Satellite image classification using Resnet_12 architecture

CH. Hemanth Kumar, G. Prasanna, Ch. Sai Pradeep, A. Yasaswi vivek, CH. Jhansi.
UG Students, Department of Electronics and Communication Engineering,
GMR Institute of Technology, Rajam,

Abstract –
There has been a rise in interest in the use of satellite image analysis for a range of purposes, including mineralogy, forestry, agriculture, military operations, mapping, urban planning, ocean surveillance, and disaster management. However, conventional algorithms for object detection and classification are not precise or dependable enough to tackle this problem. To automate this task, deep learning algorithms, specifically convolutional neural networks, have shown potential., it has achieved success in image understanding. The Hyperspectral images have been classified into four distinct categories, namely Cloudy, Water, Deserts, and green area. This classification has been performed on the satellite images used in the study. The process of manually classifying satellite images using image interpretation methods is a time-consuming task that requires the expertise of field professionals. Therefore, our research focuses on creating an effective automated system for satellite image classification. We plan to achieve this by utilizing a Deep Convolutional Neural Network (DCNN) model, specifically the ResNet-12 framework with 12 layers. The ResNet-12 model utilizes skip connections that merge the input with the convolutional layer's output to tackle the issue of vanishing and exploding gradients often encountered in traditional CNN models. Efficiency of all the models were measured using the metrics Accuracy & Precision. Resnet-12 model got highest accuracy of 97.3% maintaining the Precision same as other models.

Index Terms - Resnet-12, Hyperspectral images, Skip connections, Deep Learning, CNN.

I. Introduction

Satellite image classification is a crucial problem in remote sensing and has many uses in different industries, such as environmental monitoring, urban planning, and agriculture. To detect and classify various land uses, vegetation types, water bodies, and built-up regions, satellite photos must be analyzed. The progress made in machine learning and deep learning has resulted in the development of accurate and efficient techniques for categorizing satellite images. Convolutional Neural Networks have been particularly successful in this task because of their ability to extract complex features from images.
In order to classify satellite images into different types of land cover, a CNN model's structure usually consists of multiple convolutional and pooling layers, as well as fully connected layers. In the process of analyzing an input image, the convolutional layers utilize a collection of filters that can be learned to extract features. The pooling layers are then used to decrease the spatial dimensions of the feature maps while also providing translation invariance. Finally, the fully connected layers are responsible for the last stage of classification. To prevent overfitting, dropout layers can be incorporated into the system. The predicted class probabilities are generated by the output layer, which is the last layer of the CNN. Based on the available data and the complexity of the task, the number and size of layers can be a changed accordingly. The performance of the CNN can be optimized by adjusting hyperparameters such as the learning rate, batch size, and number of epochs.

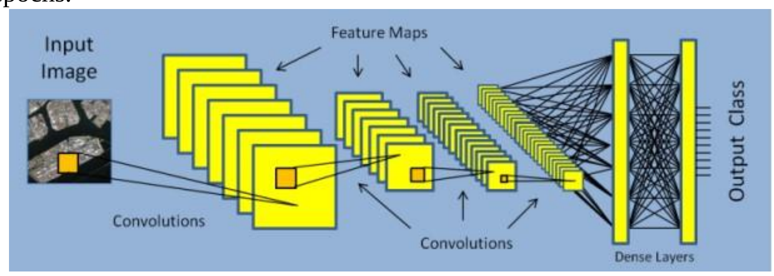


Fig 1: Convolutional Neural Network layers

As a CNN gets deeper, the gradients can become very small and the learning process can slow down. This is known as the problem of vanishing gradients. CNNs are typically designed to work with fixed-sized inputs, which can be a limitation in some applications where the input size may vary. CNNs can be difficult to interpret, as they are often treated as black boxes that are optimized for a specific task without providing much insight into how they are making their predictions.
However, the CNN model takes a better approach to satellite image classification. To address these above-mentioned issues and achieve better results, we propose a ResNet-12-based satellite image classification method. As demonstrated, this architecture outperforms other cutting-edge CNN architectures for satellite image classification. The ResNet-12 model is trained and fine-tuned on a dataset of labelled satellite images to achieve high classification accuracy.
The Residual Network is a popular CNN architecture that has been successful in a variety of image classification tasks, including satellite image classification (ResNet). ResNet is a deep neural network architecture that uses skip connections to overcome the vanishing gradients problem that occurs when training very deep neural networks.

II. Literature survey

Prior research has demonstrated an intriguing approach that combines both machine learning and deep learning methods. Some people have even combined the two approaches by employing deep learning algorithms and handcrafted features. rewrite the phrase with same meaning without best method in Satellite image classification.

Mark Pritt and Gary Chern proposed a CNN model method [1]. This paper's main goal is to assess the performance of deep learning models, specifically CNNs, for satellite image classification. This paper aims to explore a deep learning system capable of categorizing objects and facilities in the fMoW dataset. The system utilizes a satellite image and metadata that includes a bounding box. It consists of several CNNs, image processing, and neural networks (NNs). The challenge presented by IARPA Functional Map of the World (fMoW) involves developing a deep learning system that can classify satellite imagery into 62 distinct object and facility classes.

Top of Form in their work, Chong Yu, Qinfu Qiu, Yilu Zhao, and Xiong Chen [2] suggested a technique for MCA decomposition, which involves decomposing a preprocessed image into texture and cartoon layers using MCA. This is accomplished by attempting to minimize the difference between the original image and the sum of the texture and cartoon layers. To classify the different land cover types in the satellite image, a machine learning algorithm, such as SVM or CNN, is trained on the selected features.

Mengyun Shi, Fengying Xie, Yue Zi, and Jihao Yin's paper [3] proposes a creative approach on deep learning-which is based on cloud detection for remote sensing images. They went over the specifics of our deep learning-based cloud detection method. To create the cloud probability map, CNNs that are specifically designed are utilized. The cloud region is ultimately pinpointed by refining the generated cloud probability map. The authors of the paper [4], Rohit Gandikota, Radha Krishna K, Anupama Sharma, Manju Sarma M, and Vinod M Bothale, introduced a Generative Adversarial Network (GAN) that is capable of training and producing RGB satellite images from the given data. After achieving stability in GAN training, the network's encoder has the ability to extract spatial information from multi-spectral satellite data, which has been theoretically proven. This pre-trained encoder then becomes the feature extractor in the image classification network.

Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez presented a machine learning method [5] called Fully Convolutional Neural Networks for Remote Sensing Image Classification, which utilizes FCNs to classify remote sensing images. The authors proposed a two-stage training procedure to enhance the network's efficiency. In the first stage, the network learns generic features by training on a dataset of natural images. In the second stage, the network is fine-tuned on the remote sensing dataset to learn specific features for the classification task.

An approach for land cover classification of satellite images was proposed by Boshir Ahmed and Md. Abdullah Al Noman [6] in which a normalization technique and an artificial neural network were utilized. Back Propagation Algorithm, which is a dynamic and self-adaptive system with a large number of neurons, is utilized for satellite image classification. This algorithm is considered one of the top choices for this task.

III.Satellite image classification with ResNet-12

In this section, the adopted technique for classifying satellite images is outlined, employing ResNet12, one of the victorious networks in the ImageNet machine learning contest, due to its reduced parameter size. This choice enables faster model storage, weight storage, and training. The outcome includes pre-processing, training, and classification, all of which are described in the proposed method's process flow.

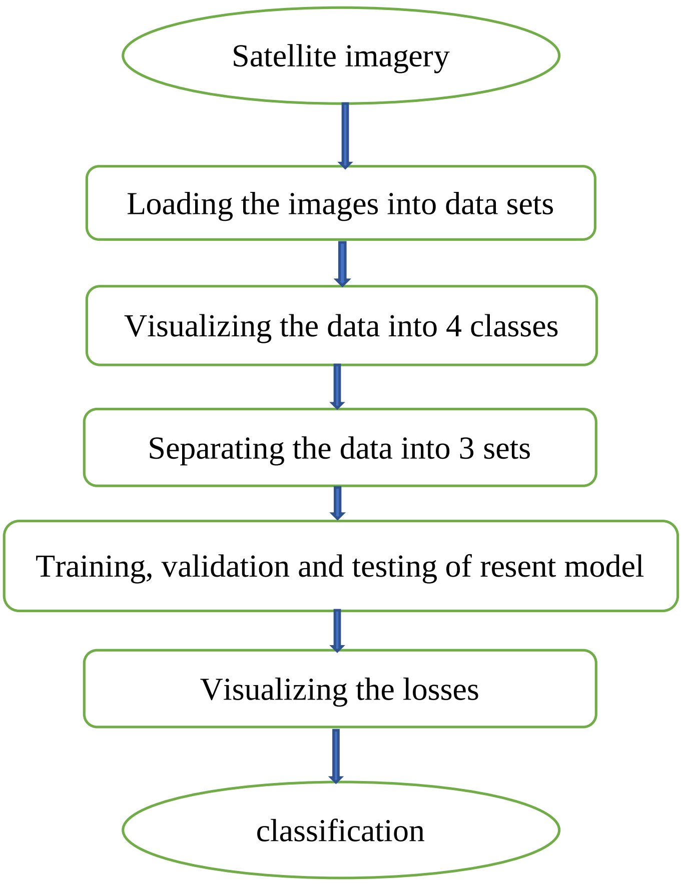


Fig 2. Project workflow

Visualizing each class

Visualizing each class in satellite image classification can help us to understand the features that distinguish one class from another. Here are some steps to creating visualizations for each class.

- Preprocessed the images to remove any noise or artefacts that could interfere with the classification results. We can also resize the images to a standard size for easier processing.
- To extract features from images, we used feature extraction techniques such as convolutional neural networks (CNNs). These characteristics can be used to categorize the images.

Training, Validation & Testing

The process of teaching a machine learning model on a marked dataset to comprehend the fundamental patterns and connections between the input features and output labels is known as training. The objective of training is to reduce the difference between the predicted output of the model and the ground truth labels found in the training data.

Fig 3. Training Data for 16 batch size

Once the model is trained on the training set, its performance is assessed using the validation set. The validation data should be a true representation of the data distribution that the model is presently encountering. The primary objective of validation is to pinpoint any potential problems with the model, such as overfitting or underfitting, and refine the model accordingly to enhance its performance on new, unseen data.

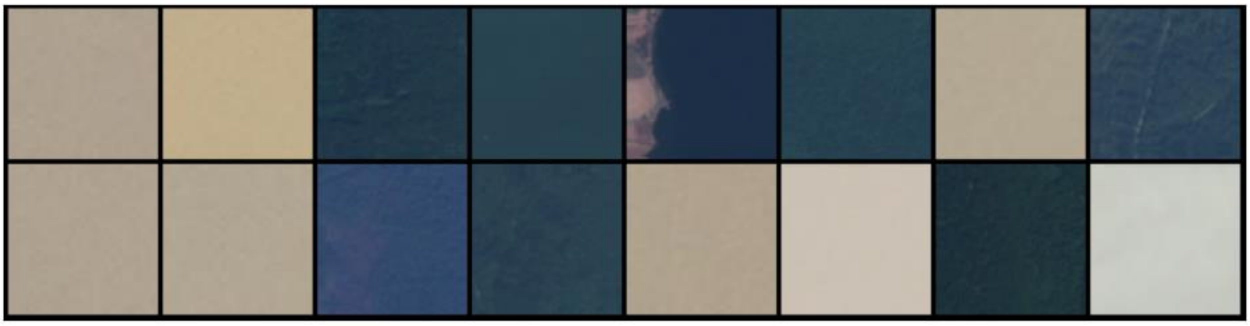


Fig 4. Validating data for 16 batch size

Satellite image classification involves a crucial step known as testing, where the performance of the classification model is evaluated on a distinct set of images that were not employed during the training phase.

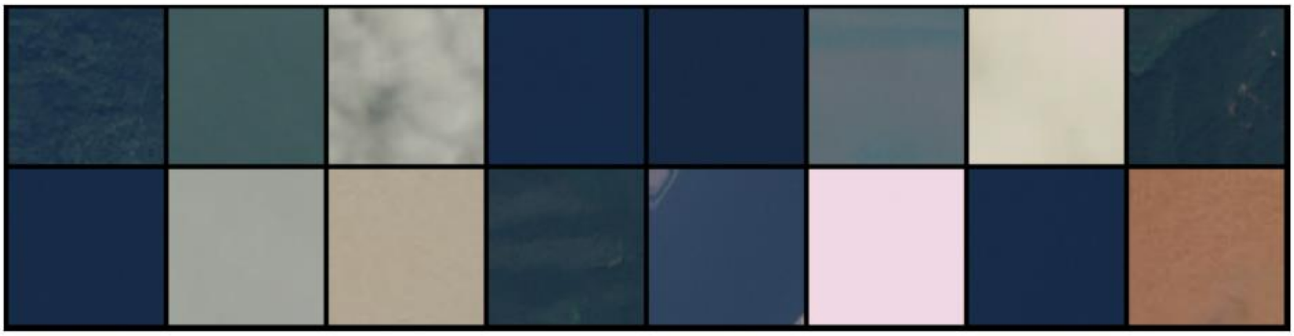


Fig 5. Testing data for 16 Batch size

Resnet-12 Architecture

ResNet-12 is a ResNet architecture variant with 12 layers, including input and output layers. ResNet-12 has 12 layers: a convolutional layer, a max-pooling layer, and ten residual blocks. Each residual block has two convolutional layers as well as a skip connections, which allows the input to avoid the convolutional layers and be summed up directly to the output. Between the residual blocks, there are transition layers that include a convolutional layer with a kernel size of 1x1, batch normalization, and max pooling of stride of 2.

These transition layers have the goal of reducing the spatial dimensions of the feature maps while increasing the number of filters, which helps to maintain the network's representational power while keeping the computation cost manageable.
The output of the fourth residual block is passed through a global average pooling layer, which reduces the spatial dimensions to a single value for each filter.
The final classification scores are produced by passing the resultant feature vector through a fully connected layer with a Softmax activation function.

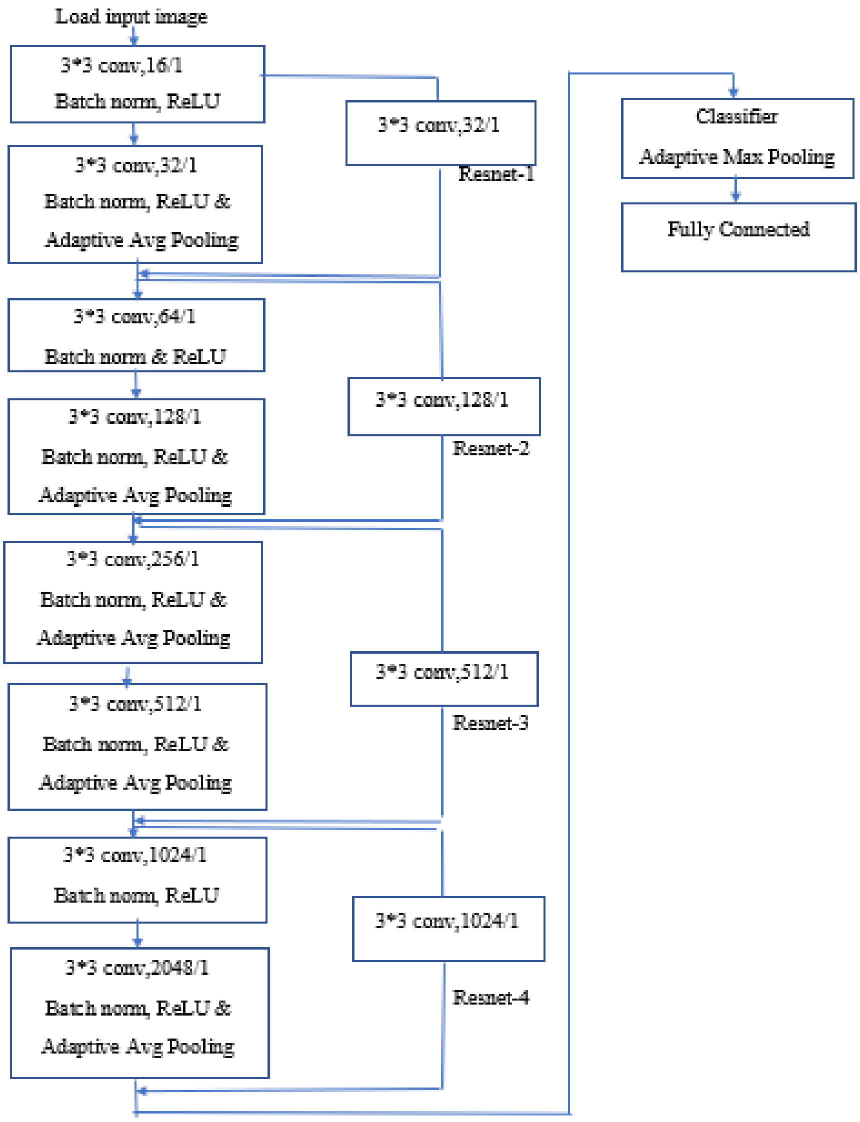


Fig 6. Architecture of Resnet-12

Convolutional layer

The ResNet-12 architecture employs a series of convolutional layers with varying filter sizes and filter counts. The first layer contains 64 7x7 filters with a stride of 2, reducing the input image size by half. The following layers have smaller filter sizes and fewer filters, allowing the network to learn more complex features at higher levels of abstraction.
An input image is processed by a convolutional layer filter to extract features. Weights with values are used to represent these characteristics as output. The weight values indicate the importance of a specific feature. The greater the importance of the feature, the greater the weight. These weights are then applied to an input image at the convolutional layer, the fully connected Multi-Layer Perceptron (MLP) layer, to classify it. Each layer's result (weights) is usually, an activation function is used. This transfer function aids the chain in its learning of complex features. Relu, short for Rectified Linear Unit, is an activation function that is frequently utilized. Relu function can be expressed as:
$g(y) = Max(0, y)$
Backpropagation is used to train the weights in the Convolutional layer. The chain rule is used in backpropagation's forward pass. The later layer passes the absolute value derivation as an input to the earlier layer.
Shortcut connections in the convolutional layers of ResNet-12 enable the network to avoid the vanishing gradients issue encountered by deep networks. By bypassing certain layers, the network can learn better representations of input images, leading to enhanced accuracy in image classification.

Pooling Layers

The pooling layer produces a single output value by sliding a fixed-size window (e.g., 2x2) over the input volume and taking the maximum, average, or some other function of the values within the window. This output value is the pooled feature for that specific window location. ResNet-12's max pooling layer also contributes to the network's resistance to small shifts and distortions in the input images. The max pooling layer can preserve the most important features while discarding irrelevant or noisy information by taking the maximum value in each region.
The average pooling layer is used to reduce the spatial dimensions of the feature maps before the fully connected layers. To generate a single output value for each channel, the average pooling layer computes the average of the values present in each 7x7 feature map.

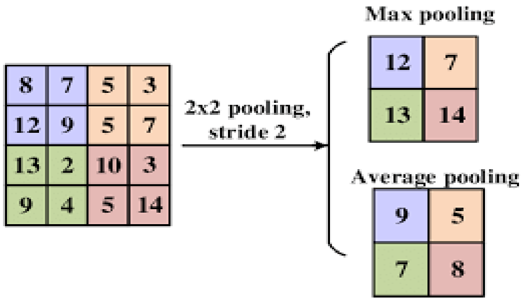


Fig 7. Max & Avg pooling Layers

The average pooling layer is used before the fully connected layers to reduce the spatial dimensions of the feature maps. The average pooling layer then averages the values in each 7x7 feature map to generate a single output value for each channel.

Model Evaluation

The metrics for model evaluation may include Accuracy, Precision and Loss, for example, and continuous evaluation might include customer lifetime value. Furthermore, the distribution of data may differ between historical and live data. Continuous model monitoring is one technique to classify distribution drift.

IV.Results

The experimental results and analysis of the proposed solution on the dataset are presented in this portion. These datasets are pre-processed as previously described.

Dataset

Dataset used to classify cloudy, desert, water, and green areas from satellite imagery is compiled from all publicly available Kaggle Datasets titled Remote sensing Satellite image classification. The dataset contains images from satellites in four different situations: cloudy, desert, water, and green. It contains over 5600 plus images in total.
The dataset is split into three parts for training, validation, and testing, in the ratios of 70:20:10. The validation set plays a crucial role in fine-tuning the model's hyperparameters, as the trained model is periodically assessed for improvements. The testing set is used for the model's final evaluation. 70% of the total dataset is used for model training, 10% for model testing, and the remaining 20% is used for model validation.

Experimental results

DCNN method proposed examine the model. Table I displays the DCNN's parameter settings. The calibrated ResNet-12 model only accepts 64x64 images as input due to the weight used during the model's pre-training. To be fair, the paths were tested with various epoch counts, batch sizes, and sizes of dataset. This assures that the model's training time is fairly measured. This method employs model checkpoints to achieve the best accuracy in the time of training. Table II displays the accuracy and loss of the proposed solution's result.

Table I. Initial Dataset Evaluation Criteria Setup

S.no	Criteria	Value
1.	Image size	64x64
2.	No. of training images	3942
3.	No. of testing images	563
4.	No. of classes	4
5.	Batch size	64,32 ,24 &16
6.	Learning rate	0.001
7.	Optimizer	SGD

Table II. Accuracy and loss % of proposed method

S.no	Batch Size	No. of epochs	Accuracy (%)	Loss (%)
1	16	10	91.07	8.93
		25	92.72	7.28
		100	93.43	5.33
2	24	10	90.41	9.59
		25	91.65	8.35
		50	94.67	5.33
3	32	10	91.83	8.17
		25	93.25	6.75
		50	94.67	5.33
4	64	10	93.43	6.57
		25	95.20	4.80
		100	97.34	2.66

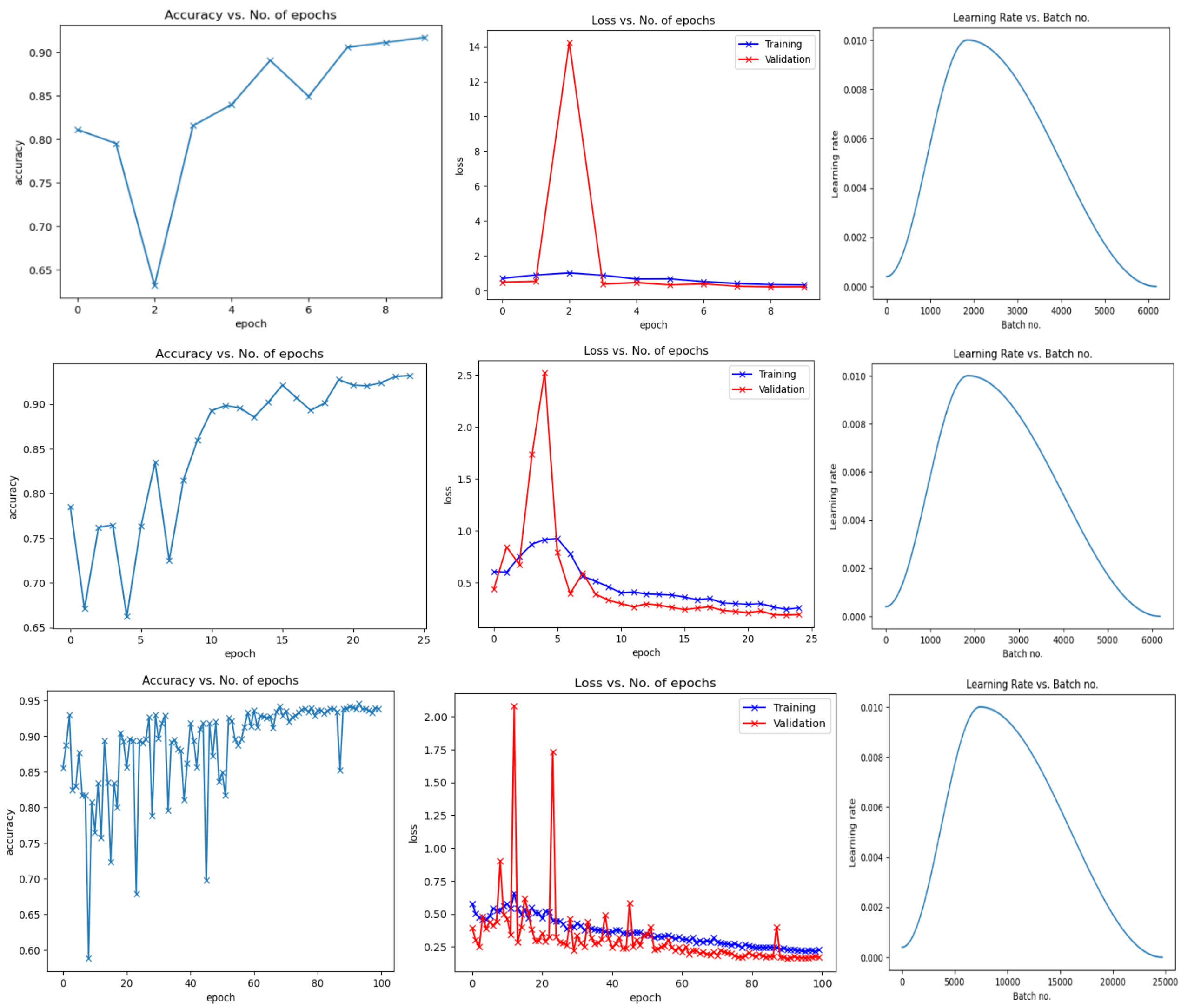
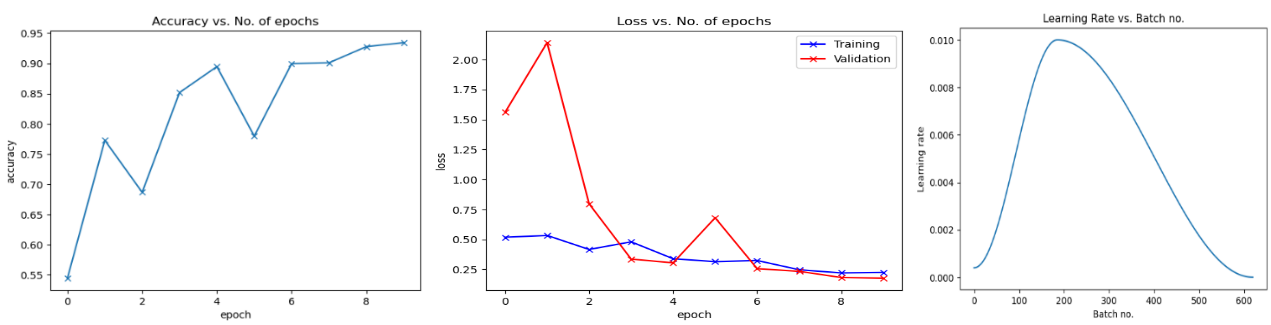


Fig 8. Accuracy, Loss & Learning rate of different epochs of Batch size 16



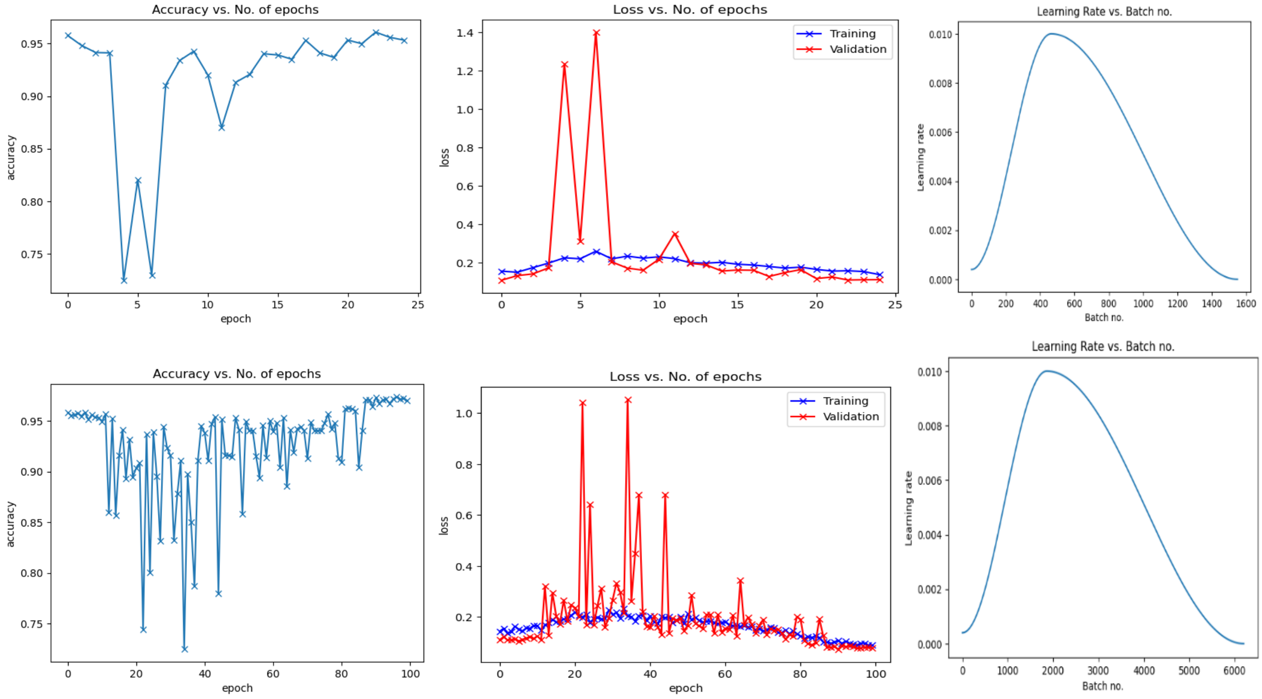


Fig 9. Accuracy, Loss & Learning rate of different epochs of Batch size 64

The proposed method outperforms increasing accuracy on increasing epochs of different batch sizes, as shown in Fig 8,9 &Table II. On the dataset, the implemented method had a best accuracy of 97.34% to the batch size 64. Especially, accuracy decreases as the number of classes increases, whereas sample size per class and dataset epochs increases directly. This means that the more epochs there are, the greater the accuracy, and the more classes there are, the lower the accuracy. This demonstrates the need for a large dataset for deep learning methods to perform well.

V. Conclusion

Human vision cannot always comprehend the regions that are present in satellite images. As a result, satellite image classification is extremely useful for gathering information about a specific region. We proposed a Deep Learning System for classifying high-resolution multi-spectral satellite imagery into four different classes. By running the model with different number of epochs and batch sizes, a perfect model for satellite image classification is developed using ResNet 12 architecture. We used a Kaggle Dataset of nearly 5600 images for training, validation, and testing. We created a system for satellite image classification that is 97.6% accurate. We observed that as we increase the number of epochs the better the output is and if more amount of data is fed for the model, then the model will be more intelligent. There is a huge scope for Deep learning Algorithms in the field of Satellite image detection.

VII. Future Scope

Satellite imagery can provide complete project planning oversight. Satellite imagery is used to map out the surrounding environment and oversee the project's entire development. It is much easier to account for potential delays and plan asset management in this manner. To summarize the potential trends in deep learning, unsupervised learning methods have the potential to create models that closely mimic human behavior with sufficient time and research. The tension between consumer data protection laws and the needs for high-volume consumer data research is expected to persist.

VII. References

[1] M. Pritt and G. Chern, "Satellite Image Classification with Deep Learning," *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Washington, DC, USA, 2017, pp. 1-7, doi: 10.1109/AIPR.2017.8457969.

[2] C. Yu, Q. Qiu, Y. Zhao and X. Chen, "Satellite Image Classification Using Morphological Component Analysis of Texture and Cartoon Layers," in *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1109-1113, Sept. 2013, doi: 10.1109/LGRS.2012.2230612.

[3] S. R. Surya and M. A. Rahiman, "Cloud detection from satellite images based on Haar wavelet and clustering," *2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2)*, Chennai, India, 2017, pp. 163-167, doi: 10.1109/ICNETS2.2017.8067921.

[4] R. Gandikota, R. K. K, A. Sharma, M. M and V. M. Bothale, "RTC-GAN: REAL-TIME CLASSIFICATION OF SATELLITE IMAGERY USING DEEP GENERATIVE ADVERSARIALNETWORKS WITH INFUSED SPECTRAL INFORMATION," *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, Waikoloa, HI, USA, 2020, pp. 6993-6996, doi: 10.1109/IGARSS39084.2020.9323363.

[5] E. Maggiori, Y. Tarabalka, G. Charpiat and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, China, 2016, pp. 5071-5074, doi: 10.1109/IGARSS.2016.7730322.

[6] I. C. Duta, L. Liu, F. Zhu and L. Shao, "Improved Residual Networks for Image and Video Recognition," *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2021, pp. 9415-9422, doi: 10.1109/ICPR48806.2021.9412193.
[7] H. Jung, M. -K. Choi, J. Jung, J. -H. Lee, S. Kwon and W. Y. Jung, "ResNet-Based Vehicle Classification and Localization in Traffic Surveillance Systems," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 2017, pp. 934-940, doi: 10.1109/CVPRW.2017.129.
[8] S. Akshay, T. K. Mytravarun, N. Manohar and M. A. Pranav, "Satellite Image Classification for Detecting Unused Landscape using CNN," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 215-222, doi: 10.1109/ICESC48915.2020.9155859.