International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Leveraging Deep Hierarchies in CNNs for Enhanced Satellite Image Classification

Deepika Pahuja[1], Sarika Jain[2], Shishir Kumar[3]

Submitted: 03/03/2024 **Revised**: 25/04/2024 **Accepted**: 03/05/2024

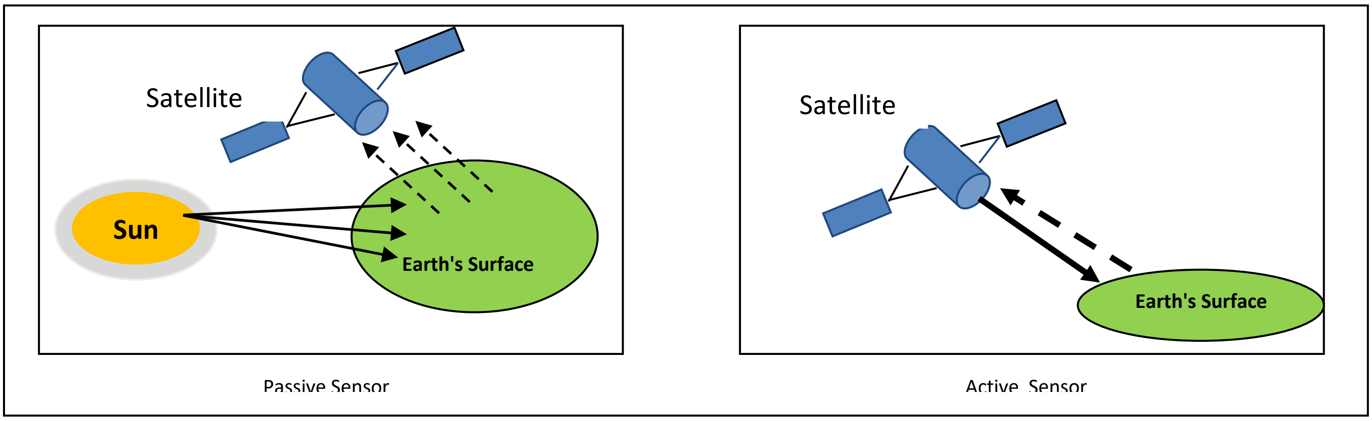
Abstract: Satellite imagery has been transforming our understanding and prediction of global economic activity with evolution in hardware and in much less cost for rocket launching enabling near real time and high resolution coverage across the earth. It is impractical to analyse petabytes of satellite imagery manually, requiring automated solutions with higher accuracy and prediction speed, essential for latency sensitive industrial applications. Enhancement in recognition model design, training and complexity regularization, along with a Multi-Level Convolutional neural network architecture optimized for satellite imagery, the proposed model has made it possible to deliver fivefold improvement in training time and rapid prediction of 20 classes which is necessary for real-time applications. We have substantiated the proficiency through reviewing algorithmic trading environments and release a proprietary annotated satellite imagery dataset for further research. Satellite imagery play influential roles in disaster response, law enforcement, and environmental monitoring. Demanding automated object identification because of its coverage across the globe's geography. The Multi-Level Convolutional Neural Network, as proposed, underwent training with a training dataset comprising 7000 images (350 images for each of the classes), achieving a training accuracy of 98.14%. Upon evaluation on a separate test dataset consisting of 3000 images (150 images per class), the model demonstrated an overall accuracy of 95%. Moreover, each class is predicted with an accuracy of 99% when tested individually. The whole implementation is carried out in Python using Keras, TensorFlow, and Gocolab with GPU and High RAM.

Keywords: *Multi-level CNN, Satellite imagery, Remote Sensing, Artificial Intelligence, Deep learning, Classification.*

1. Introduction

The technology for remote sensing photos and satellite imagery was widely available throughout the previous ten years. They were widely used to track changes over time and detect urban sprawl (Zhang et al.(2018). Freely available satellite images have poor high spatial and spectral resolution. Hence there is a great need for satellite imaging with high spatial and spectral resolution. Different sensors have been introduced to address the demand of photos with extremely high spatial resolution, such as IKONOSN with 4-m multispectral(MS), Quick bird's 2.6-m multispectral, and panchromatic image of 1-m and 0.6-m (Birk et al.,2003). Images taken by satellites orbiting the planet, whether they are operated by the government or a private company, are referred as Satellite images. Satellite's active and passive sensors depicted in Figure 1, evaluates reflected radiation to collect data about the Earth's atmosphere and surface. Harnessing solar energy, passive sensors take precise measurements of the electromagnetic radiation that has been reflected off the Earth. The visible, infrared, thermal infrared and microwave regions of the spectrum are within which the overwhelming majority of passive systems are operated. As a result, passive sensors in the visible and near-infrared have limitations in tropical and wet regions. Active sensors produce their own energy by emitting and absorbing a signal that escapes after being reflected off of the Earth. Altimeters, LIDAR sensors, and other kinds of radar sensors are the examples of active sensors.

1,2AIIT, Amity University, Noida, India
3School of Information Science & Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, India
**Corresponding author:im.deepikapahuja@gmail.com*



Passive Sensor

Active Sensor

Fig.1. Active and Passive sensors

The majority of active sensors work in the electromagnetic spectrum's microwave range, which can pass across clouds. These sensors are capable of being utilized to measure a variety of things, notably aerosols, the structure of forests, precipitation, wind, and sea surface topography.

At first, satellite imagery was primarily intended for military use. They received a lot of commercialization in 1984 intending to use them for numerous non-military uses. Due to their plethora of applications, remote sensing pictures have captured the interest of numerous researchers in recent times.

Only the visible portion of the electromagnetic spectrum can be seen by humans, although other bands, such as infrared, ultraviolet, and even microwaves, can be captured by light sensors. The three most popular satellite image types or bands are VIS, IR, and WV.

1) Visible Imagery: The most prevalent sort of satellite imagery, visible imagery relies on the idea of reflected sunlight. Since clouds reflect sunlight, it can only be seen during the day. Clouds, the earth, and the water are depicted in this type of photograph as white, grey, or dark. In the winter, it is challenging to differentiate between clouds and snow-covered land using topography features.

2) Infrared imagery: Since this kind of imagery doesn't rely on sunlight to bounce off of clouds, it may be viewed both during the day and at night. Sensors in this case use heat radiation from the earth's surface to determine the presence of clouds. They can easily detect clouds given that they are colder than land and water. Also, they can be used to spot thunderstorms, low clouds, and fog.

3) Water Vapour Imagery: The aforementioned type of imagery utilizes an array of grayscale tones, with warm effective layers delivering dark and cold effective layers, and certain color advancements are employed to draw attention of the features with exceptionally low temperatures. To emphasize warm effective layers and very cold effective layers, orange and red have been encompassed as warm shades, and purple, blue, and green have been used as cool tones.

Resolution, which determines how much detail an image may contain, is the primary characteristic of satellite imagery. The degree of detail that can be incorporated into an image is its resolution. Film images, digital raster pictures, and other sorts of images can all be referred to use this expression. Higher resolution optimizes image detail. The resolution of remote sensing images can be summarized into four distinct groups: spatial, spectral, radiometric, and temporal.

1. Spatial Resolution: The most subtle feature that a satellite sensor can recognize or that may be seen in a satellite image has been defined as the spatial resolution. It can often be expressed as a single value that symbolizes the length of a square's diagonal. A pixel, for instance, can represent a space on the ground that is 250 by 250 meters when the spatial resolution is 250m.

2.Spectral Resolution: The potential of a satellite sensor to monitor distinct electromagnetic spectrum wavelengths is commonly referred to as spectral resolution. Considering that spectral resolution, it has an inverse correlation with the wavelength range for a given channel or band, spectral resolution increases as the wavelength range is shrunk. Higher spectral resolution maximizes the ability to exploit differences in spectral fingerprints.

3.Temporal Resolution: The period that elapses between images have been referred to as temporal resolution. Since the dawn of the space age, satellites' capacity utilization to frequently provide images of the same geographic area has significantly improved. Monitoring land use and urban growth is greatly aided by it.

4. Radiometric resolution: It can distinguish between two diffuse targets based on comparable backscatter indices. It counts the three different types of noises that reap map speckles:

a. Additive noise, such as thermal noise, is the noise that is not related to the strength of the received signal.

b. Multiplicative Noise - This kind of noise consists of target fluctuation, frequency-dependent phase noise, and noise that is also dependent on the strength of the received signal.

c. Encoding and quantization, third Noise—a noise that neither has an additive nor a multiplicative component.

Image classification is a crucial activity that seeks to comprehend an image as a whole. The goal is to give the image a label so that it can be categorized. Image categorization is the term used to describe images with only one object visible and under examination. Object detection refers to the study of more realistic scenarios when multiple objects may be visible in a photograph and involves both classification and localization tasks. Satellite image classification is performed exclusively at the pixel level. In several fields, especially in change detection and urban areas, the categorization of pictures serves as essential and is used to improve object detection accuracy. Different applications that employ classification are shown in Figure 2.

For the sorting of many terrestrial activities, an array of classifiers, spanning supervised, unsupervised, hard, soft, and object-oriented classifiers, are readily accessible in satellite imagery. The training process for supervised and unsupervised methodologies differentiates; image analysts opt for which pixels to train, and features are subsequently recognized using the trained pixels. In contrast to supervised classification, no picture analyst is involved in choosing the training pixels. Unknown pixels are examined and grouped into classes in an unsupervised classification.

In the Thematic map's hard categorization technique, each pixel is exclusively assigned to one class(Campell,1996). When mixed pixels are allocated to one class during the usage of a hard classification technique, information may be lost, even though remote sensory images contain mixed pixels due to the real landscape present at the time of capture(Schowengerdt,1997). Estimating the fraction of classes within each pixel is done using soft classification algorithms. Soft classification approaches are used in algorithms to provide more informative images than conventional hard classification techniques. The effectiveness and accuracy of classification results are greatly influenced by the choice of classifier. A system of classification must be extensive, thorough, and standalone(Landgrebe,2003). Recent remote sensing photos are being used in numerous different contexts (Rai et al.,2020),

Earlier numerous classification techniques include Naive-Bayesian, Decision Tree, K-means, Support Vector Machine (SVM), and FCM have been applied for land cover classification but with some limitations like FCM, a traditional clustering algorithm suffers from noise due to non-consideration of spatial information (Bezdek et al.,1984). Over the past years, deep learning architectures/models have been considered the best choice for classification. Deep learning is one of the evolving branches of machine learning which has originated from the neural network that utilizes layers of algorithms for processing data to have abstraction and employed to address a range of applications including pattern recognition, face recognition, feature extraction, object detection, etc. some of the most widely used and popular deep learning models are Restricted Boltzmann machines(RBMs), Deep Auto-encoder(DA), Deep neural network (DNN), Recurrent Neural Network, and Convolutional Neural Network (CNN).

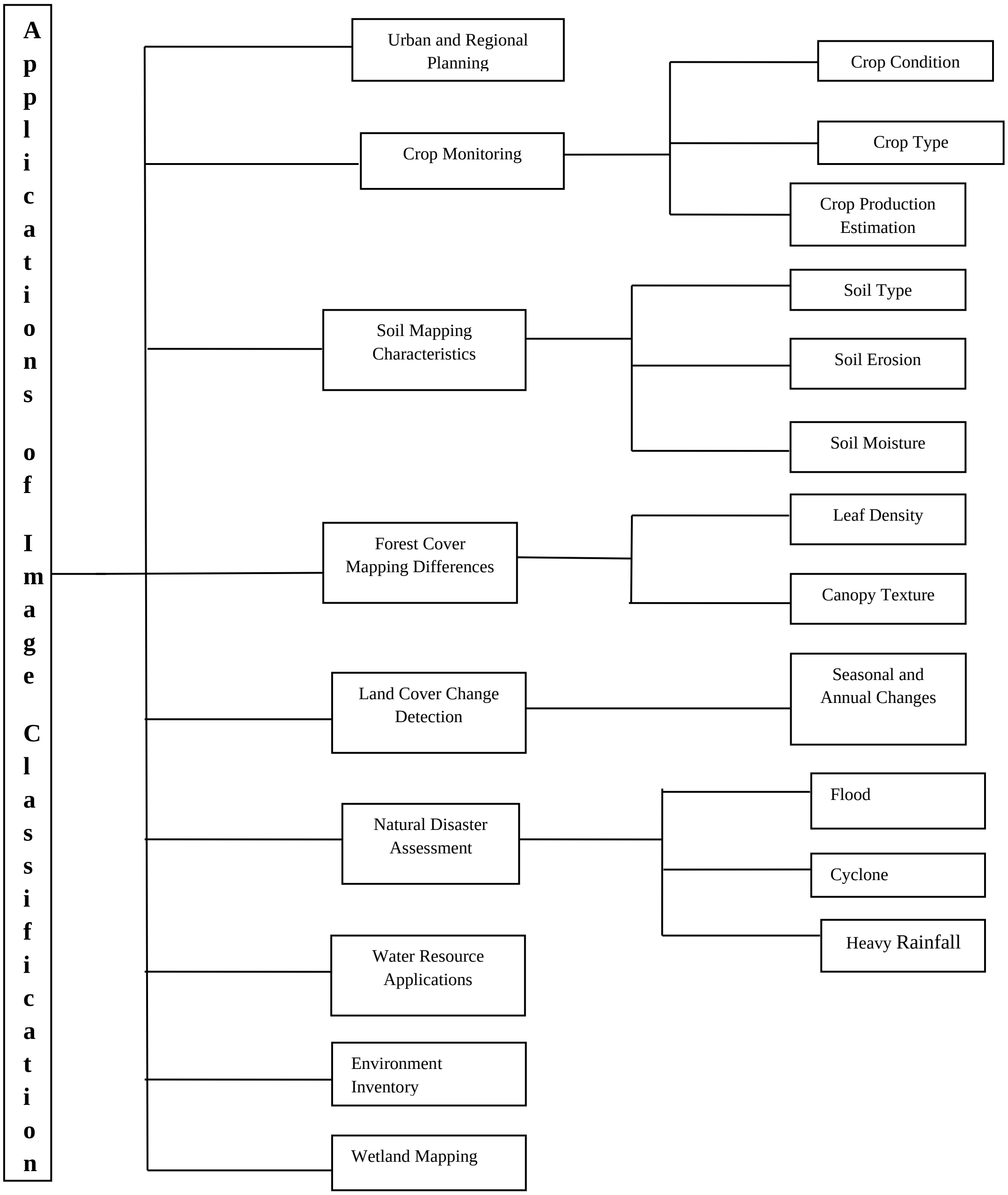

Applications of Image Classification
Urban and Regional Planning
Crop Monitoring
Crop Condition
Crop Type
Crop Production Estimation
Soil Mapping Characteristics
Soil Type
Soil Erosion
Soil Moisture
Forest Cover Mapping Differences
Leaf Density
Canopy Texture
Land Cover Change Detection
Seasonal and Annual Changes
Natural Disaster Assessment
Flood
Cyclone
Heavy Rainfall
Water Resource Applications
Environment Inventory
Wetland Mapping

Fig. 2. Application of Image Classification

Generative and unsupervised learning model architectures are also employed for solving various real-life applications like Generative Adversarial Network(GAN), and Variation Auto-Encoder(VAE) (Pouyanfar et al.,2018). In the past few years, extensive use of deep learning techniques has proven to be an optimiser for both supervised and unsupervised learning and also solved the issue of over fitting. The table1 shows the evolution of deep learning.

Year	Model
1943	Computer Model Based on Neural Network of the Human Brain
1950	Thinking Machine
1952	Hodgkin Huxley Model of the brain

1960	Shallow Neural Network
1960-1970	Back Propagation Model
1974-1980	First Artificial Intelligence Winter
1980's	Emergence of Convolution
1987-1993	Second Artificial Winter
1990's	Unsupervised Deep Learning
1990-2000	Supervised Deep Learning
2000-2010	Vanishing Gradient Problem, ImageNet
2011-2014	Alex-Net, Cat Experiment and Generative Adversarial Network
2014 to present	Modern Deep-learning methods

Table 1. Evolution of Deep Learning

Identifying objects, facilities, and alarming activities in satellite imagery from space are the intricate challenges for spotting any unlawful activities, fishing vessels, unauthorized border crossings, agricultural surveillance, and a shift in the geographical marking of the land from wilderness to urban planning. As a by-product of global growth, the risk to humans at borders has increased, making analysts scarce and automation indisputable. That's why leveraging conventional protocols for object designation and categorization to solve the issue would be erroneous and misleading. Automatic entity identification and labeling are possible with deep learning. Preprocessing satellite imagery for use as an input in contrast to conventional neural networks is the primary hurdle when implementing deep learning techniques.

CNN needs relatively modest fixed-size images to enable them to operate at the justified processing times. For instance, Inception (Szegedy et al.,2014,2015) facilitates images as large as 255x255 pixels, but ResNet and DenseNet only function with 224x224 pixel images. Cropping and warping the images to the imperative size is a customary procedure in deep learning. These procedures preserve significant visual elements for typical images. However, this is not legitimate for satellite images because structures and objects could potentially be much larger than they tend to be in ordinary photographs. Small details are lost when large-size images are downsized to 224x224 or 255x255 pixels. The viewing circumstances for items in satellite photographs taken by satellite are more severe than those in regular photos. The above explained challenges are faced by numerous researchers to perform image classification and object detection in satellite images.

This paper proposes a novel approach to perform satellite image classification using a ***mutli-level CNN architecture***. Earlier literature depicts the dearth of substantial datasets of labeled imagery for the instructive process of algorithm could be the main contributing factor to the restricted performance of utilizing deep learning techniques on satellite imagery. For deep learning algorithms to be capable of identifying objects in photos, they typically need thousands of labeled images for every classification. The proposed framework is applied on ***customized aggregated dataset of 10000 satellite images of 20 classes***. The architecture of prospective work utilizes a multi-level CNN architecture and executing the planned work on the satellite imagery dataset leads to **99% accurate predictions** for each class.

Further, the paper is structured in the following way: Section 2 provides the literature review, Section 3 explains the proposed methodology, Section 4 shows the customized dataset and experimental investigations, results with performance metrics are demonstrated in Section 5. The conclusion and future scope is discussed in Section 6.

2. Literature Review

Deep learning architectures support automatic feature extraction thus they can recognize concealed data patterns without requiring an explicit feature extraction mechanism, which solves the problem faced by traditional techniques. It is one of the machine learning models that has multiple processing layers that produce data at different levels of cogitation (LeCun et al., 2015). By fusing enormous neural network models, known as convolutional neural networks, it has achieved astounding performance in object detection and classification. Initially, CNNs with less than 10 layers and LeNet with 5 layers (LeCun et al.,1989) have successfully performed handwritten zip code recognition. Since 2012, CNN-based algorithms have led ImageNet's annual large-scale visual recognition contest for identifying and categorizing objects in pictures. The Table 2 shows the implementation of CNN-based algorithms by various authors to reduce the error rate for the imagined dataset. The use of CNN-based goods and services has already been adopted by the main technological firms, including Google, Microsoft, and Facebook, this success has sparked a revolution in image interpretation.

Authors	Model	No. of Layers	Dataset	Top-5 error rate
Krizhevsky et al.,2012	Alex Net	8	ILSRVC-2010	18.20%
Szegedy et al.,2014	Google Inception-Google Net	22	ILSRVC-2014	6.67%
Simonyan, et al.,2015	VGG (2 nets) VGG(1 net)	16	Multi-crop	6.8% 7.1 %
He et al.,2015	ResNet	152	ILSRVC-2015	3.57%
Huang,2017	DenseNet	161	CIFAR C10+ C100+	3.46% 3.74%

Table 2. Use of CNN based Algorithms for classification

In image classification, there are various driving forces. Researchers have applied distinct machine learning models for classifying images automatically of distinguished datasets. The Table 3 shows a comparison between the use of machine learning models on various datasets and depicts accuracy also.

Authors	Machine learning Model used	Dataset	Task	Applications	Accuracy
Wang et al.,2019	Alexnet	Alcoholism Dataset	Image Classification	Identifying Alcoholism	97%
Nascimento et al.,2018	CNN	RADIO SIGNALS	Image Classification	Classification of Radio Access Technologies over Signal Noise Ratio	100%
Patil et al., 2018	CNN	Military and Space Images	Object Detection	Detecting objects of military and space	95.6%
Yang et al.,(2018)	Deep transfer learning	Military images	Object detection and Image Classification	Discriminate military objects and ordinary objects	95% approx.
Remjee et al.,2019	Convolution Long Short-term Deep neural Networks (CLDNN), and a deep Residual Network(ResNet)	Signals	Image Classification	Automatic Modulation Classification	90%
Wu et al.,2016	constrained deep transfer feature learning	cross-view facial expressions in the NIVE database and MAHNOB laughter database	Object Detection	thermal eye detection	81.6 % and 87.2%respectively
Singh et al.,2018	VGG16	Yelp Restaurants Datasets from Kaggle	Image Classification	Classifying yelp restaurants	60%
Smith et al., 2018	VGG19	HUMANS DATASET	Image Classification	Classifying GENDER based on estimation of Age	98%
Hussain et	INCEPTION V3	CIFAR-10	Image	Image Classification	70%

al.,2018			Classification		
Kwan et al.,1993	XCEPTION	Microsoft Malware Dataset	Object detection and Image Classification	Detecting Pulse Radar	99%
Talo et al., 2018	ResNet	MRI Images	Image Classification	Detecting Brain abnormalities	100%
Almisreb et al., 2018	AlexNet	Ear Image dataset	Object detection	Ear Recognition	100%
Signoroni et al., 2019	ResNet	Military objects	Image Classification	Image Classification	90%
Pahuja and Jain, 2022	Multi-level CNN	Military Weapons datasets from Kaggle	Image Classification	Classifying images into 5 military weapons	97.7%
Pritt and Chern,2020	Deep Learning System- Ensemble of CNN's	Functional Map of the World (fMow)	Image Classification	Classifying satellite imagery into 15 classes accurately	83%
Zhao et al., 2015	Deep Learning	Pavia datasets	Image Classification	Classifying hyperspectral satellite imagery into 9 classes	Upto 100%
Liang et al., 2016	Fine-tuned CNN	Aerial images	Image Classification	classify the images from the land use land cover (LULC) image dataset	93-96%
Rai et al.,2020	CNN	LANDsAT8 OLI satellite image	Image Classification	MULTISPECTRAL Satellite images into 5 classes	94.5%

Table 3. Machine Learning models for Image classification

3. Proposed Methodology

For the recognition of objects in the given image, the proposed methodology is Multi-Level Convolution Neural Nets which is depicted in the Figure 3.

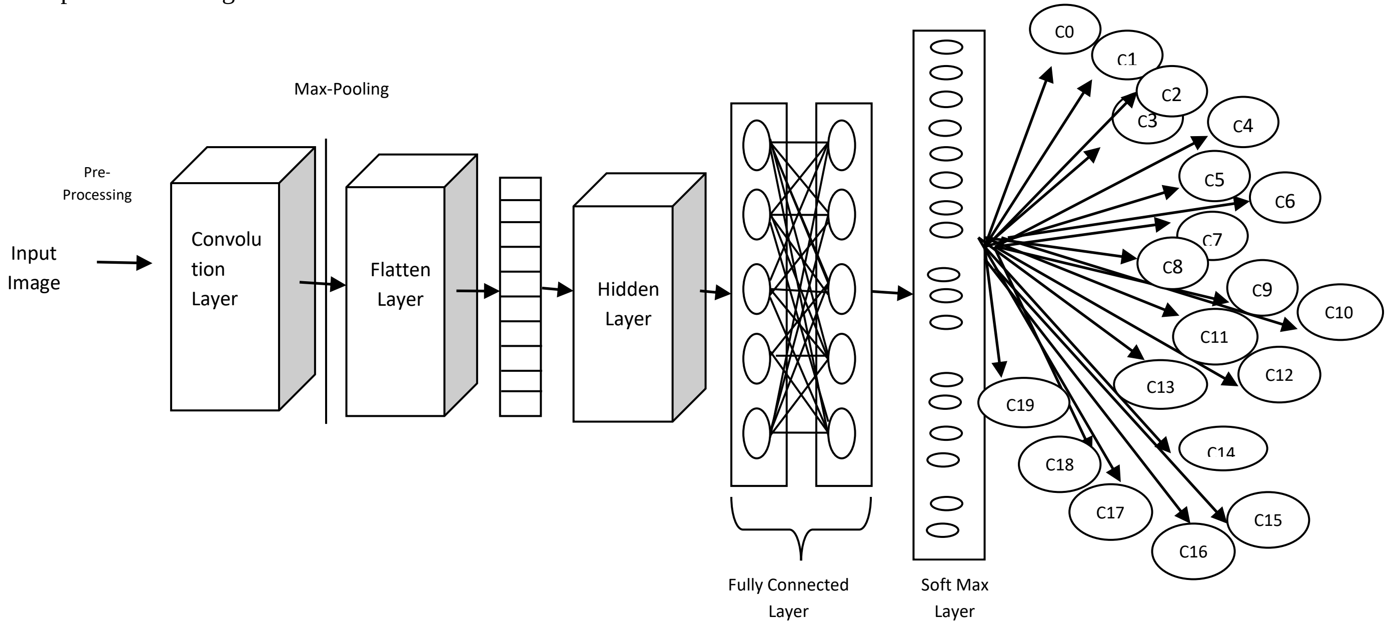


Fig. 3. Architecture of Proposed Methodology

Twenty distinct objects namely are given as an input to the proposed architecture. To improve the contrast and eliminate noise from the input photos, preprocessing is performed on images. For providing the input to multi-level CNNs, images go through the preprocessing depicted in the Figure 4.

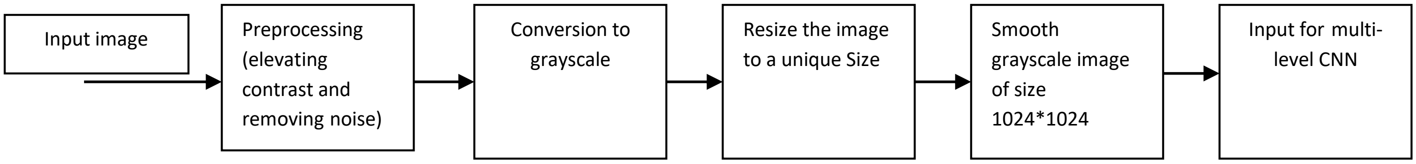

Input image
Preprocessing (elevating contrast and removing noise)
Conversion to grayscale
Resize the image to a unique Size
Smooth grayscale image of size 1024*1024
Input for multi-level CNN

Fig. 4. Input Preprocessing

3.1 Convolution Layer

The core component of CNN is the convolution layer. In this kernels are applied to an input image to extract features. It involves multiplying the set of weights by input which is accomplished by multiplication of a 2D array of weights with a 2D array of input data. Since the input will consist of a 3D matrix of pixels, assuming it is a colored image, it will have 3 dimensions' height, width, and depth corresponding to RGB image. Additionally, it has a feature detector known as kernel or filter which performs the convolution process by moving across the image's respective fields in a direction from left to right and then top to bottom to verify the existence of the feature.

The purpose of using the same kernel across a picture has a significant role as it can detect a specific type of feature in a given image. Applying kernel repeatedly in a 2D array of values produces filtered input. Convolution operation in a convolution layer yields a feature map. To identify the image, a single feature is not sufficient, many more features are needed. To achieve the same, several kernels are applied to an input image as a result distinct feature map is provided by each kernel. Several convolution filters can be applied in a pattern by adding deeper layers to have feature maps.

To design a network in CNN four hyper parameters, need to be adjusted and they are the size of a Kernel, kernel count, padding, and stride. Commonly used kernel sizes are 3x3, 5x5, 7x7, and 9x9 are similarly employed in accordance with the application input image 2D kernels are adequate for black and white but for RGB image 3D kernels are essential since the input's depth at a given layer corresponds to the kernel's depth at that layer. Kernel range varies in terms of power of 2 from 32 to 1024. To extract more features its mandate is to increase the number of kernels. Stride: the kernel's movement on the input image is indicated by the stride. By default, value of stride is 1, which denotes shifting just one cell to the right and bottom. The convolution operation is applied to reduce the size of an image but with a limitation of applying the number of times. The result of convolution operations needs padding to the same size, to overcome this limitation. As the number of convolution layers' increases, so does the number of weights. To reduce the number of parameters, its sharing is conceptualized so that weights can be shared among all neurons in one feature map which is best to generalize it.The working of Convolution layer for proposed model with its feature maps is depicted in Figure 5.

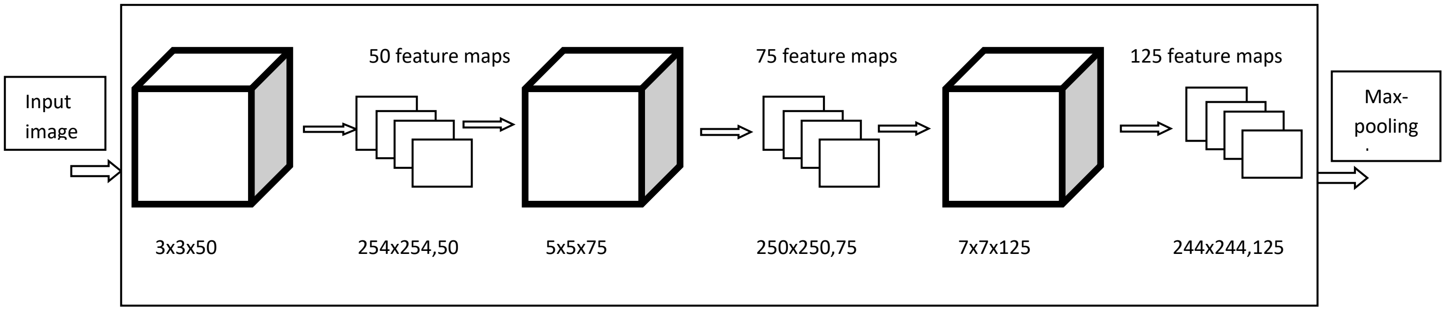


Fig. 5. Working of Convolution Layer

Convolution with several photos was accomplished in this study by utilizing the Conv2D function. The preprocessed input image is of size 256x256 to the convolution layer. For the first layer(Conv2D) the input is of size 256x256 with a kernel size of 3x3 and stride 1 resulting in a feature map of size[$\frac{256-3}{1}+1$]i.e. [254x254,50] which is considered as input to the next layer(Conv2D_1).The kernel in (Conv2D_1) is 5x5 so the resulting feature map of this layer is of size [$\frac{254-5}{1}+1$] i.e. [250x250,75] which becomes input to the next layer(Conv2D_2). As the third layer Conv2D_2 the kernel size is 7x7 yielding a feature map of size[$\frac{250-7}{1}+1$] i.e. [244x244,125]. So for the next layer, a final input is [244x244,125]. A total of 256,684,595 parameters are tuned in this multi-level CNN architecture. The dropout for the second and third layers is 0.25%.

3.2 Maxpooling Layer

This layer usually comes after a Convolutional Layer. The main goal of pooling is to reduce the convolved feature map's size to save computing expenses. This is accomplished by reducing the connections between layers and performing independent operations on every feature map. There are various kinds of pooling processes, depending on the technique employed. It essentially provides an overview of the features produced by a convolution layer. The largest element in max pooling is extracted from the feature map. The average of the components in a predetermined-sized Image segment is determined by average pooling. In sum pooling, the total sum of the items in the predefined section is calculated. Typically, the convolutional Layer and the fully connected Layer are connected by the Pooling Layer.

In this architecture max-pooling function is also utilized to provide location-invariant feature detection to reduce dimension and solve the problem of overfitting. The result of the max-pooling layer at Conv2D_1 and Conv2D_2 yields 128x128x75 (1228800) and 64x64x125 (512000) feature maps respectively. The result of the last layer Conv2D_2 i.e. 512000 feature maps are passed as an input to flatten the layer.

3.3 Flatten Layer:

Data is transformed into a 1-dimensional array by the flattening layer so that it can be fed into the subsequent layer. A single, lengthy feature vector was created by flattening the convolution layer's output. It is also known as the completely linked layer and is connected to the final classification model.

The feature map, or collection of features expressed as a 2-D array from the output of the convolution layer above must be transformed into a 1-D array for additional processing. It flattened the results of the Conv2D_2 layer from 64x64x125 into a single array of 1x512000 vectors shown in Figure 6.

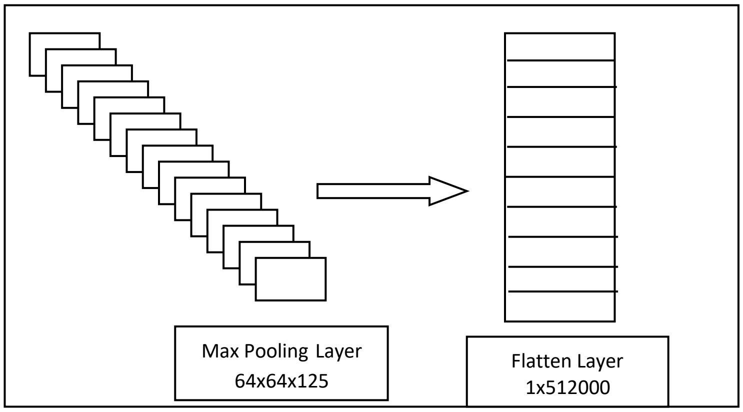


Fig. 6. Max-pooling Layer

3.4 Hidden Layer

A layer of neurons that is neither the input layer nor the output layer is referred to as a hidden layer in the context of artificial neural networks. Neural networks may learn complex data representations because of hidden layers, which give them their "deep" characteristic. Deep learning models rely on them as their computational workhorse, which enables neural networks to extract patterns and approximate functions from incoming data. Hidden layers' main responsibility is to convert inputs into something that the output layer can utilize. They accomplish this by giving the inputs weights and putting them through an activation function. Through this approach, non-linear correlations between the input and output data can be learned by the network. Every neuron in a hidden layer takes in information from every other neuron in the layer above, multiplies it by its weights, adds a bias term, and then runs the outcome past an activation function.

The dense function utilized in this architecture was implied to create a fully linked CNN for the hidden layer and a dense function is utilized to create a fully linked CNN for the hidden layer. To prevent overfitting, two hidden layers with ReLU as an activation function are constructed with 500 and 250 nodes, respectively.

3.5 Fully connected Layer

A collection of interdependent non-linear functions makes up a neural network. A single neuron (or perceptron) performs each distinct function. The input vector is subjected to a linear transformation in fully connected layers using a weights matrix in the neural network. In this layer, each neuron's output from the hidden layer is utilized. To increase Neural networks' learning ReLU, a non-linear activation function was applied to provide the network with non-linear characteristics, that enable it to intricate patterns increasingly.

Once the image dimension is reduced, the next layer is a fully connected convolutional layer with 500 filters each of size 3×3. In this layer, each of the 500 units in this layer will be connected to the 1250 (7x7x250) units from the previous layers. The other FC layers are also fully connected layers with 250 and 125 units respectively.

Overfitting in the training dataset may result from connecting every feature to the FC layer. It is the phenomenon where a certain model performs poorly when applied to new data since it performed so well on the training set.

In order to solve this issue, a dropout layer is used, in which a small number of neurons are removed from the neural network during training, while reducing the size of the model. On passing a dropout of 0.4, 40 % of the nodes are dropped out randomly from the neural network. This model is made simple by using regularization techniques like dropout of 20%, 40%, and 30% dropout in layers respectively.

3.6 Output Layer

The final layer in this architecture is the output layer. Depending on the number of classes in the dataset, the final layer will be a Softmax function in the output layer that has the potential to predict "20" distinct classes.

4. Experimental Investigations

The proposed model's potential is illustrated with a dataset of 10000 satellite images. The proposed model using multi-level CNN architecture is done in Python due to the wide availability of libraries and frameworks for deep learning. The experiment is performed using Google Colab. Colab is a hosted Jupyter Notebook service that offers free access for computation resources, such as GPUs and TPUs and doesn't require any setup.

4.1 Dataset

For validating the proposed multi-level CNN-based model, the datasets used in the experiments are self-built set with 20 distinct satellite images (agricultural, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, tennis court). The created dataset contains 10000 images collected from the GitHub repository and the internet. This dataset contains 500 images for each category of distinct objects. Based on the category, these images are labeled to their respective classes. The details of satellite images in the collected dataset are given in the Table 4. The sample images of 20 classes are shown in the Figure 7.

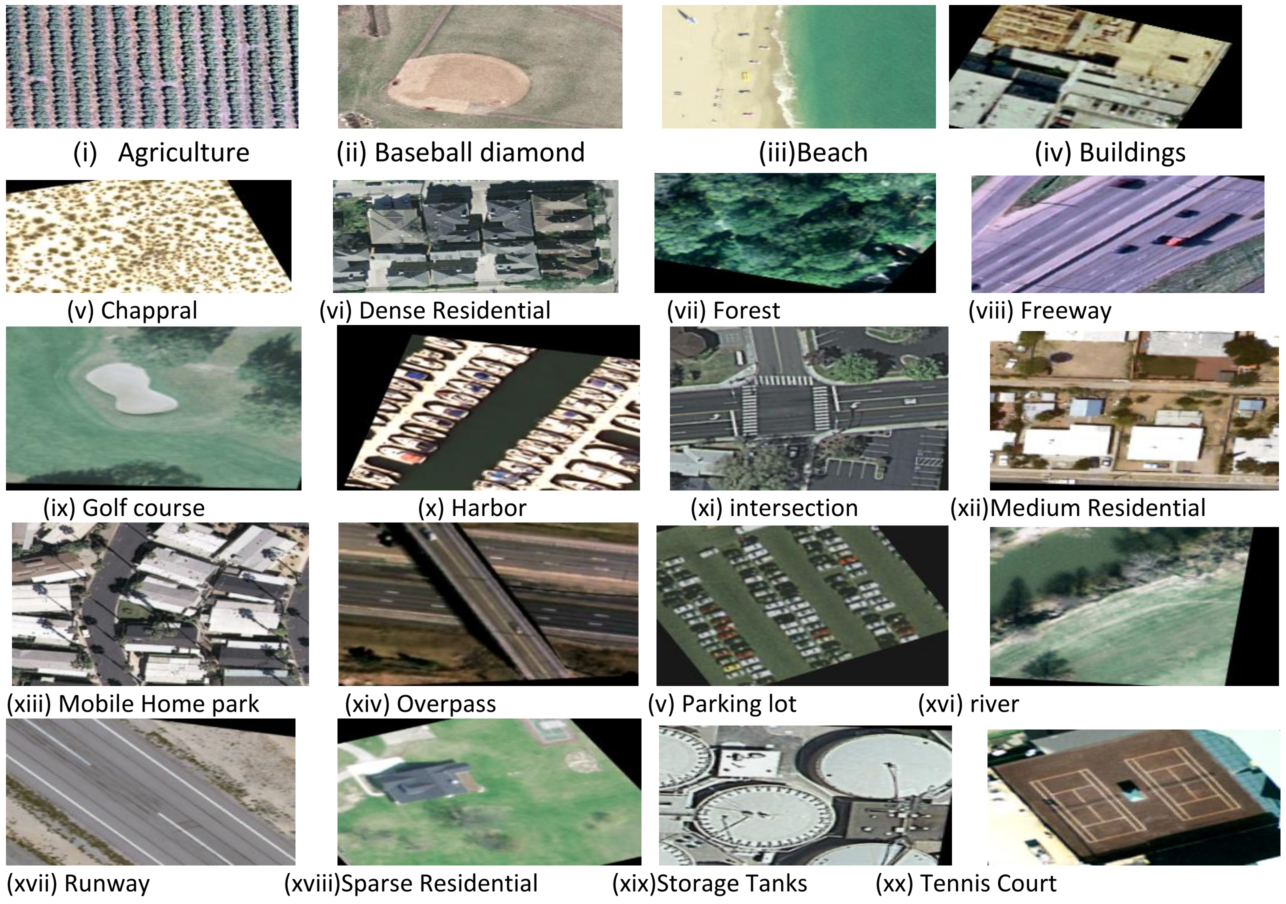
Image category	Class Label	No. of Images
Agriculture	C0	500
Base Ball Diamond	C1	500
Beach	C2	500
Buildings	C3	500
Chaparral	C4	500
Dense Residential	C5	500

Forest	C6	500
Freeway	C7	500
Golf Course	C8	500
Harbor	C9	500
Intersection	C10	500
Medium Residential	C11	500
Mobile Home Park	C12	500
Overpass	C13	500
Parking Lot	C14	500
River	C15	500
Runway	C16	500
Sparse Residential	C17	500
Storage Tanks	C18	500
Tennis Court	C19	500
	Total	10000

Table 4. Distribution of Sample images

5. Performance Metrics

The proposed system is experimented with the train-test split method. Splitting the dataset is essential for an unbiased evaluation of prediction performance. The dataset is randomly divided into three subsets: training, testing and validation set. Among the dataset of 10000 images, its split into training and testing data, of 7000 images, 3000 images respectively.



(i) Agriculture (ii) Baseball diamond (iii)Beach (iv) Buildings

(v) Chappral (vi) Dense Residential (vii) Forest (viii) Freeway

(ix) Golf course (x) Harbor (xi) intersection (xii)Medium Residential

(xiii) Mobile Home park (xiv) Overpass (v) Parking lot (xvi) river

(xvii) Runway (xviii)Sparse Residential (xix)Storage Tanks (xx) Tennis Court

Fig.7. Image dataset illustration

To construct any classification model, precision in predicting the correct class is highly significant but not only precision some additional measures also become an integral factor for estimating the performance of the proposed model.

The performance metrics of a model are based on a confusion matrix having different measures to quantify the quality of the model which are accuracy, precision, recall, and F1-score. For classifier models, accuracy is a crucial parameter. For both binary and multi-class classification, it is straightforward to implement and easy to understand. Accuracy represents the percentage of real outcomes across all tested records. It gives better results when a classification model is built using balanced datasets.

Precision indicates the percentage of genuine positives in anticipated positives. Recall is another crucial metric that provides additional information if all potential benefits are captured. Recall shows that a proper prediction of the percentage of all positive samples is made. Recall is 1 if all positive samples are predicted as positive. The F-1 score is a new metric that can be created by combining these two metrics if the ideal combination of recall and precision is needed. The F-1 score is the harmonic mean of these two, precision and recall lie between 0 and 1. A high F1 score denotes both high recall and precision. It achieves good results on imbalanced classification issues and delivers a good balance between precision and recall.

The formulas to evaluate all these measures are depicted in equations 1 to 4.

$$\boldsymbol{Accurcay} = \frac{\mathbf{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \qquad (1)$$

$$\boldsymbol{Precision} = \frac{\mathbf{TP}}{\boldsymbol{TP+FP}} \qquad (2)$$

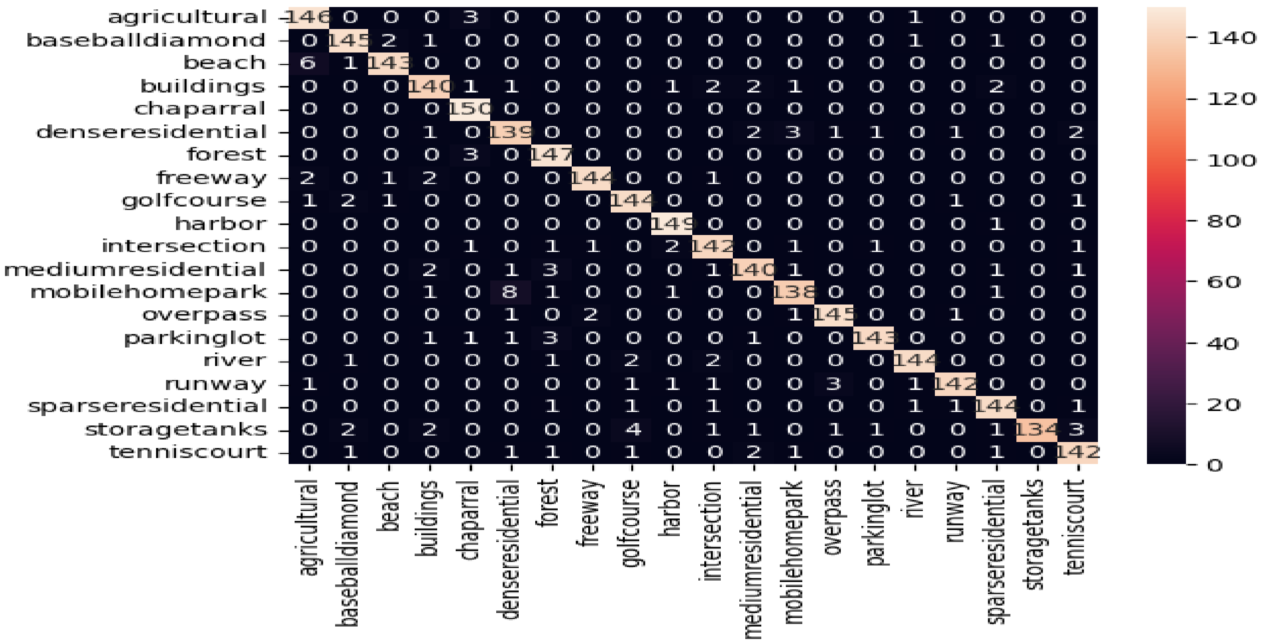
$$\boldsymbol{Recall} = \frac{\mathbf{TP}}{\mathbf{TP+FN}} \qquad (3)$$

$$\boldsymbol{F1 - Score} = \frac{\mathbf{2xPrediction\ x\ Recall}}{\mathbf{Prediction+Recall}} \qquad (4)$$

It is not possible to achieve 100% accuracy in a classification test when building a model with precision and recall of 1, which yields an F1-score of 1. Therefore, a greater recall value and higher precision should be expected from the classification model.

5.1 Results

The experiment is conducted with the proposed Multi-level CNN architecture using sequential model. Model is compiled with 'sparse_categorical_crossentropy' in loss with SGD optimizer having learning rate of 0.01. Model is trained in batch size of 16 with 20 epochs and validation split of 1% and thus model has achieved an accuracy of 98.14%. The obtained confusion matrix is obtained using seaborn library in python shown in Figure 8.



	agricultural	baseballdiamond	beach	buildings	chaparral	denseresidential	forest	freeway	golfcourse	harbor	intersection	mediumresidential	mobilehomepark	overpass	parkinglot	river	runway	sparseresidential	storagetanks	tenniscourt
agricultural	146	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
baseballdiamond	0	145	2	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
beach	6	1	143	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
buildings	0	0	0	140	1	1	0	0	0	1	2	2	1	0	0	0	0	2	0	0
chaparral	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
denseresidential	0	0	0	1	0	139	0	0	0	0	0	2	3	1	1	0	1	0	0	2
forest	0	0	0	0	3	0	147	0	0	0	0	0	0	0	0	0	0	0	0	0
freeway	2	0	1	2	0	0	0	144	0	0	1	0	0	0	0	0	0	0	0	0
golfcourse	1	2	1	0	0	0	0	0	144	0	0	0	0	0	0	0	1	0	0	1
harbor	0	0	0	0	0	0	0	0	0	149	0	0	0	0	0	0	0	1	0	0
intersection	0	0	0	0	1	0	1	1	0	2	142	0	1	0	1	0	0	0	0	1
mediumresidential	0	0	0	2	0	1	3	0	0	0	1	140	1	0	0	0	0	1	0	1
mobilehomepark	0	0	0	1	0	8	1	0	0	1	0	0	138	0	0	0	0	1	0	0
overpass	0	0	0	0	0	1	0	2	0	0	0	0	1	145	0	0	1	0	0	0
parkinglot	0	0	0	1	1	1	3	0	0	0	0	1	0	0	143	0	0	0	0	0
river	0	1	0	0	0	0	1	0	2	0	2	0	0	0	0	144	0	0	0	0
runway	1	0	0	0	0	0	0	0	1	1	1	0	0	3	0	1	142	0	0	0
sparseresidential	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	1	1	144	0	1
storagetanks	0	2	0	2	0	0	0	0	4	0	1	1	0	1	1	0	0	1	134	3
tenniscourt	0	1	0	0	0	1	1	0	1	0	0	2	1	0	0	0	0	1	0	142

Fig.8. Confusion Matrix

Among the 150 images of each class agricultural, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks and tennis court the correctly identified classes are 146, 145, 143,140, 150,139,147,144,149,142,140,138,145,143,142,144,134 and 142 respectively.

Table 5 shows the True-Positives (TP), True-Negatives (TN), False Positives (FP) and False Negative (FN) for each class which is estimated on the basis of the confusion matrix shown in figure 8. And also, the performance measures of proposed model: accuracy, precision, recall and f-1score are evaluated for each class separately. The overall accuracy is calculated as the number of right predictions divided by the total number of predictions made across all classes. The proposed model achieved an overall accuracy of 95% on test data of 3000 images.

S. No.	Image category	TP	TN	FP	FN	Accuracy	Precision	Recall	F-1 Score
1	Agriculture	146	2840	10	4	0.99	0.94	0.95	0.95
2	Base Ball Diamond	145	2843	7	5	0.99	0.95	0.96	0.96
3	Beach	143	2846	4	7	0.99	0.97	0.96	0.96
4	Buildings	140	2840	10	10	0.99	0.93	0.93	0.93
5	Chaparral	150	2841	9	0	0.99	0.94	0.97	0.97
6	Dense Residential	139	2837	13	11	0.99	0.91	0.92	0.92
7	Forest	147	2839	11	3	0.99	0.93	0.95	0.95
8	Freeway	144	2847	3	6	0.99	0.98	0.97	0.97
9	Golf Course	144	2841	9	6	0.99	0.94	0.95	0.95
10	Harbor	149	2845	5	1	0.99	0.97	0.98	0.98
11	Intersection	142	2841	9	8	0.99	0.94	0.94	0.94
12	Medium Residential	140	2842	8	10	0.99	0.95	0.94	0.94
13	Mobile Home Park	138	2842	8	12	0.99	0.95	0.93	0.93
14	Overpass	145	2845	5	5	0.99	0.97	0.97	0.97
15	Parking Lot	143	2847	3	7	0.99	0.98	0.97	0.97
16	River	144	2846	4	6	0.99	0.97	0.97	0.97
17	Runway	142	2846	4	8	0.996	0.97	0.96	0.96
18	Sparse Residential	144	2842	8	6	0.99	0.95	0.95	0.95
19	Storage Tanks	134	2850	0	16	0.99	1.00	0.94	0.94
20	Tennis Court	142	2841	9	8	0.99	0.94	0.94	0.94

Table 5. Performance Metrics

6. Conclusion

A deep learning model has been developed utilizing a dataset of 10,000 photos divided into 20 groups, with 500 images in each class, to provide a novel method for classifying satellite images through Multi-level CNN architecture. It has achieved an overall accuracy of 95%. It has classified all 20 classes with an accuracy of 99%. With Reference to future work, it can be combined with the detection system for large datasets of satellite images, which can be deployed on real-time satellite imagery training and predicting the alarming situations in defense and in military surveillance. With object detection module, it can be utilized in numerous applications of remote sensing like land cover detection, urban planning, forest cover detection, water pollution, Crop Monitoring etc.

Declaration Of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment
This research would like to thank the support of Amity Institute of Information Technology, Amity University, India.

References

[1] Almisreb, A.A., Jamil, N. and Din, N.Md., 2018. Utilizing AlexNet deep transfer learning for ear recognition. Fourth International Conference on Information Retrieval and Knowledge Management (CAMP). doi: 10.1109/INFRKM.2018.8464769.

[2] Birk, R., Stanley, T., Snyder, G., Henning, T., Fladeland, M. and Policelli, F., 2003.Government programs for research and operational uses of commercial remote sensing data. *Remote Sensing of Environ.*, vol. 88, no.1-2, pp.3-16. doi:10.1016/j.rse.2003.07.007.

[3] Bezdek,C., Ehrllich,R. and Full,W., 2018. FCM:Fuzzy c-means clustering Algorithm. *Computers and Geosciences*, vol.2,pp.191-203.

[4] Campell, J., 1996. *Introduction to Remote Sensing*, 2nd edition, London, U.K.: Taylor & Francis.

[5] He, K., Zhang, X., Ren, S. and Sun, J.,2015. Deep residual learning for image recognition, *Computer Vision and Pattern Recoginition*(*CVPR*).doi: https://doi.org/10.48550/arXiv.1512.03385.

[6] Huang, G., Liu, Z., Maaten, L. and Weinberger,K.Q., 2017. Dense connected convolutional neural networks, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). doi: https://doi.org/10.48550/arXiv.1608.06993*

[7] Hussain, M., Bird, J.J. and Faria, D.R.,2018. A study on CNN transfer learning for image classification. *18th Annual UK Workshop on Computational Intelligence (UKCI), Nottingham*.

[8] Krizhevsky,A., Sutskever,I., and Hinton,G., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012).*

[9] Kwan, H.K. and Lee, C.K., 1993. A neural network approach to pulse radar detection. *IEEE Trans. Aerosp. Electron. Syst.* vol. 29(1), pp.9–21. doi: https://doi.org/10.1109/7.249109

[10] LeCun, Y., Bengio, Y., and Hinton, G., 2015.Deep learning. *Nature*, vol. 521, pp. 436-444. doi: https://doi.org/10.1038/nature14539.

[11] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D.,1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, vol. 1(4), pp.541-551. doi: 10.1162/neco.1989.1.4.541.

[12] Landgrebe, D.A., 2003.Signal Theory Methods in Multispectral Remote Sensing. NJ, Hoboken: Wiley.

[13] Liang, Y., Monteiro, S and Saber, E., 2016.Transfer learning for highresolution aerial image classification. *IEEE Workshop Applied Imagery Pattern Recognition (AIPR)*, doi: arXiv:1510.00098v2

[14] Nascimento, I., Flavio,M., Marcus, D., Andrey, S. and Aldebaro, K.,(2018). Deep learning in RAT and modulation classification with a new radio signals dataset, Simpósio brasileiro de telecomunicações e processamento de sinais - sbrt, 16–19 de setembro de. doi:10.14209/sbrt.2018.144.

[15] Patil et al.,2018. Object detection in military and space image by deep learning with a convolutional neural network", *Int. J. Comput. Sci. Eng.*, vol. 6(11), pp.363–368.

[16] Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P. and Iyengar, S. S.,2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* Vol. 51(5): 92, pp. 1–36. doi: https://doi.org/10.1145/3234150.

[17] Pahuja, D. and Jain, S.,2022 An Automatic Detection of Military Weapons Using Multi-Level CNN Architecture.*10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India*, pp. 1-4, doi: 10.1109/ICRITO56286.2022.9964878.

[18] Pritt, M. and Chern, G.,2020. Satellite Image Classification with Deep Learning. *Computer Vision*

and Pattern Recognition(CPVR), doi: https://doi.org/10.48550/arXiv.2010.06497.

[19] Rai, A.K., Mandal, N., Singh, A. and Singh, K.K., 2020. Landsat 8 OLI Satellite Image Classification using Convolutional Neural Network. *Procedia Computer Science vol. 167*, pp.987-993. doi: https://doi.org/10.1016/j.procs.2020.03.398.

[20] Remjee, S., Ju, S., Yang, D., Liu,X. and Gamal, A.E. and Eldar, Y.C., 2019. Fast deep learning for automatic modulation classification. *Electrical Engineering and Systems Science: Signal Processing*. doi: https://doi.org/10.48550/arXiv.1901.05850.

[21] Schowengerdt, R., 1997. *Remote Sensing, Models and Methods, for Image Processing*, 2nd Edition, Academic Press, Cambridge.

[22] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A, 2014. Going deeper with convolutions. *CoRR*, abs/1409.4842.

[23] Szegedy,C., Vanhoucke,V., Loffe,L., Shlens,J. and Wojna, J.,2015.Rethinking the Inception Architecture for Computer Vision. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).* doi: https://doi.org/10.48550/arXiv.1512.00567.

[24] Simonyan, K. and Zisserman, A., 2015. Very deep convolutional networks for large-scale image recoginition(ICLR), *Computer Vision and Pattern Recoginitio*n(*CVPR*).

[25] doi: https://doi.org/10.48550/arXiv.1409.1556..

[26] Singh, D. and Garzon, P.,2018. Using Convolutional Neural Networks and Transfer Learning to Perform Yelp Restaurant Photo Classification.

[27] Smith, P. and Chen, C.,2018. Transfer learning with deep CNNs for gender recognition and age estimation. *IEEE international conference on Big Data*. doi: https://doi.org/10.1109/BigData.2018.862189.

[28] Signoroni, A., Savardi, M., Baronio, A. and Benini, S.,2019. Deep learning meets hyperspectral image analysis: a multidisciplinary review. *J. Imaging* 5, 52, doi: https://doi.org/10.3390/jimaging5050052

[29] Talo, M., Baloglu,U.B., Yıldırım,O. and Acharya,U.R., 2018. Application of deep transfer learning for automated brain abnormality classification using MR images. *Cognitive Systems Research*, vol.54, pp.176–188.doi: https://doi.org/10.1016/j.cogsys.2018.12.007

[30] Wang, S.H., Xie, S., Chen, X., Guttery, D.S., Tang, C., Sun, J. and Zhang, Y.D., 2019. Alcoholism Identification Based on an AlexNet Transfer Learning Model, *Front Psychiatry*. doi: 10.3389/fpsyt.2019.00205. PMID: 31031657; PMCID: PMC6470295.

[31] Wu, Y. and Ji, Q., 2016. Constrained deep transfer feature learning and its applications. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5101–5109. doi: https://doi.org/10.48550/arXiv.1709.08128

[32] Yang, Z., Yu, W., Liang, P. Guo, H., Xia, L., Zhang, F., Ma, Y., and Ma J, 2019. Deep transfer learning for military object recognition under small training set condition. Neural Comput Appl, vol. 31(10), pp.6469–6478. doi: https://doi.org/10.1007/s00521-018-3468-3

[33] Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J. and Atkinson, P.,2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing vol. 140*, pp.133-144. doi: https://doi.org/10.1016/j.isprsjprs.2017.07.014.

[34] Zhao, W., Gou, Z., Yue, J , Zhang, X. and Luo, L, 2015.On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *Int. Journal of Remote Sensing*. vol.36 pp., 3368–3379 doi: 10.1080/2150704X.2015.1062157.