The Pharma Innovation Journal 2022; SP-11(3): 517-521

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(3): 517-521

www.thepharmajournal.com
Received: 15-01-2022
Accepted: 19-02-2022

Govind Madariya
M.Sc. (Ag) Horticulture (Fruit Science), Department of Horticulture, College of Agriculture, JNKVV, Jabalpur, Madhya Pradesh, India

Dr. SK Pandey
Professor and Head, Department of Horticulture, College of Agriculture, JNKVV, Jabalpur, Madhya Pradesh, India

Dr. SK Sharma
Professor, Department of Soil and Water Engineering, College of Agricultural Engineering, JNKVV, Jabalpur, Madhya Pradesh, India

Corresponding Author
Govind Madariya
M.Sc. (Ag) Horticulture (Fruit Science), Department of Horticulture, College of Agriculture, JNKVV, Jabalpur, Madhya Pradesh, India

A comparative study on supervised and unsupervised techniques of land use and land cover classification

Govind Madariya, Dr. SK Pandey and Dr. SK Sharma

Abstract
This study reviewed a number of supervised and unsupervised techniques of Land Use and Land Cover classification. Land use and land cover is a basic concept of the Earth's system that is associated with the relationship between humans' activities and the environment. Frequently updated land use/land cover information is necessitated for many socioeconomic and environmental activities, such as urban and regional planning, natural resource conservation and management, etc (Homer *et al.*, 2007; Lu and Weng, 2007; Jensen, 2009). Earth observation technology such as remote sensing offers a useful source of data that covers the earth surface at a spatial and spectral resolution that allows easier land use and land cover classification than traditional approaches. LULC map is the most significant remote sensing products, which is achieved by a technique called image classification. Classification is a very important statistical method for obtaining information from satellite images. LULC maps are essential for monitoring of numerous earth observation activities such as land use planning, natural resources inventories, agricultural forecasting, deforestation, environmental disasters monitoring and estimation of soil erosion. As enough platforms are deployed in orbit and new applications arise in various fields, so many researchers in LULC have examined in various sections of subjects during the last two decades. The purpose of the study is to provide consistent information on all-encompassing algorithms of supervised and unsupervised technique for LULC mapping.

Keywords: LULC classification, classification techniques, supervised and unsupervised techniques, merits and demerits

Introduction
The word "Land use" refers to how humans interact with the environment, whether for development or conservation whereas the word 'Land cover' refers to how much forest, wetlands, agriculture, other land covers a region. Remote sensing provides data with a medium to high spectral, spatial, and temporal resolution, as well as constant and repeatable monitoring of the earth's surface. For mapping specific regions or areas of the earth's surface, remote sensing data together with Geographic Information Systems (GIS) proved to be effective. Regional Land Use and Land Cover (LULC) maps are still needed for a range of applications, including landscape planning, disaster monitoring, change detection, resource management, site suitability assessments, and ecological investigations (Jensen, 2005) [7]. Remotely sensed images offer quantitative and qualitative information, reducing the complexity and time required for field work, and may be utilised to create LULC maps using an image classification process (Chaichoke *et al.*, 2011) [3]. The process of obtaining valuable information from a large satellite imagery by classifying image pixel values into different categories or land cover classes is known as image classification.
In recent decades, remote sensing has become one of the most often utilised sources for LULC analysis. LULC information and its spatial distribution patterns are critical for a wide range of research themes, especially in urban studies with different classes, as well as maintenance and development plans (Stefanov *et al*. 2001) [29]. Modern classification techniques have recently gained the interest of researchers, due to advances in high resolution remotely sensed images. LULC monitoring is required for three primary purposes:

- To know the current availability of each land cover classes.
- To determine the changes that occurred in different land cover classes throughout the course of two or more time periods.
- To make it possible to take preventative measures against the depletion and destruction of natural resources.

Image Classification Process

There are seven major steps of image classification according to Lu and Weng (2007) [4].

1. Selection of remotely sensed data
2. Selection of a classification system and training samples
3. Data pre-processing
4. Feature extraction and selection
5. Selection of a suitable classification method
6. Post-classification processing
7. Evaluation of classification performance

LULC classification techniques and approaches

To generate LULC maps, a number of classification approaches have been developed and widely utilised. Many modern classification techniques for image classification, such as artificial neural networks, and expert systems fuzzy sets have been widely used in recent years. According to the methodology and technology employed, image classifications may be categorized into several groups such as supervised and unsupervised, hard and soft, parametric and nonparametric, per-pixel, and sub pixel. Unsupervised image classification, supervised image classification, and object-based image analysis are the three main methods for classifying remote sensing images. The unsupervised and supervised classification approaches are the two most used methodologies. On the other hand, classification based on objects, has risen in popularity as it works well with high-resolution images.

Unsupervised Classification

It is the process of automatically recognising similar pixels based on their spectral values within the satellite image. The image examiners do not need to know the names of any classes, nor do they need any prior information of the land cover types. In this classification method the clustering techniques are used to group the image pixels. The image computes clusters and classifies according to the system based on knowledge. After automatic clustering of pixels, the researcher needs to select a land-cover class manually for each cluster in order to get a well-classified satellite image. The researcher merges clusters into a single land cover class. When there are no training images available for experiments, unsupervised classification is used. K-means and Iterative Self-Organizing Data Analysis Technique (ISODATA) are the two common clustering algorithms that are used for unsupervised classification without supervision.

K-MEANS

It is one of the simplest and often used unsupervised machine learning algorithms. The K-means algorithm divides each pixel into groups based on a set of initial mean values. The repeated re-definition of groups continues until the means reach a point where they do not change any longer. In other words, the algorithm identifies k number of centroids, and then assigns each data point to the closest cluster while keeping the centroids as minimal as possible.

Hence, the K-means clustering algorithm supports the division of unidentified dataset into a specified number (k) of user-defined clusters with the purpose of minimizing the variability within the cluster. Based on the mean distance between classes, the algorithm has been used in several study to identify groups or classes in unlabelled datasets. The technique generates the classes that are later employed in other prospective analyses. Until convergence, this algorithm performs the three phases listed below.

- Determine the coordinates of the centroids
- Determine the distance between each object and the centroids.
- The object is divided into groups based on the minimum distance between them.

Merits

- The key benefit of this technique is that it is simple and fast, allowing it to operate on large datasets.

Demerits

- Each run does not provide the same outcome, because the final clusters are dependent on the initial random assignments.
- One of the key drawbacks of k-means is that the number of clusters must be specified as an input to the algorithm.

ISODATA

The term ISODATA stands for Iterative Self-Organizing Data Analysis. It is a combination of algorithmic procedures that have been incorporated into an iterative classification system. Many of the processes included in the algorithm are the result of experimentation and experience. It calculates class means that are evenly distributed throughout the data space, then clusters the remaining pixels using minimum distance approaches in an iterative manner. It is iterative because it runs over the remote sensing dataset several times until the desired results are reached. When all bands in a remote sensing image have similar data ranges, good results are obtained. It allows for the automated merging of identical clusters as well as the splitting of dissimilar clusters. The following algorithm is used to execute ISODATA:

- The cluster means of k arbitrary are established.
- By computing the distance between pixel and cluster, all pixels are replaced into the nearest clusters.
- All cluster centroids are recalculated, and the above step is repeated till the threshold convergence is reached.
- Clustering is considered complete only if the number of clusters is within a certain range and the distances between them meet a certain criterion.

Merits

- It plays an important role in identifying "true" clusters in data.
- It is not biased towards the image's top pixels.
- It is not necessary for image data to be distributed normally.
- Cluster signatures can be saved, and supervised spectral signatures can be simply merged and altered with them.

Demerits

- It takes time.
- As an input to the algorithm, it requires the maximum number of clusters, the convergence threshold, and the maximum number of iterations.

Supervised Classification

In supervised classification training samples are required from the image to be classified for each class. As the name implies, supervised classification requires human intervention who selects a continuous set of pixels of the area of interest from an image to collect the spectral signature value. This technique is time-consuming since it requires specialized knowledge. The collected training samples are saves as signature file which is further used for classification of image. The training sites determine the quality of a supervised classification. Supervised classify cation involves three basic steps:

- Collection of training samples
- Signature extraction
- Image classification

The most widely used supervised classification techniques are

Maximum likelihood
Support vector machine (SVM)
Minimum distance,
Artificial neural network,
Decision tree classifier.
Parallelepiped classifier

Maximum likelihood classification

The maximum likelihood (MXL) classification method is one of the most extensively used supervised classification technique in remote sensing in which training samples are taken from the image to be classified. In this method a pixel with the maximum likelihood is classified into the corresponding class. This classification technique estimates the means and variances of the classes using training data, which is then used to estimate the probability of pixels belonging to distinct classes.

Merits

- It is one of the most precise methods of image classification.
- It fixes the issue of unclassified pixels (subject to threshold values).
- It establishes a standardised method for separating pixels in overlap zones between classes.

Demerits

- The algorithm requires adequate pixels in each training region to describe a normal population.
- Classes that aren't assigned to training sets are more likely to be misclassified, which is especially problematic for mixes.
- It is dependent on the accuracy of the training data. Changes in a class's training set can have an impact on other classes.
- It’s not practical with data from image spectrometers.

Minimum distance

In multi-feature space, the minimum distance classification method is used to classify unknown image data into classes that minimise the distance between the image data and the class. The distance is considered as an index of similarity where minimum distance is equivalent to the maximum similarity. Minimum distance classifiers are simple in concept and execution, but they are not extensively employed in remote sensing applications. Minimum distance classification is not always correct in its most basic form; there is no mechanism for addressing differences in class variability, and some classes may overlap at their margins.

Merits

- It is a simple approach in which a class's training set is represented as a centre point based on information about the average of all pixels in the sample class.
- Another advantage of this method is that it runs quickly, time saving depending on the training set and no pixels are unclassified.

Demerits

- The main disadvantage of the minimal distance classifier technique is that it leads to pixel misclassification since it is prone to mistakes when all pixels are assessed, even if the shortest distance is long.
- The mean value for all classes will be calculated using spectral distance classification, and the result will show that all pixels were classified, but they were assigned to the unclassified pixel class with the lowest distance.
- A maximum likelihood classifier is a preferable option in most of cases.

Artificial neural network

The ANN algorithm is a supervised non-parametric method. It uses training data to estimate data attributes. The spectral data of the time of change is utilised to train the neural network. ANN is a type of artificial intelligence that simulates some of the functions of the human brain in order to assign meaningful labels to image pixels. Because ANN-based classification employs a nonparametric technique, it is simple to incorporate additional data into the classification process to increase classification accuracy.

Merits

- It's a nonparametric strategy that makes no assumptions about the data and depends on the number of inputs and network size to determine correctness.
- Capable of expressing functions like OR, AND, and NOT.
- It is a data-driven self-adaptive approach that effectively handles noisy inputs and has a high computation rate.

Demerits

- This technique requires training which is expensive and time-consuming
- Because ANN involves network design, it might be difficult to determine which network is best for an over approach.
- Over fitting is a problem.

Support Vector Machine (SVM)

SVMs are based on a concept of statistical learning theory and are one of the more recent advancements in the field of machine learning (Vapnik, 1979) [31]. It's non-parametric, and no data distribution assumptions are made. It can handle small training data sets and frequently outperforms traditional approaches in terms of classification accuracy (Mantero *et al*, 2005) [12]. It can theoretically handle large data sets with higher dimensionality, and it is especially useful for hyperspectral image classification (Melgani and Bruzzone, 2004) [13].

Merits

- It gets flexibility in the kind of threshold it may use.
- It has a nonlinear transformation.
- It provides a high generalisation capability
- The issue of over fitting is no longer an issue.

- Computational complexity is reduced.
- Easy to control the complexity of decision rules and the frequency of errors.

Demerits
- Training is required, which takes time.
- The final result has very little transparency.
- The algorithm's structure is difficult to understand.
- If there is nonlinearly separable training data, determining optimal parameters is difficult.

Decision tree classifier

The decision tree classification techniques are non-parametric with no assumptions on distribution of data or independency. These methods establish a flowchart-like tree (hierarchical) structure, where each node shows a test on a set of attribute values, each branch represents a test result, and tree leaves indicate classes or distribution of classes (Han *et al*, 2011; Larose, 2005) [19]. The classification rules at the decision tree's node are based on attribute value analysis. Change vs. no-change can be regarded as a binary classification problem, or a post-classification analysis can be used to quantify differences.

Merits
- Nonparametric training data can be handled.
- It does not need substantial design or training.
- Provides a set of rules that are easy to interpret and provides hierarchical correlations between input variables to determine class membership.
- It's simple, and efficient in terms of calculation.

Demerits
- The use of hyperplane decision boundaries parallel to feature axes may limit their application in situations when classes are clearly identifiable.
- The computation becomes difficult when several values are undetermined or when numerous outcomes are correlated.

Parallelepiped Classifier

The parallelepiped classifier is one of the most extensively used supervised classification techniques for multispectral images. It is very simple supervised classification algorithm that is formed by evaluating histograms of the distinct spectral components of the training data. The classifier utilizes the class limits and specified in each class signature to detect whether a given pixel lies within the class or not. In feature space, the class limits define the dimensions (in standard deviation units) of each side of a parallelepiped enclosing the class mean.

Merits
- It's an easy and computationally inexpensive technique.
- It comprises class variance and does not assume a statistical distribution for the classes.

Demerits
- It is the least precise approach.
- It struggles to adapt to large (high-covariance) clusters.
- It frequently results in overlapping classes, requires a second classification process.
- It also becomes more difficult as the number of channels increases.
- Pixels that do not fall within the designated parallelepiped are left unclassified.

Advantages and disadvantages of supervised and unsupervised techniques

Strategy	Advantages	Disadvantages
Supervised technique	▪ The classification is within the control of the analyst. ▪ Errors may be identified and corrected in many cases.	▪ The Special or unique categories that is not represented in the training data are unable to recognise. ▪ Unsuitable for dealing with massive data because each area needs the expertise knowledge ▪ Extremely time consuming; identifying pre-labelled samples takes a long time.
Unsupervised technique	▪ Scientists spend less time in classifying the area. ▪ This method is ideal for classifying large datasets. ▪ There is no need of prior knowledge of the study area. ▪ Human mistake is reduced to a minimum. ▪ Unique or special classes are identified as distinct units.	▪ This technique does not require any form of training; thus, it needs extensive knowledge of the region or the technique that is suitable for the targeted area. ▪ Large data sets take a long time to process, resulting in a poor classifier. ▪ Analyst has a limited amount of control over classes.

Conclusion

This paper attempts to examine and offer a basic understanding of the various image classification techniques and methods. The most common image classification techniques may be categorized as supervised and unsupervised, object-oriented, parametric and non-parametric, subpixel, per-pixel, and hard and soft classification. Among which supervised and Unsupervised Classification techniques are the two most used classification techniques. Image classification in remote sensing is continuously evolving, and a number of techniques are available in various software applications. Classification techniques are either selected subjectively for their ability to classify specific features or proposed by researchers to explore their capabilities. Many land use and land cover classifications have been found in research papers and on the Internet. These studies need a level of understanding of the classification target in order to ensure that the technique is carried out correctly. This review covers theoretical information about several supervised and unsupervised classification methods, as well as the merits and demerits of each method.

References

1. Abd Mukti SN, Tahar KN. Current Status of Classification Method for Mapping Application: A Review. International Journal of Emerging Technologies. 2021;12(1):122-128.
2. Andrade R, Alves A, Bento C. Exploring different combinations of data and methods for urban land use

analysis: A survey, CEUR Workshop Proceedings. 2019;2492:55-65.
3. Chaichoke V, Supawee P, Tanasak V, KS Andrew. A Normalized Difference Vegetation Index (NDVI) time-series of idle agriculture lands: A preliminary study. Engineering Journal. 2011;15(1):9-16.
4. D Lu, Q. Weng. A survey of image classification methods and techniques for improving classification performance, International Journal of Remote Sensing. 2007;28:5:823-870.
5. Han J, Pei J, Kamber M. Data mining: concepts and techniques, 2011.
6. Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, *et al*. Completion of the National Land Cover Database for the Conterminous United States. Photogrammetric Engineering and Remote Sensing. 2007;73: 337-341.
7. Jensen J. Introductory Digital Image Processing, Third Edition, Prentice Hall, 2005, 526.
8. Jipsa Kurian, V Karunakaran. "A survey on image classification methods. IJARECE. 2012;1:4
9. Jensen JR. Remote Sensing of the Environment: An Earth Resource Perspective 2/e. Pearson Education India, 2009.
10. Larose DT. Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley & Sons Inc., Hoboken, New Jersey, 2003.
11. Maneela Jain, Pushpendra Singh Tomar. Review of Image Classification Methods and Techniques, International Journal of Engineering Research & Technology (IJERT). 2013;2:8.
12. Mantero P, Moser G, Serpico SB. Partially supervised classification of remote sensing images through SVM-based probability density estimation. IEEE Transactions on Geoscience and Remote Sensing. 2005;43:559-570.
13. Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing. 2004;42:1778-1790.
14. Pooja Kamavisdar, Sonam Saluja, Sonu Agrawal. A survey on image classification approaches and techniques. SSCST, Bhilai, India, IJARCCE, 2013, 2.
15. Prem Chandra Pandey, Nikos Koutsias, George Petropoulos P, Prashant Srivastava K, Eyal Ben Dor. Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers-a review of the state of the art, Geocarto International. 2021;36(9);957-988.
16. Salah Mahmoud. A survey of modern classification techniques in remote sensing for improved image classification. Journal of Geomatics. 2017;11:21.
17. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier, 2011.
18. Hartigan JA. Clustering Algorithms. John Wiley & Sons, Inc., New York, 1975.
19. Han J, Pei J, Kamber M. Data mining: Concepts and techniques. Elsevier, 2011.
20. Hartigan JA. Clustering Algorithms. John Wiley & Sons, Inc., New York, 1975.
21. Han J, Pei J, Kamber M. Data mining: Concepts and techniques. Elsevier, 2011.
22. Hartigan JA. Clustering Algorithms. John Wiley & Sons, Inc., New York, 1975.
23. Han J, Pei J, Kamber M. Data mining: Concepts and techniques. Elsevier, 2011.
24. Hartigan JA. Clustering Algorithms. John Wiley & Sons, Inc., New York, 1975.
25. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier, 2011.
26. Hartigan JA. Clustering Algorithms. John Wiley & Sons, Inc., New York, 1975.
27. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier, 2011.
28. Hartigan JA. Clustering Algorithms. John Wiley & Sons, Inc., New York, 1975.
29. Stefanov WL, Ramsey MS, Christensen PR. Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. Remote Sens Environ. 2001;77(2):173-185.
30. Sunayana G, Domadia, Tanish, Hemalbhai, Zaveri. Comparative Analysis of Unsupervised and Supervised Image Classification Techniques, 2011.
31. Vapnik V. Estimation of dependences based on empirical data [in Russian]. Nauka, Moscow. (English translation: Springer Verlag, New York, 1979-1982.