

Research Article

# Leveraging Pre-trained Deep Learning Models for Remote Sensing Image Classification: A Case Study with ResNet50 and EfficientNet

**Srivani Bobba** \* 

Department of Information Technology, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

## Abstract

The procedure of categorizing images from remote sensing is also another application of machine learning not just ground-based platforms (for instance satellites), aerial platforms become platforms sometimes in aviation either. They erase the counterparts that were based on individual categories and are portrayed on a specific part of the image. Geospatial Supply of gravel mainly is used for producing railway track, road and concrete surface. Data by analyzing their buildup, dams, bridges, extraordinary open spaces, reservoirs and canals. It targets to be specific and exact as possible in a different specific area of the land. Aspects of the enlarged portrait or distinctions weaved into the completed arts. This might have aspects such as mapping of the trees, plants, rivers, cities, farms and woodlands, and other items. Geospatial image classification is necessary for the identification and real-time analysis of different hazards and unrests. Provide numerous applications, including waste management, water resources, air quality, and traffic control in the urban contexts. Planning, monitoring the environment, land cover, mapping, as well as post-disaster recovery. Management team, traffic control, and situation assessments. In the past, human experts situated in a selected area classified geographical images by means of manual processing. One that involved the allocation of too much time. As this is one of the two broad categories, how to get rid of it is consequently. Applying machine learning and deep learning methods we analyze and interpret the data in order to reduce the time required to provide feedback which allows the system to reach a higher accuracy. The procedure will also be more reliable and the outcome will hopefully be more efficient CNNs are one of the deep learning subclasses in which the network learns and improves without the need for human intervention. It extracts features from images. They are main for the performance and metrics to help the organization to decide on whether they have accomplished their goals, using visual imagery.

## Keywords

CNN, Machine Learning, Deep Learning, ResNet50, EfficientNetv2

## 1. Introduction

Urban planning, environmental reconnaissance, and land cover mapping are just a few of the numerous uses for re-

mote sensing that have made it a vital technology. Because of their complexity and large amount of data, geographic

\*Corresponding author: [srivaanib@gmail.com](mailto:srivaanib@gmail.com) (Srivani Bobba)

**Received:** 26 June 2024; **Accepted:** 22 July 2024; **Published:** 15 August 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

photographs can be challenging to understand and classify, even if they provide a wealth of information about the Earth's surface. Convolutional neural networks are one type of deep learning technology that has recently shown impressive results in image classification tests. Compared with all the other cnn models, ResNet is more advanced.

ResNet solves the disappearing gradients issue and permits deeper training through adding remaining connections into the network. Better environmental monitoring as well as more precise mapping of land cover could result from the ResNet-50 model's ability to improve the effectiveness of geospatial image processing. The quality of the geographical image categorization procedure is determined by Resnet50's standard functionality [11].

Two strategies have been used recently to tackle the problem of envision classification: (i) feature extraction to a classification model. (ii) Feature extraction using deep learning architectures. The first technique can be further separated into unsupervised as well as supervised learning. However, fully unsupervised techniques lack the higher-order relationships inherent in the classification task. Supervised instruction requires a large amount of labeled data as well as carefully selected, manually created features. Satellite classification of pictures has made the most of both learning algorithms. Another method is also state-of-the-art in many classification tasks because it extracts the features from the images.

Res-Net (Residual Network) and Efficient-Net are two CNN designs that have emerged as the best in the field of computer vision for picture classification. Res-Net adds remaining connections to remedy the vanishing gradient problem and allow for deeper training of the network, which improves performance. On the other hand, In this work, we examine the application of the Res-Net and Efficient-Net systems to remote sensing photo categorization. Our aim is to leverage these models' capabilities to refine the accuracy and efficiency of the analysis of remote sensing imagery. This will make it possible to map land cover more precisely and monitor the environment better. Here benchmark dataset is used to compare experiment results.

## 2. Related Work

The research [1] aimed at lending sectors which comprise land use and land cover landscapes (LCLU) shall be assessed based on satellite images by utilizing deep learning techniques. The three models of deep learning under thorough examination will be. Into consideration were incorporated the Techniques of Fine-tuning, Transfer Learning (TL), and the others.

Convolutional Neural Network Partial Network Feature Analyzer (CNN-FE). The CNN-FE model performed sub optimally compared to the TL only. Accuracy of solutions, the former ones possess the most prominent. Performance at 88%. Scientific analysis has further demonstrated the great importance of the number of the parameter size in the dis-

tinctive model that is playing the prime role. Refining the sequels of fine-tuning, and using pre-trained layers goes a long way in attaining the intended affairs improved performance. This research establishes that women affected by pregnancy depression are notified that due to background factors like such as social background, personal history level of economic development, the severity of their depressive symptoms are related to. Innovative approaches, such as model simplifications and small data sizes, as well as TL and training strategies well.

Bo Feng and coworkers' work [2] based on residual generative models. Automated plants monitoring also generally involves hyperspectral remote sensing images classification for identifying plant species and adversarial architectures for networks. This paper presents a new network architecture in which each layer of the network has its own adversarial classifier. A photorealistic CNN model integrated with residual blocks GAN to produce state of art hyperspectral images of classification method. Here is what will try to do give a piece of holistic research in which labeled data problem, a common issue among system trainers, is one of the major factors. The Hyperspectral image processing is the suggested approach seeks by the way to get higher classification precision and not rely only on the exactness of the prediction exclusively to explain their decision making, and with annotated data sets.

The researchers have come up with the 18-layer residual convolutional network discriminator along with the 6-layer residual network generator as the new structure of the GAN. In this way, such modifications become a means of increasing the efficiency of feature extraction and data representation. A combination of forward propagation, Skip connection routing, and other components such as Batch Normalization, ReLU Activation function, Conv3D, and UpSampling3D helps the Generator achieve real-time and better spectral information rendering together with the actual sample. Non-Convolutional Residual Block (NCRB) and Convolutional Residual Block are the two residual block structures that the discriminator is built on. Such distributed architecture automatically evolves to correctly flag produced samples and classified them by this double-layer structure.

Convolution neural network (CNN) and residual generative adversarial network (ResGAN) architecture in tandem help in an increase in a hyperspectral image classification accuracy. Experiments performed on common benchmark hyperspectral datasets demonstrate that the presented approach produces superior results in remote sensing and earth observation leading to performance enhancements from as small as 0.7% up to 2.3% in Overall Accuracy (OA) against the other approaches. Kumar Ajay et al. [3] Labeling of aerial views from a distance, seen in the satellite pictures, with different labels that could help in unique identification of these images.

The material which is presented addresses the problem of how rapidly-changing images are classified accurately from

such a dataset as a new Sat-2 picture is given. Through this data set we present two approaches, different in classification of land covers and/or land use. The main technique is the Deep Learning model for ResNet50 that was recently upgraded to classify multiple categories from Sentinel-2 photos. The advantage of ResNet50 over traditional machine learning algorithms which rely heavily on shallow feature extraction is discussed. Contrary to the light technique that receives an F1-score of 0.87, the deep method, ResNet50, is the most appropriate and gives 0.924 F1-score, further confirming its acceptability when it comes to segregating multi-label satellite scenes.

The proposed model architecture utilizes batch normalization, dropout and the ReLU activation function as well as a convolution layer and a dense layer to handle the problem of overfitting. Model can be considered as a very effective one with 200 epochs on 10-folds without any overfitting within minimum range. The ResNet50 model shows higher outputs for a flight direction, as well as for other characteristics, such as accuracy and F1-score, which is better than that of other top-notch algorithms used for image classification. On the other side, shallow learning algorithms, for instance Random Forest and AdaBoost, are highly favored in the cumulative computation expenses over ResNet50, especially for the classes with unusual color histograms. As well as sharing the strengths of the both shallow and deep approaches to image classification tasks, the study suggested a necessity of more sophisticated designs which would bring in the concepts of data augmentation in order to produce efficient solutions for real-time application like early fire detection, and also include useful components from the environment.

Teng Wenxiu et al. [4] Deep Adversarial Domain Adaptation for Cross-Domain Semi-supervised Classification in Remote Sensing Images: Classifier-Directed Constrained Approach. In this paper, a novel method is introduced. Classifier-constrained deep domain adapted semi-supervised learning model with adversarial training aimed at transferring knowledge across different domains. The mechanism consists of an underlying DCNN layer which would feed the semantic content of RS scenes in a feature representation process, before adaption.

After that, adversarial domain adaptation is used in order to map the feature distributions of source and target domains and consequently, to enhance the model's ability of generalizing across domains. What is specifically characteristic about CDADA is that it is a twofold land-cover choice divider. These classifiers unravel land-cover categories and therefore transfer them from their certain boundaries into more generalized and transposable representations. Three benchmark RS scene datasets (AID, Merced, and RSI-CB) were used, where six instances were available in each dataset, and which were then evaluated experimentally. The results indicate that of CDADA the classification accuracy is 92.17% of the average for the six tested domain cross-scenarios. This accuracy notably beats that of the revised VGG16 model,

and it even outshines the earlier domain customization techniques. The advantage of increased data accuracy using CDADA, as it is reported by Minetto Rodrigo et al.

Wang Junjue et al. [5] The Search for Deep Neural Networks in Remote Sensing for Recognition Tasks: RSNet. The authors of this research introduce the concept of the Remote Sensing Deep Neural Network (ResNet) architecture, which improves the recognition of high spatial resolution remote sensing (HRS) photographs through the application of deep learning techniques, namely convolutional neural networks (CNNs). While CNNs have been effectively used for a range of HRS identification tasks, the authors note that existing architectures often stem from natural image processing and neglect to consider the unique challenges of images. The RSNet framework automatically finds and refines CNN designs for HRS image recognition through a two-step method.

A hierarchical search space is constructed in the Gradient-Based (GB) architecture search stage, allowing customization of both spatial resolution changes and fundamental structure blocks. Gradient descent is used to optimize this search space in order to determine the model parameters and suitable architecture. The searched RSNet is then retrained with the identification module in the Task-Driven Architecture Train stage to further optimize model parameters. Experimental results show that the RSNet performs better than state-of-the-art deep learning architectures such as AlexNet, GoogLeNet, VGG16, FCN8S, PSPNet, DeepLabv3, DeepLabv3+, and UNet, including benchmark data sets for land-cover and picture categorization. The balance between accuracy and computational effectiveness is consistently greater with RSNet. Several methods have been applied to the UC Merced dataset, and each approach has achieved its accuracy. At 96.78%, RSNet (UCM) has the highest overall accuracy, followed by VGG16 (96.76%) and AlexNet (90.76%). The RSNet (UCM) likewise has the highest Kappa values, at 0.9670, followed by VGG16 at 0.9669 and AlexNet at 0.8963. Not only is RSNet more precise than handcrafted CNNs, but it is also more efficient. Lightweight deep neural network designs that are more appropriate for remote sensing images recognition are made possible by the RSNet framework. The versatility and promise of RSNet for the field of remote sensing are highlighted by the authors' intention to extend its use to in-orbit satellite data processing and hyperspectral picture analysis.

Kumar Ajay et al. [6] Classification of remotely sensed satellite imagery utilizing several labels. The multi-label scene classification problem for very high-resolution (V.H.R.) remote sensing photos is the focus of this research. Although there are several well-established single-label scene classification methods, little research has been done on the prospect of detecting multiple objects in a single V.H.R. satellite image. The study suggests a method for optimizing cutting-edge Convolutional Neural Network (CNN) architectures in order to close this gap. Following pre-training on the

ImageNet dataset, multi-label modifications are made to the CNNs.

Using the completed models on the U.C.M.E.R.C.E.D. image dataset—which includes twenty-one terrestrial land use categories at sub-meter resolution—the efficacy of this approach is assessed. These models are compared to one based on Graph Convolutional Neural Networks in terms of efficiency. The findings show that the suggested models produce findings that are similar. The findings demonstrate that the suggested models need a notably smaller number of training epochs in order to provide comparable outcomes. The multi-label classification aspect of remotely sensed V.H.R. picture classification, which has not received much attention in the literature, is greatly advanced by this work. It takes into consideration sixteen deep learning architectures on ImageNet that have pre-trained relationships. It finds that after fine-tuning, these architectures outperform others on the multi-label. With an F1-score of 0.7549, InceptionV3 had the best score, while Exception came in second with a score of 0.754138. These are noteworthy accomplishments. Particularly in the deeper variants (50, 101, and 152 layers), ResNet's performance was not as good. MobileNet and MobileNetV2 received F1-scores of 0.707787 and 0.608429, respectively, in relation to their lightweight designs. Both DenseNet and ResNet's F1-scores were below 70%, indicating comparable performance. According to the study, further optimization techniques could raise the F1-score. The authors also recommend examining low-resolution image models and the reasons behind the wide range of metrics among different network configurations. The ratio of non-trainable parameters to total parameters in each design is another potential area of research.

Thirumaladevi, S., et al. [7] Transfer learning for remote sensing image scene classification to improve accuracy.

This paper focuses on the effective use of high spatial quality remote sensing imagery for image classification, which is essential for scene classification. Achieving high-performance scene classification requires effective feature representation, which is crucial considering the difficulties in processing sensor data. In particular, the use of convolutional neural networks (CNNs) and transfer learning to extract spatiotemporal information for classification is discussed, along with the advantages of combining deep learning with remote sensing. In order to increase scene classification accuracy, the study looks into image categorization methods. Transfer learning using pre-trained networks like AlexNet and Visual Geometry Group (VGG) networks is the main method used. Two primary methods are taken into consideration.

To categorize fresh datasets, features are taken from the second fully connected layer of pre-trained networks, followed by SVM classification and the replacement of the final layers of pre-trained networks. The evaluation takes use of the UCM and SIRI-WHU datasets. The recommended approaches produce significantly higher accuracy, with 93% for the SIRI-WHU Dataset and 95% for the UCM Dataset.

The study's graphic depictions of the training stage, batch accuracy, loss, and baseline learning rate over time demonstrate how the training process evolved. The confusion matrix and accuracy metrics show how effective the approaches utilized were.

Transfer learning has been shown to be crucial for achieving high accuracy; pre-trained network performance is benchmarked at 91-93% for the SIRI-WHU Dataset and 93-95% for the UCM Dataset. SVM is used as a comparison, and transfer learning is discovered to be a dependable method for achieving the best results. The findings suggest that future work may involve fine-tuning each pre-trained network separately.

Cheng Gong and others [8] Taking everything into account, the research demonstrates how combining deep learning with remote sensing can increase the accuracy of picture classification. Remote sensing picture scene categorization meets deep learning: potential, difficulties, methodologies, and benchmarks. The crucial topic of identifying remote sensing images based on their content is examined in this research, which delves into the realm of remote sensing picture scene classification. There are several applications for this classification in many different fields. The study highlights how deep neural networks have altered this field because of their powerful feature learning powers, and how this has attracted a lot of interest and produced innovations. The research points out a significant gap in the field despite the significant advancements: there isn't a thorough analysis of recent achievements in deep learning for remote sensing picture scene classification. The study does a thorough analysis of more than 160 research publications that explore deep learning techniques in order to fill this gap. Researchers have specifically addressed the main difficulties in classifying remote sensing image scenes and have divided the surveyed techniques into three groups: Techniques based on autoencoders, convolutional neural networks, and generative adversarial networks. The publication also presents benchmark datasets that are frequently used to assess algorithms for classifying remote sensing image scenes. The UC-Merced dataset is one among the datasets that is detailed. There are 2100 scene images in total, divided into 21 scene classes, each of which has 100 land-use photos. These photos were taken from the United States Geological Survey's National Map and have a pixel quality of 0.3 meters. Scene classification still makes considerable use of the UC-Merced dataset [12]. When evaluating algorithms, two often used training ratios are 50% and 80%, with the remaining 50% and 20% reserved for testing. The report concludes with several interesting future ideas for the subject of remote sensing picture scene categorization. It offers a comprehensive overview of the developments and methods in deep learning for this goal, shedding light on the challenges and potential paths. The table uses the NWPU-RESISC45 data set to compare the overall accuracy of the five scenario simulation approaches. 10% to 20% is the range of training ratios. In terms of over-

all accuracy, HW-CNNs perform best, with 94.38% at 10% training ratio and 96.01% at 20% training ratio.

Gladima Nisia T et al. [9] Value-added ensemble to reach the high-resolution and precise classification accuracy of the images in remote sensing classification. With the pictured images of high-resolution remote sensing, feature extraction is of great importance, which impacts the preciseness and the speed of final results. The latest phase is the neuronal networks automated feature extraction technique that cuts out human involvement and uses neural networks. This method implies utilization of comprehensive convolutional neural networks (CNNs) that are capable to extract deep information from remote sensing pictures [10]. Moreover, this method further combines these extracted deep features with other special types of features, such as localized binary pattern (LBP) characteristics, and Gabor features, to get the best results from these features set. Subsequently, the sets of hybrid characteristics are involved to produce a classifier.

Experiments using the hierarchical stacked sparse autoencoder (SSAE) networks [13] and LISS IV Madurai image

validate the system's performance and demonstrate that the proposed methodology outperforms the state-of-the-art techniques. These findings are graphically represented to highlight the advantages of the chosen approach over alternative methods. The study demonstrates how image augmentation improves classification precision. In order to demonstrate the positive effects of image enhancement on categorization, a comparison of how well it performs both with and without image enhancement is provided. Additionally, the system's performance is compared to established techniques, such as CNN with Gabor and CNN with LBP combinations, Inception, ResNet, VGG, and AlexNet, with a particular focus on overall classification accuracy levels.

### 3. Proposed Work

The model architecture is shown in figure 1, which represents input and output of the model.

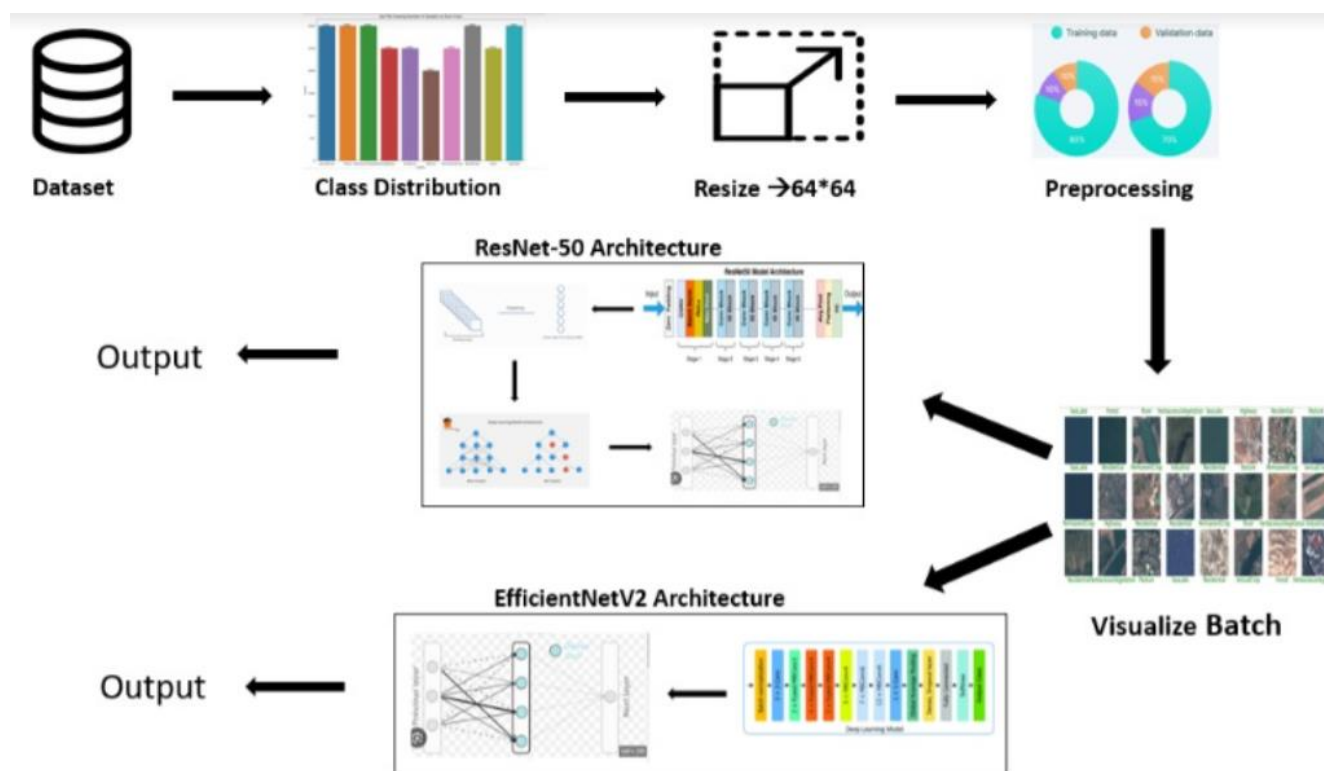


Figure 1. Architecture of the proposed work.

#### 3.1. Dataset Description

Using the EuroSAT high-resolution images remote sensing dataset, land cover and usage are categorized. A surveillance satellite took this image. Deep learning and computer vision techniques can be compared and evaluated using this standard

benchmark, which was developed for that purpose. A high-resolution classification of land cover and land use in remote sensing pictures is made easier with the help of the EuroSAT dataset. This dataset allows for a comparison and contrast between deep learning and computer vision techniques.

Image Content: The 27,000 satellite photos that the Sentinel-2 spacecraft took are part of the EuroSAT collection.

These images are in RGB (Red, Green, Blue) format and measure 64 by 64 pixels. The collection offers comprehensive coverage of both rural and urban places throughout multiple European regions.

**Land Cover Classes:** The images in the collection display ten different land cover classes, each of which represents a certain type of land use or land cover. Annual Crop, Permanent Crop, Pasture, Forest, Herbaceous Vegetation, Highway, Residential, Industrial, River, Sea/Lake are all included in these courses.

**Class Distribution:** The carefully chosen photos in the EuroSAT collection correspond to the number of each land use class. There are approximately 2,700 images in each class, for a total of 27,000 images in the collection, based on this equilibrium.

**Annotations:** Among the pictures of the EuroSAT statistics collection each image has been these phenomena should

be observed and selected as they are assigned a class name that. Is related directly to the type of land cover due to which the signal is sent back. These monitoring the annotations helps supervised learning methods, which are based on the labeled data sets. Provides a platform for algorithms to identify and classify the area via land cover categories with accuracy. The EuroSAT data set is meant as most useful for researchers, practitioners, policymakers, and advocates and developers, as part of a team of other professionals who affect the land. Surveys and land classification duties. It makes deep designing learning algorithms and image classification becoming more convenient to design and assess. The set of data being comprised of an equiprobable distribution is another valuable aspect of the dataset. Diverse cover of land classes makes it very good for the generation of the reliable and accurate models thus for its utility. Figure 2 depicts the sample image of the dataset.

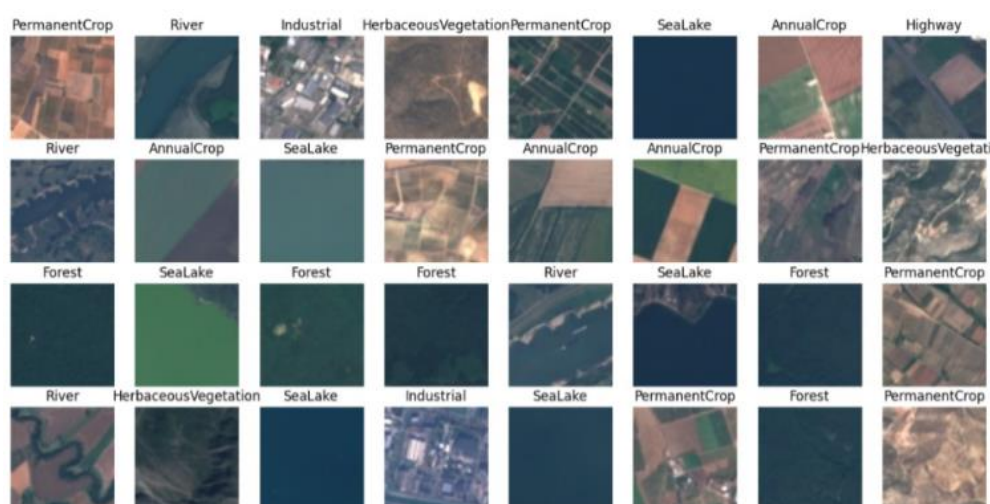


Figure 2. Sample Images of Dataset.

### 3.1.1. Class Distribution

In this dataset the proportion of the ground among ten land use types constitutes the quality of the data. Nowadays, the balance of land cover categories across Europe is quite narrow. However, upon. On the other hand, it was discovered that Forest fin and Annual crops fin regions have relatively low MOI values. The video documentary seems to incorporate many on-site pictures, i.e. from rulers' quarters, the Common Council chamber, the Meat Market, Shire Hall, and the Cathedral. Lowest carbon footprint is for "Pasture" biomass with this one. This variation ensures diverse identifies the group's experience and expands the training set, which makes the model work better in any case.

### 3.1.2. Image Resizing

Secondly, once the expansion has been achieved, these images are consequently formatted into 64 x 64 pixels. These

features also contribute to the quick operations of calculations and model learning. It does so by the way of choosing these things as opposed to the precision. Alternative operational mode cutting the computational load and simplify the workflow of the process. By outlining the difference between the land type's appearances on various satellite imageries for the hundreds of terabytes of data. Use classification endeavors.

### 3.1.3. Preprocessing: This Step Involves

#### (i). Caching

Within TensorFlow, the use of the cache means putting those datasets to memory or disk. Instead of recurring access to the data on the disk, which would otherwise take longer, the training could be fastened by simply retrieving the data. The data will then be kept; hence, process it whenever need and get the report as if it is happening in a split second. Scal-

able training with multitude data sets. They are beneficial for datasets which can fit into the random-access memory (RAM) or databases which are reused across different jobs.

### (ii). Shuffling

No matter which data pre-processing is under question, shuffling entails a randomness of the data. sorting the data samples. The model is retrofitted for sure. That can generalize, outperform the baseline model over new samples as well, which leads to the reduction of risks. The problem of the model overfitting which in turn results in the model not being able to learn. Sequence of the data.

### (iii). Repeat

Besides, one of the noticeable options of TensorFlow datasets is that the repeat function helps the whole data set to be applied for data training. Therefore, era and time could be extended through this parameter and the data collection length could be of any length training. This brought us the opportunity to perform the test with every sample and replicate it several times. By which machine learning is able to have the basis of optimum optimization and progress.

### (iv). Batch

Batch a sort of value has the TensorFlow datasets organize the data points. What machines do is that they combine all the components one by one is what allows the machines to take care of the processing process simultaneously. Giving more examples as well as repetition during class. Undoubtedly, the learning process is improved as a result. The speed and efficiency of computations along with convergence conditions become the main aspects to pay close attention to.

Neural networks are created on the grounds of large datasets in a fast run.

### (v). Prefetch

Mostly, TensorFlow accepts multiple samples in place of a single cardinal sample. While utilizing batch functions which group multiple trials we hence accomplish that target. By feeding dataset pieces into batches, it becomes possible to reduce the time of convergence. Computing speed and stability could not be ignored while the main decisions are being made about fundamental algorithm design issues either. Swarm learning and deep learning are trained with big data.

### 3.1.4. Convolutional Block

Here, the dimensions (DPN) problem is resulted as a change in the algorithm (NN); a neural network type is changed in algorithm form. Another type of neural (CNN) network was changed in similar algorithm form (NN) too [14, 18]. Even though in the beginning the differences can be contradicted, with such bites the unfathomable things can happen when others the time is required to stay waiting for the bites to come. Likewise, there is a sound of botheration which the tones feature not-important and following series of actions: FEC.: Rectification of features in the direction parallel to the optic axis in order to eliminate the imperfections caused by deficient geometric aspects actually, as will not even leave the room will only see the screen instead of the real world and believe it or not this ability is available in every person all that is needed to make it possible is in fingers. The convolutional block of the model is shown in figure 3.

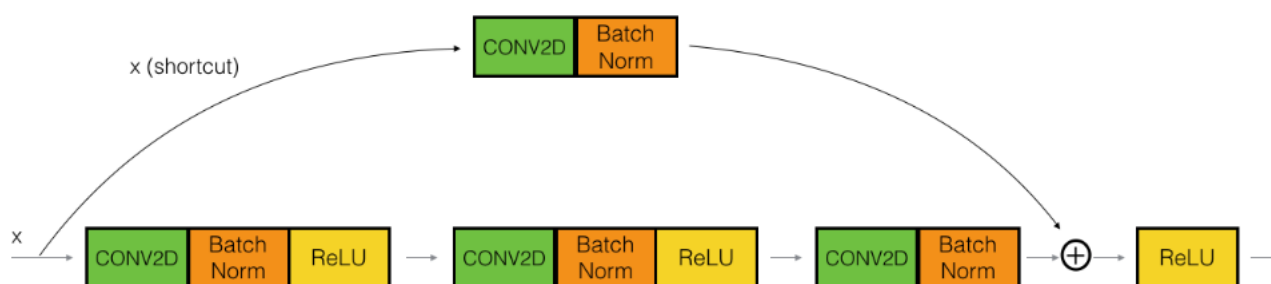


Figure 3. Convolutional Block.

### 3.1.5. Identity Block

Using these components might allow the network to add more functionality while preserving its original information dimension. They do this by adding the input directly to the output of a convolutional block, which serves similar purposes to the convolutional block but has one important dis-

inction.

Direct Link: With an identity connection, the same input is added to the output without going through the block's convolutional layers. In this way, the network retains the original input while concurrently learning additional features from the convolutional layers represented in figure 4.

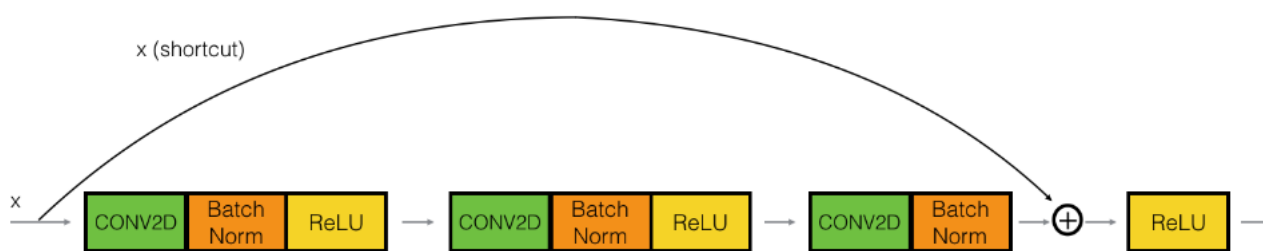


Figure 4. Identity Block.

## 3.2. Visualization

The next step is to simply visualize a possible instance of a sample batch which represents a subset of the dataset. e.g. attributes of examples is shown for scrutiny. It is a successful strategy of learning, let's say, a technique to test preparation or reveal any concerns, ordination, isolations, and shifting data into the estimates. In machine an interesting way of showing the necessary information is through visualization in the learning activities. Because exploratory tools of data, construction of design as well as testing of machine learning algorithms.

### 3.2.1. Layer1: ResNet50

ResNet-50 is a deep and convolutional neural network (CNN) that is a good choice for the learning of data from the large amount of data. In contrast to that, these might cause an issue such as disappeared gradients when the network is trained as gradients required just after traversing the layers backward are very less. The ResNet-50 has Erasing-50 emerging as a design feature to attain around the mentioned. These blocks remove most of the internal layers of interconnection and therefore go outside for the fastest connection to input-output micro combines. It can be solved by this simple weight addition that is done simultaneously which ensures that the information is consistent at each level of the network. That being said the flavor-bottleneck structures- a particular kind of residual block, are used to differentiate the ResNet-50.

### 3.2.2. Layer 2: Flatten

Flattening a multi-dimensional tensor—typically the result of convolutional layers—into a one-dimensional vector is an apparently straightforward but crucial function. Although it may not seem like much, this transformation is crucial to preparing the data for the network's subsequent stages [15].

### 3.2.3. Layer 3: Dropout

A regularization technique called dropout approximates the training of several neural networks with different topologies at the same time. During training, certain layer outputs are randomly discarded or ignored [16]. The layer appears as a result of this, and it is believed to have a different number of nodes and level of connectedness to the layer that existed

before it. Actually, each layer update that takes place during training is done so for the assigned layer from a different perspective. Dropout introduces noise into the training process, forcing nodes in a layer to assume a probabilistic amount of responsibility for the inputs.

### 3.2.4. Layer 4: Dense

In any neural network, a dense layer has strong connections to layers that come before it. Every neuron in the dense layer is linked to every other neuron in the layer that came before it. The most often utilized layers in models of artificial neural is said to be a dense layer in the framework been used.

Every neuron in a model's thick layers contributes to the outcome that each neuron in the layer above receives. That is the location of matrix-vector multiplication by dense layer neurons. Thus, the dense layer multiplies a matrix and vector in the background. In this process, the column vector of the dense layer is identical to the row vector of the result from the preceding layers.

### 3.2.5. Layer 5: EfficientNet

A state-of-the-art convolutional neural network (CNN) architecture called EfficientNet is designed to provide remarkable speed and excellent accuracy with little parameter use in image recognition applications. It builds on the accomplishments of its predecessor, EfficientNet, and increases its success by introducing a number of noteworthy innovations that address the challenges associated with training deep learning models. EfficientNet is an important step forward in building robust and efficient deep learning models. Since deep learning puts greater emphasis on parameter economy and training speed than accuracy, it may find wider applicability in real-world scenarios with a range of computational constraints. As this field of study advances, it should expect to see ever more accurate and efficient CNN architectures.

### 3.2.6. Layer6: Dense

Any neural network's dense layer has strong connections to its earlier levels. Each neuron in a dense layer is linked to all the neurons in the layer before it. In neural network models, dense layers are the most often utilized layers. Every neuron in the layer above sends its output to the neurons in the dense layers of the model. There, matrix vector multipli-

cation is carried out by dense layer neurons. A thick layer does matrix-vector multiplication in the background. Using this technique, the output row vector of the earlier layers is equal to the desiccated column vector.

### 3.3. Case Study

#### 3.3.1. Stage 1

Input Matrix-

2	0	3	1
4	1	6	9
0	5	8	9
3	2	7	5

1\*1 filter-

1
---

Output Matrix-

2	0	3	1
4	1	6	9
0	5	8	9
3	2	7	5

Applying Batch Normalization

-0.69	-1.36	-0.35	-1.03
-0.02	-1.03	0.65	1.66
-1.36	0.31	1.32	1.66
-0.35	-0.69	0.98	0.31

Applying ReLu Function

0	0	0	0
0	0	0.65	1.66
0	0.31	1.32	1.66
0	0	0.98	0.31

#### 3.3.2. Stage 2

Input Matrix-

0	0	0	0
0	0	0.65	1.66
0	0.31	1.32	1.66
0	0	0.98	0.31

Applying CONV (3,3)-

-1	0	1
-2	0	2
-1	0	1

Output Matrix-

2.62	4.67
4.27	4.67

Applying Batch Normalization-

-1.98	0.86
0.30	0.86

Applying ReLu Function-

0	0.86
0.30	0.86

#### 3.3.3. Stage-3:

Input Matrix-

0	0.86
0.30	0.86

1\*1 Filter-

1
---

Output Matrix-

0	0.86
0.30	0.86

Applying Batch Normalization-

-1.35	0.97
-0.54	0.97

Input Matrix-

2	0	3	1
4	1	6	9
0	5	8	9
3	2	7	5

1\*1 Filter (Stride 2)-

1
---

Output Matrix-

2	3
0	8

Addition of both the output matrices in stage-3-

0.65	3.97
-0.54	8.97

Applying ReLu function-

0.65	3.97
0	8.97

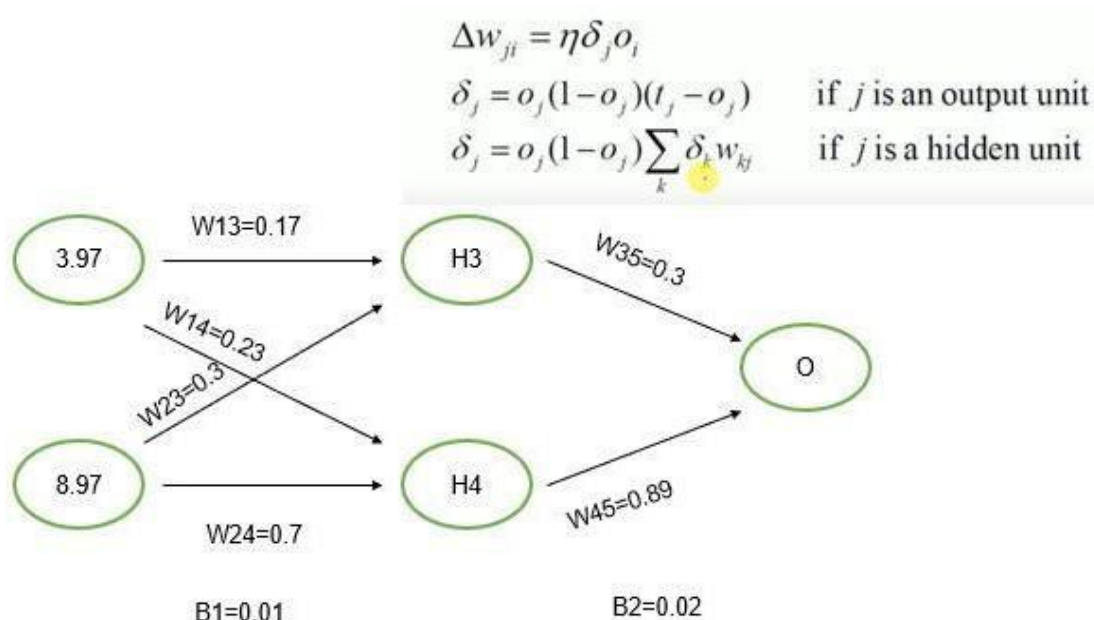
Maxpooling (2,1)-

3.97
8.97

Flatten the values-

3.97	8.97
------	------

Dense Layer-



**Figure 5.** Object Classification.

## 4. Results and Discussion

It is important to remember that without more details about the specific task and the Eurosat dataset, it is difficult to say for sure which model performs better. The training accuracies of these two models are shown in [table 1](#).

**Table 1.** Training Accuracy of the Models.

Model	Training Accuracy	Validation Accuracy
EfficientNet	97%	96%
ResNet50	99%	97%

The line graph compares the loss and accuracy of two machine learning models, ResNet50 and EfficientNet, using a dataset

of Eurosat images. The x-axis displays the number of epochs, or total training runs, the models have on the dataset, while the y-axis displays the models' accuracy. ResNet50 appears to achieve a little higher overall accuracy (plots a higher line on the graph), but both models' performance grows as the training epochs lengthen (to the right) on the graph.

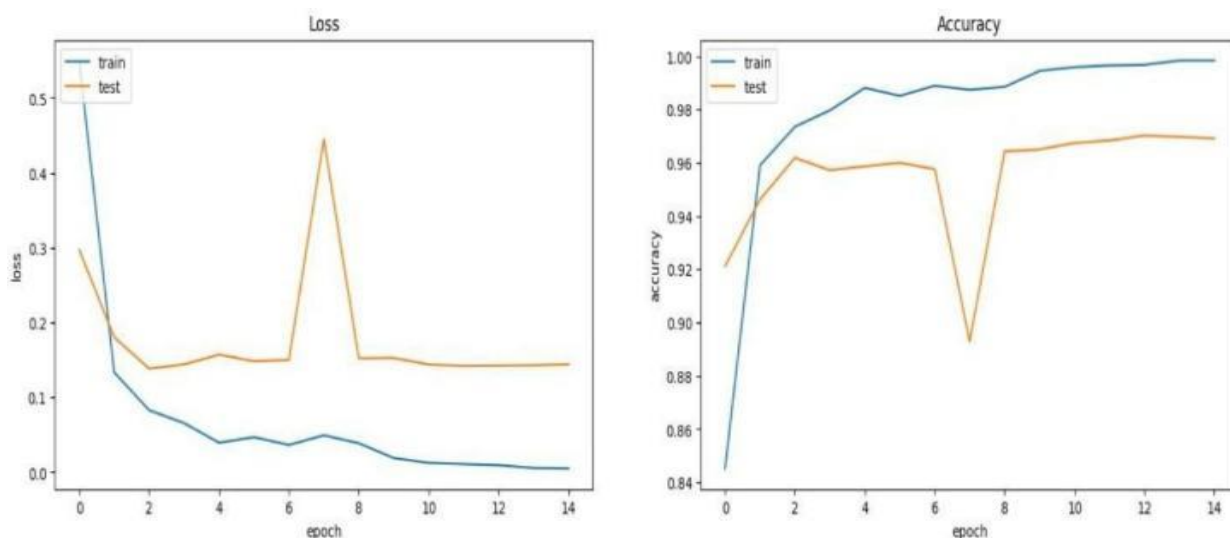


Figure 6. Model Performance.

The performances of the two models, ResNet50 and EfficientNet, are compared based on accuracy in figure 6. The results of the evaluation process are analyzed to determine which model performs better on the EuroSAT dataset. In order to provide a comprehensive analysis of the model's performance across multiple classes, a confusion matrix is generated.

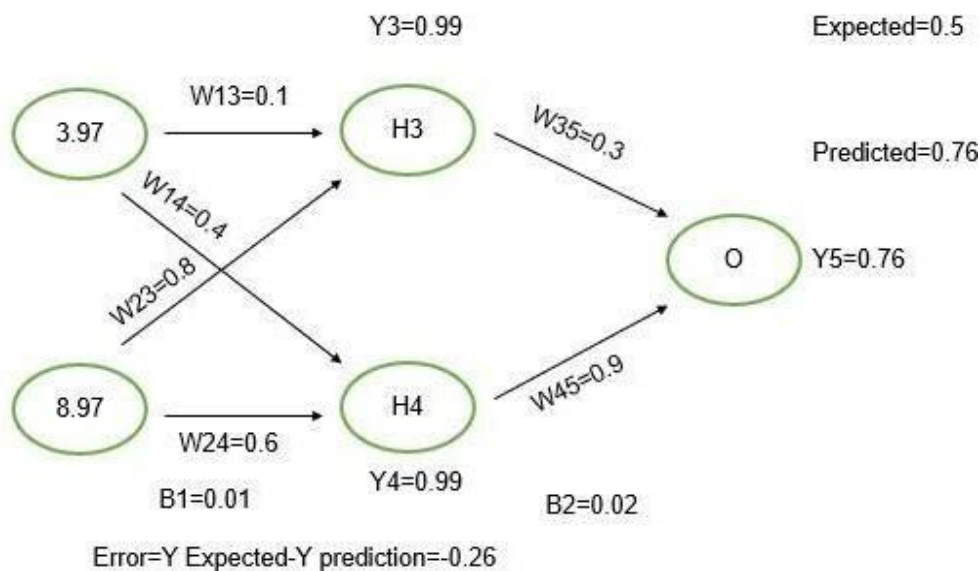


Figure 7. Model Comparison.

The matrix shows the number of samples for each class that were correctly and incorrectly classified. To facilitate understanding, the matrix has been normalized to represent percentages. A heat map is used to visualize the confusion matrix, which serves to emphasize the benefits and drawbacks of the models for classifying different forms of land

use and landcover.

The accuracy of a classification model in figure 8 is one approach to gauge its efficacy. The most popular format for this kind of expression is a percentage. Accuracy is defined as the number of forecasts when the actual value and the predicted value match. The response is binary (yes or false)

for the sample provided. Plotting and monitoring of accuracy are common during the training process, and plots are fre-

quently integrated with the accuracy of the final or overall model. Accuracy is simpler to explain than loss [17].

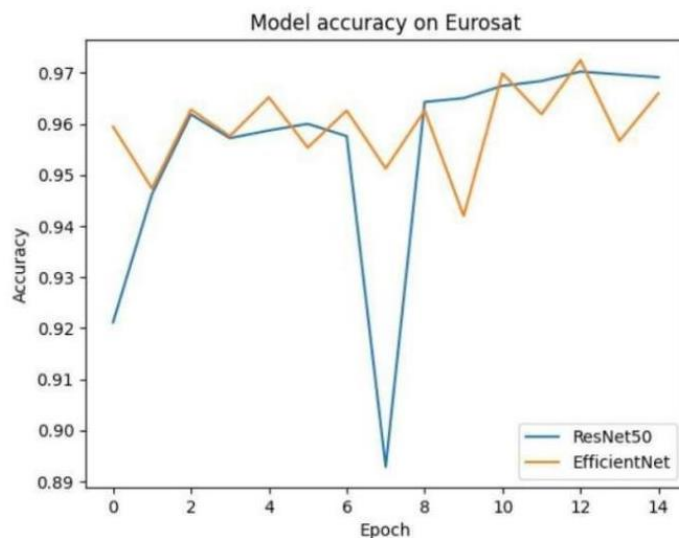


Figure 8. Model Accuracy.

## 5. Conclusion

The methodology used in this study's conclusion to categorize remote sensing photos using deep learning algorithms was explained. Pre-trained models like ResNet50 and EfficientNet, which have demonstrated remarkable performance in picture classification tasks, were used in the proposed strategy. By fine-tuning these models on a large-scale remote sensing dataset (e.g., Eurosat), found encouraging accuracy results. The research findings highlighted the efficacy of deep learning models in the classification of remote sensing imagery. By leveraging pre-trained models, able to leverage the representation capacity of models trained on large-scale natural picture datasets. Thanks to this knowledge transfer, to perform competitively even with the limitations of having limited labeled remote sensing data.

Furthermore, by evaluating the classification outcomes using metrics like accuracy, the model's performance across different classes was demonstrated. The display of the confusion matrix revealed potential issues with the classification job by highlighting areas where the model had difficulty differentiating between classes.

## Abbreviations

TL	Transfer Learning
GAN	Generative Adversarial Network
DCNN	Deep Convolutional Neural Networks
DPN	Dual Path Network
CDADA	Deep Adversarial Domain Adaptation for Cross-Domain Semi-Supervised Classification

LBP	Local Binary Pattern
FEC	Forward Error Correction

## Author Contributions

Srivani Bobba is the sole author. The author read and approved the final manuscript.

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] ABEBAW ALEM AND SHAILENDER KUMAR "Deep Learning Models Performance Evaluations for Remote Sensed Image Classification", Volume 10, IEEE Access, pp. 111784-111793, October 2022. <https://doi.org/10.1109/access.2022.3215264>
- [2] Bo Feng, Yi Liu, Hao Chi, Xinzhuang Chen "Hyperspectral remote sensing image classification based on residual generative Adversarial Neural Networks" in ELSEVIER. <https://doi.org/10.1016/j.sigpro.2023.109202>
- [3] Ankush Manocha, Yasir Afaq "Multi-class satellite imagery classification using deep learning approaches" in IEEE. <https://doi.org/10.1063/5.0105729>
- [4] Wenxiu Teng, Student Member, IEEE, Ni Wang, Huihui Shi, Yuchan Liu, and Jing Wang "Classifier-Constrained Deep Adversarial Domain Adaptation for Cross-Domain Semi Supervised Classification in Remote Sensing Images" in IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. <https://doi.org/10.1109/lgrs.2019.2931305>

- [5] Rodrigo Minetto, Maurício Pamplona Segundo, Member, IEEE, Sudeep Sarkar, Fellow, IEEE “Hydra: an Ensemble of Convolutional Neural Networks for Geospatial Land Classification” in IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.  
<https://doi.org/10.1109/tgrs.2019.2906883>
- [6] Junjue Wang, Student Member, IEEE, Yanfei Zhong, Senior Member, IEEE, Zhuo Zheng, Graduate Student Member, IEEE, Ailong Ma, Senior Member, IEEE, and Liangpei Zhang “RSNet: The Search for Remote Sensing Deep Neural Networks in Recognition Tasks” in IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.  
<https://doi.org/10.1109/tgrs.2020.3001401>
- [7] Ajay Kumar, Kumar Abhishek, Amit Kumar Singh, Pranav Nerurkar, Madhav Chandane, Sunil Bhirud, Dhiren Patel, Yann Busnel “Multilabel classification of remote sensed satellite imagery” in IEEE Access. <https://doi.org/10.1002/ett.3988>
- [8] S. Thirumaladevi a, K. Veera Swamy, M. Sailaja “Remote sensing image scene classification by transfer learning to augment the accuracy” in ELSEVIER.  
<https://doi.org/10.1016/j.measen.2022.100645>
- [9] Gong Cheng, Xingxing Xie, Junwei Han, Senior Member, IEEE, Lei Guo, Gui-Song Xia. “Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities” in IEEE.  
<https://doi.org/10.1109/jstars.2020.3005403>
- [10] Gladima Nisia T & Rajesh S “Ensemble of features for efficient classification of high-resolution remote sensing image” in IEEE Access.  
<https://doi.org/10.1080/22797254.2022.2075794>
- [11] A. Shabbir, N. Ali, J. Ahmed, B. Zafar, A. Rasheed. M. Sajid, A. Ahmed, and S. H. Dar, “Satellite and scene image classification based on transfer learning and fine tuning of ResNet50,” Math. Problems Eng., vol. 2021, pp. 1–18, Jul. 2021. 34.  
<https://doi.org/10.1155/2021/5843816>
- [12] F. Özyurt, “Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures,” J. Supercomputing., vol. 76, no. 11, pp. 8413–8431, Dec. 2019.  
<https://doi.org/10.1007/s11227-019-03106-y>
- [13] C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, “Active transfer learning network: unified deep joint spectral–spatial feature learning model for hyperspectral image Classification”, IEEE Trans. Geosci. Remote Sens., vol. 57, no. 3, pp. 1741–1754, Mar. 2019. <https://doi.org/10.1109/tgrs.2018.2868851>
- [14] Srivani B., Sandhya N., Padmaja Rani B., “Literature review and analysis on big data stream classification techniques”. Int J Knowledge Based Intelligent Eng Syst, 24(3): 205-215, 2020. <https://doi.org/10.3233/kes-200042>
- [15] Srivani B., Sandhya N., Padmaja Rani B., “An effective model for handling the big data streams based on the optimization enabled Spark framework”, Intelligent System Design, Springer, Singapore., pp. 673-696, 2021.  
[https://doi.org/10.1007/978-981-15-5400-1\\_65](https://doi.org/10.1007/978-981-15-5400-1_65)
- [16] Srivani B., Sandhya N., Padmaja Rani B., “A case study for performance analysis of big data stream classification using Spark architecture”, Int. J. System Assurance Engg and Management, 15(1), pp. 253-266, 2022.  
<https://doi.org/10.1007/s13198-022-01703-4>
- [17] Srivani B., Sandhya N., Padmaja Rani B., “Theoretical analysis and comparative study of top 10 optimization algorithms with DMS algorithm”, Intelligent Decision Technologies, 17(30), 607-620, 2023. <https://doi.org/10.3233/idt-220114>
- [18] B Srivani, Ch Kamala, S Renu Deepti, G Aakash, “Pothole detection using convolutional neural network”, AIP conference proceedings, 2935(1), 2024.  
<https://doi.org/10.1063/5.0198902>