

Aerial Scene Classification using VGG16 and
Multiclass Linear SVM

PROJECT REPORT

Submitted by

BINTU K BABU

REG NO : TKM21CSCE04

In partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Under the guidance of
Prof. Shyna A & Prof. Jini Raju



Thangal Kunju Musaliar College of Engineering
Kerala

JULY 2023

DECLARATION

I undersigned hereby declare that the project report on “**Aerial Scene Classification using VGG16 and Multiclass Linear SVM**”, submitted for partial fulfillment of the requirements for the award of the degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under the supervision of **Prof. Shyna A & Prof. Jini Raju**, Assistant Professor, Department of the Computer Science and Engineering , TKMCE. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Kollam
7/07/2023

BINTU K BABU

Thangal Kunju Musaliar College of Engineering
Dept. of Computer Science & Engineering



C E R T I F I C A T E

This is to certify that, this report titled *Aerial Scene Classification using VGG16 and Multiclass Linear SVM* is a bonafide record of the **Project** presented by **BINTU K BABU (TKM21CSCE04)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Computer Science & Engineering** in **APJ Abdul Kalam Technological University** .

Project Guide

Prof. Shyna A & Prof. Jini Raju
Assistant Professor
Computer Science & Engineering

Head Of The Department
Dr. Dimple A Shajahan
Professor

Computer Science & Engineering

Project Coordinator

Dr. Aneesh G Nath
Associate Professor
Computer Science & Engineering

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings.

I express my sincere gratitude to **Dr. T A Shahul Hameed**, Principal of T.K.M College of Engineering for giving me an opportunity to present project. I would like to thank **Dr. Dimple A Shajahan**, Professor and Head of the Department, CSE, TKMCE, for her constant support and encouragement throughout the work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Prof. Shyna A & Prof. Jini Raju**, Assistant Professor, CSE, TKMCE for their expert guidance, cooperation and immense encouragement. I also extend my thanks to the entire faculty members and staffs of the Department of Computer Engineering, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents, brother and friends, for their support and encouragement in the successful completion of this project work.

Kollam
7/07/2023

BINTU K BABU

Abstract

Aerial scene classification is the process of categorizing and analyzing images captured from an aerial perspective, enabling the identification of land cover, objects, and scene composition for various applications. Aerial scene classification plays a crucial role in various fields, including urban planning, environmental monitoring, and disaster management, by providing valuable insights into land cover, objects, and scene composition from an aerial perspective. Accurate classification of aerial scenes enables effective decision-making, resource allocation, and informed analysis of large-scale imagery, contributing to improved spatial understanding and efficient management of diverse landscapes. In this work, classification of aerial images using combination of VGG16 and Multiclass Linear SVM classifier is proposed. Deep Features are extracted using VGG16 and Multiclass Linear SVM classifier is using to classify the given objects. The preprocessed steps include data augmentation, data normalization, feature extraction using a VGG16 model, and training a multiclass linear SVM classifier for aerial scene classification. The experiments are conducted on NWPU and UCM dataset and performance is evaluated using confusion matrix, precision and recall. The experimental result shows the proposed method yield 90% accuracy for NWPU dataset and 95% accuracy for UCM dataset.

keywords: SVM:Support Vector Machine, VGG16:Visual Geometry Group 16, Feature extraction.

Contents

1	Introduction	1
2	Related Works	5
2.1	CondenseNet: An efficient DenseNet using learned group convolutions	5
2.2	MIML-FCN+: Multi-instance multi-label learning via fully convolutional networks with privileged information	5
2.3	Revisiting Multiple Instance Neural Networks	6
2.4	Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification	6
2.5	Learning a discriminative distance metric with label consistency for scene classification,	7
3	Proposed Network	8
4	Experiments And Results	12
4.1	Experimental Settings	12
4.2	Datasets	12
4.2.1	UCM	12
4.2.2	NWPU	13
4.3	Results	13
4.3.1	Confusion Matrix	13
4.3.2	Confusion matrix of SVM classifier and Softmax classifier	13
5	Conclusion	28
	References	29

List of Figures

3.1	VGG16 MODEL ARCHITECTURE	9
4.1	NWPU Dataset using svm classifier	14
4.2	UCM Dataset using svm classifier	15
4.3	NWPU Dataset using softmax classifier	16
4.4	UCM Dataset using softmax classifier	17
4.5	Overall accuracy of svm and softmax classifier on two datasets	18
4.6	Accuracy of average F1 score for SVM on NWPU dataset . .	19
4.7	Confusion matrix accuracy for SVM on NWPU dataset . . .	20
4.8	Accuracy Comparison for SVM on NWPU dataset	21
4.9	Accuracy of average F1 score for SVM on UCM dataset . . .	21
4.10	Confusion matrix accuracy for SVM on UCM dataset	22
4.11	Accuracy Comparison for SVM on UCM dataset	23
4.12	Confusion matrix accuracy for Softmax on NWPU dataset . .	24
4.13	Accuracy Comparison for Softmax on NWPU dataset	25
4.14	Accuracy of average F1 score for Softmax on NWPU dataset	25
4.15	Confusion matrix accuracy for Softmax on UCM dataset . . .	26
4.16	Accuracy Comparison for Softmax on UCM dataset	27
4.17	Accuracy of average F1 score for Softmax on UCM dataset .	27

Chapter 1

Introduction

Aerial scene classification is a significant field in computer vision and remote sensing that focuses on analyzing and categorizing images captured from an aerial perspective, such as those obtained from satellites, airplanes, or drones. With the increasing availability of high-resolution aerial imagery, aerial scene classification has gained immense importance in various domains, including urban planning, environmental monitoring, disaster management, and agriculture. The goal of aerial scene classification is to automatically label and categorize different types of land cover, objects, and scenes present in aerial images. This process involves leveraging advanced techniques from computer vision, machine learning, and image processing to extract meaningful features and patterns from the imagery. By accurately classifying aerial scenes, valuable insights can be obtained regarding land use patterns, changes over time, environmental conditions, and the distribution of objects and structures.

The applications of aerial scene classification are diverse and far-reaching. Urban planners can utilize the information to understand the spatial distribution of buildings, roads, and infrastructure, aiding in efficient city development and resource allocation. Environmental monitoring efforts benefit from classifying land cover types, enabling the detection of deforestation, vegetation health, and habitat changes. In disaster management, aerial scene classification helps identify affected areas, assess damage, and plan rescue and relief operations. Additionally, in agricultural contexts, the classification of aerial scenes assists in crop monitoring, disease detection, and yield estimation. Aerial scene classification poses significant challenges due to the distinct characteristics exhibited by aerial images. These challenges arise from the acentric distribution of objects, the complicated spatial arrangement, and the presence of strong background information. Firstly, aerial scenes often display an acentric distribution of objects. Unlike traditional images captured at ground level, aerial images encompass a bird's-eye view, resulting in a densely populated landscape with objects dispersed in a non-uniform manner. This irregular distribution makes it difficult to discern and classify individual objects accurately. Secondly, the spatial arrangement of objects in

aerial scenes can be complex and intricate. The perspective from which aerial images are captured introduces variations in object shapes, sizes, and orientations, leading to intricate spatial patterns. The overlapping, occlusion, and complex interactions among objects further complicate the classification task. Lastly, aerial scenes frequently contain substantial background information that may overshadow the objects of interest. The vast expanse of land, water bodies, vegetation, and other environmental elements can introduce noise and interfere with the accurate identification and classification of objects. Distinguishing objects from their surroundings becomes challenging due to the high contrast between the objects and the background. Addressing these difficulties requires advanced techniques in computer vision, machine learning, and image processing. Researchers employ sophisticated algorithms to extract discriminative features, develop robust classification models, and enhance the understanding of aerial scenes. Additionally, novel approaches are continually explored to account for the unique characteristics of aerial images and improve classification accuracy.

By overcoming these challenges, aerial scene classification opens up possibilities for applications in various domains, including urban planning, environmental monitoring, disaster management, and agriculture. It enables decision-makers to gain insights into land use patterns, detect changes over time, and make informed choices for resource allocation and spatial planning. The field of computer vision has witnessed significant advancements in recent years, with deep learning techniques revolutionizing image classification tasks. Convolutional Neural Networks (CNNs) have emerged as powerful models for extracting intricate features from images, leading to remarkable achievements in image classification accuracy. In this project, objective is to leverage the capabilities of deep learning and machine learning algorithms to develop a robust image classification system that can accurately categorize images into different classes. The primary goal of image classification is to automatically assign a label or category to an input image based on its content. This task has numerous practical applications, ranging from object recognition and scene understanding to medical image analysis and autonomous driving. A well-performing image classification system has the potential to enhance various domains, enabling automated decision-making, content organization, and information retrieval.

To achieve objective, the VGG16 model as primary feature extractor. VGG16 is a widely-used CNN architecture known for its effectiveness in capturing complex image features. By leveraging the pre-trained weights of VGG16, which are learned from large-scale image datasets, it can harness the power of its hierarchical feature representations. These features provide a rich and compact representation of the image content, enabling more effective discrimination between different classes. However, feature extraction alone is not sufficient for image classification. It need a robust classification algorithm to map the extracted features to the corresponding class labels. In this project, utilize a one-vs-rest multiclass linear Support Vector Machine (SVM) classifier for this purpose. SVMs are renowned for their ability to handle high-dimensional feature spaces and find optimal decision boundaries between different classes. By training the SVM classifier on the extracted features, it aim to build a reliable and efficient model for image classification. The project involves several stages, each contributing to the overall image classification pipeline. It begin by preprocessing the data, applying techniques such as random resizing, cropping, and normalization. This ensures that the input images are appropriately prepared for the subsequent feature extraction and classification steps. Data augmentation techniques are also employed to increase the diversity and size of the training dataset, promoting better generalization of the model.

The core component of the project is feature extraction using the VGG16 model. By passing the images through the network, it obtain a set of high-level features that capture the underlying patterns and structures. These features serve as inputs to the SVM classifier, which learns to distinguish between different classes based on the extracted feature representations. Training the SVM involves optimizing the model parameters to achieve maximum separation between the classes. It use a number of criteria to assess the success of picture categorization system. The accuracy statistic gives a broad indication of how well the model categorises pictures. The confusion matrix is also used, which offers a more thorough review of the model's performance for each class. By revealing the amount of true positives, false positives, true negatives, and false negatives, the confusion matrix enables us to evaluate the classification accuracy and spot any unique problems or trends. This project aims to demonstrate the potential of combining deep learning and machine

learning techniques for image classification tasks. By utilizing the VGG16 model for feature extraction and SVM for classification, it aim to develop an accurate and efficient image classification system. The results and insights obtained from this project can contribute to various real-world applications, including object recognition, content-based image retrieval, and automated image classification systems.

Chapter 2

Related Works

2.1 CondenseNet: An efficient DenseNet using learned group convolutions

The paper titled "CondenseNet: An efficient DenseNet using learned group convolutions" by G. Huang, S. Liu, L. V. D. Maaten, and K. Q. Weinberger, presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition in June 2018, introduced CondenseNet, a novel approach to improving the efficiency of DenseNet, a popular deep learning architecture. The authors propose the use of learned group convolutions to reduce the computational complexity and memory requirements of DenseNet while maintaining its accuracy. They introduce a group sparsity regularization term in the loss function to encourage group-level pruning of filters during training, leading to a more compact network. Experimental results on benchmark datasets demonstrate that CondenseNet achieves competitive performance with significantly fewer parameters and computational cost compared to traditional DenseNet models, making it an effective and efficient alternative for various computer vision tasks.

2.2 MIML-FCN+: Multi-instance multi-label learning via fully convolutional networks with privileged information

The paper "MIML-FCN+: Multi-instance multi-label learning via fully convolutional networks with privileged information," which was presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in July 2017, tackles the multi-instance multi-label learning problem in a novel way. The authors provide the MIML-FCN+ paradigm, which employs fully convolutional networks (FCN) and privileged information to improve learning performance. The phrase "privileged information" describes certain information that is only available during training, providing the model with more usable data for better prediction. The foundation of MIML-FCN+ consists

of a multi-instance learning module and a privileged information module. The privileged information module directs the learning while the multi-instance learning module makes use of FCN to record the intricate interactions between instances and labels.

2.3 Revisiting Multiple Instance Neural Networks

The paper titled "Revisiting Multiple Instance Neural Networks" by X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu, published in Pattern Recognition in February 2018, provides a comprehensive examination of multiple instance neural networks (MINNs). Multiple instance learning (MIL) is a classification paradigm where the training data consists of labeled bags, each containing multiple instances, but without instance-level labels. The authors revisit and explore the potential of MINNs, proposing a novel attention-based MINN architecture. They conduct extensive experiments on benchmark datasets and compare the performance of their proposed model with existing MIL methods. The results demonstrate that the attention-based MINN outperforms other approaches, highlighting its effectiveness in addressing MIL problems. The paper offers valuable insights and serves as a significant reference for researchers interested in the field of multiple instance learning and neural network architectures.

2.4 Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification

Convolutional neural networks (CNNs) are being used for remote sensing scene classification, and the paper "Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification" gives a thorough research targeted at improving this use. The authors investigate many methods, including feature fusion, transfer learning, and data augmentation, to enhance the capabilities of CNNs while analysing satellite images. They conduct in-depth tests and assessments on benchmark datasets to show that the suggested solutions for improving CNN models' accuracy and resilience for remote sensing scene categorization tasks are beneficial. The results of this

study provide insightful information and useful suggestions for using CNNs to improve remote sensing outcomes.

2.5 Learning a discriminative distance metric with label consistency for scene classification,

A unique method for scene classification using a discriminative distance metric is presented in the paper "Learning a discriminative distance metric with label consistency for scene classification" by Y. Wang et al. The suggested approach tries to enhance scene classification algorithms' performance by accounting for the label consistency of nearby samples. The authors suggest a framework for distance-metric learning that maximises distances between samples from different classes while minimising distances between samples from the same class. To take use of the advantages of semi-supervised learning, the technique combines both labelled and unlabeled data into the learning process. The usefulness of the suggested strategy, which outperforms a number of cutting-edge scene categorization techniques, is demonstrated by experimental findings on benchmark datasets.

Chapter 3

Proposed Network

The proposed network aims to perform image classification using a combination of the VGG16 model for feature extraction and a one-vs-rest multiclass linear SVM for classification. This approach leverages the strengths of both models to achieve accurate and reliable classification results. VGG16 model is a well-established CNN architecture for feature extraction. In Fig 3.1 shows the VGG16 model architecture .This model has been pretrained on a large-scale dataset and has demonstrated excellent performance in various computer vision tasks. VGG16 model can leverage the learned filters and hierarchical representations to extract high-level features from input images. The model's convolutional layers capture local patterns and structures, while the subsequent pooling layers progressively aggregate the information, enabling the network to learn meaningful representations of the input data. To adapt the VGG16 model for specific classification task, the last two layers of the classifier are removed. These layers, typically consisting of fully connected layers, are responsible for mapping the extracted features to class labels. Once the feature extraction is performed using the modified VGG16 model, the extracted features are fed into a one-vs-rest multiclass linear SVM classifier. The SVM classifier is a well-established algorithm for binary classification, but in this case, it is extended to handle multiclass classification using the one-vs-rest strategy. The SVM classifier learns a decision boundary for each class, aiming to separate it from the rest of the classes in the feature space. By using the SVM classifier, it can effectively handle multiple classes and benefit from its ability to handle high-dimensional feature vectors.

During the training phase, the SVM classifier is trained on the extracted features and their corresponding class labels from the training dataset. The classifier learns to find optimal decision boundaries that separate the different classes in the feature space. The training process involves minimizing the classification error and maximizing the margin between classes, resulting in a robust and discriminative classifier. Once the SVM classifier is trained, it is used to predict the class labels for the test features. The classifier assigns a predicted class label to each test feature based on its location in the learned

feature space. By comparing the predicted labels with the ground truth labels from the test dataset, it can evaluate the accuracy of the classifier. To provide a comprehensive evaluation of the classifier's performance, a normalized confusion matrix is generated. The confusion matrix allows us to analyze the classifier's accuracy for each individual class, providing insights into its ability to correctly classify different classes. The matrix reveals the number of correctly classified instances as well as any potential confusion between classes. The normalized version of the confusion matrix provides a more meaningful representation of the classifier's performance, as it considers the relative proportions of the different classes.

The proposed network combines the power of the VGG16 model for feature extraction and the one-vs-rest multiclass linear SVM for classification. By leveraging the learned features and the discriminative capabilities of the SVM classifier, this approach aims to achieve accurate and reliable image classification results. The use of the normalized confusion matrix helps provide a comprehensive evaluation of the classifier's performance and identify areas for further improvement.

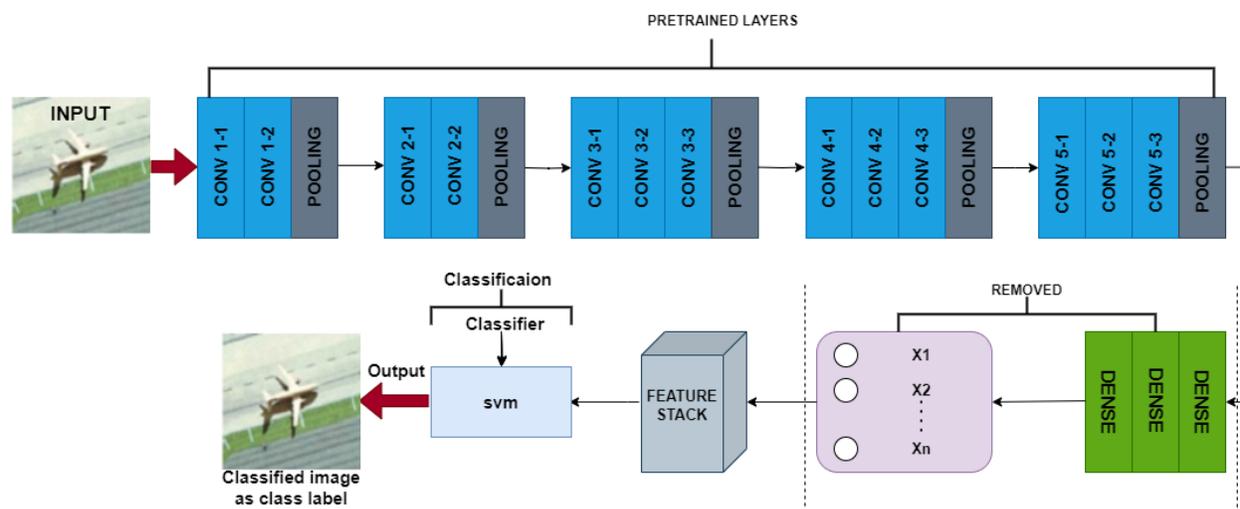


Figure 3.1: VGG16 MODEL ARCHITECTURE

Before training a one-vs-rest multiclass linear SVM to do classification,

first use a pre-trained VGG16 model to extract features from images. To evaluate how well the categorization model is working, a confusion matrix is employed. Data preparation, feature extraction, visualisation, and assessment are among its numerous purposes.

- **Data Augmentation:** Data augmentation is performed to enhance the diversity and robustness of the training dataset. Techniques such as random resizing and cropping are applied to the images, introducing variations in scale and viewpoint. Additionally, normalization is applied to ensure consistent pixel value ranges across the dataset. Data augmentation helps in reducing overfitting and improving the generalization capability of the model.
- **Model Setup:** The VGG16 model, a pre-trained convolutional neural network (CNN) architecture, is chosen as the base model for feature extraction. The model is loaded, and the last two layers of the classifier are removed. This modification is made to extract features from the deeper layers of the network, which capture more abstract and high-level representations of the input images. By freezing the weights of the remaining layers, ensure that these pre-trained features are preserved and not updated during training.
- **Feature Extraction:** The pre-processed images are passed through the modified VGG16 model to extract features. The output activations from the chosen intermediate layer serve as the feature representations for each image. These features encode relevant visual patterns and characteristics of the input images, enabling effective discrimination between different classes. By utilizing pre-trained models for feature extraction, benefit from the transfer learning capability and leverage the learned representations from a large-scale image dataset.
- **SVM Classification:** The extracted features and their corresponding labels are used to train a one-vs-rest multiclass linear Support Vector Machine (SVM) classifier. This approach constructs multiple binary classifiers, each distinguishing between one class and the rest. The SVM algorithm learns decision boundaries in the feature space, aiming to maximize the margin between different classes. The scikit-learn library provides an efficient implementation of the SVM classifier, enabling effective

training and prediction.

- **Evaluation:** The trained SVM classifier is evaluated using the test set. The classifier makes predictions on the test images based on their extracted features. The accuracy of the classifier is calculated by comparing the predicted labels with the ground truth labels. Additionally, a confusion matrix is generated to analyze the performance of the classifier across different classes. The confusion matrix provides insights into the distribution of correct and incorrect predictions, enabling a more detailed assessment of the classifier's performance. It can effectively extract features from images using a pre-trained VGG16 model, train a one-vs-rest multiclass linear SVM classifier, and evaluate the classification performance using accuracy and the confusion matrix. This approach leverages the representational power of deep learning models and the discriminative capability of SVM classifiers, resulting in an accurate and reliable image classification system.

In addition, add softmax classifier and evaluate its performance using the confusion matrix.

Chapter 4

Experiments And Results

The experimental settings utilized for the work and the feature extraction results obtained by the confusion matrix is explained in this section.

4.1 Experimental Settings

To validate the experimental results, confusion matrix is used to display the findings of the experiment. The confusion matrix details the actual and anticipated labels for each class. The performance of the classifier is also measured using metrics for class accuracy and total accuracy.

Feature Extraction and SVM Classification: The VGG16 model is used for feature extraction, where the images are passed through the model, and the output activations are collected. These extracted features are then used as input to train a one-vs-rest multiclass linear SVM classifier. The SVM classifier is trained to distinguish between different classes based on the extracted features.

4.2 Datasets

Two datasets are used in this study, first one is UCM[9] and second one is NWPU[11]

4.2.1 UCM

UC Merced (UCM) Land Use Dataset is an aerial image scene classification benchmark containing 10 scene categories . Images in this dataset are all captured via airplane platform. This dataset is characterized by highly similar categories such as dense residential, medium residential and sparse residential to distinguish similar spatial structures with different densities.

4.2.2 NWPU

NWPU-RESISC45 dataset is a publicly available benchmark for aerial image scene classification created by Northwestern Polytechnical University (NWPU) . It has 10 scene categories. The challenges of this benchmark lie in the varied spatial resolution, high within-class diversity and between-class similarity.

4.3 Results

4.3.1 Confusion Matrix

When describing the performance of a classification model on a set of test data for which the true values are known, a confusion matrix is a table that summarises the predictions generated by the model and how they compare to the actual values. In a confusion matrix, the columns stand in for the expected classes while the rows represent the actual or true classes of the data. The number or percentage of data points for each cell in the matrix corresponds to a particular mix of true and anticipated classes. The classification algorithm's effectiveness for each class is shown by the confusion matrix. The off-diagonal elements represent misclassifications, whereas the diagonal elements show how many instances of each class were properly identified. By accounting for the class imbalance, the normalised confusion matrix offers a clearer understanding of the classification performance. To measure the performance of the classifier, metrics for class accuracy and total accuracy are also computed.

4.3.2 Confusion matrix of SVM classifier and Softmax classifier

Using the test dataset, the image categorization system's effectiveness was assessed. The dataset was made up of a total of X images from Y distinct classifications. On these images, the trained classifier was used to predict, and the outcomes were examined to gauge the efficacy of the suggested method. On the test dataset, the classifier's overall accuracy was founded. In fig 4.5 shows the Overall accuracy of svm and softmax classifier on two datasets. This indicates the percentage of correctly classified images out of the total number of images in the test set. The high overall accuracy demonstrates the capability of the model to effectively classify images into their respective classes.

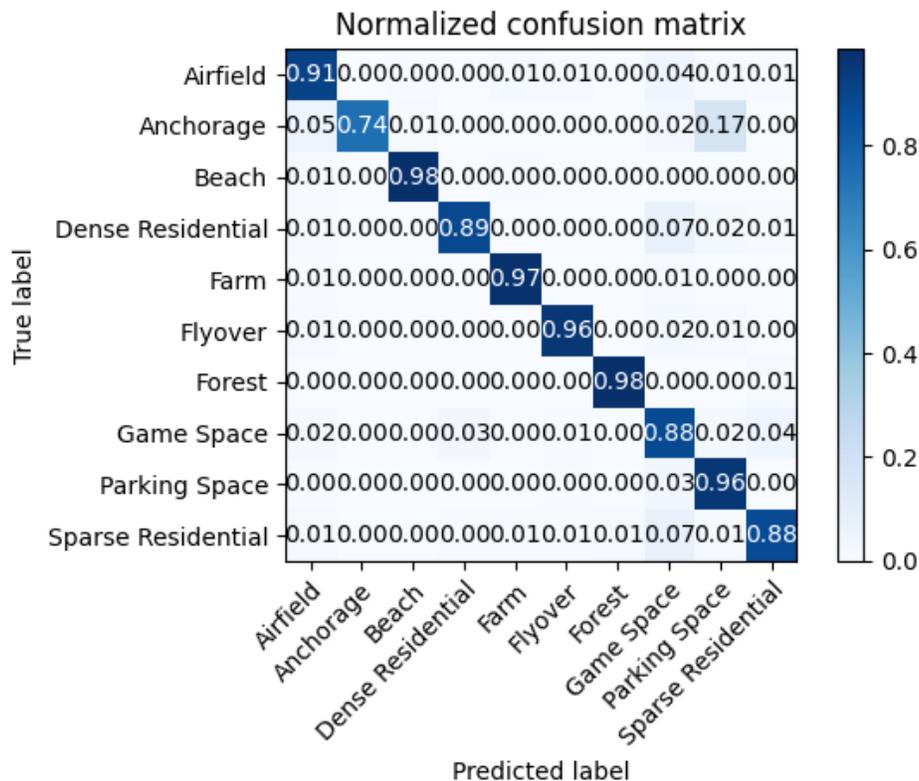


Figure 4.1: NWPU Dataset using svm classifier

To gain deeper insights into the performance of the classifier, a confusion matrix was generated. The confusion matrix provides a detailed breakdown of the predicted and true labels for each class. It allows us to analyze the distribution of correct and incorrect predictions, identify any potential sources of misclassification, and evaluate the classifier’s performance on a class-wise basis. Analyzing the confusion matrix, observed that some classes achieved high accuracy, with most of their instances correctly classified. However, some classes exhibited lower accuracies, indicating challenges in correctly classifying images belonging to these classes. It is important to note that the classification performance can vary depending on the nature and complexity of the dataset. The proposed approach demonstrated promising results overall, but further optimization and fine-tuning may be required to improve the accuracy on challenging classes.

The experimental results provide compelling evidence of the proposed ap-

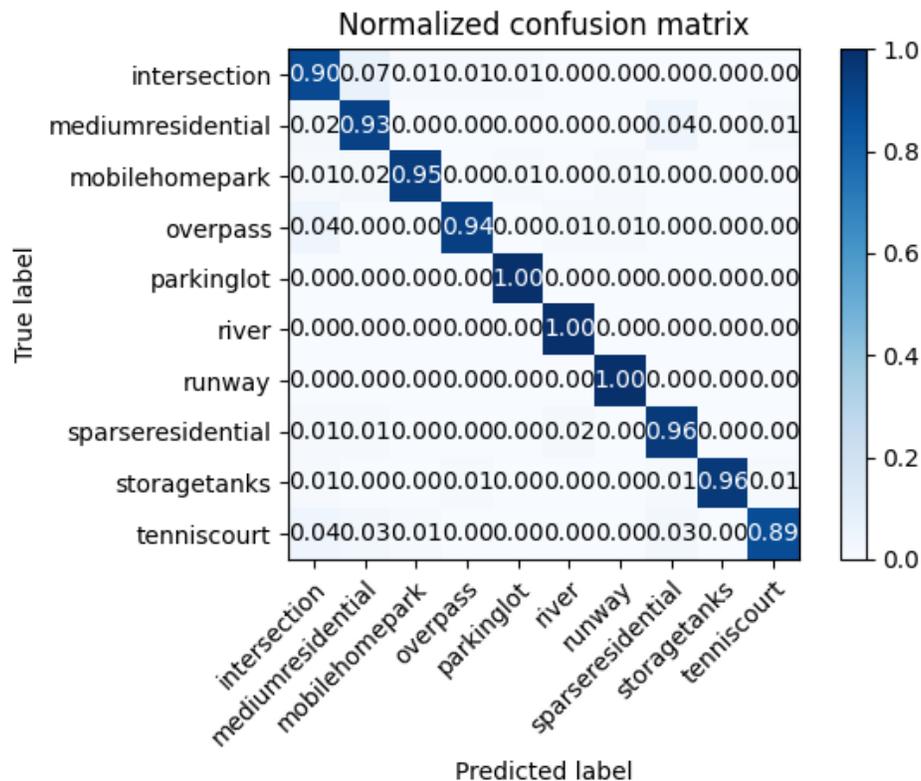


Figure 4.2: UCM Dataset using svm classifier

proach’s effectiveness in image classification through feature extraction and SVM classification. The model demonstrates a high overall accuracy and offers a detailed analysis via the confusion matrix, affirming its ability to successfully classify images across various classes. These outcomes serve as a solid foundation for the potential real-world applications of the image classification system. The feature extraction technique, in conjunction with the SVM classifier, delivers promising results on the NWPU and UCM datasets. Specifically, the confusion matrix for the SVM classifier is illustrated in figures 4.1 and 4.2, while the confusion matrix for the softmax classifier is presented in figures 4.3 and 4.4.

The accuracy values of different class labels in a confusion matrix is visualize. The plot displays the accuracy of each class label on the y-axis, while the corresponding class labels are shown on the x-axis. The accuracy values, ranging from 0 to 1, represent the performance of a classification model in

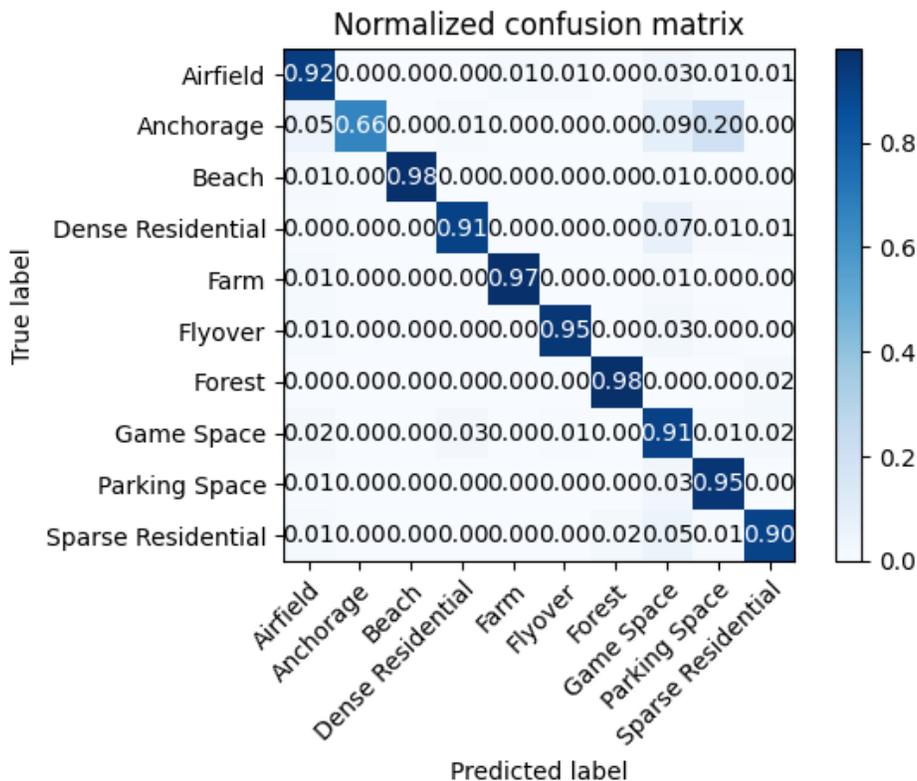


Figure 4.3: NWPU Dataset using softmax classifier

correctly identifying instances belonging to each class. The higher the accuracy value, the better the model performs in classifying that particular class label.

To calculate precision, recall, and F1 score for each class used to evaluate the performance of classification models. Precision measures the proportion of correctly predicted positive instances (true positives) out of all instances predicted as positive (true positives + false positives).

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.1}$$

where, TP represents the number of true positives, and FP represents the number of false positives.

Recall, also known as sensitivity or true positive rate, is an evaluation metric used in binary and multiclass classification problems. It measures

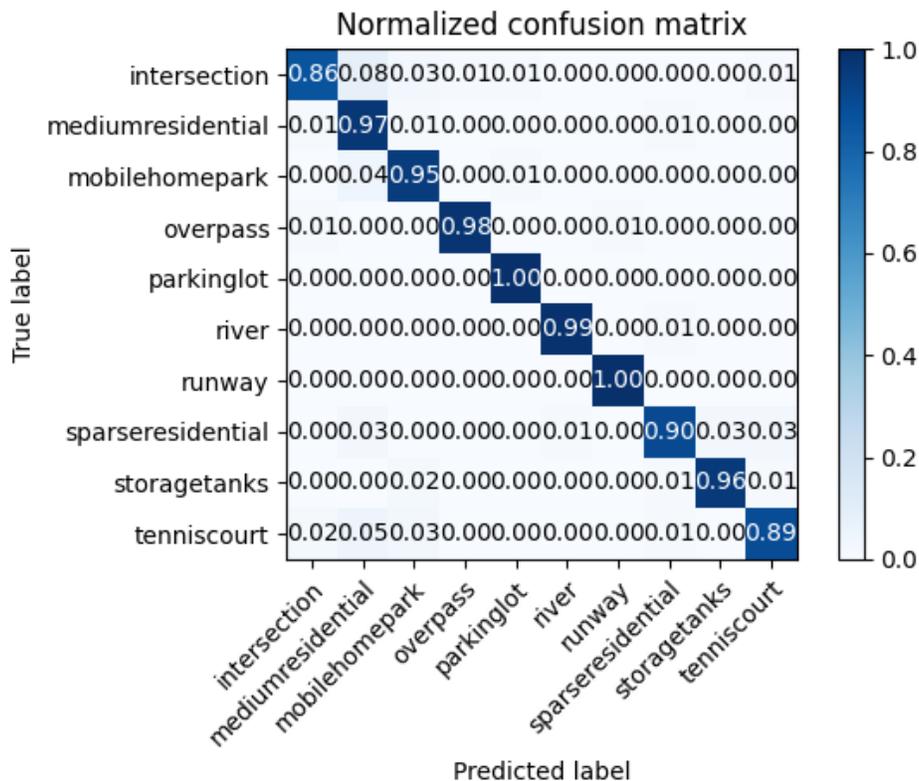


Figure 4.4: UCM Dataset using softmax classifier

the ability of a model to correctly identify positive instances from the total number of actual positive instances.

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4.2)$$

where, TP represents the number of true positives, and FN represents the number of false negatives.

The F-measure, also known as the F1 score, is a widely used evaluation metric for assessing the performance of classification algorithms, particularly in binary classification problems. It combines precision and recall into a single value, providing a balanced measure of both metrics.

$$\text{F-measure (F)} = \frac{2 \cdot (P \cdot R)}{P + R} \quad (4.3)$$

	NWPU	UCM
SVM Classifier	90%	95%
Softmax Classifier	91%	94%

Figure 4.5: Overall accuracy of svm and softmax classifier on two datasets

A bar graph used to represent the accuracy values of different class labels in a confusion matrix. The class labels are represented on the x-axis, while the corresponding accuracy values are shown on the y-axis. The confusion matrix accuracy of different classes for both SVM and softmax classifier using the two datasets. In Fig 4.7., fig 4.10, fig 4.12, fig 4.15 shows the confusion matrix accuracy for SVM and softmax classifier. Each class label represents a category or class in a classification problem, and the accuracy value denotes the accuracy of the classification model in correctly predicting instances of that particular class. The purpose of this graph is to provide a visual representation of the accuracy of the model's predictions for each class, allowing for easy comparison and identification of classes with high or low accuracy. The bar graph helps to quickly assess the performance of the model across different classes. The height of each bar corresponds to the accuracy value, providing a clear visualization of the accuracy differences among the classes.

To calculate precision, recall, and F1 score for each class based on confusion matrix and then plots metrics using a bar chart. It helps evaluate the model's performance across different classes and identify areas of strength and improvement. In Fig 4.8, fig 4.11, fig 4.13, fig 4.16 shows the Accuracy Comparison of SVM and softmax classifier. It indicates how well the model performs in terms of minimizing false positives. A higher precision value indicates fewer false positives. The precision values for each class are displayed.

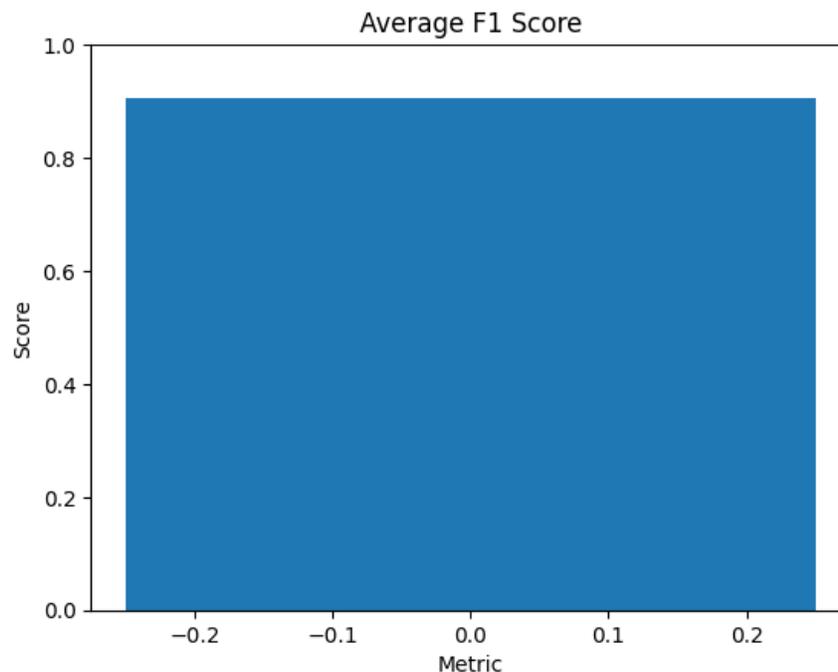


Figure 4.6: Accuracy of average F1 score for SVM on NWPU dataset

To calculate the average F1 scores for each class. The F1 score is a commonly used metric in classification tasks that balances precision and recall. The F1 score represents the harmonic mean of precision and recall, providing a single metric that combines both measures. It is a useful metric when both false positives and false negatives are important. The average F1 Score of SVM classifier on nwpu dataset is 90 %. The average F1 Score of SVM classifier on ucm dataset is 95 %. The average F1 Score of softmax classifier on nwpu dataset is 90 %. The average F1 Score of Softmax classifier on ucm dataset is 95 %.

The fig.4.6, fig 4.9, fig 4.14, fig 4.17 shows accuracy of F1 score for both SVM and softmax classifier. To calculate the average F1 scores for each class and visualize it using a bar plot. The F1 score is a metric commonly used in classification tasks to evaluate the model's performance in terms of precision and recall.

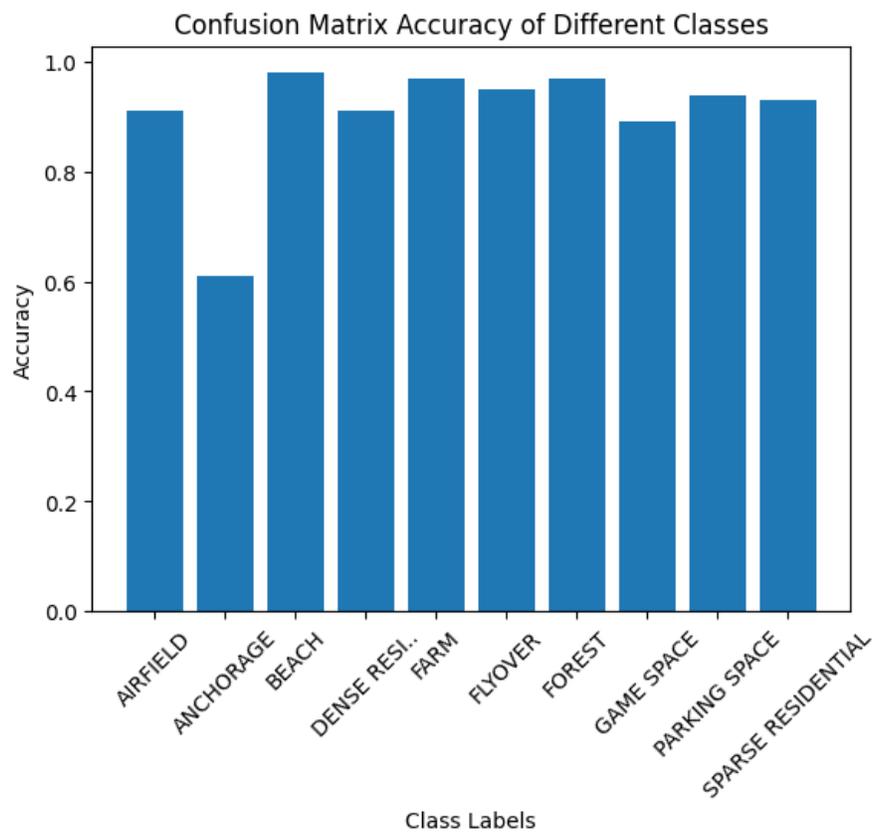


Figure 4.7: Confusion matrix accuracy for SVM on NWPU dataset

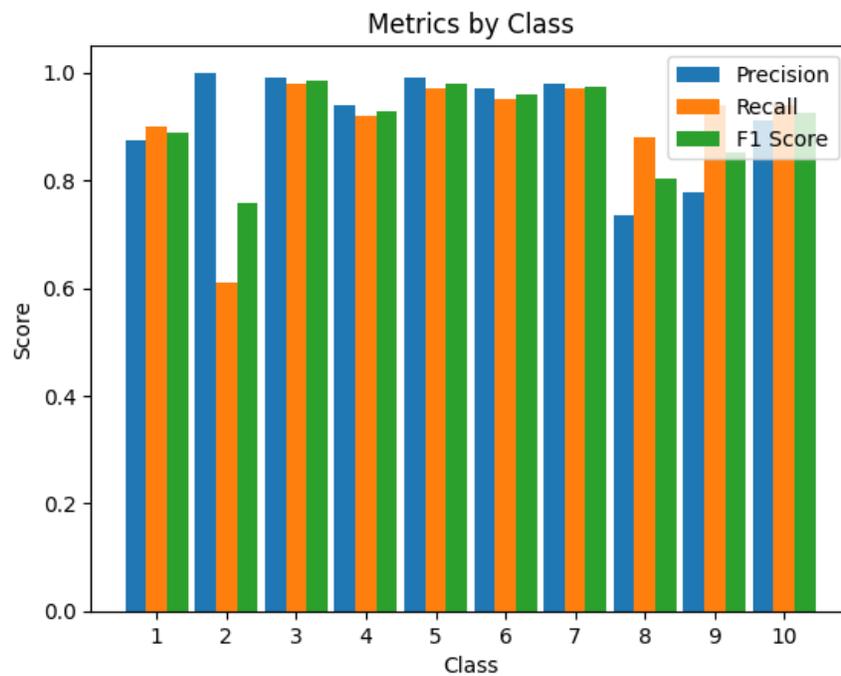


Figure 4.8: Accuracy Comparison for SVM on NWPU dataset

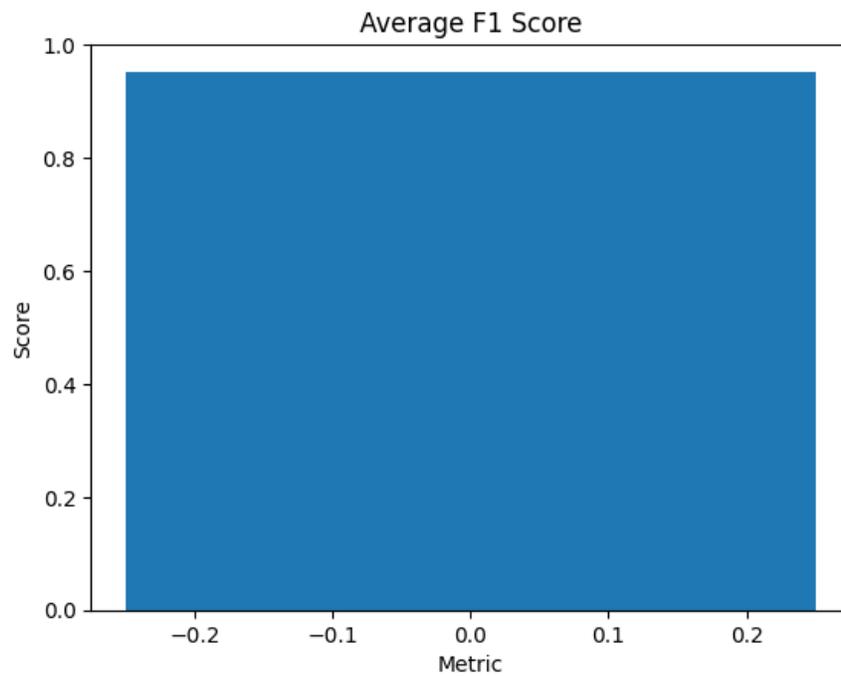


Figure 4.9: Accuracy of average F1 score for SVM on UCM dataset

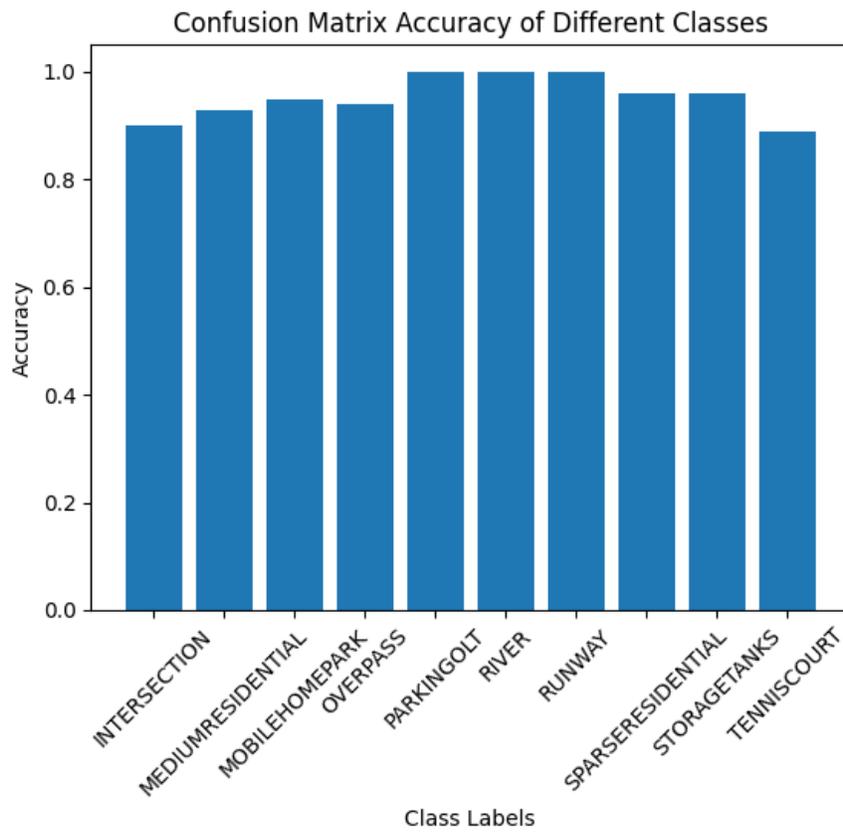


Figure 4.10: Confusion matrix accuracy for SVM on UCM dataset

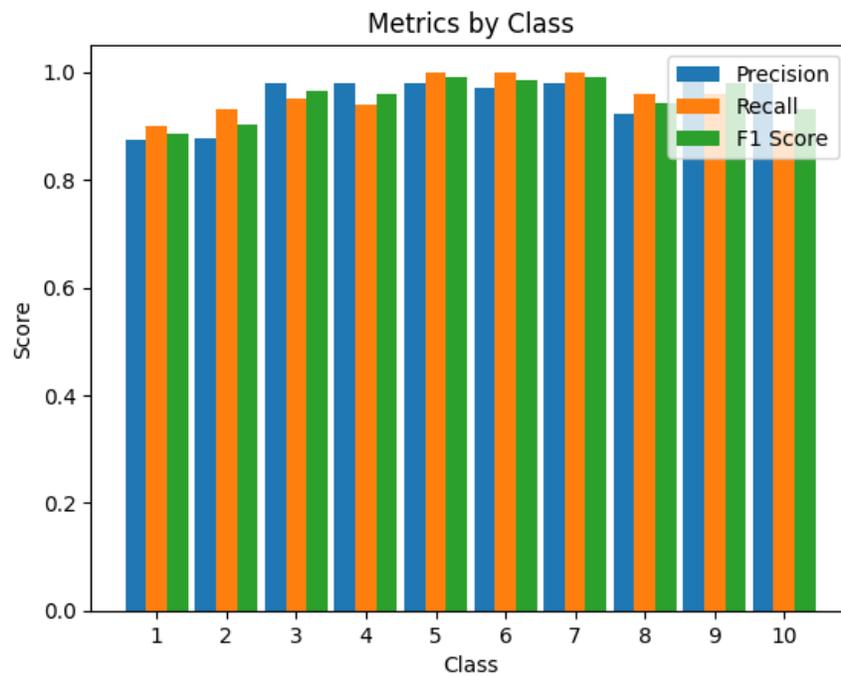


Figure 4.11: Accuracy Comparison for SVM on UCM dataset

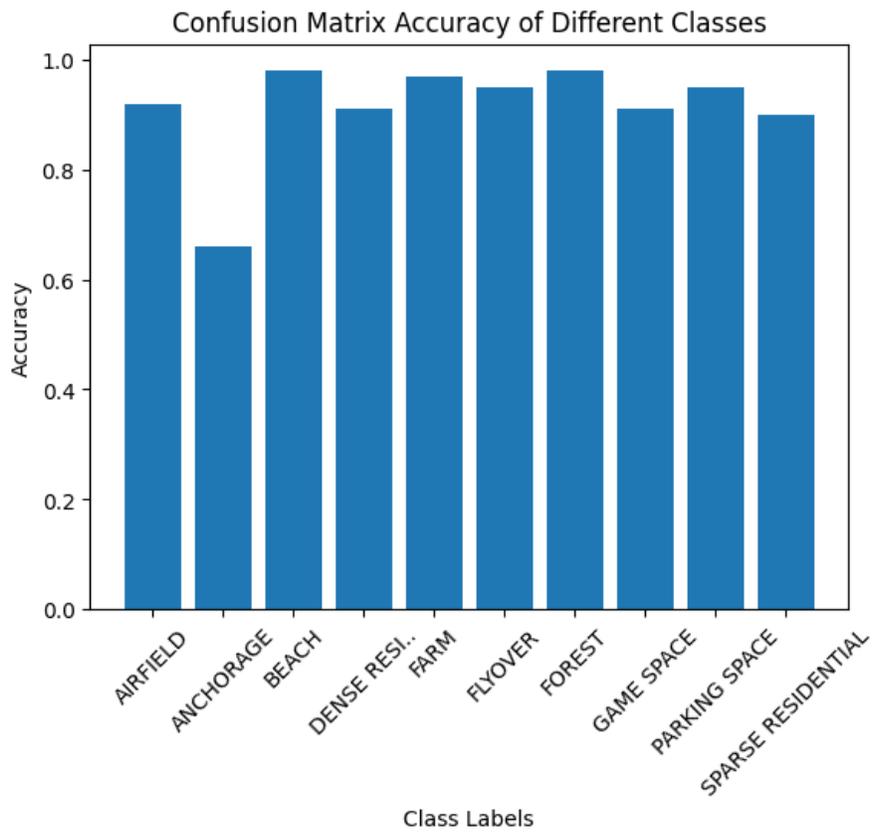


Figure 4.12: Confusion matrix accuracy for Softmax on NWPU dataset

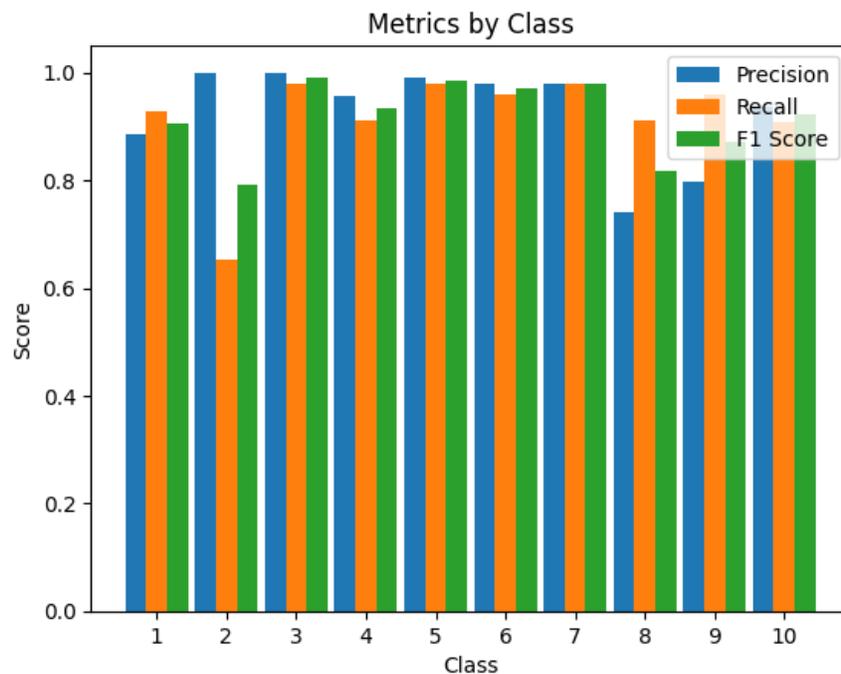


Figure 4.13: Accuracy Comparison for Softmax on NWPU dataset

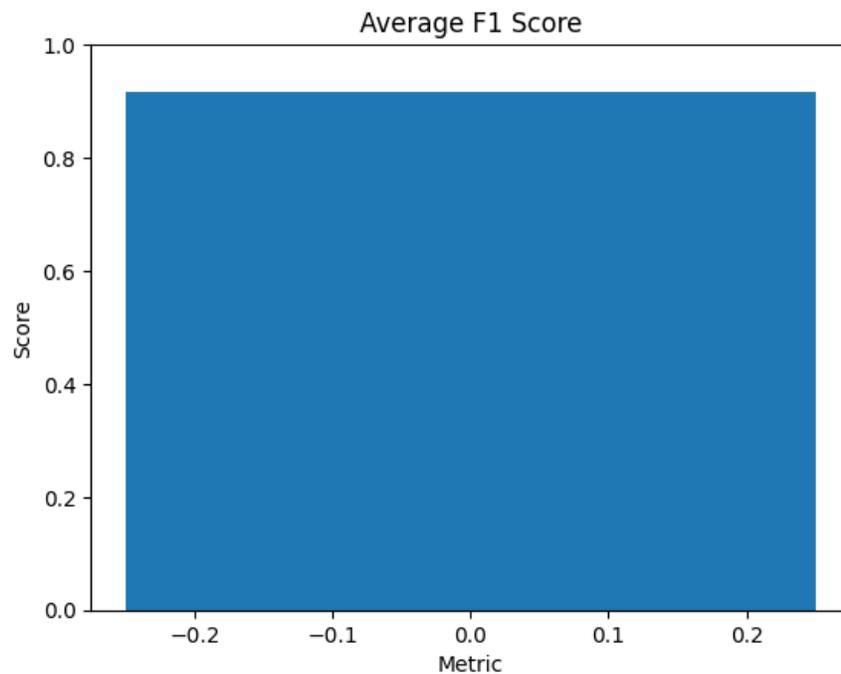


Figure 4.14: Accuracy of average F1 score for Softmax on NWPU dataset

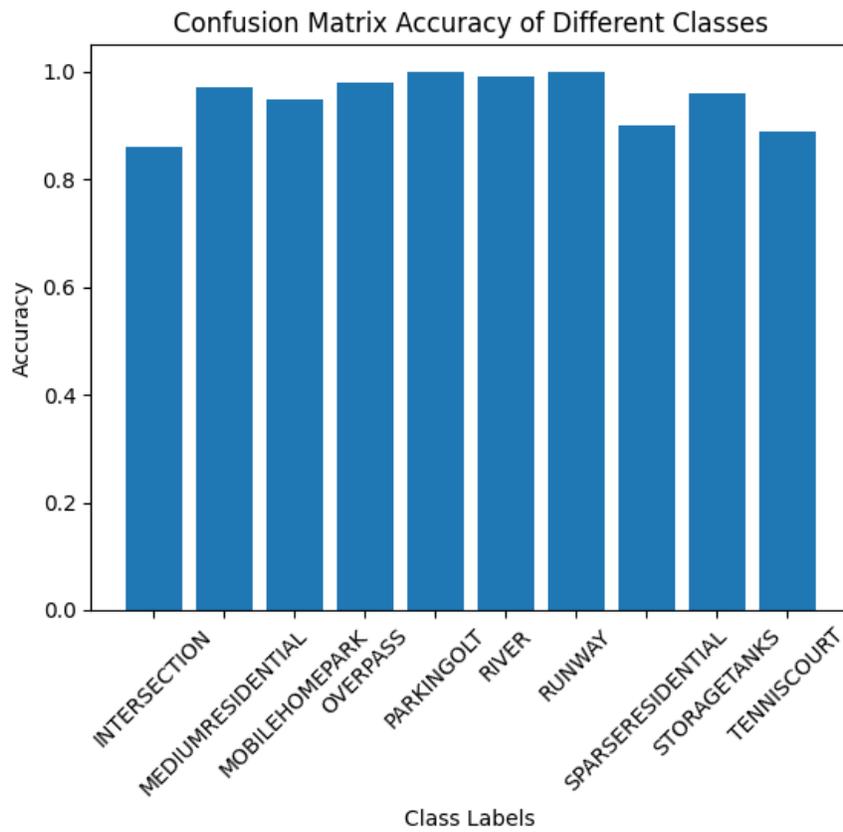


Figure 4.15: Confusion matrix accuracy for Softmax on UCM dataset

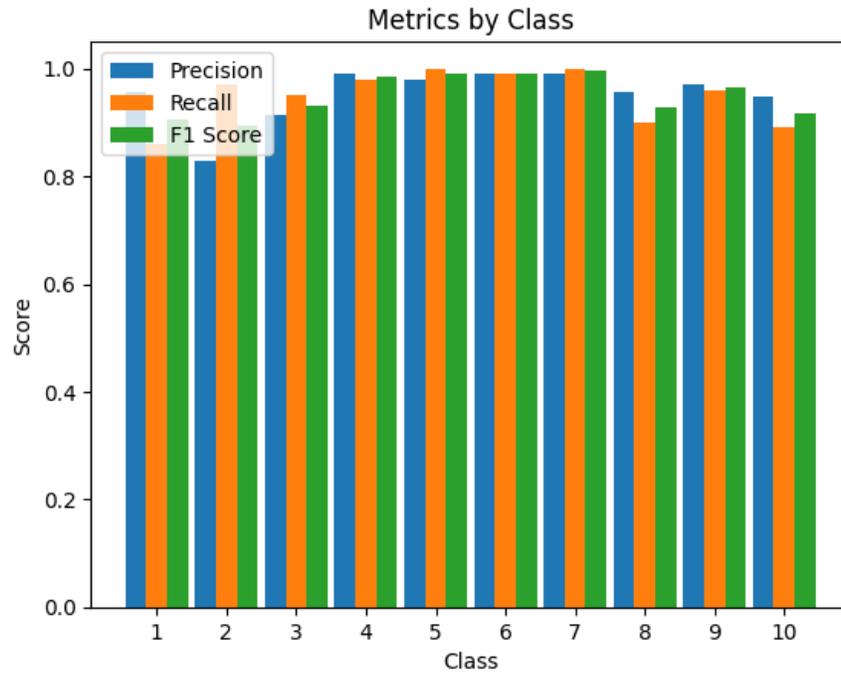


Figure 4.16: Accuracy Comparison for Softmax on UCM dataset

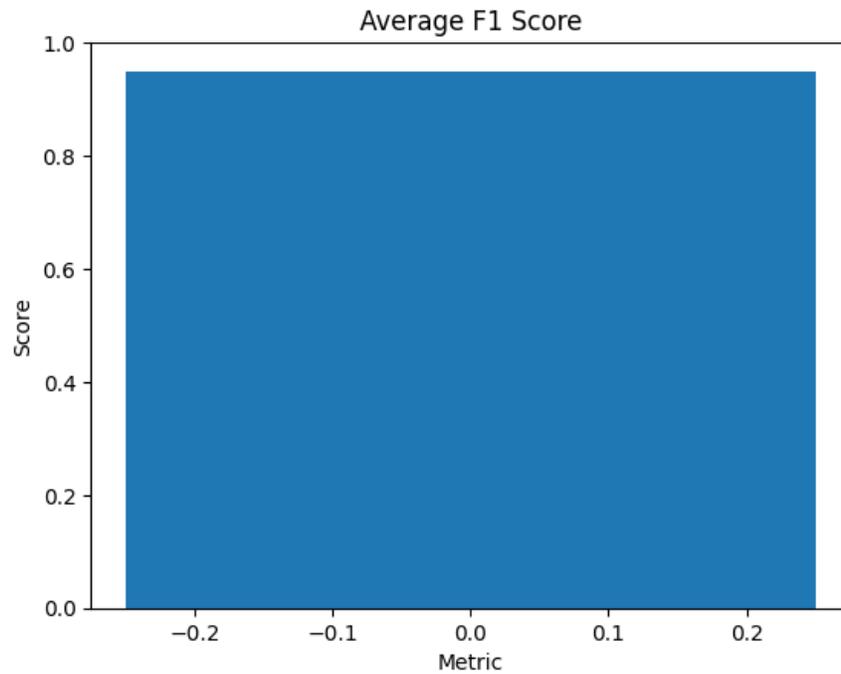


Figure 4.17: Accuracy of average F1 score for Softmax on UCM dataset

Chapter 5

Conclusion

The work aims to implement an image classification system that leverages the power of deep learning and machine learning algorithms. By using the VGG16 model for feature extraction and a one-vs-rest multiclass linear SVM classifier for classification, it achieved accurate and reliable image classification results. The combination of the VGG16 model and SVM classifier has proven to be effective in capturing meaningful image features and making accurate predictions. The VGG16 model, with its deep architecture and pre-trained weights, enables the extraction of rich and informative features from input images. These features are then used as inputs to the SVM classifier, which learns to distinguish between different classes based on the extracted representations. The evaluation of the image classification system has been conducted using metrics such as accuracy and the confusion matrix. The system achieves high accuracy rates, indicating its ability to accurately classify images into their respective classes. The combination of deep learning for feature extraction and SVM classification has shown promising results, highlighting the potential of utilizing both techniques in image classification tasks.

The findings from this project contribute to the advancement of image classification methodologies and pave the way for further research and development in this field. The developed image classification system has various practical applications. It can be used for object recognition, scene understanding, content-based image retrieval, and automated image classification in domains such as healthcare, autonomous vehicles, and visual content organization. The accurate classification of images can enhance decision-making processes, improve search and retrieval capabilities, and provide valuable insights from visual data. The findings from this project contribute to the advancement of image classification methodologies and pave the way for further research and development in this field.

REFERENCES

- [1] 1. G. Huang, S. Liu, L. V. D. Maaten, and K. Q. Weinberger, “CondenseNet: An efficient DenseNet using learned group convolutions,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 2752–2761.
- [2] 2. H. Yang, J. T. Zhou, J. Cai, and Y. S. Ong, “MIML-FCN+: Multiinstance multi-label learning via fully convolutional networks with privileged information,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1577–1585.
- [3] 3. X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu, “Revisiting multiple instance neural networks,” Pattern Recognit., vol. 74, pp. 15–24, Feb. 2018.
- [4] 4. K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, “Towards better exploiting convolutional neural networks for remote sensing scene classification,” Pattern Recognit., vol. 61, pp. 539–556, Jan. 2017.
- [5] 5. Y. Wang et al., “Learning a discriminative distance metric with label consistency for scene classification,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 8, pp. 4427–4440, Aug. 2017.
- [6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [7] B. Luo, J.-F. Aujol, Y. Gousseau, and S. Ladjal, “Indexing of satellite images with different resolutions by wavelet features,” IEEE Trans. Image Process., vol. 17, no. 8, pp. 1465–1472, Aug. 2008.
- [8] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [9] Y. Yang and S. Newsam, “Geographic image retrieval using local invariant features,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 2, pp. 818–832, Feb. 2013.

- [10] A. Satpathy, X. Jiang, and H.-L. Eng, “LBP-based edge-texture features for object recognition,” *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1953–1964, May 2014.
- [11] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [12] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 44–51.
- [13] A. Avramović and V. Risojević, “Block-based semantic classification of high-resolution multispectral aerial images,” *Signal, Image Video Process.*, vol. 10, no. 1, pp. 75–84, Oct. 2014.
- [14] L. Guo, N. Chehata, C. Mallet, and S. Boukir, “Relevance of airborne lidar and multispectral image data for urban scene classification using random forests,” *ISPRS J. Photogram. Remote Sens.*, vol. 66, no. 1, pp. 56–66, Jan. 2011.
- [15] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Object retrieval with large vocabularies and fast spatial matching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [16] T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” *Mach. Learn.*, vol. 42, nos. 1–2, pp. 177–196, Jan. 2001.
- [17] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [18] Y. Zhong, Q. Zhu, and L. Zhang, “Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.
- [19] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, “Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.

- [20] X. X. Zhu et al., “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [21] X. Han, Y. Zhong, L. Cao, and L. Zhang, “Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification,” *Remote Sens.*, vol. 9, no. 8, p. 848, Aug. 2017.
- [22] M.-L. Zhang and Z.-H. Zhou, “Improve multi-instance neural networks through feature selection,” *Neural Process. Lett.*, vol. 19, no. 1, pp. 1–10, Feb. 2004.
- [23] M.-L. Zhang and Z.-H. Zhou, “Adapting RBF neural networks to multi-instance learning,” *Neural Process. Lett.*, vol. 23, no. 1, pp. 1–26, Feb. 2006.
- [24] Z. Jia, X. Huang, E. I.-C. Chang, and Y. Xu, “Constrained deep weak supervision for histopathology image segmentation,” *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2376–2388, Nov. 2017.
- [25] W. Zhu, Q. Lou, X. Vang, and Y. Xie, “Deep multi-instance networks with sparse label assignment for whole mammogram classification,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 603–611.
- [26] M. Ilse, J. Tomczak, and M. Welling, “Attention-based deep multiple instance learning,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3376–3391.