Deep Learning-Based Satellite Image Classification for Land Use and Land Cover Analysis

Kartik Rodagi
IT dept
211IT029
NITK, Surathkal 57025

Manjunath Harale
IT dept
211IT039
NITK, Surathkal 57025

Adish Kerkar
IT dept
211IT004
NITK, Surathkal 57025

***Abstract*—Satellite image analysis plays a crucial role in understanding and monitoring Earth's land use and land cover changes. In this project, we present a comprehensive analysis of land use and cover classification using deep learning models applied to the EuroSAT dataset. The dataset consists of 27,000 labeled images covering 10 land use and cover classes, acquired from the Sentinel-2A satellite. Our main objective is to demonstrate the effectiveness of deep learning models, specifically VGG19 and a custom CNN, in classifying satellite images for real-world applications.**

***Index Terms*—Satellite Image Analysis, Deep Learning, VGG19, Convolutional Neural Networks (CNN), EuroSAT Dataset, Land Use and Land Cover Classification, Earth Observation, Sentinel-2A Satellite**

I. INTRODUCTION

Economic development and population growth have been significant drivers of rapid changes to Earth's land cover over the past centuries. Moreover, there are strong indications that the pace of these changes will continue to accelerate in the future. Hence, it is essential to conduct a comprehensive analysis in this field to understand and address the implications effectively. We are currently at the edge of having public and continuous access to satellite image data for Earth observation. With access to such data, applications in the domains of agriculture, disaster recovery, climate change, urban development, or environmental monitoring can be realized One type of such fundamental semantics is Land Use and Land Cover Classification. The aim of land use and land cover classification is to automatically provide labels describing the represented physical land type or how a land area is used Although the terms "land cover (LC)" and "land use (LU)" are sometimes used interchangeably, they are actually different. Simply put, land cover is what covers the surface of the earth and land use describes how the land is used. This report focuses on the classification and analysis of high-resolution satellite images of land cover to examine patterns in land use and land cover.

In recent times, Deep Learning, a sophisticated tool in the field of machine learning, has demonstrated its effectiveness in the realm of computer vision.The conventional machine learning tools such as Support Vector Machine (SVM) and Random Forest (RF) which are shallow-structured, have major limitations that are addressed by these advanced machine learning algorithms.The Prominent deep learning model such as deep Convolutional Neural Network (CNN) have shown promising results in classification. In this report we present deep learning approaches based on Convolutional Neural Networks for classification of satellite imagery dataset of land use and land cover.

II. DATASET

In this section, we present the EuroSAT dataset, a novel satellite image dataset designed for land use and land cover classification. [14]The dataset consists of 27,000 labeled images covering 10 different land use and land cover classes. Unlike previous datasets, EuroSAT is multi-spectral, encompassing 13 spectral bands in the visible, near infrared, and short wave infrared parts of the spectrum. Additionally, the dataset is georeferenced and based on openly and freely accessible Earth observation data, enabling a wide range of applications.

A. Dataset Acquisition

The European Space Agency (ESA) has intensified its Earth observation efforts through the Copernicus program, which includes the operation of the Sentinel satellites. For this study, we utilized multispectral image data from the Sentinel-2A satellite to tackle the challenge of land use and land cover classification. Sentinel-2A is part of a two-satellite constellation, along with Sentinel-2B, both launched successfully in June 2015 and March 2017, respectively. These satellites capture the Earth's land surface using a Multi-spectral Imager (MSI) that covers 13 spectral bands. While some bands are used for atmospheric correction, the others are primarily for land use and land cover classification. The satellites provide imagery with a spatial resolution of up to 10 meters per pixel, covering mainland, large islands, inland, and coastal waters. Each satellite's operational life is expected to be at least 7 years, with the potential for extension.

The Sentinel satellites offer a short repeat cycle, capturing each point in the covered area approximately every five days. This, coupled with their expected operational life of 20-30 years, allows for continuous monitoring of the Earth's land surface. Importantly, the data from Sentinel satellites are openly and freely accessible for both commercial and non-commercial use.

To construct our image classification dataset, we undertook the following steps:

- Satellite Image Acquisition: We obtained satellite images of European cities from the Sentinel-2A satellite via Amazon S3. These cities are distributed across 34 European countries.
- Dataset Creation: Using the acquired satellite images, we created a dataset consisting of 27,000 georeferenced and labeled image patches, each measuring 64x64 pixels. The dataset underwent manual checks to ensure accuracy.

This EuroSAT dataset was created to address the lack of labeled ground truth data for supervised machine learning applications using satellite data. It consists of 10 different classes, each with 2,000 to 3,000 images, totaling 27,000 images. The dataset includes classes such as annual crop, permanent crop (e.g., fruit orchards, vineyards), pastures, built-up areas (highway, residential buildings, industrial buildings), water bodies (river, sea lake), and undeveloped environments (forest, herbaceous vegetation). This dataset is intended to be a valuable resource for various Earth observation applications, particularly in conjunction with powerful machine learning methods.

B. Class Distribution

The dataset is balanced, with an equal distribution of images across the 10 land use and land cover classes. This balance ensures that the classification model is trained on a representative sample of each class, avoiding bias towards any particular category.

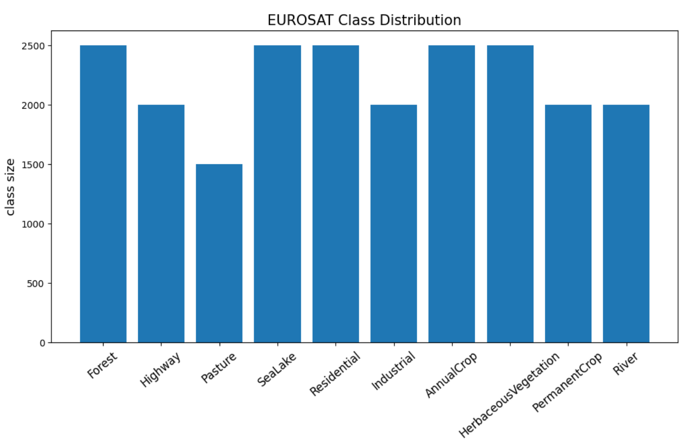


Fig. 1. EUROSAT Class Distribution

C. Data Access

The EuroSAT dataset is publicly available and can be accessed through the project's GitHub repository link[link]. This open access facilitates reproducibility and allows other researchers to use the dataset for their own experiments and applications.

D. Potential Applications

The EuroSAT dataset enables a wide range of applications, including:

- Land use and land cover classification
- Detection of land use or land cover changes over time
- Improvement of geographical maps through automated classification

E. Sample Data images

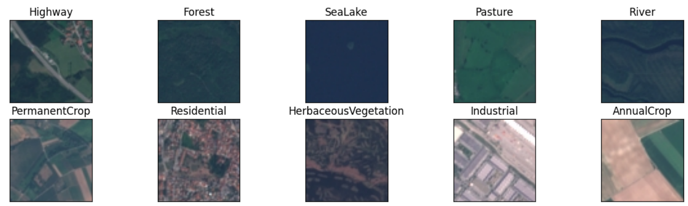


Fig. 2. Example images

III. Related Works

Recently, deep learning techniques have gained popularity for remote sensing image classification. Bosco et al. [1] introduced a neural network encoding architecture that uses pre-trained CNN models like InceptionV3, InceptionResNetV2, VGG16, and DenseNet201. They incorporated activation functions and ensemble learning to extract features effectively.

Mahdianpari et al. [2] also used deep learning tools based on CNNs for classifying complex wetland classes in Canada using multispectral RapidEye optical imagery. They explored seven deep convnets, including DenseNet121, InceptionV3, VGG16, VGG19, Xception, ResNet50, and InceptionResNetV2 for wetland classification and mapping. InceptionResNetV2, ResNet50, and Xception were identified as the top performers, providing state-of-the-art classification accuracies for complex remote sensing scenes like wetlands.

Alhichri et al. [13] proposed a Deep Convolutional Neural Network (CNN) with an attention mechanism for scene classification in remote sensing. The approach computes a new feature map by weighting the original feature maps. The CNN, named EfficientNet-B3-Attn-2, was developed by enhancing the pre-trained EfficientNet-B3 CNN with an attention mechanism. The study demonstrated the effectiveness of the approach on six remote sensing datasets, showing strong performance in accurately classifying remote sensing images and scenes.

The progress of classification in remote sensing has been hindered by the lack of reliably labeled ground truth datasets. One popular and extensively studied dataset in this area is the UC Merced (UCM) land use dataset introduced by Yang et al. [7]. This dataset contains 21 land use and land cover classes, with each class comprising 100 images.

The images are 256x256 pixels in size, with a spatial resolution of approximately 30 cm per pixel. All images are in the RGB color space and were extracted from the USGS National Map Urban Area Imagery collection, sourced from aircraft. However, a dataset with only 100 images per class is considered small-scale. To address this limitation, various efforts have used commercial Google Earth images to manually create new datasets [8], [6], [9], [10], such as the two benchmark datasets PatternNet [11] and NWPU-RESISC45 [12]. These datasets are based on very-high-resolution images with a spatial resolution of up to 30 cm per pixel.

IV. Methodology

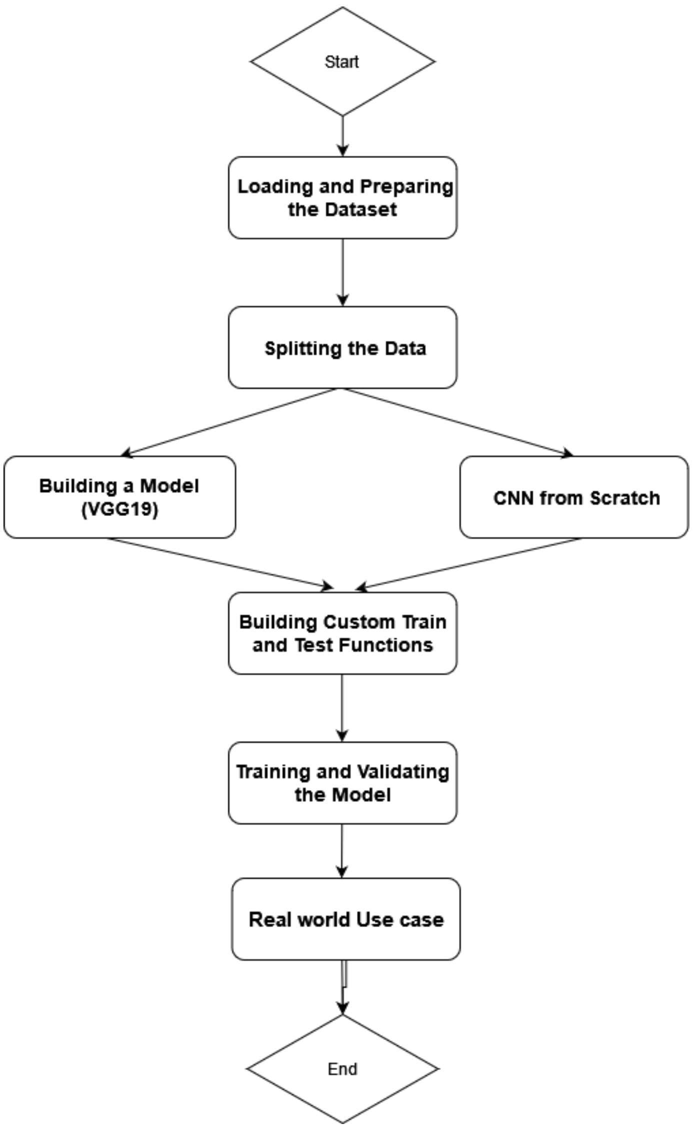


Fig. 3. Methodology Diagram

1) Loading and Preparing the Dataset: We begin by loading the EuroSAT dataset, which consists of 27,000 labeled images covering 10 land use and land cover classes. These images are preprocessed by resizing them to 64x64 pixels and normalizing the pixel values to ensure consistency. The dataset is then sorted into their respective classes for easy access during training and testing. Exploratory data analysis is performed to gain insights into the dataset's structure and content, which is crucial for subsequent steps.

2) Splitting the Data: To evaluate the model's performance, we split the dataset into training (80%) and testing (20%) sets using a stratified approach to maintain class balance in both sets. This ensures that the model is trained and tested on a representative sample of the entire dataset. The data is shuffled to introduce randomness in the split, preventing any bias in the training and testing sets.

3) Building a Model (VGG19): We utilize the VGG19 model, pre-trained on the ImageNet dataset, as the base model for our classification task. The pre-trained weights are retained in the initial layers, which are frozen to preserve the learned features. We then customize the model by adding a Global Average Pooling layer and a Dense output layer with 10 units (one for each class) and softmax activation. This configuration allows the model to learn class-specific features and make predictions accordingly.

4) Building Custom Train and Test Functions: For efficient training and testing of the model, we develop custom functions tailored to our dataset. The training function implements data augmentation techniques such as random rotation, horizontal and vertical flips, and zoom to increase the diversity of the training samples. It iterates over the training dataset in batches, applies data augmentation, and updates the model weights based on the computed gradients during backpropagation. The testing function evaluates the model's performance on the testing dataset without updating the weights.

5) Training and Validating the Model: The model is trained on the training dataset using the custom training function for a specified number of epochs. We monitor the training process by tracking metrics such as accuracy, loss, and validation accuracy. The model's performance is evaluated on the testing dataset using the custom testing function to assess its generalization ability. This iterative process helps us fine-tune the model and improve its performance over time.

6) Building Another CNN from Scratch: In addition to the VGG19-based model, we design a custom CNN architecture from scratch for comparison purposes. This CNN consists of multiple convolutional and pooling layers followed by fully connected layers. We compile the model with the same configuration as the VGG19 model and train it on the training dataset. The model is then validated on the testing dataset to evaluate its performance.

7) Processing Sample Data for a city (Mangalore): To demonstrate the model's applicability to real-world scenarios, we select satellite images of Mangalore. These images are processed into 64x64 pixel formats to match the input requirements of the models. The goal is to showcase how the models can be used to process and classify satellite images of specific regions of interest.

8) Processing and Classifying Images: Using the VGG19-based model, we process and classify the selected images of Mangalore. The models classify the images into the 10 land use and cover classes, providing valuable insights into the land use patterns of the region. By comparing the classifications from different models, we can evaluate their effectiveness in accurately classifying satellite images.

9) Visualizing Changes in the Terrain: To visualize changes in the terrain over time, we select pairs of images of Mangalore taken at different times, such as 4 years apart. These image pairs are processed and classified using the trained models to identify changes in land use and land cover. By visualizing these changes, we can gain a better understanding of how the region has evolved over time and assess the impact of various factors such as urbanization, agriculture, and natural disasters.

V. Models

A. VGG19 Model

1) Model description: VGG19 is a convolutional neural network (CNN) model belonging to the VGG (Visual Geometry Group) family of models. It was introduced by the Visual Geometry Group at the University of Oxford in 2014. VGG19 is named after the number of layers it has, specifically 19 layers, including 16 convolutional layers and 3 fully connected layers.

1) **Input Layer:** The input to the VGG19 model is an RGB image with a size of 224x224 pixels.
2) **Convolutional Layers:** VGG19 consists of 16 convolutional layers, each followed by a ReLU activation function and a 3x3 filter size. The convolutional layers are designed to extract features from the input image.
3) **Max Pooling Layers:** After some of the convolutional layers, max pooling layers with a 2x2 filter and stride 2 are used to downsample the spatial dimensions of the feature maps, reducing computational complexity and controlling overfitting.
4) **Fully Connected Layers:** The last three layers of VGG19 are fully connected layers. The first two fully connected layers have 4096 units each, followed by a dropout layer with a dropout rate of 0.5 to reduce overfitting. The final fully connected layer has 1000 units, corresponding to the 1000 classes in the ImageNet dataset (for which VGG was originally designed).
5) **Output Layer:** The output layer uses a softmax activation function to produce the final class probabilities for the input image.

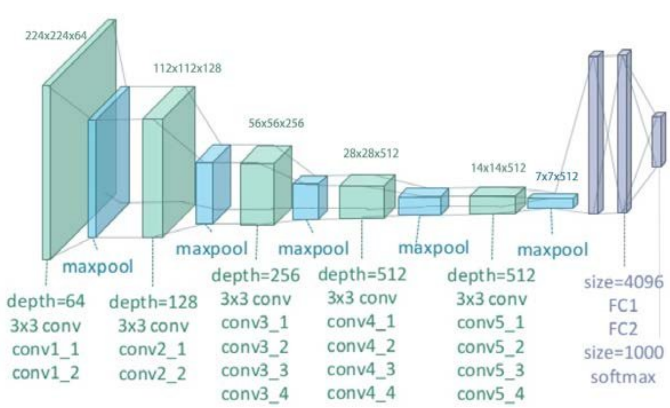


Fig. 4. Model Architecture(VGG19)

VGG19 is known for its simplicity and effectiveness in image classification tasks.

2) Customizing the VGG19 Model: The buildmodel function modifies a pre-trained VGG19 model for a specific image classification task. It first freezes the pre-trained layers to retain their learned features. Then, it replaces the average pooling layer with an adaptive one and modifies the classifier layers to suit the new task. The classifier is customized with a sequence of operations including flattening, linear transformations, ReLU activation, dropout for regularization, and a final linear layer for classification. Additionally, it defines the loss function as cross-entropy loss and initializes the Adam optimizer for training. Overall, these modifications help the VGG19 model to efficiently classify images for the proposed task.

B. Custom CNN

1) Custom CNN Model Architecture: The model is designed for land use and land cover classification using the EuroSAT dataset. Here's a detailed explanation of the model architecture:

- CNN with three layers (32, 64, 128)
- Kernel size: 5x5 (padding: 2)
- Max pooling after each layer
- Fully connected NN with three layers (2048, 2048, 10)
- Activation function: ReLu (last layer: Softmax)

1) **Input:** The model takes as input satellite images represented as tensors with shape '(batchsize, 13, height, width)', where 'batchsize' is the number of images in a batch, and 'height' and 'width' are the spatial dimensions of the images.
2) **Convolutional Layers:**
   - The model starts with a convolutional layer 'conv1' that has 13 input channels (corresponding to the 13 spectral bands) and 32 output channels. It uses a kernel size of 7x7 and padding of 3 to preserve spatial dimensions.
   - This is followed by a ReLU activation function to introduce non-linearity.
3) **Max Pooling:**
   - After the first convolutional layer, the model applies max pooling ('pool1') with a kernel size of 2x2 and a stride of 2. This reduces the spatial dimensions of the feature maps by half.
4) **Convolutional and Pooling Layers:**
   - The model then adds another set of convolutional ('conv2') and max pooling ('pool2') layers with similar configurations to further extract features and reduce spatial dimensions.
5) **Fully Connected (FC) Layers:**
   - Next, the model flattens the features and passes them through two fully connected layers ('linear1' and 'linear2') with ReLU activation functions. These layers help in learning high-level features relevant to land use and land cover classification.

- The output of the last FC layer is passed through a LogSoftmax activation function (‘logsoftmax‘) to obtain the final class probabilities.

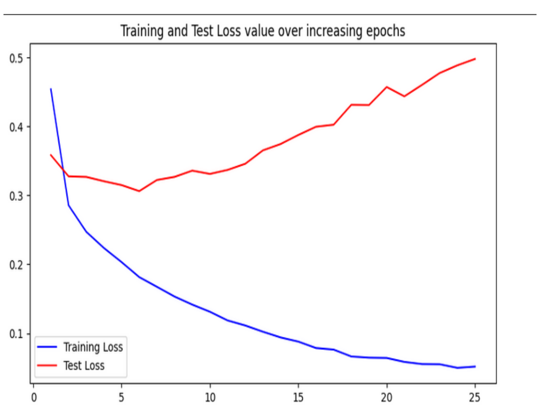
6) **Output:**
   - The output of the model is a tensor of shape ‘(batchsize, numclasses)‘, where ‘numclasses‘ is 10 (the number of land use and land cover classes in the EuroSAT dataset). Each value in the output tensor represents the predicted probability of the corresponding class.

Overall, the ‘EurosatNet‘ model uses convolutional and fully connected layers to extract features from satellite images and classify them into different land use and land cover classes. The model architecture is designed to effectively capture spatial and spectral information from the input images to make accurate predictions.

VI. Results

A. VGG19 Model

Epoch	Train Loss	Train Acc	Eval Loss	Test Acc
1	0.4537	90.48%	0.3582	87.26%
2	0.2855	92.42%	0.3275	88.94%
3	0.2472	93.09%	0.3267	88.78%
4	0.2236	93.78%	0.3204	89.22%
5	0.2032	94.80%	0.3148	89.58%
6	0.1814	95.41%	0.3061	90.08%
7	0.1674	95.85%	0.3222	89.94%
8	0.1532	96.35%	0.3267	89.96%
9	0.1416	96.36%	0.3358	90.08%
10	0.1312	97.33%	0.3310	90.48%
11	0.1186	97.65%	0.3368	90.24%
12	0.1113	97.75%	0.3457	89.72%
13	0.1023	98.01%	0.3652	89.96%
14	0.0940	98.26%	0.3744	90.40%
15	0.0880	98.34%	0.3874	90.34%
16	0.0786	98.44%	0.3994	89.90%
17	0.0764	98.72%	0.4023	90.12%
18	0.0664	98.86%	0.4312	89.90%
19	0.0646	99.04%	0.4309	90.10%
20	0.0642	98.84%	0.4570	89.74%
21	0.0585	99.23%	0.4434	90.32%
22	0.0554	99.32%	0.4601	90.26%
23	0.0551	99.39%	0.4772	89.90%
24	0.0499	99.47%	0.4884	89.86%
25	0.0517	99.32%	0.4977	89.52%



Overall, the model shows a consistent decrease in both training loss and evaluation loss over the epochs, indicating

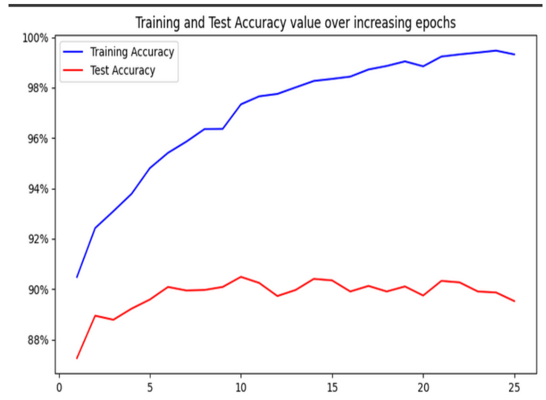


Fig. 5. Accuracy of VGG219 Model

that it is learning from the data. The training accuracy increases steadily, reaching a high of 99.47% by the end of training. However, the test accuracy fluctuates slightly but remains relatively stable, reaching 90.52% at the end of training. This indicates that the model is generalizing well to unseen data.

B. Custom CNN

Epoch	Train Loss	Eval Loss	Train Acc	Test Acc
180	0.0271	0.1188	0.9926	0.9718
181	0.0253	0.1168	0.9923	0.9713
182	0.0313	0.1329	0.9917	0.9689
183	0.0255	0.1438	0.9926	0.9641
184	0.0273	0.1219	0.9933	0.9705
185	0.0259	0.1338	0.9924	0.9676
186	0.0252	0.1237	0.9929	0.97
187	0.0244	0.1447	0.9934	0.9649
188	0.0219	0.1207	0.994	0.9722
189	0.0245	0.1213	0.9926	0.9707
190	0.0211	0.1313	0.9941	0.9686
191	0.0192	0.1161	0.9951	0.9736
192	0.0211	0.1275	0.9942	0.97
193	0.0217	0.1234	0.9941	0.9715
194	0.0193	0.1394	0.995	0.9683
195	0.0385	0.1177	0.9919	0.9726
196	0.0202	0.1241	0.9944	0.9707
197	0.0177	0.1416	0.9953	0.9675
198	0.0183	0.1158	0.9952	0.9728
199	0.0182	0.1229	0.9955	0.972

The model trained on the dataset achieved impressive results, with a high training accuracy of around 99% and a test accuracy of around 97%. These results indicate that the model has learned to generalize well to unseen data, achieving a good balance between fitting the training data and generalizing to new data. The evaluation loss, which measures how well the model is performing on the validation set, also remained relatively low throughout training, indicating that the model's performance is consistent. These results demonstrate the effectiveness of the model in accurately classifying remote sensing images, making it suitable for real-world applications.

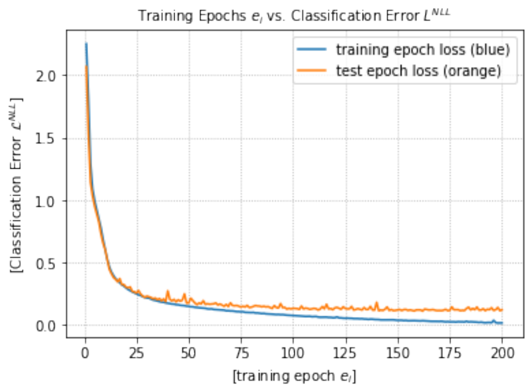


Fig. 6. Training Epochs vs.Classification Error(custom CNN

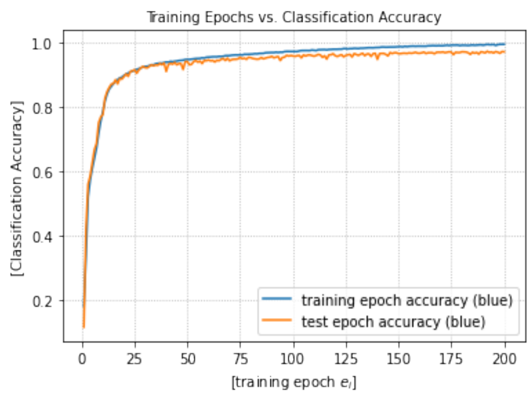


Fig. 7. Training Epochs vs. Classification accuracy(custom CNN)

VII. Real World Use case

A. Demonstrating Model's Functionality with Real-world Satellite Images

In the context of the project, the following sections describe the process of demonstrating the model's functionality and effectiveness using real-world satellite images of Mangalore.

1) Processing Sample Data for a City (Mangalore): Satellite images of Mangalore city is processed into a standard format (64x64 pixels) suitable for input to the models. The goal is to prepare the data for classification and showcase the model's ability to handle images from specific regions.

2) Processing and Classifying Images: The processed images of Mangalore are used as input to the VGG19-based model for classification. The models classify the images into 10 land use and cover classes, providing insights into the region's land use patterns. Comparing classifications from different models helps evaluate their effectiveness in accurately classifying satellite images.

3) Visualizing Changes in the Terrain: To visualize changes in the terrain over time, pairs of images taken at different times (e.g., 4 years apart) are selected. These image pairs are processed and classified using the trained models to identify changes in land use and cover. This visualization helps understand how the region has evolved and assess the impact of various factors such as urbanization, agriculture, and natural disasters.

Fig. 8. Snapshot of the city

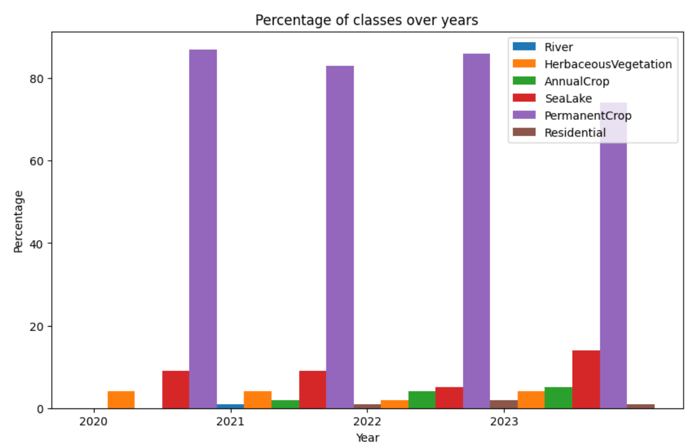


Fig. 9. Land cover over the years

VIII. Conclusion

In this project, we have developed and implemented satellite image classification models for land use and land cover analysis using deep learning techniques. We have utilized the EuroSAT dataset, which consists of 27,000 labeled images covering 10 different land use and land cover classes. Our goal was to demonstrate the effectiveness of these models in real-world scenarios and showcase their potential applications.

A. Models Used

- **VGG19-based Model:** We first utilized a pretrained VGG19 model and customized it to our needs by replacing the final fully connected layer with a new layer for our classification task. This model achieved an accuracy of around 90percent on the EuroSAT dataset.

- **Custom CNN Model:** Additionally, we built a custom CNN model from scratch, which also achieved competitive accuracy on the dataset.nm

B. Real-World Applications

Our models have various real-world applications, including:

- **Urban Planning:** Analyzing urban growth patterns, infrastructure development, and monitoring changes in land use for sustainable planning.
- **Disaster Management:** Assessing the impact of natural disasters and aiding in rescue and relief operations.
- **Agricultural Monitoring:** Monitoring crop health, soil conditions, and water usage for improved agricultural practices.
- **Environmental Conservation:** Monitoring deforestation, habitat loss, and environmental changes for conservation efforts.

C. Implementation Example: Mangalore

To demonstrate the practical use of our models, we implemented a time analysis of the city of Mangalore using satellite images. We processed and classified sample satellite images of Mangalore to identify changes in land use and cover over time. This analysis provides valuable insights into the city's development and can aid in urban planning and environmental management.

D. Future Scope

- **Enhanced Model Performance:** Further optimization and fine-tuning of the models can improve their accuracy and efficiency.
- **Integration with Remote Sensing Technologies:** Integrating satellite data with other remote sensing technologies can provide more comprehensive insights into land use and cover dynamics.
- **Application in Other Regions:** Expanding the application of our models to other regions can provide valuable insights into global land use patterns and environmental changes.

In conclusion, our project demonstrates the potential of deep learning models in satellite image analysis for land use and cover classification. These models have significant implications for various real-world applications and can contribute to informed decision-making for sustainable development and environmental conservation.

REFERENCES

[1] Bosco JM, Wang G, Hategekimana Y. Learning multi-granularity neural network encoding image classification using DCNNs for Easter Africa Community Countries. IEEE Access. 2021;9:146703–18.

[2] Mahdianpari M, Salehi B, Rezaee M, Mohammadimanesh F, Zhang Y. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sens. 2018;10(7):1119.

[3] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva. Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv:1508.00092, 2015.

[4] K. Nogueira, O. A. Penatti, and J. A. dos Santos. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61:539–556, 2017.

[5] O. A. Penatti, K. Nogueira, and J. A. dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 44–51, 2015.

[6] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, and L. Zhang. Aid: A benchmark dataset for performance evaluation of aerial scene classification. arXiv preprint arXiv:1608.05167, 2016.

[7] Y. Yang and S. Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, pages 270–279. ACM, 2010.

[8] G. Sheng, W. Yang, T. Xu, and H. Sun. High-resolution satellite scene classification using a sparse coding based multiple feature combination. International journal of remote sensing, 33(8):2395–2412, 2012.

[9] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maˆıtre. Structural high-resolution satellite image indexing. In ISPRS TC VII Symposium-100 Years ISPRS, volume 38, pages 298–303, 2010.

[10] L. Zhao, P. Tang, and L. Huo. Feature significance-based multibag-ofvisual-words model for remote sensing image scene classification. Journal of Applied Remote Sensing, 10(3):035004–035004, 2016

[11] Weixun Zhou, Shawn Newsam, Congmin Li, and Zhenfeng Shao. Patternnet: a benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS Journal of Photogrammetry and Remote Sensing, 2018.

[12] Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE, 105(10):1865–1883, 2017.

[13] Alhichri H, Alswayed AS, Bazi Y, Ammour N, Alajlan NA. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. IEEE Access. 2021;9:14078–94.

[14] Helber, P., Bischke, B., Dengel, A., Borth, D. (2019). EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7), 2217-2226.