\right) \tag{22}$$

Here, F_j^l denotes convolutional filters, η represents the activate function, b_j^l denotes the additive bias of the jth point of l^{th} layer, whereas q_i^{l-1} denotes the subsequent layer. The max-pooling rule is used for pooling after a convolutional layer. The rule states that the pooling layer selects the highest co-efficients from each sample cell to cut the dimensions in half. The feature map for this pooling layer was built using

$$q_i^l = \eta\left(\beta_i^l \, down\left(q_i^{l-1}\right) + b_i^l\right) \tag{23}$$

where, β_i^l indicates the multiplicative bias of ith point of lth layer. The same techniques are used to create further convolutional and pooling layers. The last pooling layer's output is imputed to the completely connected layer, which has an output of hundred dimensions. Furthermore, this output is used to extract characteristics from real signals. The layer-by-layer network's learning process might be repeated numerous times depending on the necessity. The feature map would then be imputed to the RSVM classifier because it has more important information. The cost function is defined in conjunction with the weight update regulation.

$$E^i = \frac{1}{2}\sum_{k=1}^{m}\left(Z_k^i - t_k^i\right)^2 \tag{24}$$

$$\frac{\partial E}{\partial F^L} = S^{l-1}\left(\delta^l\right)^T \tag{25}$$

Where, Z_k^i indicates the kth corresponding target label of sample i, δ^l signifies the gradients for the lth layer, and t_k^i specifies the similar values of the kth output-layer unit.

2.5.1 Detection Results

We divide the dataset into training samples (70%) and testing samples (30%) once it has been gathered, and then we employ four metrics to assess our performance. These metrics are made up of the confusion matrix's definitions of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). False alarms and misses are denoted by FP and FN, respectively. The following definitions apply to the four metrics Accuracy, Precision, Recall, and F1-score.

The number of times diverse samples or images are evaluated using the same algorithm and the machine or system provides correct findings is referred to as accuracy. It is the proportion of correctly classified cases to the total number of occurrences.

$$Accuracy = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)} \tag{26}$$

Where, true positive (TP) represents the number of correctly classified attack instances, true negative (TN) represents the number of correctly classified normal instances, false positive (FP) is the number of normal instances wrongly classified as attack instances, and false negative (FN) is the number of attack instances wrongly classified as normal instances.

Sensitivity (Recall or True positive rate) is defined as the number of correctly classified attack instances over the total number of attack instances.

$$Recall/sensitivity = \frac{T_P}{(T_P + F_N)} \tag{27}$$

Precision (Positive predictive value) is defined as the fraction of relevant samples among the selected samples.

$$Precision = \frac{t_p}{(t_p + f_n)} \tag{28}$$

The harmonic mean of accuracy and recall will be provided by F1 Score. F1 score is calculated as the weighted average of accuracy and recall. F1's greatest value is 1 and its lowest value is 0.

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{29}$$

The accuracy metric displays the proportion of accurate detections to all detections. Precision identifies the percentage of notifications that are genuine alarms. Recall displays the detected portion of land cover. A measure to balance FP and FN is the F1-score. A model will perform better if its F1 score is greater.

3 Implementation Result

The performance of the proposed method is based on a set of datasets. Figure 3 shows the dataset's randomly selected samples from each category. This dataset collection has 21 separate scene categories, each having 100 images with size of 256 X 256 pixels.

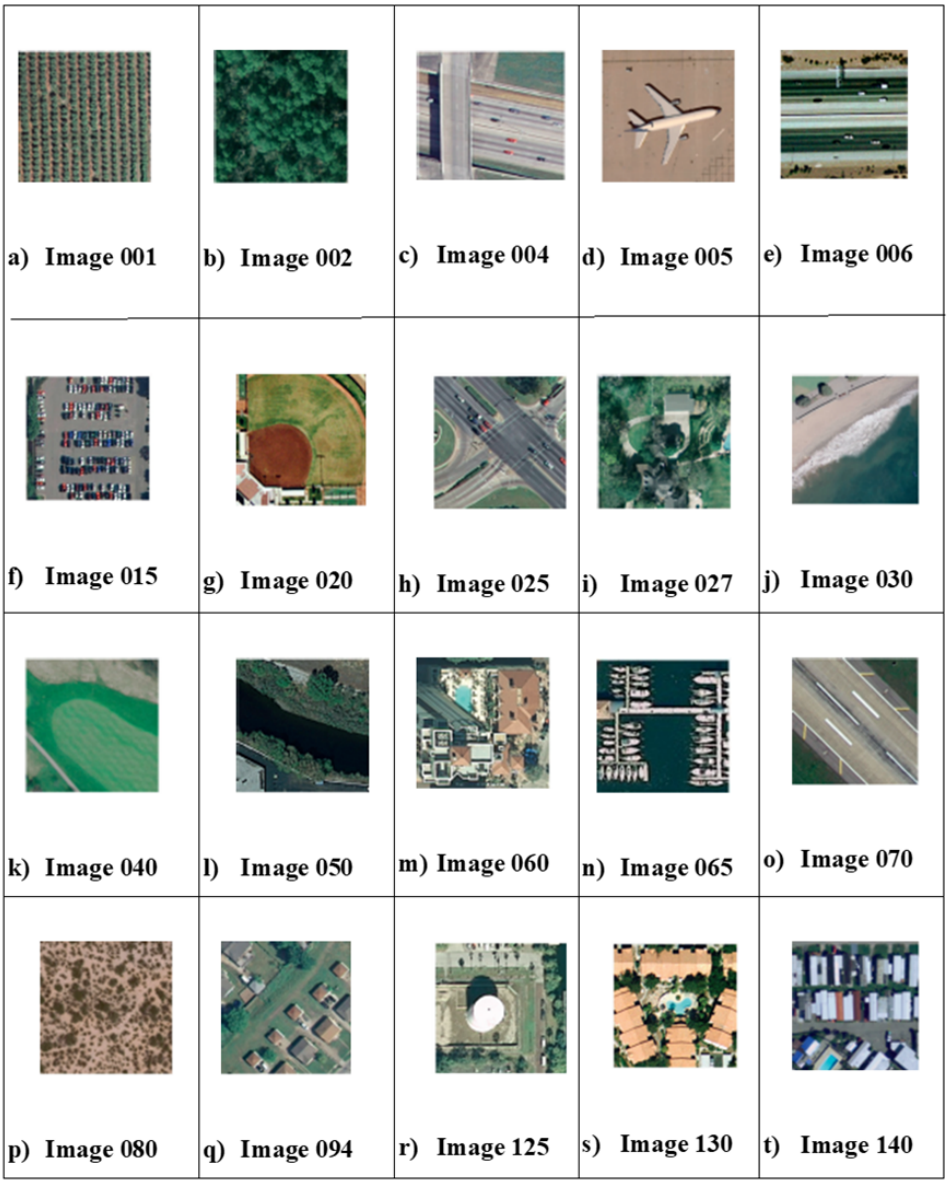

Fig 3. Sample image

The preprocessing includes noise elimination by adaptive median filter. This pre-processed image was shown in **figure 4**.

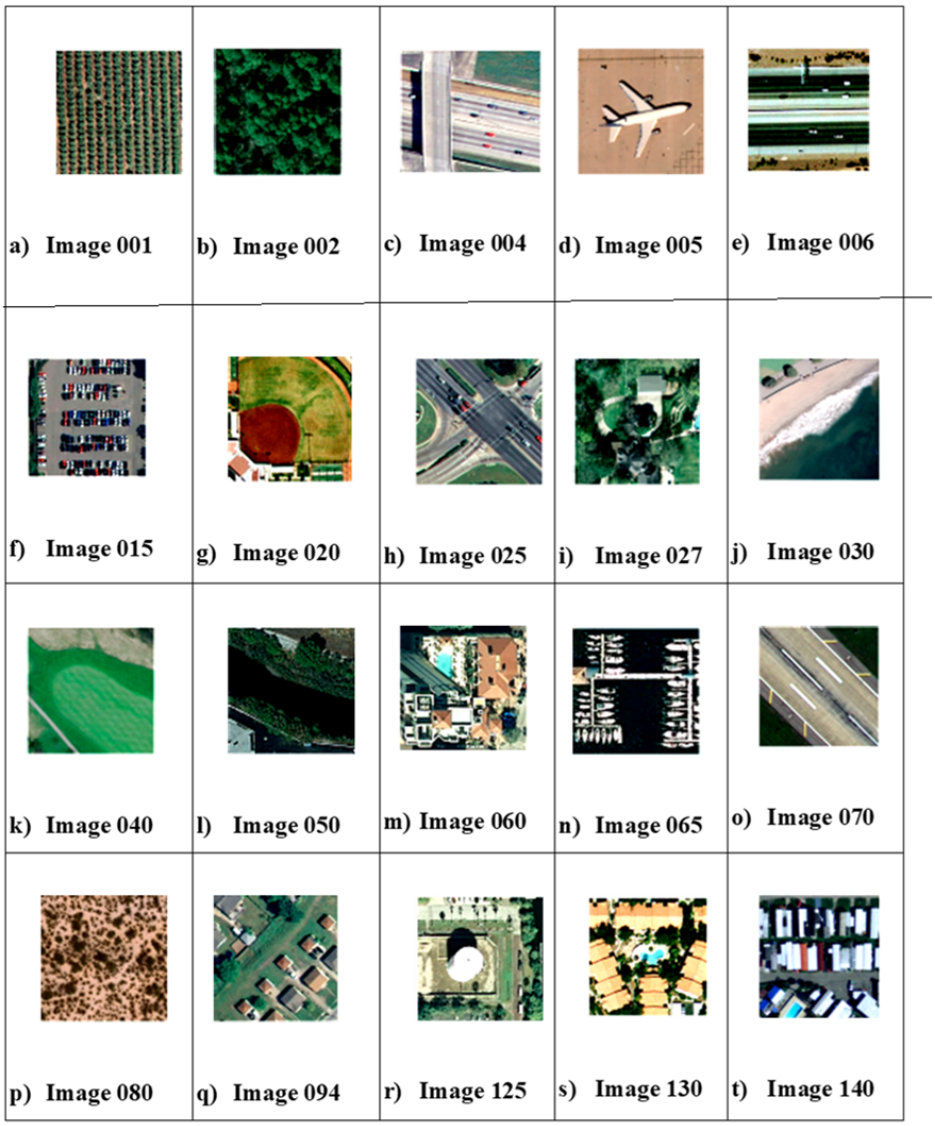

Fig 4. Pre-processed output

To segment the pictures and identify the land cover, the proposed Transform and Improved Chan-Vese Model was utilized. Figure 5 depicts the segmented output.

The sensitivity, specificity, accuracy, and similarity index values of the recommended GLCM approach were evaluated. As stated in Table 1, those parameters were used to create a performance matrix from the Confusion matrix.

Making a confusion matrix may assist in determining where your classification model is succeeding and where it is failing. Table 2 displays the classification result and confusion matrix of the Hybrid HRSVM-CNN Classifiers. Except for marshes, all Land Use classifications are identified with satisfactory results. In this, AI stands for Airplane, BE stands for Beach, CA stands for Commercial Area, DE stands for Desert, FO stands for Forest, LK stands for Lake, and OP is for Overpass. RI stands for River, TC is for Tennis Court, and WL stands for Wetland.

The detection results of Hybrid HRSVM-CNN classifier were shown in Table 3 and illustrated in Figure 6. From this result the overall classification accuracy of proposed Hybrid HRSVM-CNN is 98.83%.

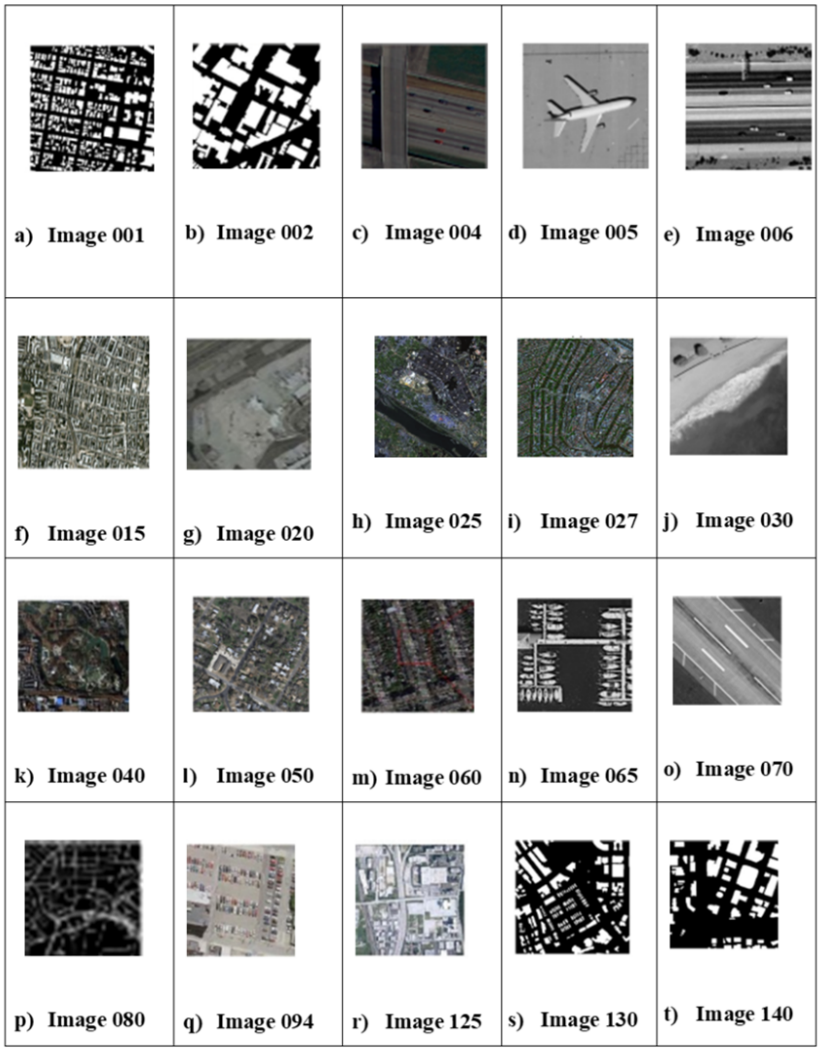

Fig 5. Segmented output

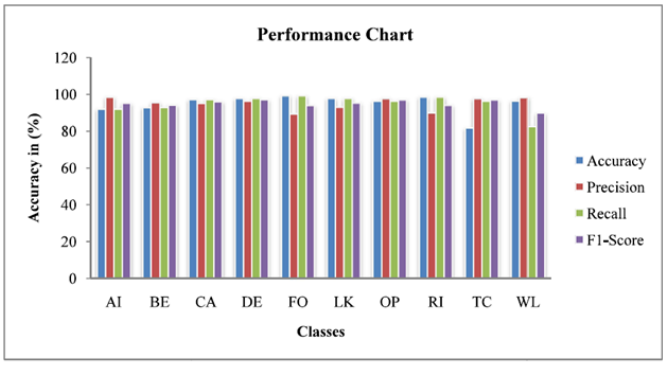

Fig 6. The performance chart of Hybrid HRSVM-CNN

Table 1. Image features

Image	Entropy	Contrast	Correlation	Homogeneity	Dissimilarity	Sum Of Average	Energy
001	1.5698	0.894	0.563	0.948	1.359	10.340	0.86
002	2.453	0.8674	0.675	0.8734	1.643	10.854	0.8502
004	2.5130	0.873	0.629	0.958	1.054	10.610	0.9643
005	3.5232	0.8432	0.6290	0.9159	0.9705	9.837	0.9100
006	2.8963	0.9643	0.5954	0.8793	1.0052	9.65089	0.9482
015	1.558	0.84	0.5357	0.988	1.369	10.420	0.886
020	2.464	0.864	0.653	0.8343	1.6343	10.9674	0.802
025	2.130	0.7354	0.6987	0.987	1.754	10.630	0.9323
027	3.5222	0.825	0.6296	0.992	0.972	9.520	0.913
030	2.8313	0.963	0.5549	0.678	1.742	9.6359	0.9253
040	1.987	0.94	0.5735	0.887	1.997	10.070	0.8766
050	2.338	0.674	0.5674	0.434	1.352	10.642	0.8852
060	2.105	0.73	0.769	0.898	1.475	10.663	0.6243
065	3.226	0.432	0.86290	0.699	0.965	9.337	0.9600
070	2.3673	0.643	0.7594	0.883	1.2032	9.6922	0.9622
080	1.98	0.4003	0.5433	0.984	1.975	10.487	0.8436
094	2.365	0.746	0.69775	0.877	1.365	10.655	0.8862
125	2.030	0.7387	0.6549	0.867	1.0783	10.10	0.9973
130	3.5532	0.8232	0.6650	0.9943	0.95645	9.372	0.9745
140	2.9327	0.9673	0.58464	0.83654	1.826	9.8964	0.9722

Table 2. Confusion matrix of Hybrid HRSVM-CNN

	AI	BE	CA	DE	FO	LK	OP	RI	TC	WL
AI	92.1	0.7	0.7	0.7	0.0	0.0	0.7	4.2	0.0	0.7
BE	0.7	92.8	0.7	0.0	0.7	4.2	0.0	0.0	0.0	0.7
CA	0.0	0.0	97.1	1.4	0.0	0.0	0.0	0.0	1.4	0.0
DE	0.0	0.0	1.4	97.8	0.0	0.0	0.0	0.0	0.7	0.0
FO	0.0	0.0	0.7	0.0	99.2	0.0	0.0	0.0	0.0	0.0
LK	0.0	0.7	0.7	0.7	0.0	97.8	0.0	0.0	0.0	0.0
OP	0.0	0.0	0.0	0.0	0.0	0.0	96.4	3.5	0.0	0.0
RI	0.0	0.0	0.7	0.7	0.0	0.0	0.0	98.5	0.0	0.0
TC	0.7	2.1	0.0	0.0	0.0	0.0	0.0	0.7	96.4	0.0

Table 3. The detection results of Hybrid HRSVM-CNN

Class	Accuracy	Precision	Recall	Fl-Score
AI	98.3	97.77	94.28	96.0
BE	98.4	95.52	91.42	93.43
CA	99.4	90.78	91.42	91.10
DE	99.4	93.75	96.42	95.07
FO	98.7	97.81	95.71	96.75
LK	99.7	95.71	95.71	95.71
OP	98.8	97.01	92.85	94.89
RI	98.8	86.16	97.85	91.63
TC	98.5	88.02	89.28	88.65
WL	99.3	97.77	94.28	96.0

The classification accuracy of the proposed classification method was compared with other classification algorithms like Support Vector Machine (SVM), Artificial Neural Networks (ANN), Enhanced Probabilistic Neural Network (EPNN), Convolutional Neural Network (CNN) and Deep Convolutional Neural Network (DCNN). The Classification comparison result was shown in Table 4 and illustrated in Figure 6.

Table 4. Classification comparison result

Number of Samples	Classification Accuracy					
	SVM	ANN	EPNN	CNN	DCNN	Proposed
300	95.9	84	84	84	84	84
400	95.9	87.5	87.8	88.6	88.9	90.2
500	95.9	90	90.3	90.8	91.2	91.7
600	95.9	92.7	92.9	93.4	93.7	93.9
700	95.9	93.5	93.7	93.8	94.1	94.6
800	95.9	93.6	93.8	94.2	94.5	94.9
900	95.9	94	94.1	94.4	94.7	95

From Table 4 and it can be observed that the Proposed Hybrid HRSVM-CNN model performed better than Classification algorithms for all classes such as airplane, beach, desert, commercial area, forest, lake, overpass, river, tennis court, and wetland.

4 Conclusion

Efficient classification and image processing techniques are required to extract the relevant information from diverse satellite images. Data received from the imaging sensor on satellite platforms has inadequacies and errors due to image format, thus further actions are required to improve image quality. The proposed work uses an Improved Chan-Vese Model and Bendlet Transform technique to segment the input image, and a Hybrid HRSVM-CNN to classify the high resolution remote sensing image. The proposed Hybrid HRSVM-CNN has an overall classification accuracy of 98.83%. The HRSVMCNN produces superior outcomes than the other approaches. This method might be extended to operate with real-time images in the future.

References

1) Hamida AB, Benoit A, Lambert P, Amar CB. 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*. 2018;56(8):4420–4434. Available from: https://doi.org/10.1109/TGRS.2018.2818945.
2) Vali A, Comai S, Matteucci M. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sensing*. 2020;12(15):1–31. Available from: https://doi.org/10.3390/rs12152495.
3) Zafar B, Ashraf R, Ali N, Ahmed M, Jabbar S, Naseer K, et al. Intelligent image classification-based on spatial weighted histograms of concentric circles. *Computer Science and Information Systems*. 2018;15(3):615–633. Available from: https://doi.org/10.2298/CSIS180105025Z.
4) Zafar B, Ashraf R, Ali N, Ahmed M, Jabbar SA, Chatzichristofis SA. Image classification by addition of spatial information based on histograms of orthogonal vectors. *PLOS ONE*. 2018;13(6):1–26. Available from: https://doi.org/10.1371/journal.pone.0198175.
5) Devi NB. Satellite image retrieval of random forest (rf-PNN) based probablistic neural network. *Earth Science Informatics*. 2022;15(2):941–949. Available from: https://doi.org/10.1007/s12145-021-00759-3.
6) Devi NB, Kavida AC, Murugan R. Feature Extraction and Object Detection Using Fast-Convolutional Neural Network for Remote Sensing Satellite Image. *Journal of the Indian Society of Remote Sensing*. 2022;50(6):961–973. Available from: https://doi.org/10.1007/s12524-022-01506-x.
7) Cabrera D, Cabrera L, Cabrera E. Perspectives Organize Information in Mind and Nature: Empirical Findings of Point-View Perspective (P) in Cognitive and Material Complexity. *Systems*. 2022;10(3):1–39. Available from: https://doi.org/10.3390/systems10030052.
8) Lary DJ, Alavi AH, Gandomi AH, Walker AL. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*. 2016;7(1):3–10. Available from: https://doi.org/10.1016/j.gsf.2015.07.003.
9) Wang D, Zhang C, Han M. MLFC-net: A multi-level feature combination attention model for remote sensing scene classification. *Computers & Geosciences*. 2022;160:105042. Available from: https://doi.org/10.1016/j.cageo.2022.105042.
10) Wang D, Zhang C, Han M. MLFC-net: A multi-level feature combination attention model for remote sensing scene classification. *Computers & Geosciences*. 2022;160(No. C). Available from: https://doi.org/10.1016/j.cageo.2022.105042.
11) De Luca G. A Survey of NISQ Era Hybrid Quantum-Classical Machine Learning Research. *Journal of Artificial Intelligence and Technology*. 2022;2(1):9–15. Available from: https://doi.org/10.37965/jait.2021.12002.
12) Song H, Yang W. GSCCTL: a general semi-supervised scene classification method for remote sensing images based on clustering and transfer learning. *International Journal of Remote Sensing*. 2022;43(15-16):5976–6000. Available from: https://doi.org/10.1080/01431161.2021.2019851.
13) Pijush KDP, Pal S, Mukhopadhyay M, Singh SP. Big data classification: techniques and tools. *Applications of Big Data in Healthcare*. 2021;p. 1–43. Available from: https://doi.org/10.1016/B978-0-12-820203-6.00002-3.
14) Khan MA, Ali M, Shah M, Mahmood T, Ahmad M, Jhanjhi NZ, et al. Machine learning-based detection and classification of walnut fungi diseases. *Intelligent Automation & Soft Computing*. 2021;30(3):771–785. Available from: https://doi.org/10.32604/iasc.2021.018039.

15) Awais M, Li W, Hussain S, Cheema MJM, Li W, Song R, et al. Comparative Evaluation of Land Surface Temperature Images from Unmanned Aerial Vehicle and Satellite Observation for Agricultural Areas Using In Situ Data. *Agriculture*. 2022;12(2):1–19. Available from: https://doi.org/10.3390/agriculture12020184.