



Land use land cover classification of remote sensing images based on the deep learning approaches: a statistical analysis and review

Monia Digra¹ · Renu Dhir¹ · Nonita Sharma²

Received: 16 February 2022 / Accepted: 28 April 2022 / Published online: 17 May 2022
© Saudi Society for Geosciences 2022

Abstract

Over the last few years, deep learning (DL) techniques have gained popularity and have become the new standard for data processing in remote sensing analysis. Deep learning architectures have drawn significant attention due to their improved performance in a variety of segmentation, classification, and other machine vision applications. In remote sensing, land use and land cover (LULC) are critical components of a wide variety of environmental applications. Changes in land use on a spatial and temporal scale occur due to accuracy, the capacity to develop, flexibility, uncertainty, structure, and the capability to integrate available models. Therefore, LULC modeling's high performance demands the employment of a wide variety of model types in remote sensing, which include dynamic, statistical, and DL models. In this study, we first analysed several key findings and research gaps in traditional technology while discussing various software applications used for LULC analysis. Second, the fundamental DL and ML concepts applicable to LULC are introduced with their merits and demerits. We employ a comprehensive review of distinct DL architectures and a custom framework to handle the challenging task of detecting changes in LULC. Subsequently, a detailed statistical analysis is conducted on the "Scopus database" to ascertain current trends in LULC utilising DL methods. This overview encompasses practically all applications and technologies in the field of LULC, from preprocessing to mapping. Finally, we conclude with a proposal for researchers to perform future potential using state-of-the-art methodologies.

Keywords Deep learning (DL) · Land use land cover (LULC) · Machine learning (ML) · Remote sensing · Statistical-analysis

Introduction

The growth of the earth's environment is driven by the growth of human beings. The level of resources and their consumption are increasing rapidly because of the human population. The Earth's surface has experienced a variety of

changes in the past 50 years due to human beings' exploitation. The changing of land areas from arable to built-up and the extension of urbanisation change the pattern of Land Use and Land Cover (LULC). Land Cover (LC) specifies the spatial variation information of the surface of planet Earth such as vegetation, soil, and water, whereas Land Use (LU) specifies the changes made by human activities or the physical changes on the earth's surface such as deforestation, urbanization, built-up areas, drought and floods etc. LULC change is an essential part of remote sensing by extracting valuable information, image processing, and classification of spectral signs of Land Cover.

The spatial-temporal analysis of physical surveys conducted in large-scale landscapes is the most difficult task to complete. To alternate the physical survey, modelling techniques were a replacement that could provide the framework by understanding the spatial pattern under various conditions. Although the physical model is more reliant on the prior knowledge of model parameters, this contributes to the model's poor accuracy. An enormous effort has been made over the last few decades to automate the LULC

Responsible Editor: Biswajeet Pradhan

Renu Dhir and Nonita Sharma contributed equally to this work.

✉ Monia Digra
monia13ite@gmail.com

Renu Dhir
dhirr@nitj.ac.in

Nonita Sharma
nonitasharma@igdtuw.ac.in

¹ Department of Computer Science and Engineering,
Dr. B. R. Ambedkar National Institute of Technology,
Jalandhar 144011, Punjab, India

² Department of Information Technology, IGDTUW, Delhi,
India

classification. Recent advances in remote sensing images allow for large data analysis, image classification, processing, and prediction for future changes. Many modelling techniques have been used, such as dynamic, statistical, and Neural Network (NN), which provide realistic simulations that include spatial–temporal, economic, and social aspects (Yuan et al 2020). Machine Learning (ML) modelling has the capability to solve the problems of classification, anomaly-detection, and prediction in remote sensing images. Traditionally, ML algorithms have been used to classify images using maximum likelihood classifiers, MarkovChain model, k-nearest neighbour, Artificial Neural Network (ANN), Support Vector Machine (SVM) (Aburas et al 2019) etc. With the growing number of earth data and the advancement of ML modelling techniques, a novel modelling technique has been presented that can handle enormous volumes of data and better predictive analysis of spatial–temporal aspects via deep learning (DL) (LeCun et al 2015). The DL model has outperformed traditional models in extracting spatial multilevel features extracted from remote sensing images and allows them to provide high performance in image processing and classification (image classification and object classification using Convolutional Neural Network (CNN)) (Zhang et al 2016). Our ultimate goal is to develop a methodical procedure that includes DL methods and produces a reliable result to detect LULC change. The motivation of our study was to conduct an exhaustive survey of DL applications in remote sensing images, including LULC analysis. Through an exhaustive review, we have analysed the research papers of DL approaches in LULC and summarised the main scientific advances in the related work.

Some key findings and research gaps

The main purpose of the review article was to determine the gaps in traditional approaches and analyse the new opportunities in LULC classification. Although in the past years, image classification in LULC using machine learning has made remarkable progress, there are still certain issues that need further study. 1) The ML community has used various algorithms to classify images in LULC classification, but as of now, the data is increasing tremendously and new technology with new datasets has introduced these complexities of classification that cannot be resolved by machine learning. A variety of socioeconomic data is readily available, providing vital material for urban growth. DL is competent to handle that data, which is integrated with socioeconomic data. 2) The feasibility and actual use of the methods in both the image classifications LU and LC have not been explored before, and the features of land use have yet to be resolved due to extremely high intraclass heterogeneity and inter-class similarity. 3) Most of the truth inference algorithms are domain-dependent, so there is a certain scope

for creating a domain-independent algorithm. 4) Achieving real-time or near-real-time LULC monitoring systems has become more complex due to changes in the components involved. 5) Forecasting urban land expansion is far more difficult than image analysis. By identifying the driving mechanisms of urban land cover change, significant factors such as the economy, transportation, population, and growth provide important insight into how human activities modify the urban environment. Therefore, setting the benchmark framework in ML model is a challenging task.

Software application for LULC analysis

In the precise study of LULC analysis, we have identified some software tools used in LULC analysis for pre-processing images, classification of images, analysing, and prediction analysis using spectral images. Table 1 contains a list of the software applications that have been identified and are as follows: Google Earth, Pro-ArcGIS, QGIS, ENVI, ERDAS IMAGINE, IDRISI, etc.

whereas the structure of the paper consist the “[Background](#)” section which follows the background of this work and discusses remote sensing applications in various areas using ML and DL in LULC and the most widely used Machine Learning model with merits and demerits of ML model in LULC. Section 3 highlights the “[How DL approaches outperformed ML approaches in LULC classification](#)” and the performance evaluation of traditional classification was improvised by using DL approaches, which outperformed classic ML approaches in terms of performance. furthermore, the deep learning architecture as well as models that can implement an image understanding task for LULC classification. The “[Deep Learning framework for LULC classification](#)” section introduces the framework for LULC classification using DL approaches and cutting-edge techniques in remote sensing applications. The “[Discussion](#)” section emphasises the statistical analysis of the LULC classification by providing a conclusion and future work outlook.

Scope of the study

In this article, we examine state-of-the-art approaches for LULC analysis with DL techniques. The main outline of the paper is as follows:

- The ultimate goal of this article is to provide a roadmap for future trends in LULC analysis using DL techniques.
- Discuss how DL approaches improve performance over the traditional ML approaches.
- A detailed, comprehensive review of existing DL approaches in remote sensing.

Table 1 Software applications used in remote sensing

Software	Features										
	Cloud Based	MSI and HSI data	3D image visualization	Clipping and masking	Anomaly detection	Pre-classification	Post-classification	3D map viewer	Symbology interface	Real-time analysis	Accuracy assessment
Google Earth Pro (https://www.google.com/earth/versions/)	✓	✓	✓	✓	✓			✓		✓	
Google Earth Engine (GEE) Code Editor. (https://code.earthengine.google.com/)	✓	✓	✓	✓		✓	✓			✓	✓
ArcGIS (https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview)	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
QGIS (https://qgis.org/en/site/)		✓	✓	✓	✓	✓	✓	✓	✓		✓
ENVI (https://www.l3harrisgeospatial.com/)	✓	✓		✓	✓	✓			✓	✓	✓
ERDAS IMAGINE (https://www.hexagongeospatial.com/)	✓		✓	✓	✓	✓		✓	✓	✓	✓
IDRISI (https://clarklabs.org/)		✓		✓	✓	✓					✓

- A generic framework of LULC change analysis using DL.
- Finally, we outline the statistical analysis of LULC and provide a conclusion with future research in LULC analysis.

According to the review analysis in Table 6, DL models achieved the best results in terms of classification or prediction.

Finally, some new perspectives on how the DL approaches can provide efficient work for LULC analysis and insights for future research are presented.

Background

Database for earth observation

The selection of image acquisition through a database is the most crucial step in LULC analysis. The comprehensive LULC data repository has been expanded to facilitate the implementation of many policies related to natural resources, food scarcity, deforestation, climate change, agriculture, etc. (Barker et al. 2020) (Xu et al 2018). A Big Earth Observation (EO) dataset is applied to provide the LULC change analysis, and time-series satellite images provide a better understanding of agricultural expansion, deforestation during the particular time period (Petitjean et al 2013). Various datasets used by researchers in LULC analysis are shown in Table 2.

LULC classification in remote sensing using ML and DL models

LULC classification are labelling the pixels in the remote sensing images for creating the classified images. LULC changes are divided into: a) preprocessing, b) change detection approach, and c) accuracy assessment. Atmospheric corrections, multi-temporal radiometric corrections, topographic corrections, geometrical rectification, and image registration are addressed at the preprocessing step. A correction is required to minimise the impact of these features (Song and Woodcock 2003). It is important to evaluate the changing reliance of temporal elements when collecting the remote data for LULC (Lunetta et al 2004). Selecting the appropriate change detection method is the essential step, although several pixel-based and object-based classification techniques that give a wider selection range have been employed (van der Meer 2011). The pixel-based approach provides the classification of a single pixel without considering the spatial context. It is based on the spectral reflectance of a particular LULC category. While comparing medium resolution imagery, it has some limited accuracy, which

Table 2 Database and its sources

S.No	Dataset	Equipment/Sensor	Spatial Resolution	Sources
1	Landsat 8 1	Operational Land Imager (OLI)	VNIR:30 m TIR:100 m	https://earthexplorer.usgs.gov/
2	Landsat 7	Enhanced Thematic Mapper (ETM+) Sensor	VNIR: 30 m TIR: 60 m	https://earthexplorer.usgs.gov/
3	Landsat 4 and 5	Thematic Mapper (TM)	VNIR: 30 m TIR: 120 m	https://earthexplorer.usgs.gov/
4	Sentinel 2	Sentinel 2A, sentinel 2B	VNIR:B2, 3,4,8: 10 m, B5,6,7,8A,11,12: 20 m, B1, 9, 10: 60 m	https://scihub.copernicus.eu/
5	Digital Elevation Model (DEM)	Space Shuttle Radar Topography Mission (SRTM) and ASTER(Advanced Spaceborne Thermal Emission)	SRTM: 30 m and ASTER: B1,2:15 m, B4,5,6,7,8,9:30 m, B10,11,12,13,14: 90 m	http://www.bhuvan.com/ and https://earthexplorer.usgs.gov/
6	LISS-III (Linear Imaging SelfScanning System—III)	Resourcesat1/ Resourcesat-2	23.5 m; 70 m	http://www.bhuvan.com/
7	MODIS (Moderate Resolution Imaging Spectroradiometer)	NASA Terra	Band 1–2: 250 m, Band 3–7: 500 m, Band 8–36: 1000 m	http://earthdata.nasa.gov/ and https://earthexplorer.usgs.gov/
8	REIS (Rapid Eye Earth Imaging System)	Rapid Eye Satellite Sensors (5 m)	6.5 m and 5 m	https://earth.esa.int
9	Quick Bird	BGIS-2000 (Ball Global Imaging System 2000)	2 to 8 m	http://www.digitalglobe.com/
10	IRS-P3	Wide Field Sensor (WIFS)/ Modular Opto-electronic Scanner (MOS)	188 m(WiFS); 1500 m/520 m/550 m(MOS)	https://www.isro.gov.in/Spacecraft/irs-p3

leads to the noisy output and maximum interclass variance (McRoberts 2014). However, numerous alternative ways have been proposed to overcome the pixel-based technique's drawbacks. Over the last decade, object-based image classification in LULC has been popular to provide identification through some physical classes (shape, spectra, and texture). It allows the extraction and segmentation of spatial features with the integration of vector and raster based processing. Image segmentation and extraction work on stacked multi-temporal images that include one or more than one spectral transform, multi-temporal images, multi-spectral waveband and texture. Statistical approaches have been used to identify the changes in LULC (Hussain et al 2013a). Accuracy assessment is a conclusive step to measure the remote sensing image classification in LULC.

The Kappa Index is the most commonly used technique for accuracy assessment to provide the correctness of the image classification. Whereas the overall accuracy assessment is used to validate the classification of images (Fan et al 2008). However, other statistical techniques are used to test or validate the performance of the model, such as the fuzzy similarities measure (FSM). Receiver operating characteristic (ROC) analysis is used to assess the simulation of change detection approach with prediction, and average spatial deviation distance (ASDD) is used to evaluate the

model's performance (Almeida et al 2008) (Pal and Ghosh 2017).

Until now, most of the studies have reviewed the articles in different areas, e.g., medical image recognition (Litjens et al 2017), prediction in autonomous vehicles (Migliani and Kumar 2019), speech recognition (Hinton et al 2012) etc. Although several review papers have been published using DL application in remote sensing for image classification, (Ma et al 2019) (Li et al 2018), data fusion, (Liu et al 2018a), atmospheric aerosol, (Di Noia 2018) etc., they ignored the other areas of remote sensing, i.e., LULC. Therefore, this study explored how DL application on the earth's surface changes the pattern of LULC. Due to the rapid growth in the number of related publications, it is required to conduct a comprehensive review and have a thorough understanding of the DL application in LULC. As shown in the Table 3 the discussion of various ML/DL models in remote sensing applications of LULC.

An overview of ML model merits and demerits

The archive of current remote sensing data is growing at an exponential rate in terms of quantity, and the planned satellite launches are expected to keep this trend going in the future (dlr.de, 2018). The remote sensing sector has

Table 3 Remote sensing applications of LULC using ML and DL model

Applications of remote sensing	Remote sensing in various computer visions	Dataset/source of dataset	Architecture/Model of ML and DL	Result / Performances	References
Object detection	Object detection	Pascal VOC 2007 and Pascal VOC 2012	G-CNN (Grid based CNN)	G-CNN is faster than CNN	(Najjibi et al 2016)
	Object detection (real-time)	VOC 2007, Picasso, People-art dataset	YOLO	Quantitative result (Picasso AP 53.3, F1 0.59, VOC 2007 AP 59.2, People-art AP 45)	(Redmon et al. 2016)
	Road network segmentation and vectorization	Massachusetts and ottawa road imagery	RoadVecNet (integrated with UNet)	RMS 1.47 and 0.63 m	(Abdollahi et al 2021a)
	Object detection and classification	Pascal VOC 2007 and MS COCO	DSOD (Deeply supervised object detector)	mAP is 73.6% with 5.9 M parameters	(Shen et al 2017)
Urban change detection	Urban area feature extraction	Aerial Images of Massachusetts	BCL-UNet and MCG-UNet	Both provide better performance than other DL models	(Abdollahi et al 2021c)
	Detection of building objects	AIRS	MultiRes-UNet	Average (F1 is 0.9698, MCC is 0.9573, IOU is 0.9413)	(Abdollahi and Pradhan 2021a)
	Extraction of roads	Aerial High Resolution Images	CNN	Precision is 92.75%, Recall is 96.08%, F1 is 94.39% and OA is 94.48%	(Abdollahi et al 2021b)
	Urban change detection and prediction	Land maps (SPOT 2,4 and 5)	CA-Markov chain model	Model predicted the urban area will increase 177 to 195.5 km sq and non-urban will decrease from 774.87 to 756.37 km sq	(Aburas et al 2021)
Vegetation mapping	LULC in urban area	Landsat images	Stochastic Markov chain model	OA is 87.5%	(Devi et al 2022)
	Urban vegetation mapping	AIRS	DNN model with SHAP technique	OA is achieved 94.4% and F1 is achieved 92.66%	(Abdollahi and Pradhan 2021b)
	Prediction of yield before harvesting	MODIS-NDVI	SNN	Accuracy of 95.64% and error prediction is 0.236 t/ha with correlation coefficient of 0.801	(Bose et al 2016)
	Detection of bug weed in the forest	Airbone Image Spectrometer for application	SVM-RFE	Accuracy rate is 97%	(Atkinson et al 2013)

Table 3 (continued)

Applications of remote sensing	Remote sensing in various computer visions	Dataset/source of dataset	Architecture/Model of ML and DL	Result / Performances	References
Climate change detection	Drought prediction	EM-DAT	CNN-BiLSTM model	SHAP values used to correlate with actual prediction valuable information	(Dikshit and Pradhan 2021)
	Soil erosion	SRTM DEM	GWR and ANN	AUC of GWR is 87.13, ANN is 89.57 and GWR + ANN is 91.67	(Chakraborty et al 2020)
	Landslide forecasting	Geological survey of India	SIGMA + tilt model	ROC curve max area under curve value is 0.976	(Abraham et al 2021)

quickly adopted machine learning for a variety of applications. Furthermore, there is a continuing attempt to build an automated system for mapping LULC. The majority of research so far has supported supervised learning techniques, with the notion that LULC is more likely to happen in situations similar to those that produced previous occurrences. Most of the ML algorithms are classified into two groups: supervised, unsupervised, or reinforcement learning. These classifications are based on data types and the requirements of the project, respectively. While working with labelled data, supervised learning methods are often performed in order to forecast the values whereas the values of a continuous set could be predicted using regression, while the category of a discrete set can be predicted using classification. A sample's value or classification can be predicted using the K-nearest neighbours (kNN) algorithm, which uses the sample's nearby neighbours in the feature space. In the case of regression, prediction results are calculated by taking an average of the k nearest neighbours' values. For classification applications, take the class with the highest number of appearances obtained (Altman 1992). For each class, the goal of parametric classification should be to characterise the usual subspace values or distribution associated with that class. Instead, SVM concentrates entirely on the training samples that seem to be closest to the ideal boundary between two classes in terms of their location in the feature space. In SVM, the goal is to determine the ideal border that maximises the distance, or margin, between the support vectors while minimising the support vectors numbers. SVMs were first developed for the purpose of determining a linear class limit (i.e. a hyperplane) (Cortes and Vapnik 1995) (Pal and Foody 2012). One of the most basic and simple classifiers is the Decision Tree (DT). A DT is a mechanism for recursively splitting the input data. The tree like structure is used to illustrate the general framework of tress splits into the branches while splits are represented by branches that show the paths between them, and leaves that represent the final objective values. Classification trees have leaf values that indicate categories of data, whereas regression trees have leaf values that represent one continuous variable after another. Segmentation can be done based on the frequency in a given band which exceeds or falls below a predetermined threshold (Pal and Mather 2003). The weakness of DT is that it decreases the accuracy of the classified training data while pruning the tree. To overcome the limitations of DT, a random classifier (RF) is used to give the final class to each unknown parameter (Belgiu and Drăguț 2016). While a single tree may not be the ideal solution, integrating many trees can result in a global optimal solution that overcomes the DT problem. The concept is further developed: each tree is trained by randomly selecting a subset of the training

data and employing the corresponding subset of variables. The conjunction of decreasing training data as well as a decreasing number of variables individually can provide the least accuracy of the tree. So the less correlated, the better, making the group more dependable as a whole. The relative relevance of each band may be calculated by comparing the evaluation of trees. In RF, tree pruning isn't required because of the existence of multiple trees. Regression techniques such as linear, polynomial, and so on are widely used in other areas, but when it comes to classification problems, the logistic regression (LR) and naive Bayesian (NB) classifier have been widely used for a longer time. The NB classifier is used to compute conditional probabilities based on previous probability and the probability is updated based on the ability to do the subsequent task. To normalise the anticipated values, the LR employs the sigmoid function, which calculates the likelihood of an event occurring and compares it to a pre-determined value (typically 0.5) can create the projected binary results (Ng and Jordan 2001).

Unsupervised learning methods are frequently used to identify the inherent properties and principles of unlabeled sample data. It has been used in the reduction of dimensionality, grouping, and detection of anomalies. Principal component analysis (PCA) is a technique for generating uncorrelated variables from correlated data. PCA seeks to uncover the most fundamental characteristics of a dataset or to construct a new feature which can represent the novel dataset, hence reducing the dataset dimensionality and increasing its generalization ability while keeping information loss to a minimum level (Jolliffe and Cadima 2016). The basic PCA technique might be used as a simple framework for developing a more operative feature extraction technique. The claim is made that PCA may not be applicable to HSI categorization (Cheriyadat and Bruce 2003; Uddin et al 2021). Due to the HSI's global variance, it may be unable to extract subtle information from some data distributions. KMeans clustering analysis is a popular technique used widely. It separate the dataset into K distinct, non-overlapping subgroups (clusters), each of the K clusters has a single data point in each of them. It aims to create the data points within a cluster and also to make it as distinct as feasible (Likas et al 2003). The non-linear clustering algorithm that have been used in spatial and non-spatial data are known as Self Organizing Map (SOM). An n-dimensional feature vector is assigned to a neuron in the output layer of this neural network, which has no hidden layers and n weight. The input feature factor is first measured with the similarity index to find out the most similar neurons, and then the nearby and activated neurons' weights are adjusted with the input vector which is identical. Each feature vector in the input set is subjected to this method. Lastly, it organises the neurons spatially in a one, two, or three-dimensional region where different units are

further apart whereas K-means use the nearest neighbour distance, whereas SOM employs the distances between all coupled neurons (Kohonen 2012). Table 4 summarize the merits and demerits of machine learning models in LULC.

How DL approaches outperformed ML approaches in LULC classification

Deep learning image classification

Pixel-based classification task involves the semantic segmentation of images which assign the classes to the individual pixels in an image (For example, road, grass, built-up area, etc.). The objective of pixel-based classification is to cluster the pixels of the image that correspond with specific perceptual items are included in that image, hence providing context for the pixels. According to the great degree of similarity in spectral across classes and the heterogeneity within classes, the pixel-based technique does not provide a desirable outcome. In traditional schemes, remote sensing images using pixel-based classification and consider the pure label pixel among the natural targets. On the other hand, object-based classification is a novel paradigm for segmenting remotely sensed images that outperforms pixel-based classification. The spectral information about object is aggregated, whereas textural and contextual information is gathered for classification of image using object-based (Hussain et al 2013b). In the remote sensing domain, new DL models have gained significance over older models. DL approaches outperform almost all other remote sensing techniques in a wide range of applications.

Deep neural network architecture in LULC

Deep neural network architectures like VGGNet, GoogleNet, AlexNet, ResNet, and DensNet have attained tremendous popularity in image classification and semantic segmentation. Using feature extraction in DL techniques, these architectures are very popular and often used for image classification in Table 5.

AlexNet

(Krizhevsky et al 2012) proposed the AlexNet, which is the first deep CNN architecture for image classification and recognition tasks. The learning capacity of AlexNet has been increased by performing different strategies of parameter optimization. For diverse categories of image dataset, the AlexNet depth has been increased from 5 to 8 layers, which improves the resolution of images. To improve the performance and solve the problem of gradient vanishing, a ReLU activation function has been employed. To increase

Table 4 Merits and demerits of ML in LULC

Types of Machine Learning	ML approach	Algorithms	Merits	Demerits
Supervised Machine Learning (SML)	Regression	Linear Regression, logistic Regression	Less complicated, outputs are always probabilistic in nature, Normalized to avoid overfitting	Real-time apps are not recommended since the patterns in real-time applications are quite complex, correlated errors, collinearity between two or more variables
	Classification	NB, DT and RF, SVM, kNN	To acquire the possible optimal classification performance in various land cover Capable of producing classified features, using an RF classifier; The feature selection may be easily modified to ensure classification accuracy, low-cost construction model and more flexibility They are more effective in generalization and less responsive to the imbalance of class	Interpretation of reference objects manually for distinct land cover, classification doesn't include the spatial information of surrounding pixels, Mixed-Pixel problem; where features from several classes combined together in a single pixel, input features that are irrelevant to the training data could result in erroneous outcomes, data preparation and preprocessing are perpetually challenging tasks, for high-resolution image; variations in texture features within an object area might result in variances in pixels inside that area which can lower the accuracy, for land cover classes the spatial and temporal resolution of satellite data is tremendous, the complexity of extraction and segmentation cannot handled by SML
Unsupervised Machine Learning (USML)	Dimension Reduction	PCA	Retain the most important information in 2-3 components, reduces the correlation to a minimum	It makes an assumption about a vast range of possibilities, during PCA transformation, it removed an excessive amount of spectral information, Maximum noise fragment maximize the SNR and reduces the image quality
	Clustering	K-means, SOM	Simple to implement, increases speed, the algorithm determines the boundaries of the back scattering intensity inter values to classify pixels, The k – means classification technique does not require any prior data processing, saving substantial time SOM is the best practice in visual datamining	Ground objects with identical spectra might readily impact the process; resulting in significant noise in the extraction results. It does not guarantee the uniformity and integrity of line characteristics SOM doesn't taken into consideration when cartographically representing of data is mapping

Table 5 Deep neural network architecture in LULC

References	Architecture	Year	Main Finding	Parameter	Input Size	Depth	Error Rate
(Krizhevsky et al 2012)	AlexNet	2012	Uses ReLu, dropout and overlap pooling	60	227×227×3	8	16.4
(Zeiler and Fergus 2014)	ZfNet	2014	Middle-layer visualisation concept	60	224×224×3	8	11.7
(Simonyan et al 2014)	VGGNet	2014	Reduce the filter size, increased the depth	138	224×224×3	16, 19	7.3
(Ioffe and Szegedy 2015)	GoogLeNet	2015	Increased the depth, introduce the block concept and various filter sizes	4	224×224×3	22	6.7
(He et al 2016)	ResNet	2016	Residual learning, symmetry mapping-based skip linkages	25.6	224×224×3	152	3.57
(Huang et al 2017)	DensNet	2017	Layers are connected with each other	25.6	224×224×3	201	3.46
(Woo et al 2018)	Convolutional Block Attention Module	2018	Uses both spatial and feature-map data	48.96	32×4d	–	5.69
(Arun et al 2019)	CapsuleNet	2018	Certain connections between characteristics	6.8	28×28×1	3	0.00855
(Wang et al 2020)	HRNetV2	2020	High-resolution representations	–	224×224×3	–	5.4

generalisation by avoiding over-fitting, overlapping sub sampling and local response normalisation were also used.

ZfNet

(Zeiler and Fergus 2014) proposed multi-layer de-convolution neural network, which is known as ZfNet. It was created to analyse the network performance statistically. ZfNet demonstrated that only a limited number of neurons are active, in the first layer some of the neurons are in dormant phase, while in the second layer, the filter size and stride are lowered to the optimum amount of features. It resulted in the improvement of CNN topology to enhance the performance.

VGGNet

(Simonyan et al 2014) suggested a simple and comprehensive design paradigm for CNN architectures that reduced the number of parameters and resulted in a 19-layer deep and 3×3 filter architecture with the added benefit of low computing complexity. It achieves superior outcomes when used to solve image classification and localisation challenges.

GoogleNet

(Ioffe and Szegedy 2015) proposed architecture, called Inception-VI, was designed with the primary purpose of providing great accuracy at a minimal computing cost. In GoogleNet, convolutional layers were replaced by small neural network layers in each layer. These small layers have different filters (1 X 1, 3 X 3, 5 X 5) to gather the spatial information, whereas it uses the sparse connection to avoid the problem of redundant information and remove the featured map if it is not important. However, rather than employing

a fully linked layer as the final layer, global average pooling was employed to decrease the connection density.

ResNet

(He et al 2016) developed the notion of residual learning in CNN, a highly effective technique for deep network training. The computational complexity of ResNet is lower than that of prior proposed networks. ResNet required less computational time and its depth is 20 and 8 times that of AlexNet and VGG, respectively. ResNet excels at image identification and localisation problems. To visualize the recognition task, spatial depth has been demonstrated in ResNet.

DenseNet

(Huang et al 2017) presented a solution to the problem of vanishing gradients. DenseNet overcame this issue by repurposing cross-layer connectivity. It connected each preceding layer to the subsequent layer in a feed-forward fashion; hence, as specified in Eqs. 1 and 2, the feature-maps of all preceding layers were used as inputs to all successive layers.

$$Fm_2^k = f_c(I_c, k_1) \tag{1}$$

$$Fm_l^k = f_k(I_k, \dots, Fm_{l-1}^k) \tag{2}$$

whereas, Fm_2^k and Fm_l^k are the resultant feature map for 1st and $l-1$ _{th} layer respectively, and f_k is a function that enables the cross-layer connection by concatenating the information from preceding layers before to assigning it to the new transformation layer l . Due to this reason, it gains the ability to explicit on distinguishing between information which is contributed to the network.

Convolutional block attention module

(Woo et al 2018) proposed a new type of CNN that is based on attention, termed the Convolutional Block Attention Module (CBAM). CBAM combines average and maximum pooling operations, resulting in a robust spatial attention map. The author has demonstrated that max-pooling may reveal information on object properties that differentiate them, whereas global average pooling can infer feature-map attention. These revised featuremaps improved a feature-capacity map's ability to be expressed. Due to the protocol's simplicity, it can be simply integrated into any CNN design.

CapsuleNet

(Arun et al 2019) the proposed technique involves a specific neuron called a capsule that has the ability to determine the face as well as other related information. Many specific capsules combine to create a capsule network called CapsuleNet which has three layers of capsule nodes at the each encoding part. Whereas, 28×28 images with 256 filters, and a size of image is 9×9 with stride 1. This input is given to the first layer of capsule to produce the vector image rather than a scalar image. Since then, CapsuleNet has performed the accumulation of the preceding layer's weighted features, which is significantly important in the detection and segmentation processes.

HRNetV2

(Wang et al 2020) proposed architecture, which represents the high resolution for vision tasks. HRNet has two main features. First, a parallel connection is made between the convolution series of high-to-low resolution. Second, information is transmitted frequently throughout resolutions. The benefit attained is a more exact representation in the geographical domain and an extraordinarily rich semantic domain.

DL approaches outperformed ML approaches in LULC classification

Table 6 highlights many examples of DL algorithms for simulating the LULC that outperformed in picture classification, object recognition, semantic analysis, and image segmentation. Allowing for multidimensional analysis in the LULC classification may be important to meet the expanding number and accessibility of remote sensing data. The current studies of remote sensing applications evaluate the effectiveness of DL approaches that employ

a variety of data sets with a high spatial resolution and a large number of parameters to achieve a higher degree of accuracy than ML models.

Deep learning framework for LULC classification

Deep learning for remote sensing is actively being studied and has a lot of potential. Between 2016 and 2021, significant improvements in DL performance were often observed in Fig. 7. This graph illustrates the growth of published journals of DL in remote sensing. As there is no requirement for human aid in modelling the future LULC analysis, the basic framework for LULC modelling using the DL model performs modelling automatically as shown in Fig. 1. When it comes to learning hierarchical characteristics, DL models offer a wide variety of advantages. LULC categories are primarily expansions or abstractions of the current terrain or landscape. Traditional ML models have been replaced by DL models because they outperform standard models in terms of performance, interpretability, data interpretation, and processing.

The design of the overall DL model divides the problem into different modules. 1) Data acquisition: Selecting an appropriate dataset is the most critical step in LULC analysis. The quality of data is necessary for generating a precise result when simulating the LULC. In general, the most relevant data for analyzing land-use change are physical, statistical, dynamical and spatiotemporal data. The HSI/MSI includes aerial images, satellite images, ancillary data, Google Maps, topographical maps, and maps for urban planners and land use. 2) Preprocessing the data-set: The preprocessing stage contains sub-tasks like feature engineering and classifier training, where the input data is prepared for denoise, eliminating irrelevant information from the data, synchronize, fusion of data, reducing its dimensionality, image re-sampling, clipping vector and raster images, buffering and geo-referencing. 3) Train model: After obtaining high-quality training data, it is possible to use this data to train a DL model using the feature extraction technique. 4) Validation and Evaluation: In order to ensure that the trained model is accurate, the model is evaluated and updated as needed. 5) Labelled sub-images and post-processing: Following the labelling of the sub-images, post-classification is a procedure that eliminates noise, corrects misclassifications, and improves overall accuracy. 6) LULC maps: Predicting LULC maps can assist urban planners and land resource management in taking appropriate action on the land cover.

In this section, we discuss the most commonly used networks like Convolutional Neural Networks (CNN), Fully Convolutional Network (FCN), and Autoencoder (AE) are the three major framework for LULC classification in remote sensing.

Table 6 DL approaches for simulating the remote sensing applications

References	DL Model	Dataset	Spatial Resolution	Parameters	Accuracy Metrics
(Li et al 2014)	DBN	Houston	2.5 m	14.7 K	Overall Accuracy
(Castelluccio et al 2015)	CNN	UC Merced Land Use	–	5 M	Overall Accuracy
(Hu et al 2015b)	CNN	Indian Pines	20 m	80.6 K	Overall Accuracy
(Zhou et al 2015)	SAE (Stacked Autoencoders)	UC Merced	1ft	51.6 K	Average Accuracy
(Ghamisi et al 2016)	CNN	Indian Pines and Pavia University	20 m	188 K	Overall Accuracy
(Zhao et al 2017)	SAE	Indian Pines	20 m	30.2 K	Overall Accuracy
(Sun et al 2017)	SAE	Indian Pines	20 m	107 K	Overall Accuracy
(Ding et al 2017)	CNN	Indian Pines	20 m	380 K	Overall Accuracy
(Paoletti et al 2018)	CNN	Indian Pines	20 m	96 M	Overall Accuracy
(Zhang et al 2018a)	CNN	ISPRS Postdam	5 cm	17 K	Overall Accuracy
(Liu et al 2018b)	CNN (Transfer Learning)	ISPRS (International Society for Photogrammetry and Remote Sensing) Postdam	5 cm	481 M	Overall Accuracy
(Marcos et al 2018)	CNN	Vaihingen and Zeebruges data	5 cm	430 K	Overall Accuracy
(Huang et al 2018)	CNN	other	1.24 m	39 M	Overall Accuracy
(Rezaee et al 2018)	CNN	other	5 m	53.9 M	Overall Accuracy
(Pashaei et al 2020)	DL (Deep Learning) Networks	UAS hyper-spatial imagery	–	–	Overall Accuracy + Training, Overall accuracy + validation
(Zhang et al 2019)	JDL (Joint DeepLearning) Model	Southampton and Manchester	50 cm	10 K	Overall Accuracy
(Alhassan et al 2020)	FCN (Fully Convolutional Network)	Landsat 5/7 Manitoba, Canada	1.32 m	2 M	Global Accuracy
(Papadomanolaki et al 2019)	FCN	(ISPRS WGII/4)	–	1.1 M	Overall Accuracy, Precision, Recall and F1 score
(Azarang et al 2019)	AE	QuickBird, Pleiades-1A, GeoEye-1	2.44–2.88 m, 0.5 m, 0.46 m	0.5 M	Average Performance
(Liu and Lee 2019)	3DADCNN (3-D atrous denoising convolution neural network)	HSI SPECIM Finland Image	1.34 m	–	Overall Performance
(Zhong et al 2019)	1D-Conv, LSTM and XGBoost	USGS	30 m	–	Overall Accuracy, F1-score
(Simonyan and Zisserman 2014)	CNN	ILSVRC	–	144 M	Validation, Error rate
(Liu et al 2021)	Multi-task Deep Learning	Pavia University	1.3 m	–	Overall Accuracy and Mapping Error
(Bittner et al 2018)	FCN	WorldView-2	58.37 m, 24.66 m (Height)	–	Overall Accuracy
(Hu et al 2021)	AE	Viareggio 2013	0.6 m	127 M	ROC
(Shi et al 2021)	CNN	Howland Dataset	1 m	15 M	Kappa Coefficient
(Mei et al 2019)	3D Convolutional Autoencoder	Indian Pine	17 m	–	Overall Accuracy
(Li et al 2021)	Deep CNN	Gaofen-2 satellite	3.2 m	–	Average Accuracy
(Sun et al 2020)	DL Model	AVIRIS images	1–4 m	–	Overall Accuracy
(Xu et al 2021)	CNN	ISPRS (Vaihingen and Potsdam) P	–	–	Overall Accuracy, Kappa Coefficient and F1-Score
(Yang et al 2019)	CNN	ISPRS (Vaihingen and Potsdam)	9 cm	–	Overall Accuracy

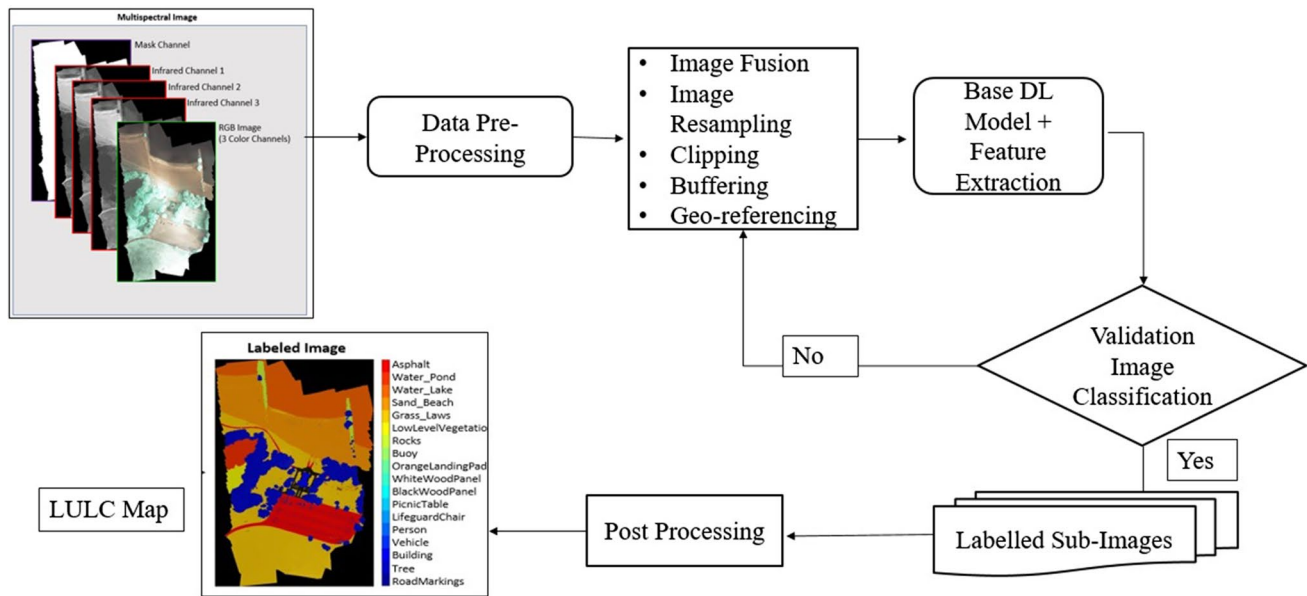


Fig. 1 An overall framework of DL model in LULC

Convolutional neural network

Among deep learning methodologies, the convolutional neural network (CNN) is the most effective and powerful framework network. CNNs have been frequently utilised to classify remote sensing data due to their ability to classify complicated contextual images. These techniques are usually not needed to complete an output image prediction. CNNs are feed-forward neural networks that employ spatially local correlation to make decisions by imposing a local connection pattern between neurons in neighbouring layers of the network. Their structure is comprised of a variety of convolutional layers, a maximum pooling layer, and fully linked layers (Zhang et al 2017). Each layer of convolution computes the weighted sum of the preceding feature, calculated using a filter, and then sends the result via an activation function to obtain the final result. When using this approach, the kernel size is calculated in order to find local correlations while maintaining invariance for each location inside the data array. The resultant feature map is generated with invariance down to the lowest feasible units. Finally, a fully-connected neural network is used to link all of the various phases of convolution or pooling layers together in a cohesive unit (LeCun et al 2015). The following is an example of a convolution operation:

$$f_n^k(x, y) = \sum_c \sum_{a,b} X_{ic(a,b)} e^k_{i(s,t)}$$

$$F_l^k = [f_n^k(1, 1), \dots, f_n^k(x, y), \dots, f_n^k(X, Y)]$$

Once the features are extracted, next is pooling or down-sampling operation used to extract the combination of features that are insensitive to translational shifts and minor distortions.

$$P_l^k = g_p(F_l^k)$$

Similarly, P_{kl} denotes the pooling feature-map of the l th layer for the K th input feature-map, and g_p denotes the pooling operation. In CNN (He et al 2015) pooling formulas include max, average, L2, overlapping, and spatial pyramid pooling. To increase the learning process and provide a decision function for a convolved feature-map are called as activation function. These activation functions speed-up the learning rate and also provide the non-linearity of features. Activation function like ReLu, sigmoid, tanh, maxout and SWISH has same functionality to provide non-linearity and overcome the problem of vanishing gradient.

$$t_l^k = g_a(F_l^k)$$

In the above equation, g_a denotes the activation function and F_{kl} denotes the convolution output, whereas t_{kl} denotes the transformed output. (Nwankpa et al 2018).

Training and optimization of CNN are the major design choices that provide the best performance and address the over-fitting problem. As the volume of data increases, the number of additional challenges for training the data tends to grow as well. It is challenging for the model when an unseen or new dataset is introduced. This problem causes overfitting, which can be addressed by dropout and batch normalization. At the end of each round of the training phase, the dropout mechanism is used to deactivate many nodes. The primary goals of batch normalisation are to enforce a zero mean and a one standard

deviation for all activation functions in the specified layer and for each small batch, in order to increase overall accuracy, make the network more resistant to overfitting, and accelerate the convergence of the gradient descent process. Finally, the fully connected layer connects each layer with another layer to classify, which is the end part of the CNN model as shown in Fig. 2. It collects information from the feature extraction stage and performs analysis on the output of all previous levels. As a result, data classification is achieved by connecting selected features in a nonlinear manner (Rawat and Wang 2017).

Fully convolutional neural network

(Ronneberger et al 2015) was first introduced for biological image segmentation, but it is currently used in a variety of remote sensing applications, where it produces promising results in high resolution images (Wurm et al. 2019). Fully Convolutional Neural Network (FCNN) is a widely used network for semantic image segmentation. Various segmentation approaches involve the encoder-decoder framework in FCNN, as shown in Fig. 3, whereas the first part extracts the feature encoding information into a condensed vector known as the encoder, and the second component is the decoder, which decodes the vector data by upsampling it to the spatial

resolution (Long et al 2015a). As a result, combining completely encoded and decoded with skip connections helps to prevent the loss of accuracy as shown in Fig. 4 (Badrinarayanan et al 2017). The major operations involved in FCNN are:

- 1) Convolution Block: The current base networks are configured to accept inputs of the size $(H * W * n_{channels})$ required for remote sensing images with three channels (RGB) of Red, Green, and Blue. Each convolution layer has a kernel size and to keep the input's height and width zero-padding is used.
- 2) Pooling: By removing the feature from the feature map, the pooling function reduces the size of the input picture.
- 3) Concatenation: In this layer, the preceding layer's output encoder part is concatenated with the decoder part by up-sampling the output with the dimensions $(H * W * n_{up})$ and the concatenated output becomes $(H * W * n_{up} * n_{conv})$.
- 4) Up-sampling: This layer doubles the height and width of the image to change the number of pixels with the same value to the same number of pixels.
- 5) Transpose Convolution: This layer transposes the convolution by switching the dimensions to increase the output.

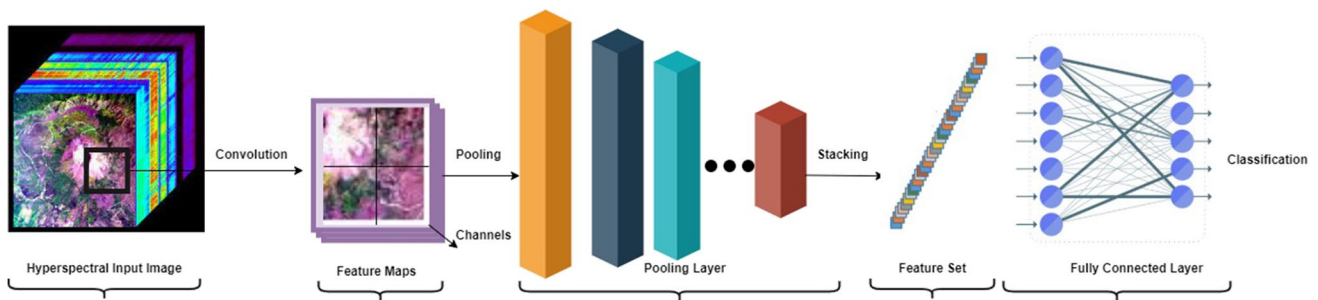


Fig. 2 Spatial feature extraction by using CNN model

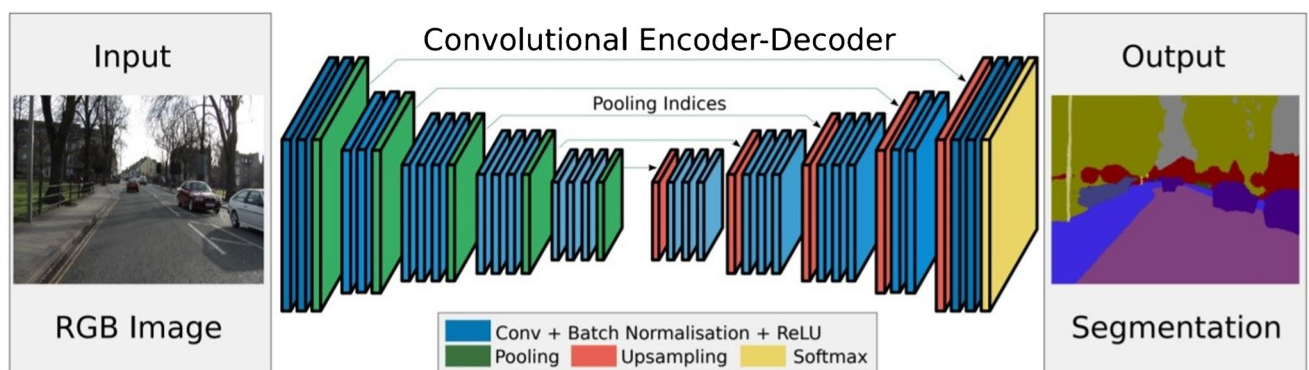


Fig. 3 Spatial feature extraction by using FCNN model (Badrinarayanan et al 2017)

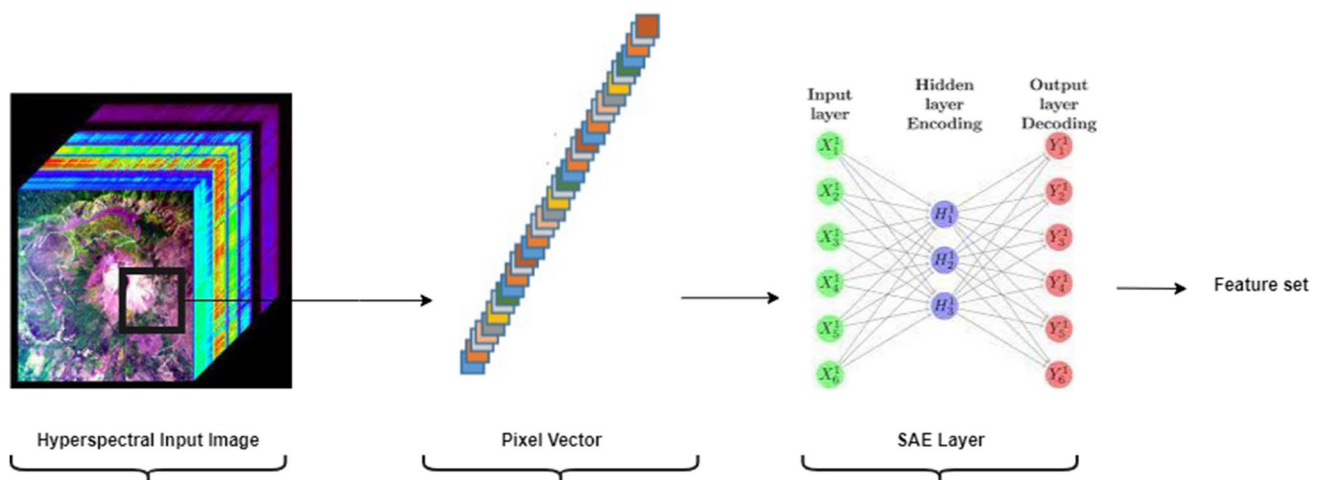


Fig. 4 Spatial feature extraction by SAE

- 6) 6) Deconvolution: This layer performs the inverse of the convolution function, whereas the deconvolutional layer's forward pass equals the convolutional layer's backward pass, and vice versa. Deconvolutions are used to drive the model to learn more accurate outputs.

Autoencoder

An autoencoder (AE) is a key approach for deep learning to feature in a hierarchical manner. Its architecture is composed of three layers: an input layer (encoded layer), a hidden layer, and a reconstruction layer (decoded layer). In comparison to the input and reconstruction layers, the hidden layer contains fewer units. Both the encoded and decoded data have the same number of units, and between each pair of layers, a non-linearity function is applied in the Fig. 4.

It converts an input layer $xinR^n$ to a hidden layer $hinR^h$ with a latent representation. whereby W is the input's weight, β is the hidden layer's bias vector, and $g()$ is the activation function.

$$h = g(Wx + \beta)$$

Following that, the latent representation h is used to reverse map $y \in R^n$ where,

$$y = g(\theta h + \gamma)$$

y denotes the output layer, θ denotes the weight matrix from the hidden layer to the output layer, and γ is the output layer's bias vector. The training procedure's objective is to reduce the reconstruction

error $j(x,y)$ between x and y . If the reconstruction error is smaller than a certain value, the latent representation can be employ to minimize the number of features. A lot of AE are piled together to lower the error rate. These hidden layers are sent into the subsequent layer, resulting in the stacked autoencoder pattern (SAE). These arrangements may gradually generate deep features and train each additional layer using a greedy technique. After each layer, a pooling process compresses the features of successively bigger input regions into smaller ones, which can aid in a variety of classification or clustering tasks (Shin et al 2013).

Discussion

Statistical analysis and meta-analysis

The LULC literature reviewed comprises research that uses the DL technique to classify land cover. A systematic literature search was conducted to locate the articles in Scopus database about LULC in image processing using DL. A systematic review has been done to analyze the research paper related to the literature and to achieve the objectives of our research based upon Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (McInnes et al 2018) and the recommendation for systematic review for prediction model (CHARM model)(Moons et al 2014).

Search strategy

SCOPUS was used as databases, whereas to verify the validity and quality of the result, we limited the search results in

journals, conference and book articles. A title and keyword search in the SCOPUS database (search date: 13 January, 2022) using the search query "remote sensing" AND "Deep Learning" to identify the 413 published articles and 25 from other database. After eliminated some articles, several form of information retrieved like "application of remote sensing", "ML and DL model used in LULC" from 56 relevant articles which were obtained by using the search query "Deep Learning" AND "LULC" which include the journals, conferences and books. A flow diagram of inclusion criteria is depicted in Fig. 5

Various inclusion factors and exclusion factors included to validate the studies based on the motivation of this paper. The exclusion factors are as follows:

- Non-English language articles.
- Remote sensing dataset were not included in the articles.
- Full-text does not provided by the publisher.
- Studies without an outcome measures.
- The inclusion factors are as follows:
- The number of articles focusing on sub-areas of remote sensing applications using ML and DL model.
- Use cases between the years 2015–2021.
- Peer-reviewed article, journals, conferences and books.

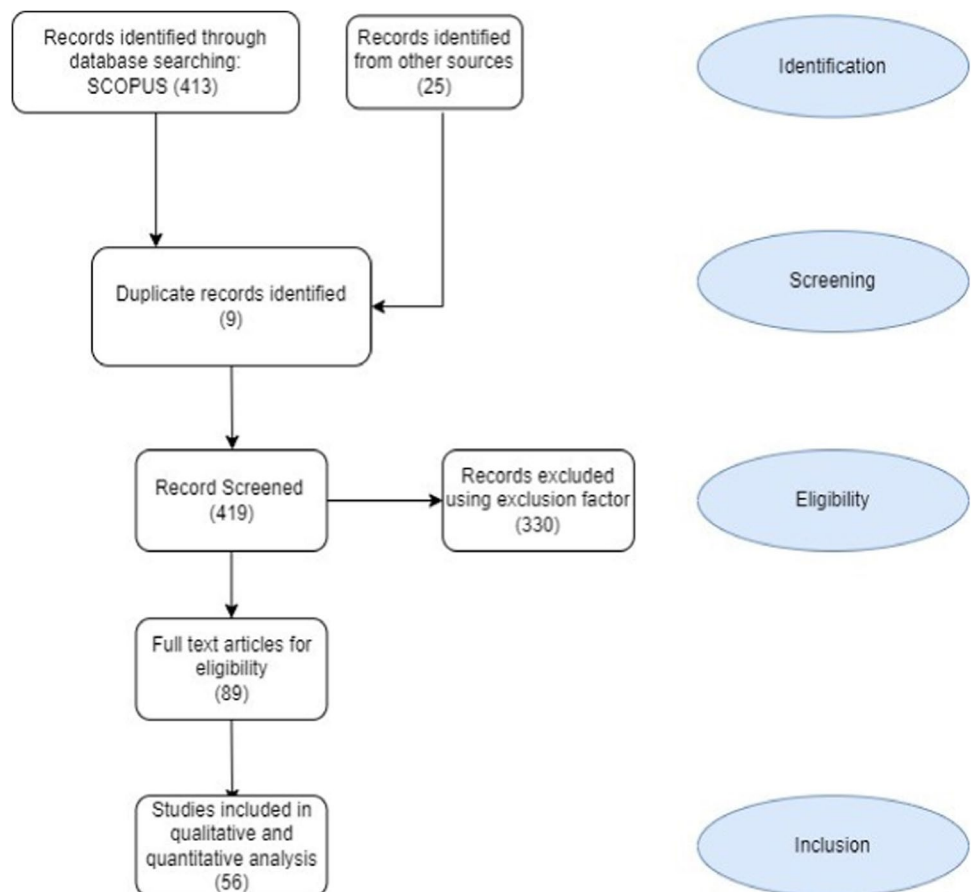
A total of 438 studies has been identified using SCOPUS and other database, 9 were identified as duplicate studies while 330 were determined as irrelevant to this meta-analysis. The database contained the record of remaining 89 articles which was further screened by using the qualitative and quantitative analysis. The final database of 56 articles were accepted in this meta-analysis.

A concise interpretation of the findings

To identify the articles in the Scopus database, type "deep learning" and "remote sensing" into the search box (search date: January 13, 2022). Based on the query, we retrieved the 438 publications from 2015 to 2021 using the Scopus database in Fig. 6 that identifies the frequency of publications in journals, which was further filtered by refining the articles in the search window by article title, keywords, and abstract. For the statistical analysis, we identified various articles in "DL" and "LULC" queries, which were refined to create the database on different DL models used in LULC, and the number of publications increased during the period (2015–2021) in Fig. 7.

As shown in Fig. 8, distribution of the publication increased during the period of 2015–2021. Most of the

Fig. 5 Flow diagram of the search and inclusion factors using PRISMA AND CHARM model



journal articles focus on remote sensing applications in various fields. As of 2021, the number of journal articles exceeds the number of conference papers, reviews, and notes, which reflects the industry’s growth in LULC. This demonstrates that DL has a wide range of applications in remote sensing. In Fig. 9, it summarize the statistical analysis of DL approach which provides the increasing frequency of articles from 2015 to 2021. It is predicted that the scope

of the article will increase in upcoming years. However, the remote sensing community has shifted its interest in recent years to DL models in light of the remarkable success of DL models in the majority of state-of-the-art approaches for a diverse range of applications.

As shown in Fig. 10, the LULC analysis using the DL model, CNN is the most often used for classifications, followed by AE, FCN, and RNN during the period of

Fig. 6 Identified journals with frequency of publications from 2015–2021

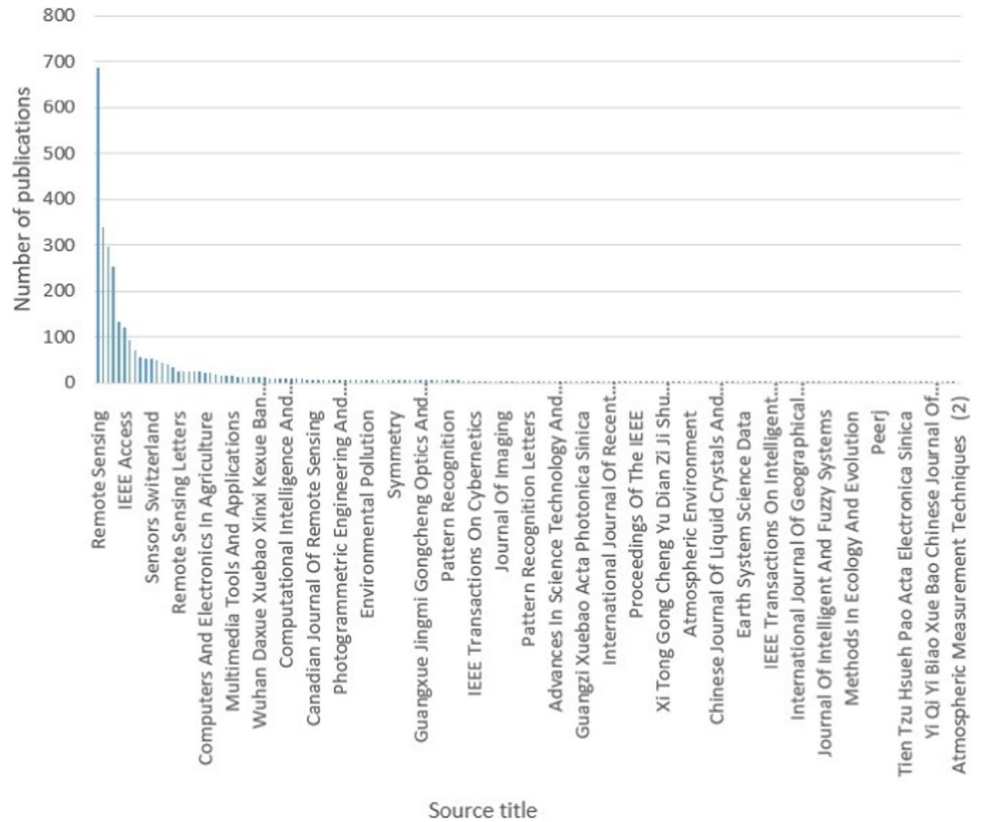


Fig. 7 Number of publications increases from (2015–2021) in the “Scopus database”

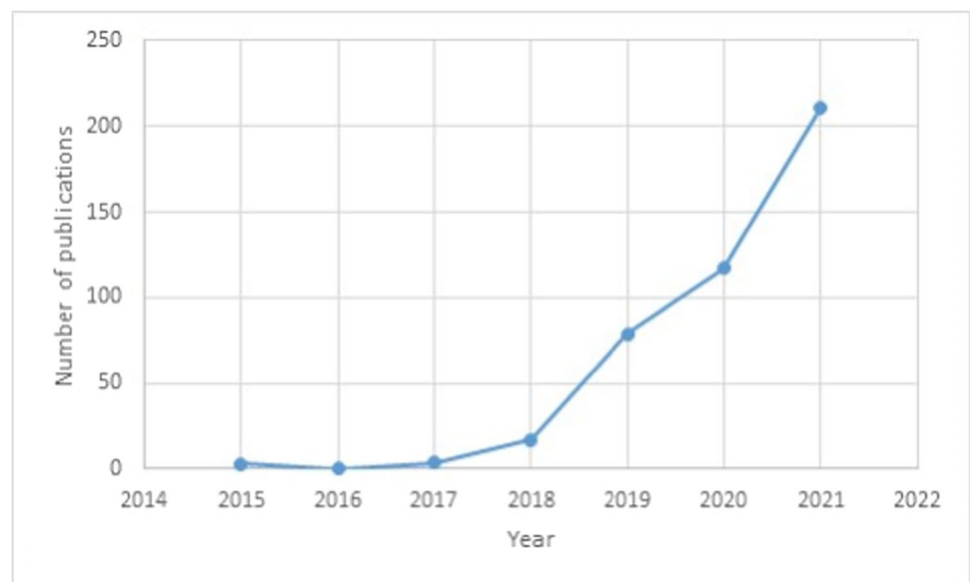


Fig. 8 Distribution of publications (Conference Paper, Article, Review and Note) increases during the period (2015–2021)

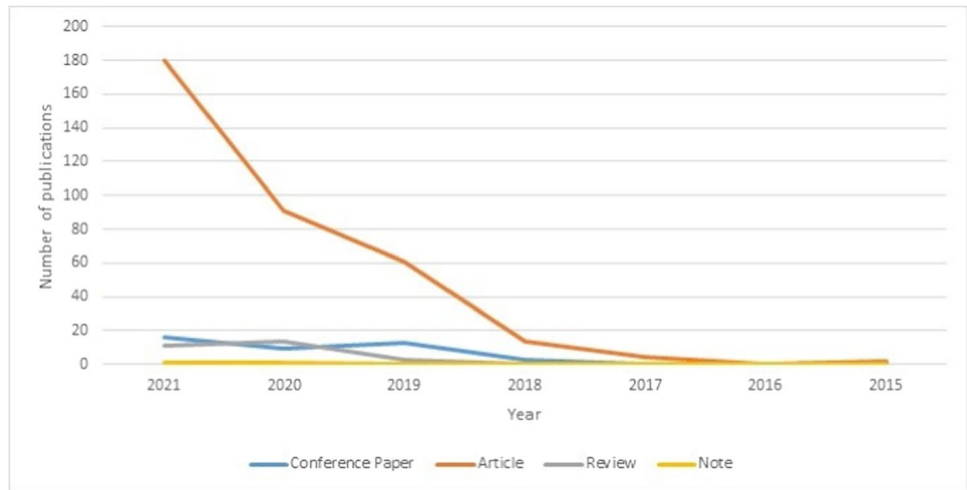


Fig. 9 Prediction and scope of the publication will increase in upcoming year

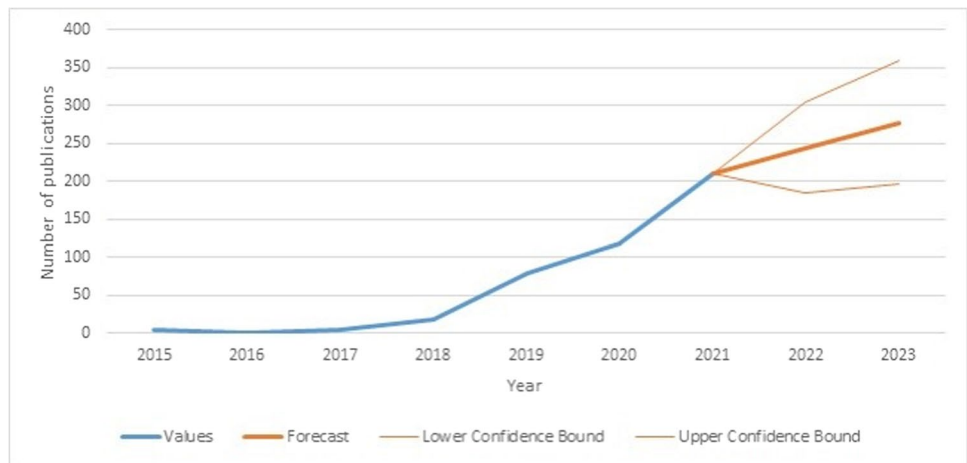


Fig. 10 Distribution of DL models used in LULC from (2018–2021)

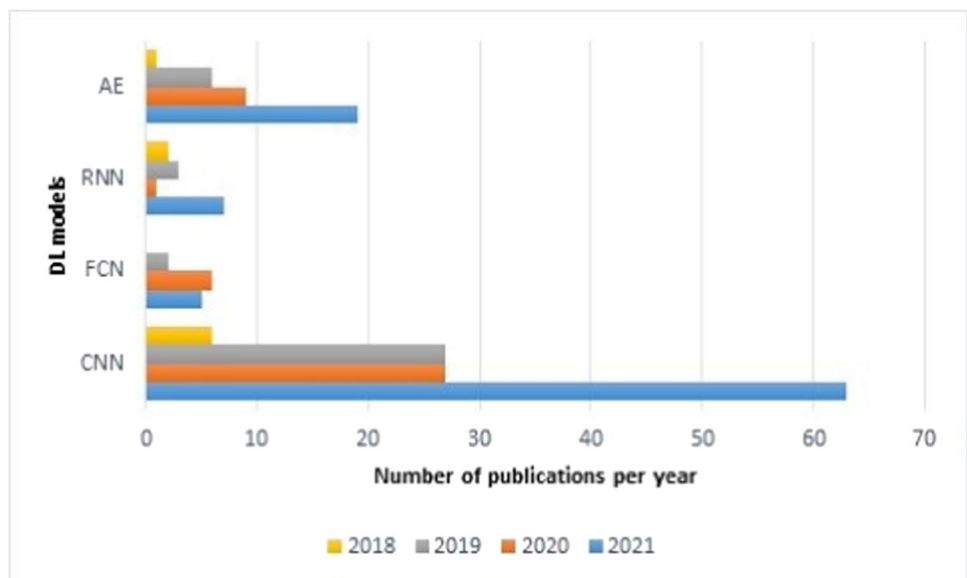


Table 7 Advantages and disadvantages of DL model in LULC

DL Model	Advantages	Disadvantages
CNN	<ol style="list-style-type: none"> 1. Weight sharing, parameter sharing capabilities 2. Good accuracy in image recognition 3. Multi-source and multi-spectrum information to the CNN based model in RS provides the better accuracy Zhang et al 2018b 4. To reducing the complexity, cost of training and efficient RS image classification can achieved by using pertained CNN model Weng et al 2017 5. CNN-ELM model provides the excellent result in scene classification 6. Relatively high detection speed, end-to-end object detection by using the Faster R-CNN model Ren et al 2017 	<ol style="list-style-type: none"> 1. Calculations will be repeated as a result of feature extraction 2. Complex training process 3. A significant number of unnecessary computations Song et al 2019 4. Lack of spatial homogeneity Wu et al 2018
FCN	<ol style="list-style-type: none"> 1. To predict the dense pixel label prediction 2. It requires less memory consumption 3. FCN used for semantic segmentation to explore the multi-level context information Long et al 2015a, b 4. In deconvolutional layer, up-sampling is used to recover the spatial information while also creating the forecast of each pixel Jiang et al 2019 5. Discontinuities caused by patch boundaries are eliminated, and the execution time at inference is reduced Maggiori et al. 2016 	<ol style="list-style-type: none"> 1. Challenging to obtain the accurate boundary localization 2. Problem of class balancing Liu et al. 2017 3. Loss of spatial information 4. Since it reduce the parameters, it may loss the flexibility of the model
SAE	<ol style="list-style-type: none"> 1. Reduce the dimensionality of data 2. They are more useful in some application where it required learning feature or extracting feature 3. Main advantage of AE are de-noising and anomaly detection 4. It takes shorter prediction time for large scale mapping in LULC Li et al 2016 5. AE do not required labelled sample to construct the model as they train unsupervised model Chalapathy and Chawla 2019 	<ol style="list-style-type: none"> 1. It takes longer sample training time 2. If the bottleneck layer is too narrow, then algorithm may loose important dimension 3. AE are specific to the data-set, so it may cause difficulties in scaling 4. Since AE capture more information due to dense layer which may not be relevant information Lawton 2020
Other DL models	<ol style="list-style-type: none"> 1. RNN model can use with CNN to extend the effective pixel neighborhood 2. In order to build an ANN, it require less formal statistical training 3. GAN approach based on RS is the novel effective way to calculate the thickness of thin clouds 4. GAN approaches are basically provide the super resolution Zhang et al 2020 	<ol style="list-style-type: none"> 1. RNN consist gradient and vanishing problem 2. ANN increased computational capacity 3. GAN based RS approach require massive data sets as well as high-power computing capabilities

2018–2021. Due to the higher popularity of CNN and its unique qualities, which make it ideal for processing HSI/MSI remote-sensing images with regularly ordered pixels. The CNN model is capable of obtaining high-level spatial characteristics, which are useful for various analysis tasks in remote sensing.

Analysis of remote sensing images in LULC utilising DL methods is summarised in this paper, which shows the graphical representation of the higher-frequency of articles in 2021. According to the analysis, Fig. 10, represents the CNN model, which is more popular than other DL models. By studying the current techniques and literature review, we conclude that DL in LULC classification of images is still at a young age and a lot of scope is available.

Advantages and disadvantages of various DL model in LULC

In remote sensing applications, sampling a large number of labeled classes of interest is challenging and error-prone, and most of the DL model is based on the number of labeled training samples that are required to optimize the weight in each iteration. Therefore, such a model requires a lot of time. (Novelli et al. 2017) has shown that a pre-trained model with fine-tuning provides better accuracy. However, many DL models are not generalized as they cannot accept more than three colors (RGB) per channel, which may not be ideal in the LULC classification. As remote sensing images always require extra information, Fu et al (2018). As a result, these models need to be rebuilt and redesigned from the initial

Table 8 Classification results of different dataset using DL models in LULC

Dataset	DL model	References	Evaluation metrics	
			OA	AA
ISPRS	CNN	Yang et al 2019	88.00	
	FCN	Papadomanolaki et al 2019	88.99	
Indian Pines	SAE	Rahimzad et al 2021	92.86	
	CNN	Ghamisi et al 2016	81.66	89.64
	FCN	Jiang et al. 2021	99.61	99.81
Pavia University	SAE	Mei et al 2019	92.35	92.04
	CNN	Chen et al 2016	99.54	99.77
	FCN	Zhang et al 2020	79.89	76.60
	SAE	Chen et al 2014	98.52	97.82

which requires sufficient training data (Novelli et al. 2017). Table 7 discusses the advantages and disadvantages of various DL models in LULC.

In this section, we compare the quantitative result of three DL model (CNN, FCN and SAE) using spectral features, spatial features respectively by comparing their result metrics overall accuracy(OA) and average accuracy(AA). The dataset from ISPRS, Indian Pines and Pavia University has been used for quantitative comparison. In terms of classification results based upon the spectral-feature achieved best performance analysis in the given Table 8 and the graphical representation of classification accuracy shown in the Fig. 11.

As seen in the Fig. 11, CNN, FCN and SAE models used for various dataset. CNN based classification model performs better than other models in Pavia university dataset. However, statistically it appears that most of the papers published using CNN model in remote sensing applications as mentioned in the Fig. 10. CNNs are the most powerful DL model for image feature extraction. In comparison to typical shallow models, DL models built using CNNs may hierarchically extract more abstract semantic features from the input images. Using scene segmentation of RS photos, pre-trained CNN models on natural image data sets such as ImageNet (Deng et al 2009) have shown amazing results (Chen et al 2014) (Firat et al 2014). To generate global feature representations for a specific application, deep features can be directly taken from the intermediate layers of a freely accessible CNN architecture, such as AlexNet (Krizhevsky et al 2012) (Simonyan et al 2014) and (Ioffe and Szegedy 2015). In (Hu et al 2015a), multi scale CNN activation functions are used as feature extractors while other coded functions are used for feature encoding method. Fine tuning is the option to provide the valuable approach when new dataset is sufficiently substantial but not large enough to fully train a new network. (Nogueira et al 2017) developed a strategy for finetuning specific high-level layers of the GoogLeNet (Ioffe and Szegedy 2015) using the UC-Merced data set (Xia et al. 2010) achieved the outstanding result. Although supervised deep learning approaches like CNNs and its variations may yield amazing image classification results, there are some drawbacks since they rely on a

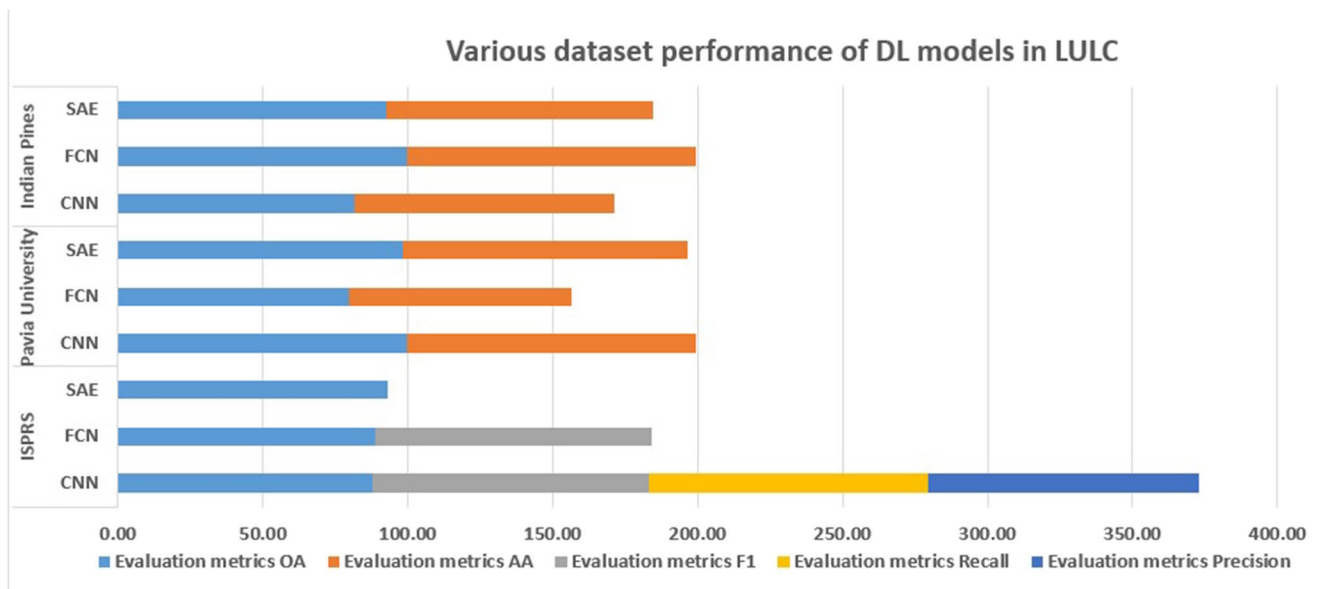


Fig. 11 Classification results of DL models in LULC

large number of labelled training data. Several feature-learning models have been successfully used in remote sensing and may be layered to create deep unsupervised models like SAE, sparse coding, RBMs etc. (Zhang et al 2014). (Romero et al 2015) proposed the use of deep CNNs for RS image classification, by using the unsupervised approach to provide the sparse feature representation to train the network. The efficiency of DL-based RS classification approaches in solving real-world situations has been demonstrated in the previous section. Because of the growing availability of RS data and computing resources, fast progress of DL in remote sensing image categorisation is projected in the future years.

Conclusion

LULC analysis is the most emerging research area in remote sensing applications like climate change, urban planning, disaster management, and ecological change etc. The study was motivated by the popularity of the DL approaches in remote sensing for land cover prediction. Due to the availability of various resources, the HSI/MSI imagery and the Landsat dataset are the most frequently used for image based classification in LULC. We have identified the various datasets which will help the researcher to analyse the LULC change and time-series satellite images. Subsequently, various remote sensing software applications has been identified for pre-processing, classification and prediction.

In this research, we employed state-of-the-art DL frameworks in our study, to explore the hierarchical characteristics of LU and LC categories and abstractions or generalizations of the actual terrain or landscape. This study examined the performance of several of the most current DL architectures that are extensively used for pixel-level labelling in a variety of remote sensing applications. Some of the key findings and gaps have been identified to analyse the new opportunities which outperformed the traditional approaches. According to the overall accuracy of DL models with different parameters, the DL models are superior to ML models in remote sensing applications. Furthermore, we have proposed an overall framework of the DL model as a solution to new challenges and discussed the most commonly used approach in LULC analysis. However, this study was motivated by the exponential growth of DL approaches in LULC, which was systematically identified through statistical analysis using the scopus database. The recommendations presented in this paper seek to greatly benefit researchers by providing a uniform approach for presenting architectural setup and DL approach in LULC analysis in the future. We conclude that DL in LULC classification of images is still at a young age and a lot of scopes are available in the future.

Declarations

Conflict of interest The author(s) declare that they have no competing interests.

References

- Abdollahi A, Pradhan B (2021a) Integrating semantic edges and segmentation information for building extraction from aerial images using unet. *Machine Learning with Applications* 6:100,194. <https://doi.org/10.1016/j.mlwa.2021.100194>, URL <https://www.sciencedirect.com/science/article/pii/S2666827021000979>
- Abdollahi A, Pradhan B (2021b) Urban vegetation mapping from aerial imagery using explainable ai (xai). *Sensors* 21(14). URL <https://www.mdpi.com/1424-8220/21/14/4738>
- Abdollahi A, Pradhan B, Alamri A (2021a) Roadvecnet: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up. *Gisci Remote Sens* 58(7):1151–1174
- Abdollahi A, Pradhan B, Shukla N (2021b) Road extraction from high-resolution orthophoto images using convolutional neural network. *J Indian Soc Remote Sens* 49(3):569–583
- Abdollahi A, Pradhan B, Shukla N, et al (2021c) Multi-object segmentation in complex urban scenes from high-resolution remote sensing data. *Remote Sens* 13(18). URL <https://www.mdpi.com/2072-4292/13/18/3710>
- Abraham M, Satyam N, Pradhan B et al (2021) Developing a prototype landslide early warning system for darjeeling himalayas using sigma model and real-time field monitoring. *Geosci J*. <https://doi.org/10.1007/s12303-021-0026-2>
- Aburas MM, Ahamad MSS, Omar NQ (2019) Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review. *Environ Monit Assess* 191(4):1–28
- Aburas MM, Ho YM, Pradhan B et al (2021) Spatio-temporal simulation of future urban growth trends using an integrated ca-markov model. *Arab J Geosci* 14(2):1–12
- Alhassan V, Henry C, Ramanna S, Storie C (2020) A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. *Neural Comput Appl* 32(12):8529–8544. <https://doi.org/10.1007/s00521-019-04349-9>
- Almeida Cd, Gleriani J, Castejon EF et al (2008) Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *Int J Geogr Inf Sci* 22(9):943–963
- Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 46(3):175–185
- Arun P, Buddhiraju KM, Porwal A (2019) Capsulenet-based spatial-spectral classifier for hyperspectral images. *IEEE J Sel Top Appl Earth Observations Remote Sensing* 12(6):1849–1865
- Atkinson JT, Ismail R, Robertson M (2013) Mapping bugweed (*solanum mauritianum*) infestations in pinus patula plantations using hyperspectral imagery and support vector machines. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(1):17–28
- Azarang A, Manoochehri HE, Kehtarnavaz N (2019) Convolutional autoencoder-based multispectral image fusion. *IEEE Access* 7:35,673–35,683. <https://doi.org/10.1109/ACCESS.2019.2905511>
- Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495

- Barker B, Humber M, Rembold F et al (2020) Strengthening agricultural decisions in countries at risk of food insecurity: The GEOGLAM Crop Monitor for Early Warning. *Remote Sens Environ* 237:111553. <https://doi.org/10.1016/j.rse.2019.111553>
- Belgiu M, Dragut L (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* 114:24–31
- Bittner K, Adam F, Cui S et al (2018) Building footprint extraction from vhr remote sensing images combined with normalized dsms using fused fully convolutional networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(8):2615–2629. <https://doi.org/10.1109/JSTARS.2018.2849363>
- Bose P, Kasabov NK, Bruzzone L et al (2016) Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. *IEEE Trans Geosci Remote Sens* 54(11):6563–6573
- Castelluccio M, Poggi G, Sansone C, et al (2015) Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv:150800092
- Chakraborty R, Pal SC, Sahana M et al (2020) Soil erosion potential hotspot zone identification using machine learning and statistical approaches in eastern india. *Nat Hazards* 104(2):1259–1294
- Chalapathy R, Chawla S (2019) Deep learning for anomaly detection: a survey. arXiv preprint arXiv:190103407
- Chen Y, Lin Z, Zhao X et al (2014) Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6):2094–2107
- Chen Y, Jiang H, Li C et al (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans Geosci Remote Sens* 54(10):6232–6251
- Cheriyadat A, Bruce L (2003) Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In: *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, pp 3420–3422 vol.6. <https://doi.org/10.1109/IGARSS.2003.1294808>
- Cortes C, Vapnik V (1995) Support-Vector networks. *Mach Learn* 20(3):273–297
- Deng J, Dong W, Socher R et al (2009) Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition. IEEE*, pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Devi AB, Deka D, Aneesh TD et al (2022) Predictive modelling of land use land cover dynamics for a tropical coastal urban city in kerala, india. *Arab J Geosci* 15(5):1–19
- Di Noia AHO (2018) Neural networks and support vector machines and their application to aerosol and cloud remote sensing: A review. In: *Kokhanovsky A (eds) Springer series in light scattering*. https://doi.org/10.1007/978-3-319-70796-9_4
- Dikshit A, Pradhan B (2021) Interpretable and explainable ai (xai) model for spatial drought prediction. *Sci Total Environ* 801(149):797. <https://doi.org/10.1016/j.scitotenv.2021.149797>
- Ding C, Li Y, Xia Y, et al (2017) Convolutional neural networks based hyperspectral image classification method with adaptive kernels. *Remote Sens* 9(6):618 dlr.de (2018)
- Fan F, Wang Y, Wang Z (2008) Temporal and spatial change detecting (19982003) and predicting of land use and land cover in core corridor of pearl river delta (china) by using tm and etm+ images. *Environ Monit Assess* 137:127–147. <https://doi.org/10.1007/s10661-007-9734-y>
- Firat O, Can G, Vural FTY (2014) Representation learning for contextual object and region detection in remote sensing. In: *2014 22nd international conference on pattern recognition. IEEE*, pp 3708–3713. <https://doi.org/10.1109/ICPR.2014.637>
- Fu T, Ma L, Li M et al (2018) Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery. *J Appl Rem Sens* 12(2):025,010
- Ghamisi P, Chen Y, Zhu XX (2016) A self-improving convolution neural network for the classification of hyperspectral data. *IEEE Geosci Remote Sens Lett* 13(10):1537–1541. <https://doi.org/10.1109/LGRS.2016.2595108>
- He K, Zhang X, Ren S et al (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 37(9):1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp 770–778
- Hinton G, Deng L, Yu D et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 29(6):82–97
- Hu F, Xia GS, Hu J et al (2015a) Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens* 7(11):14,680–14,707
- Hu W, Huang Y, Wei L et al (2015b) Deep convolutional neural networks for hyperspectral image classification. *J Sens* 2015. <https://doi.org/10.1155/2015/258619>
- Hu M, Wu C, Zhang L et al (2021) Hyperspectral anomaly change detection based on autoencoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:3750–3762. <https://doi.org/10.1109/JSTARS.2021.3066508>
- Huang G, Liu Z, Van Der Maaten L et al (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
- Huang B, Zhao B, Song Y (2018) Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens Environ* 214:73–86
- Hussain M, Chen D, Cheng A et al (2013a) Change detection from remotely sensed images: from pixel-based to object-based approaches. *ISPRS J Photogramm Remote Sens* 80:91–106. <https://doi.org/10.1016/j.isprsjprs.2013.03.006>
- Hussain M, Chen D, Cheng A et al (2013b) Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS J Photogramm Remote Sens* 80:91–106
- Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning, PMLR*, vol 37. pp 448–456. <https://proceedings.mlr.press/v37/ioffe15.html>
- Jiang X, Wang Y, Liu W et al (2019) Capsnet, cnn, fcn: comparative performance evaluation for image classification. *Int J Machine Learning Comput* 9(6):840–848
- Jiang Y, Li Y, Zou S et al (2021) Hyperspectral image classification with spatial consistence using fully convolutional spatial propagation network. *IEEE Trans Geosci Remote Sens* 59(12):10,425–10,437. <https://doi.org/10.1109/TGRS.2021.3049282>
- Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Philos Trans R Soc A Math Phys Eng Sci* 374(2065):20150,202
- Kohonen T (2012) *Self-Organization and Associative Memory*, 3rd edn. Springer, Berlin, Heidelberg, p XV–312. <https://doi.org/10.1007/978-3-642-88163-3>
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
- Lawton G (2020) How to troubleshoot 8 common autoencoder limitations. URL <https://www.techtarget.com/searchenterpriseai/feature/How-to-troubleshoot-8-common-autoencoder-limitations>

- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Li T, Zhang J, Zhang Y (2014) Classification of hyperspectral image based on deep belief networks. In: 2014 IEEE International Conference on Image Processing (ICIP), pp 5132–5136. <https://doi.org/10.1109/ICIP.2014.7026039>
- Li W, Fu H, Yu L et al (2016) Stacked autoencoder-based deep learning for remote-sensing image classification: a case study of african land-cover mapping. *Int J Remote Sens* 37(23):5632–5646. <https://doi.org/10.1080/01431161.2016.1246775>
- Li Y, Zhang H, Xue X et al (2018) Deep learning for remote sensing image classification: a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8(6):e1264
- Li X, Xu F, Lyu X et al (2021) Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images. *Int J Remote Sens* 42(9):3583–3610. <https://doi.org/10.1080/01431161.2021.1876272>
- Likas A, Vlassis N, Verbeek JJ (2003) The global k-means clustering algorithm. *Pattern Recogn* 36(2):451–461
- Litjens G, Kooi T, Bejnordi BE et al (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
- Liu W, Lee J (2019) A 3-d atrous convolution neural network for hyperspectral image denoising. *IEEE Trans Geosci Remote Sens* 57(8):5701–5715. <https://doi.org/10.1109/TGRS.2019.2901737>
- Liu S, Shi Q, Zhang L (2021) Few-shot hyperspectral image classification with unknown classes using multitask deep learning. *IEEE Trans Geosci Remote Sens* 59(6):5085–5102. <https://doi.org/10.1109/TGRS.2020.3018879>
- Liu Y, Minh Nguyen D, Deligiannis N et al (2017) Hourglass-shapenetwork based semantic segmentation for high resolution aerial imagery. *Remote Sens* 9(6):522
- Liu Y, Chen X, Wang Z et al (2018a) Deep learning for pixel-level image fusion: recent advances and future prospects. *Information Fusion* 42:158–173
- Liu Y, Fan B, Wang L et al (2018b) Semantic labeling in very high resolution images via a self-cascaded convolutional neural network. *ISPRS J Photogramm Remote Sens* 145:78–95
- Long J, Shelhamer E, Darrell T (2015a) Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(4):640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
- Long J, Shelhamer E, Darrell T (2015b) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- Lunetta RS, Johnson DM, Lyon JG, et al (2004) Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote Sens Environ* 89(4):444–454. <https://doi.org/10.1016/j.rse.2003.10.022>, URL <https://www.sciencedirect.com/science/article/pii/S0034425703002852>
- Ma L, Liu Y, Zhang X et al (2019) Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS J Photogramm Remote Sens* 152:166–177
- Maggiori E, Tarabalka Y, Charpiat G et al (2016) Fully convolutional neural networks for remote sensing image classification. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). pp 5071–5074. <https://doi.org/10.1109/IGARSS.2016.7730322>
- Marcos D, Volpi M, Kellenberger B et al (2018) Land cover mapping at very high resolution with rotation equivariant cnns: towards small yet accurate models. *ISPRS J Photogramm Remote Sens* 145:96–107
- McInnes MD, Moher D, Thombs BD et al (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the prisma-dta statement. *JAMA* 319(4):388–396
- McRoberts RE (2014) Post-classification approaches to estimating change in forest area using remotely sensed auxiliary data. *Remote Sensing of Environment* 151:149–156. <https://doi.org/10.1016/j.rse.2013.03.036>, URL <https://www.sciencedirect.com/science/article/pii/S0034425713003490>, special Issue on 2012 ForestSAT
- Mei S, Ji J, Geng Y et al (2019) Unsupervised spatial-spectral feature learning by 3d convolutional autoencoder for hyperspectral classification. *IEEE Trans Geosci Remote Sens* 57(9):6808–6820. <https://doi.org/10.1109/TGRS.2019.2908756>
- Migliani A, Kumar N (2019) Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Vehicular Communications* 20:100,184. <https://doi.org/10.1016/j.vehcom.2019.100184>, URL <https://www.sciencedirect.com/science/article/pii/S2214209619302311>
- Moons KG, de Groot JA, Bouwmeester W et al (2014) Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist. *PLoS Med* 11(10):e1001,744
- Najibi M, Rastegari M, Davis LS (2016) G-cnn: an iterative grid based object detector. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2369–2377. <https://doi.org/10.1109/CVPR.2016.260>
- Ng A, Jordan M (2014) On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. In: Dietterich T, Becker S, Ghahramani Z (eds) *Advances in Neural Information Processing Systems*. MIT Press
- Nogueira K, Penatti OA, Dos Santos JA (2017) Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recogn* 61:539–556
- Novelli A, Aguilar MA, Aguilar FJ et al (2017) Assesseg—a command line tool to quantify image segmentation quality: a test carried out in southern spain from satellite imagery. *Remote Sens* 9(1):40
- Nwankpa C, Ijomah W, Gachagan A, et al (2018) Activation functions: comparison of trends in practice and research for deep learning. arXiv preprint arXiv:181103378
- Pal M, Mather PM (2003) An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens Environ* 86(4):554–565. [https://doi.org/10.1016/S0034-4257\(03\)00132-9](https://doi.org/10.1016/S0034-4257(03)00132-9)
- Pal M, Foody GM (2012) Evaluation of svm, rvm and smlr for accurate image classification with limited ground data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(5):1344–1355. <https://doi.org/10.1109/JSTARS.2012.2215310>
- Pal S, Ghosh SK (2017) Rule based end-to-end learning framework for urban growth prediction. arXiv preprint arXiv:171110801
- Paoletti ME, Haut JM, Plaza J et al (2018) A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J Photogramm Remote Sens* 145:120–147
- Papadomanolaki M, Vakalopoulou M, Karantzas K (2019) A novel objectbased deep learning framework for semantic segmentation of very highresolution remote sensing data: comparison with convolutional and fully convolutional networks. *Remote Sens* 11(6). URL <https://www.mdpi.com/2072-4292/11/6/684>
- Pashaei M, Kamangir H, Starek MJ, et al (2020) Review and evaluation of deep learning architectures for efficient land cover mapping with uas hyperspatial imagery: a case study over a wetland. *Remote Sens* 12(6). URL <https://www.mdpi.com/2072-4292/12/6/959>
- Petitjean F, Kurtz C, Passat N et al (2013) Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recogn Lett* 33:1805. <https://doi.org/10.1016/j.patrec.2012.06.009>

- Rahimzad M, Homayouni S, Alizadeh Naeini A, et al (2021) An efficient multi-sensor remote sensing image clustering in urban areas via boosted convolutional autoencoder (bcae). *Remote Sens* 13(13). URL <https://www.mdpi.com/2072-4292/13/13/2501>
- Rawat W, Wang Z (2017) Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation* 29(9):2352–2449
- Redmon J, Divvala S, Girshick R et al (2016) You only look once: unified, realtime object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren S, He K, Girshick R (2017) andj. sun,“fasterrcnn: towards realtime object detection with region proposal networks.” *IEEE Trans Pattern Anal Mach Intell* 39(6):1137–1149
- Rezaee M, Mahdianpari M, Zhang Y et al (2018) 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* 11(9):3030–3039
- Romero A, Gatta C, Camps-Valls G (2015) Unsupervised deep feature extraction for remote sensing image classification. *IEEE Trans Geosci Remote Sens* 54(3):1349–1362
- Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Shen Z, Liu Z, Li J, et al (2017) Dsod: learning deeply supervised object detectors from scratch. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp 1937–1945. <https://doi.org/10.1109/ICCV.2017.212>
- Shi Y, Ma D, Lv J et al (2021) Actl: Asymmetric convolutional transfer learning for tree species identification based on deep neural network. *IEEE Access* 9:13,643–13,654. <https://doi.org/10.1109/ACCESS.2021.3051015>
- Shin HC, Orton MR, Collins DJ et al (2013) Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE Trans Pattern Anal Mach Intell* 35(8):1930–1943. <https://doi.org/10.1109/TPAMI.2012.277>
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for largescale image recognition. arXiv preprint arXiv:14091556. <https://doi.org/10.48550/arXiv.1409.1556>
- Simonyan K, Vedaldi A, Zisserman A (2014) Deep inside convolutional networks: visualising image classification models and saliency maps. In: In Workshop at International Conference on Learning Representations, arXiv:1312.6034. <https://doi.org/10.48550/arXiv.1312.6034>
- Song C, Woodcock C (2003) Monitoring forest succession with multitemporal landsat images: factors of uncertainty. *IEEE Trans Geosci Remote Sens* 41(11):2557–2567. <https://doi.org/10.1109/TGRS.2003.818367>
- Song J, Gao S, Zhu Y et al (2019) A survey of remote sensing image classification based on cnns. *Big Earth Data* 3(3):232–254
- Sun X, Zhou F, Dong J et al (2017) Encoding spectral and spatial context information for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 14(12):2250–2254. <https://doi.org/10.1109/LGRS.2017.2759168>
- Sun L, Yang X, Jia S et al (2020) Satellite data cloud detection using deep learning supported by hyperspectral data. *Int J Remote Sens* 41(4):1349–1371. <https://doi.org/10.1080/01431161.2019.1667548>
- Uddin MP, Mamun MA, Hossain MA (2021) Pca-based feature reduction for hyperspectral remote sensing image classification. *IETE Tech Rev* 38(4):377–396. <https://doi.org/10.1080/02564602.2020.1740615>
- van der Meer F (2011) Advances in environmental remote sensing sensors : algorithms and applications / ed. by w. qihao, crc press - taylor francis, london, 2011, 556 p., isbn 978-1-4200-9175-5 : book review. *Int J Appl Earth Obs Geoinf (JAG)* 13(5):838–839. <https://doi.org/10.1016/j.jag.2011.05.015>
- Wang J, Sun K, Cheng T et al (2020) Deep high-resolution representation learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 43(10):3349–3364. Retrieved 1 October 2021 from, <https://doi.org/10.1109/TPAMI.2020.2983686>
- Weng Q, Mao Z, Lin J et al (2017) Land-use classification via extreme learning classifier based on deep convolutional features. *IEEE Geosci Remote Sens Lett* 14(5):704–708
- Woo S, Park J, Lee JY et al (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
- Wu G, Shao X, Guo Z et al (2018) Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks. *Remote Sens* 10(3):407
- Wurm M, Stark T, Zhu XX et al (2019) Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J Photogramm Remote Sens* 150:59–69
- Xia GS, Yang W, Delon J et al (2010) Structural high-resolution satellite image indexing. In: ISPRS TC VII Symposium-100 Years ISPRS. pp 298–303. <https://hal.archives-ouvertes.fr/hal-00458685>
- Xu Y, Yu L, Zhao F et al (2018) Tracking annual cropland changes from 1984 to 2016 using time-series landsat images with a change-detection and post-classification approach: experiments from three sites in africa. *Remote Sens Environ* 218:13–31. <https://doi.org/10.1016/j.rse.2018.09.008>
- Xu Z, Su C, Zhang X (2021) A semantic segmentation method with category boundary for land use and land cover (lulc) mapping of very-high resolution (vhr) remote sensing image. *Int J Remote Sens* 42(8):3146–3165. <https://doi.org/10.1080/01431161.2020.1871100>
- Yang H, Yu B, Luo J et al (2019) Semantic segmentation of high spatial resolution images with deep neural networks. *Gisience & Remote Sensing* 56:749–768
- Yuan Q, Shen H, Li T et al (2020) Deep learning in environmental remote sensing: aAchievements and challenges. *Remote Sens Environ* 241(111):716
- Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer, pp 818–833. https://doi.org/10.1007/978-3-319-10590-1_53
- Zhang F, Du B, Zhang L (2014) Saliency-guided unsupervised feature learning for scene classification. *IEEE Trans Geosci Remote Sens* 53(4):2175–2184
- Zhang L, Zhang L, Du B (2016) Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geosci Remote Sens Mag* 4(2):22–40
- Zhang H, Li Y, Zhang Y et al (2017) Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sens Lett* 8(5):438–447
- Zhang C, Pan X, Li H et al (2018a) A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification. *ISPRS J Photogramm Remote Sens* 140:133–144
- Zhang Y, Xia W, Zhang YZ et al (2018b) Road extraction from multi-source high-resolution remote sensing image using convolutional neural network. In: 2018b International Conference on Audio, Language and Image Processing (ICALIP). IEEE, pp 201–204. <https://doi.org/10.1109/ICALIP.2018.8455367>

- Zhang C, Sargent I, Pan X et al (2019) Joint deep learning for land cover and land use classification. *Remote Sens Environ* 221:173–187. <https://doi.org/10.1016/j.rse.2018.11.014>
- Zhang K, Gu S, Timofte R (2020) Ntire 2020 challenge on perceptual extreme super-resolution: methods and results. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. pp 2045–2057. <https://doi.org/10.1109/CVPRW50498.2020.00254>
- Zhao C, Wan X, Zhao G et al (2017) Spectral-spatial classification of hyperspectral imagery based on stacked sparse autoencoder and random forest. *Eur J Remote Sens* 50(1):47–63
- Zhong L, Hu L, Zhou H (2019) Deep learning based multi-temporal crop classification. *Remote Sens Environ* 221:430–443. <https://doi.org/10.1016/j.rse.2018.11.032>
- Zhou W, Shao Z, Diao C et al (2015) High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder. *Remote Sens Lett* 6(10):775–783