A Study on the Advanced Method for Image Classification of Remote Sensing Datasets Using Deep Learning Algorithms

[1]Juvvala Bala Ambedkar, [2]Dr. Ajay Agarwal
[1]*Research Scholar, The Glocal University, Saharanpur (UP)*
[2]*Professor, The Glocal University, Saharanpur (UP)*

Abstract: Geospatial technology and advanced AI methodologies can improve the processing of enormous spatial datasets, provide accurate forecasts, rapid user-defined models, and more. Machine-Deep-Learning algorithms, a branch of artificial intelligence, are supported by powerful computing platforms and can be used with geospatial science to visualise, analyse, and predict real-time COVID-19 issues. The main aim of this study is to emphasise geospatial-analytical methods, advanced machine-deep learning algorithms in big data mining, spatial visualisation, and web-based spatial analysis that can give decision-makers new predictive models and more intuitive information. Multi-dimensional sensors simplify data collection and enable global research. This study analyses complex remote sensing datasets using free optical and microwave data. Landsat -7,8, Sentinel-2, and MODIS are optical datasets, while Sentinel-1 SAR is microwave. Machine learning and the most common deep learning models of convolutional neural network (CNN) for large-scale mapping can automatically and independently extract information without human interaction. It depicts numerous steps and the complete procedure. Data is injected and transmitted across layers to extract key features by removing picture dimensions. Google Earth Engine's cloud platform helps do the tasks. These cutting-edge approaches can use heterogeneous and complex huge data in remote sensing applications. This laid the groundwork for the new information age's geographical database.

Keywords: *Google Earth Engine, Complex Remote Sensing, Machine Learning, Real-Time Applications*

Introduction:

There have been a number of significant advancements in the field of satellite observation recently as a direct result of the development of both artificial intelligence and machine learning. Detailed understanding (DL) has lately emerged as the trend in huge data analysis that is expanding at the fastest rate, and it has been effectively applied to a wide variety of sectors (Sharma et al., 2017). When compared to other machine learning approaches, DL has recently emerged as the fastest-growing trend. The sole solution to these issues is not the implementation of deep learning techniques; in addition, it is necessary to develop new robust solutions (Yuan et al., 2020; Mahdianpari et al., 2020). Creating effective methods for the processing of challenging remote sensing datasets and an outstanding cloud computing platform is one of the most important components of producing a complete system.This is capable of handling a broad variety of issues and needs in RS applications, both now and in the future (Vali et al., 2020). The following section elaborates the past literatures related to this concept in detail.

Literature Review:

YEAR AND AUTHORS	METHODOLOGY	FINDINGS
Amani et al., (2020)	Looked at 450 journal articles that were published in a total of 150 different journals between January 2010 and May 2020.	It was discovered that users of GEE made substantial use of the Landsat and Sentinel datasets. In addition, supervised using computational learning methods like Random Forests, were used in picture categorization tasks in an increasingly widespread manner.
Wu et al. (2021)	cutting-edge techniques for processing massive data gathered through remote sensing and in-depth investigations of current parallel implementations on many well-liked platforms for computing with high speeds.	Advanced cloud computing is at processing huge data collected through remote sensing, as well as how scheduling tactics might increase computational efficiency.
Zhao et al. (2022)	Satellite imagery is currently essential to improvements in our knowledge of the Earth and its natural environments as spacecraft observational capability and the variety of Earth observation (EO) sensors increases.	(1) Links and papers that reference EO sensors are increasing swiftly, but those that indicate AVHRR, SPOT, and TerraSAR have been declining; (2) Only a few magazines mainly print works regarding EO satellites: (3) To evaluate the effects of EO sensors as well as to forecast future developments in the information's uses, from a distance sensed influence factor (RSIF), a unique impact metric, was developed.
Ezzat Salem and Hashim Al-Saedi (2023)	The article provides a thorough analysis of cloud-based malware detection technologies as well as information on how to use the cloud to safeguard key infrastructure and the Internet of Things against attacks. In addition to outlining a methodology for identifying cloud-based malware using deep learning and data extraction, this paper looks at the advantages and disadvantages of cloud environments in terms of malware detection.	The results of this study could be used in the future to draw attention to the problem that is currently being studied in malware research.
Torgbor et al. (2023)	This study is the first to test a "time series"-based sensors technique for predicting mango production. This approach uses open-source satellite images instead of manual fruit counting in the field. Annual production statistics from 2015 to 2022 were derived from 51 unique vineyard boxes across two farms (AH and MK) in Australia's northern territory	The outcome not simply gives the business a more adaptable option, nevertheless it also makes predicting for the block, farm, geographic, and general levels more automated and scalable.

As per the past literatures, Deep learning models, especially complicated CNN, are sometimes viewed as "black boxes," making their decision-making difficult to explain. Deep learning model interpretation and explanation is a research gap in distant sensing, where correct and clear reasoning is critical. Therefore, the fundamental goal of this study is to develop reliable deep learning systems that can withstand environmental variations using domain adaptability, data augmentation, and transfer learning.

Methodology:

As more remote sensing applications are needed to handle or analyse the vast volumes of remotely sensed datasets gathered, standalone mode processing is no longer sufficient. To examine different machine learning methods, this research has developed a distributed framework that is run on a cloud computing engine. An area of computing that focuses on the study of scientific data is machine learning algorithms. This research was carried out to efficiently process massive amounts of remote sensing data. The Google Earth Engine (GEE) is used to gather all of the remote sensing datasets. Following that, these datasets were filtered according to the study area, satellites, and cloud cover. The satellites whose data were used to construct the multispectral imageries were Landsat 7, Landsat 8, and Sentinel-2. The acronyms for these missions are L7, L8, and S2, respectively. The satellite that gave the information for the microwave satellite pictures is called Sentinel-1, or S1. The properties of each satellite dataset used in this inquiry, as well as its characteristics and applications. Each of these datasets has different characteristics, some of which differ between collections and could therefore have an impact on the results. Google Collab, Keras, Arcgis Software, and Google Earth Engine are the platforms and software used in this study.
The proposed architecture for advanced method for image classification of remote sensing datasets using deep learning algorithm is as follows:

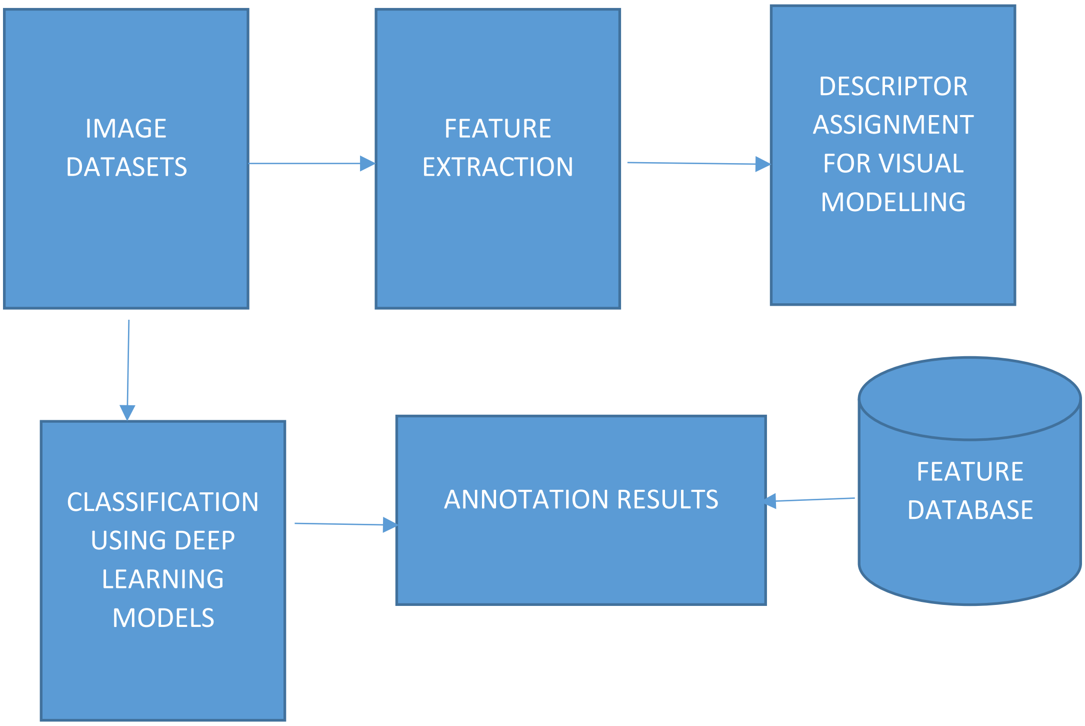

Figure 1: Proposed architecture

Results And Discussion:

To understand machine learning and deep learning, data dependence, user requirements, extracted features, a problem-solving method, the extent of the study, execution time, interpretability, data characteristics, computing, and algorithmic analysis are examined. Understanding and assessing machine learning classifiers based on these properties is the greatest way to increase processing speed. Thus, research focuses on enhancing classifiers by customising them to various applications to aid algorithm learning. This optimisation delivers the best classifiers for tough remote sensing datasets.

Characteristics/Methods	Advantages	Limitations
"Classification and Regression decision Tree" (CART)	Less effort in pre-processing the dataset. Missing values do not affect the building of the decision tree.	Accuracy might differ for large areas. Outfitting of data requires pruning of the data. This requires training samples for building a decision tree.
Support Vector Machine (SVM)	It works well when the margin of separation between the classes is clear. Highly effective in the high dimensional space.	This method requires training samples that consume a lot of time; Data with excessive noise is not suitable; the Hyperplane must be selected wisely.
Random Forest (RF)	It offers the greatest precision. The output of each tree is combined after as many trees as possible have been created on the subset of data. By doing this, it lessens the variance and the fitting difficulty for decision trees, which increases accuracy. Regression and classification issues can be resolved using Random Forest.	The chance forest technique's primary drawback is that it may become too sluggish and inefficient to use when there are many woods involved. projections in instantaneously. Algorithms such as these are often rapid to teach, but if learned, can take a while to produce predictions.
Otsu Threshold	No training samples are required for this method. Suited best for images having complex features.	Method not suited for a dynamic environment. Requires excessive pre-processing steps for datasets to remove noise and obtain accurate results.
Change Detection (Threshold)	Quicker execution of segmentation process. No mathematical calculation is required. No requirement for training samples. Do not consume much time.	User-definable threshold. Unstable results where the spectral characteristics between water and other dark pixels often get mismatched.
Deep Neural Network (Convolutional Neural Network	The residual network of CNN models helps to generate an extra number of layers for training that further minimizes the complexity and error of computation. This model generated the highest accuracy in minimum time when compared to other models	One of the major drawbacks of CNN models is that it requires large data for training. The model is not suitable for small-scale mapping. The model cannot encode the orientation of the targeted object

Table 1: Insights Obtained from Each Classifier

The findings of the study showed that the outcomes produced by machine learning classifiers are affected by a number of different factors. In addition to the differences in spatial distribution and extent, one of the factors to take into account is the features extracted for use in training from the different datasets. One of the instances taken from the Landsat-8 dataset that was presented is unique in that the same classifier produced distinct results for training samples and validation samples that were collected from the same locations. According to the results presented in table 1, RF, SVM, and CART performed significantly better when applied to the Sentinel-1 dataset, which has a greater number of features than the Landsat dataset. It has been observed that the classification accuracy of machine learning classifiers can be improved by including extra features if the sample size is greater than the number of those features.

Even while upgrading the multi-temporal satellite image dataset helped all classifiers perform better, increasing the sample size had a different effect. Kernel-based classifiers like SVM demonstrated higher accuracy even with smaller sample sizes than tree-based learners like RF and CART, but their precision only modestly increased with trial number. This is because, regardless of the experimental study size, the SVM's essential behaviors extend effectively by retraining fewer variables. Results from SVM training on smaller Landsat data sets were equivalent to those from larger training sets. Smaller samples should be examined using radiofrequency (RF) and scanning electron microscopy (SVM). However, using national and global satellite imagery, the CNN model beat the other high-resolution tagging approaches. It improved precision while using fewer computing resources, benefiting everyone.

Conclusion:

In spite of low-quality image, the CNN-based algorithm functions admirably even in challenging environments such as those including shadows and adverse weather conditions. Majority of satellites have good performances across the board for each classifier. Therefore, deep neural network models are the best solution for classification purposes on a planet-scale and nationwide scale, whereas machine learning classifiers are beneficial for monitoring specific regions in greater detail. The support vector machine and the random forest performed with the highest accuracies, and the study takes those machine learning classifiers into consideration for regional mapping.

References

[1] Amani, M. *et al.* (2020) 'Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A comprehensive review', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, pp. 5326–5350. doi:10.1109/jstars.2020.3021052.

[2] Ezzat salem, I. and Hashim Al-Saedi, K. (2023) 'Enhancing cloud security through the integration of deep learning and Data Mining Techniques: A comprehensive review', *Periodicals of Engineering and Natural Sciences (PEN)*, 11(3), p. 176. doi:10.21533/pen.v11i3.3596.

[3] Mahdianpari, M. *et al.* (2020) 'Big data for a big country: The first generation of Canadian wetland inventory map at a spatial resolution of 10-M using sentinel-1 and sentinel-2 data on the Google Earth Engine Cloud Computing Platform', *Canadian Journal of Remote Sensing*, 46(1), pp. 15–33. doi:10.1080/07038992.2019.1711366.

[4] Raparthi, M., Dodda, S. B., & Maruthi, S. (2023). Predictive Maintenance in IoT Devices using Time Series Analysis and Deep Learning. Dandao Xuebao/Journal of Ballistics, 35(3). https://doi.org/10.52783/dxjb.v35.113

[5] Sharma, A. *et al.* (2017) 'A patch-based convolutional neural network for Remote Sensing Image Classification', *Neural Networks*, 95, pp. 19–28. doi:10.1016/j.neunet.2017.07.017.

[6] Torgbor, B.A. *et al.* (2023) 'Integrating Remote Sensing and weather variables for mango yield prediction using a machine learning approach', *Remote Sensing*, 15(12), p. 3075. doi:10.3390/rs15123075.

[7] Vali, A., Comai, S. and Matteucci, M. (2020) 'Deep learning for land use and land cover classification based on hyperspectral and Multispectral Earth Observation Data: A Review', *Remote Sensing*, 12(15), p. 2495. doi:10.3390/rs12152495.

[8] Wu, Z. *et al.* (2021) 'Recent developments in parallel and distributed computing for remotely sensed big data processing', *Proceedings of the IEEE*, 109(8), pp. 1282–1305. doi:10.1109/jproc.2021.3087029.

[9] Yuan, Q. *et al.* (2020) 'Deep Learning in Environmental Remote Sensing: Achievements and challenges', *Remote Sensing of Environment*, 241, p. 111716. doi:10.1016/j.rse.2020.111716.

[10] Zhao, Q. *et al.* (2022) 'An overview of the applications of Earth Observation Satellite Data: Impacts and future trends', *Remote Sensing*, 14(8), p. 1863. doi:10.3390/rs14081863.