National College of Ireland

Aerial Scene Classification using Deep Transfer Learning Network and Tree-based Classifiers

Research Project
M.Sc.in Data Analytics

Abdul Azadh Abdul Saleem
Student ID: x18203621@student.ncirl.ie

School of Computing
National College of Ireland

Supervisor: Dr.Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Abdul Azadh Abdul Saleem
Student ID:	x18203621@student.ncirl.ie
Programme:	M.Sc.in Data Analytics
Year:	2020
Module:	Research Project
Supervisor:	Dr.Hicham Rifai
Submission Due Date:	17/12/2020
Project Title:	Aerial Scene Classification using Deep Transfer Learning Network and Tree-based Classifiers
Word Count:	7323
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	17th December 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Aerial Scene Classification using Deep Transfer Learning Network and Tree-based Classifiers

Abdul Azadh Abdul Saleem
x18203621@student.ncirl.ie

Abstract

Due to the advancements in the field of remote observation of landscapes,large memory of high-dimensional images of the different landscapes has been accumulated, which has led to the importance of developing a robust aerial scene classification model to classify the landscapes of the images. This project aims to develop a robust aerial scene classification model which could perform with better accuracy irrespective to the dimension of the aerial image. It is essential to utilize an effective feature extraction process, which could help the classifier to perform better, which led to the utilization of Inceptio V3 transfer learning model for feature extraction, because of its deep convolution network of layers. From the knowledge acquired from literature, the Tree-based classifiers such as Random Forest (RT) and Decision Tree (DT) has not been explored deeply in this field. The objective of this research is to utilize and evaluate the performance of tree-based classifiers for aerial scene classification. The models were trained with UCMerced dataset, which has 21 classes of low-resolution aerial image, and AID dataset, is a benchmark dataset with 30 classes of high-resolution aerial images, could help to prove the robustness of the classifier. The Random Forest model has performed with the accuracy of 98.52% and 98% for UCMerced and AID datasets respectively and proved its robustness to perform classification of land-use classes. But the Decision Tree performed with 94.76% accuracy on UCM dataset but failed to perform equally good with the AID dataset by performing with 62.17% accuracy.

Keywords— Aerial Scene Classification, Inception V3, Random Forest, Decision Tree.

1 Introduction

In the past decade, there has been a huge development in the field of scene classification in aerial images. Many researchers have involved in conducting research on developing an efficient way to classify the aerial images depending on the features extracted. With the development of technologies in aerial observation using Unmanned aerial drones, which has enabled the availability of data of different land terrains, which could provide way for development of Deep learning and Machine Learning models in classifying the scene in aerial images (Zhang et al. (2019)). Deep learning models extract features by combining the multiple low-level features to form more details with higher-level of detail abstraction by pixel convolution, which the traditional models fail to capture such high semantic information. The multiple layer structure of the deep learning network helps to extract the multi-scale information from the aerial images, and the complex architecture of deep

layers helps to model helps to increase the non-linear relationships between the data classes in higher resolution images, captures the multidimensional features with reduced loss.

There has been a tremendous development of numerous transfer learning approach, in which the model that the pre-trained with primary weights and re-using the network to perform different task in another model. This project involves in classification of schematic scenes of different landscapes, which are has been carried in two stages of operation. The primary task is to extract the features of the image dataset using multiple layers of convolution, which has been performed with the help of pre-trained Inception V3 model and the secondary task is to classify the features of the images into different land-use classes from aerial images, which has been done with Tree-based classifiers as Decision Tree and Random Forest, which has been mentioned to be less explored in the field of aerial scene classification (Sheykhmousa et al. (2020)). Inception V3 model is referred as the best pre-trained model for extracting the features from the high dimensional image (Wang et al. (2019)). The features obtained from different layers of convolution has been concatenated at ten different layers of the network, which helps this model to provide the low-level, mid-level and high-level features of the image data, as the output. The datasets utilized to train the classifiers are UCMerced (UCM) dataset, with 21 classes of low-resolution aerial images and Aerial Image Detection dataset (AID) database, with 30 classes of high-resolution land-use classes. AID dataset has been mentioned as the benchmark dataset (Xia et al. (2017)), which will be discussed in detail in Section 3.1.

This research has utilized the deep convolutional network for extracting high-level features and tree-based machine learning classifiers for the purpose of classification, as the classifiers can yield the result of higher accuracy and to be robust irrespective the dimension of the data (Sheykhmousa et al. (2020)). The objective of this research is to utilize the deep feature extraction network of Inception V3 model, and to adapt the tree-based classifiers to classify the different land-use classes and to evaluate the performance. And to the prove the robustness of the classifiers the process has been conducted with high resolution AID dataset and evaluated.

This study considers the following research question aimed to address the work conducted in this research.

"To what extend does the deep convolutional feature extractor and tree-based classifiers can perform in aerial scene classification of high-resolution images?"

The objectives associated with the research question is listed below,

- To conduct literature study of approaches and techniques implemented in aerial scene classification, which could assist the objective of this research.
- Utilizing a deep convolution feature extractor network, to extract deep features from high dimensional aerial images.
- Utilizing Tree-based classifiers to classify the multiple land-use classes and to evaluate the performance of each classifier.
- Utilizing a benchmark aerial image dataset, to train the classifier and to evaluate the robustness of classifiers to the high resolution data.
- Developing a Graphical User Interface and to utilize the feature extractor functions and trained classifier to classify the test aerial image.

This research contributes to development of application for aerial scene classification, utilizing the deep feature extractor and robust classifier. The below report has been organized as follows, Section 2 explores the experiment and techniques conducted in the literature which could contribute for this research. Section 3 explains the methodology followed to conduct the research. Section 4 explains the design architecture of this experiments. Section 5 explain in detail the implementation procedure of the research. Section 6 explains the evaluation results of the experiments and discuss the outcomes. Section 7 concludes the research report, and also mention the future works that could be possible in this research.

2 Literature Review

2.1 Introduction to Literature

This section reviews the research that has been conducted in the field of aerial scene classification of different classes of landscapes using different approaches and critically evaluate the limitations and contribution for this research. The section has been divided into five sub-divisions which explores the evolution of aerial scene classification models over the past, utilization of deep learning models for large-scale aerial image classification, utilization of Inception V3 model for features extraction, classifiers that has been utilized for aerial image classification and research gaps observed from the literature study, which supports the objective of this research.

2.2 Evolution of Aerial Scene Classification Models

Land-use classification is difficult because of the variety of different classes of object it covers in training the model and to utilize the model to recognize the landscapes. (Hu et al. (2018b)) has proposed concentric circle based spatial rotation invariant represent strategy in Bag-of-Visual-Words (BOVW) model which has proved to outperform the traditional model which utilize the Scale-Invariant Feature Transform (SIFT) to extract the features from the image, and it discards the spatial information, which has to be provided to the model by hard-coding. The experiment has resulted in accuracy of around 86% in classifying the landscapes in the aerial image. Likewise (Chen and Tian (2015)) has proposed Pyramid-of-Spatial relations (PSR) to overcome the performance of the previous BOVW model. This model provides the spatial relation between a group of spatial features, which could reduce the cost of storage as the amount of information extracted is reduced. This method has proved to classify the classes of landscapes with the accuracy of 89%, which is higher than the previous model.

(Cheriyadat (2014)) has proposed an unsupervised learning method for scene classification, which performs the scene classification with the help of local spatial and structural patterns. This approach involves extraction of feature, encoding and pooling the feature to classify the scene from the aerial images. This method has overcome the previous difficulties of segmentation and individual classification of the segments. This experiment is reported to perform with accuracy of 81%. However, this method has failed to perform well in the large high-resolution dataset to perform scene classification.

With the advancement in deep learning models, which has been improved to implement in different sectors, author (Penatti et al. (2015)) has developed a deep learning ConvNets model to perform the classification of different objects in the aerial images.

This model extracts the features of the aerial image by the process of generalizing it and can extract the only the low-level features from the images. However, this model has failed to perform well in datasets other than UCMerced dataset. In another research, author (Hu et al. (2015)) has overcome the previous limitation by performing the extraction of features from both low-range and mid-range with the help of deep pre-trained model, and encoding the deep features on the global features. The experiment has been conducted with different pre-trained models such as CaffeNet, VGGNet, VGG-VD, PlacesNet, which has resulted that with the feature encoding done by VGG network the model has outperformed with accuracy of 96%.

Scene Classification has been active topic that has been evolving and adopting different technologies to attain the maximum result. The author (Hu et al. (2018a)) has elaborated about the transformation of scene classification technique from Bag-of-Visual models to deep learning model recently as shown in Table 1. However, the author has mentioned the limitations such as insufficient large dataset, insufficient description the contents of the elements in the image, less domains of images to train the model, which could limit the development of this technology. This disadvantage has been rectified in this research by the utilization Aerial Image Dataset (AID), which has been proved to the benchmark dataset for evaluating the performance of the aerial scene classification model (Xia et al. (2017)). The Table 1 provides the overview of research that has been mentioned in this section.

Table 1: Evolution of Scene Classification Models

Reference	**Dataset**	**Model**	**Strategy**	**Accuracy**	**Limitation**
(Hu et al. (2018b))	UCMerced	Bag-of-Visual-Words (BOVW)	concentric circle based spatial rotation invariant represent strategy	86%	Spatial feature not considered
(Chen and Tian (2015))	UCMerced	Bag-of-Words (BOW)	Pyramid-of-Spatial (PSR) extracts spatial features	89%	Performance is not evaluated with Large Dataset
(Cheriyadat (2014))	UCMerced, ORNL-I & ORNL-II	Unsupervised feature learning	Feature Extracting, Encoding & Pooling done by hard coding	81%	Outperformed only in Low-range feature extraction.
(Penatti et al. (2015))	UCMerced & Brazilian Coffee Scenes	ConvNets model	Feature extraction using CNN	90%	Less performance in other dataset.
(Hu et al. (2015))	UCMerced & WHU-RS	CaffeNet, VGGNet, VGG-VD, PlacesNet	Feature Extraction using Transfer Learning	96%	Benchmark dataset not utilized to evaluate.

2.3 Deep Learning Models for Large-scale Aerial Image Classification

Since the development of many deep learning models for numerous purposes in technical field, the researchers has also started to utilize and develop this technique for the purpose of Large-scale aerial scene classification. (Simonyan and Zisserman (2014)) has proposed an approach of increasing the layers with 3x3 filters for convolution and has achieved increase in accuracy compared to the traditional state-of-art method. This experiment has proved that the increasing the depth of the convolutional neural network can help the model to perform better in classification of different classes in large-scale.

This article proposed by (Nogueira et al. (2015)) has highlighted the importance of improving the performance of feature descriptors utilize in a scene classification model

which could adapt or each step in real-time. The model is composed of three convolutional, two fully connected layer and one final classification layer which classifies the class of the data. This model has proved to outperform the other feature descriptors with accuracy of around 90% in extracting and describing the classes of the images. This paper has also helped in proving that the application of deep learning model as feature descriptors can help the model to perform the task with higher accuracy. The Table 2 summarizes the information obtained from the research articles explored in this section.

Table 2: Deep Learning Model for Large-Scale Scene Classification

Reference	**Dataset**	**Model**	**Strategy**	**Accuracy**	**Contribution**
(Nogueira et al. (2015))	UCMerced & Brazilian Coffee Scenes	ConvNet Model	Improving performance of feature descriptors.	90%	Feature descriptors plays important role in Large-scale classification
(Nogueira et al. (2017))	UCMerced, Brazilian Coffee Scenes & RS-19	AlexNet, CaffeNet, GoogLeNet & VGG-16	Fine tuning transfer learning model for feature extraction.	90%	Tuning the feature descriptor and usage machine learning model as classifier to improve performance.
(Li et al. (2017))	UCMerced & WHU-RS	AlexNet, CaffeNet, & different VGG models	Fusion of features extracted from different layers. & Pooling done by hard coding	VGG-VD19 model =97%	This model requires higher computation power.
(Wang et al. (2017))	UCMerced & Brazilian Coffee Scenes. & Brazilian Coffee Scenes	VGG-VD & Resnet	Encoded Mixed-Resolution (EMR) to reduce and globalize the features	VGG16–97% ResNet–98%	Globalization of features reduce the computation time.
(Qi et al. (2018))	UCMerced & WHU-RS	CCP-net	concentric circle based spatial-rotation-invariant to extract the features	97%	Benchmark dataset not utilized to evaluate.

The paper proposed by (Nogueira et al. (2017)) has explored the application of existing deep learning models such as AlexNet, CaffeNet, GoogLeNet and VGG-16 models in scene classification by fine tuning the layers to outperform the existing feature descriptors and has achieved in improved accuracy in extracting the features from the data. Further this experiment has also involved in application of machine learning model such as Support Vector Machine (SVM) and proved improved performance of around 90% accuracy. This model has helped to prove the importance of fine tuning the deep learning model and utilization of machine learning as classifier for the model. Another article proposed on the same year by (Li et al. (2017)), has implemented an approach of fusing the features extracted from the multiple layer of the pre-trained models to improve the performance of scene classification. This technique has improved the performance of pre-trained models such as AlexNet, CaffeNet, and different VGG models, and the VGG-VD19 model has performed better with around 97% accuracy. This technique has also helped the model to perform better by fusing the low, mid, and high-level of features extracted from the data, but this model requires high computation power to do the same, which has been resolved in the approach proposed by (Wang et al. (2017)), which propose the utilization of CNN for extracting features and encoded by vector of locally aggregated feature descriptors (VLAD), and further reduced to get the global feature. This method has proved to adapt to images data of different sizes and reduce the computation of the process. The VGG 16 and Resnet models has performed with the accuracy of 97% and 98% respectively.

Another research proposed by (Qi et al. (2018)) has addressed the issue of misclassification of classes in the aerial image, the author has introduced a new pooling technique

called 'Concentric-Circle Pooling' which is based on the concentric circle based spatial-rotation-invariant presentation of an high resolution image, this method had been already proved in the article proposed by (Hu et al. (2018b)), that this method has improved the performance of BOW model. This method has improved the accuracy of the model to reach 97%, however this research has failed to address the issue of handling the images with different scale and size in CCP layer. As mentioned in the table.

2.4 Feature Extraction using Inception V3

Since the development of numerous deep learning models in the recent years, Inception V3 has been well known for its ability to extract the features of the images, less computational than the other existing transfer learning models and ability to reduce the parameters of the image without degrading its features. (Szegedy et al. (2016)) has proposed a model called Inception V3, which is optimized by scaling up the convolutional networks. A traditional inception layer has 1x1 convolution, 3x3 convolution, 5x5 convolution and a max pooling layer, which helps to extract the feature from images of any size and resolution. Author has proposed an approach by factorizing the convolution with large filter size, factorization into smaller convolution, and spatial factorization into asymmetric convolution method to reduce or replace the existing convolutional layers with another convolutional layer of less size and resolution. The model has proved to perform better in extracting the features from image resolution of 79x79, 151x151 and 299x299, than the other existing models. Detailed explanation of Inception V3 model will be done in Section 3.1 of this report. Because of its ability to extract features with higher accuracy this model has been experimented in numerous medical image classification researches, like one published by (Wang et al. (2019)). The author has utilized the Inception V3 as feature extractor and replacing the final classification layer with SVM, Softmax and Logistic Regressor. The pre-trained model has proved to perform with highest accuracy of 95%, and because of lesser computation of the feature extractor, development of application with this model is feasible. Aerial Image classification is another field in which higher resolution and complex images involves and hence experimenting the classification of such images with Inception V3 can yield higher accuracy than the other networks as experimented by (Pires de Lima and Marfurt (2020)). The researcher has utilized the pre-trained Inception V3 model for the purpose of scene classification from aerial images, and it has been proved that the Inception model has outperformed all other models by performing better in all the benchmark datasets for aerial classification. However, the author has utilized the pre-trained layers for both primary and secondary tasks of feature extraction and image classification respectively, but this research will utilize the pre-trained network for extracting the features from the aerial image dataset and the feature will be classified with the tree-based classifiers.

2.5 Classification of Aerial Image Features

The research published by (Sheykhmousa et al. (2020)), has critically evaluated the performance of SVM and Random Forest in Remote Sensing Image Classification. This paper has reviewed the technical articles published in aerial image scene classification. The author has mentioned that SVM classifiers perform better in dataset of less classes, but when it comes to large datasets of many classes this model performance less, unless the kernel and parameters are fine-tuned, because of the multidimensionality in data.

Whereas, Random Forest can handle large dataset without deletion, and robust to multidimensional features of the dataset. SVM and RF are the pixel-based classifier, which means the model does not consider the spatial dependency of the nearby pixel. The author has explained that implementation of RF in aerial scene classification has not been explored deeply as compared to SVM. Additionally, the author has also mentioned that combining the feature extraction ability of deep learning network and classification of features using SVM of RF can yield higher result. (Belgiu and Drăguţ (2016)) has proved that Random forest can perform better than any other ensembled models like Adaboost, provided the parameters are tuned.

(Polat and Güneş (2009)) has published an article about classification of multi-class image using C4.5 decision tree model, which has performed with good result in recognizing the pattern of the classes and providing the result. The model of decision tree utilized for this research is Classification and Regression Tree (CART) which is similar to C4.5. CART constructs binary trees by utilizing the features and threshold and provide and provides the gain of information in each node.

2.6 Research Gap & Justification

From the research publications mentioned in this section, it has been observed Tree-based classifiers has been explored less in the field of aerial scene classification. Inception V3 model has proved to be efficient in extracting features from the complex or higher resolution images, and Random Forest classifier has the potential to process and classify the multiple classes, multidimensional and high resolution features of image (Belgiu and Drăguţ (2016)), but there was absence of experiment in utilizing the deep feature extraction network for extracting features from the aerial higher resolution image data and labeling them with the appropriate classes using Tree-based classifier, this research aims to fill that gap and determined to produce higher evaluation result and designing a robust model.

3 KDD Methodology

This research has been conducted by following the procedure of Knowledge Discovery in Database (KDD) Methodology. Figure 1 shows the steps of the KDD methodology, which has been explained in the below sections.

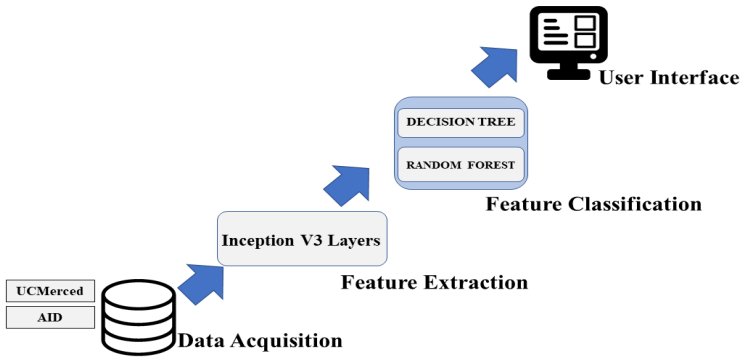

Figure 1: Research Methodology

3.1 Data Acquisition

This experiment has been conducted with two aerial image datasets of different resolution. First part of the research has been conducted with UCMerced dataset and later part has been conducted with AID dataset, to evaluate the robustness of the model.

Figure 2: UCM dataset

Figure 3: AID dataset

UCMerced is manually labeled ground truth dataset consisting of 21 different land-use classes of ground truth data, from aerial orthoimagery as shown in Figure 2. The images are with the resolution of 256 x 256 pixels, with pixel resolution of 1 foot. The dataset has been made with the large images from the US Geological Survey (USGS), the images were cropped such that it contains some land-cover and object classes. Numerous different spatial patterns, and heterogenous respect to color and texture, has been mentioned as the main reason for choosing these images for this dataset (Yang and Newsam (2010)). Every land-use class of landscape has 100 images of mentioned resolution, with different texture and color.

AID dataset on the other hand is the large-scale dataset consist of 10000 aerial images, collected, and object annotated (Xia et al. (2017)). This dataset contains 30 different land-use classes containing about 200 to 400 images of the same classes with resolution of 600 x 600 as shown in Figure 3. This dataset has the features such as higher interclass variations, smaller interclass dissimilarity and relatively large-scale dataset.

The aerial images mentioned in Figure 4 shows the variations on the texture, color and other properties of the aerial image which belongs to the same land-use class. And the aerial images in Figure 5 shows the less dissimilarity of properties such as texture and color, between two different land-use classes, which may be hard for the machine learning and shallow learning models to perform classification with accuracy. These properties of this benchmark dataset could help to train the model to distinguish the features of different landscape which could make the classifier to perform robust irrespective to the resolution of the data.

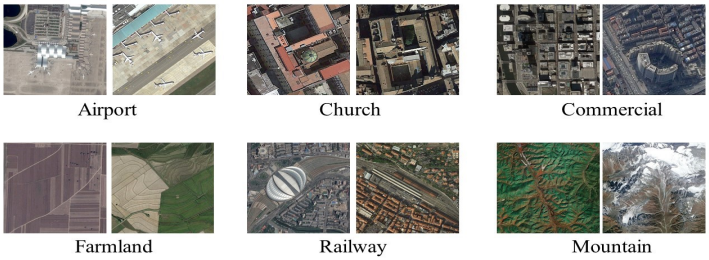

Figure 4: Aerial Images of AID dataset showing inter-class variations.

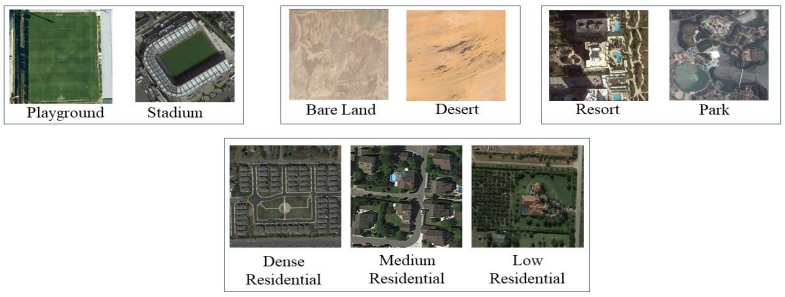

Figure 5: Aerial Images of AID dataset showing intra-class similarity

3.2 Feature Extraction

GoogLeNet, which has been declared as the winner of ILSVRC-2014 in performing classification and object detection challenge. This network utilizes the 'Inception Module', which has 22 layers and can process the data in parallel, which is called 'Inception V3'. This network has the characteristics of having filters of different sizes in the same layer, which can help the network to extract the features at different levels. The Architecture of Inception layers has been shown in the figure below.

The Figure 6 shows the schematic architecture of the layers of Inception V3 network, which has bee utilized for feature extraction. The network is 48 layers deep, which has the advantage that the shallow-level, intermediate-level and deep-level features can be extracted from the model. Since this project requires the deep features of the aerial image we have utilized the 'Mixed10' layer of the network which provides the deep features of the network as shown in the architecture. Softmax layer which performs the classification was removed from the network, because this network is utilized to extract the features of the image and classification will be performed by Random Forest and Decision Tree classifiers.

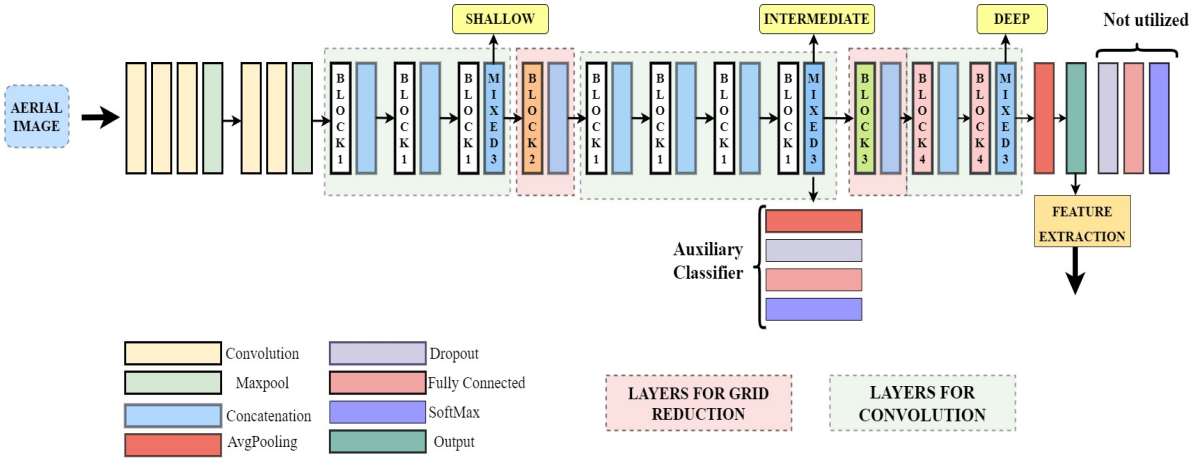

Figure 6: Schematic diagram of Inception V3 Architecture

In the Figure 7 the Block 1 and Block 4 are Inception blocks which performs the operation of extracting the features from the image data, whereas the Block 2 and Block 3 performs the operation of reducing the size of grid, by reducing the parameters of features, which has been represented in Green and Red regions in the Architecture Figure 6. This architecture has 1x1 convolution layer which performs the operation of reducing the parameters of the data, to reduce the computation time. The inception blocks contain the larger kernels of 3x3 and 5x5 convolution layers which helps the network to extraction the features of different levels of depth. Every 1x1 convolution layer has a ReLU activation layer which increase non-Linearity of the features, which could help the classifiers to distinguish the features

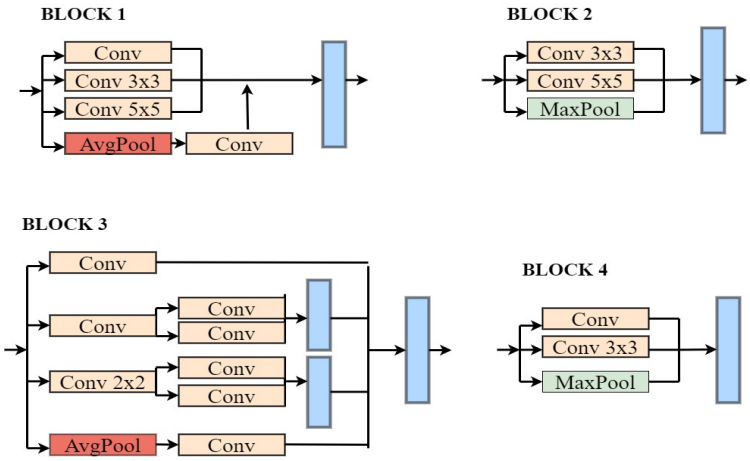

Figure 7: Architecture of different Inception blocks in Inception V3 Model

The process of extracting features proceeds as the different convolution layers extract the different features of the images and concatenates the feature at the end of the network, which helps the network to reduce the parameters of the data with losing the quality of the features. The Convolution layer in the network computes the dot product of the network by sliding the weights across the image data, and obtained the correct value by back propagation. This convolution layer has the ReLU activation function which increase the non-linearity of the data. There are two different Pooling layers utilized in this network, Maxpooling layer and Average Pooling layer, which helps to reduce the computation

strain of the model. Concatenation Layer, conflates the features extracted by different layers and reduce the dimension of the data, with high-level of feature description. This transfer learning model is utilized for Feature Extraction, the layers such as Dropout, Fully connected and Softmax classifier layers were not utilized for the research.

The features and corresponding labels extracted by the Inception V3 model has been stored as dataset in HDF5 file format[1]. HDF5 (Hierarchical Data Format) is a file designed to store and organize the huge data in organized manner.

3.3 Feature Classification

The Feature Classification of the aerial image features and labels extracted from the previous module, has been classified with Decision Tree and Random Forest. The process of classifying the classes of different land-use categories will be explained as follows.

3.3.1 Decision Tree

Decision tree has the structure that resembles the flowchart, where the internal nodes represents the features carried in the point and the branch represents the decisions made from the node. Utilization of decision tree for the purpose of scene classification is said to be non-parametric approach towards pattern recognition. The decision tree utilized for this research has been imported from the 'sklearn.tree' package [2] from the python environment. The decision tree utilized from this package is called Classification and Regression Tree (CART), which can perform classification for categorical data, and regression for continuous data. CART develops the binary trees by utilizing the features and threshold value that can provide the highest information gain at every node. The Inception layers perform the operation of extracting features and along with reducing the dimensionality, which could help the decision tree to perform better, because the decision tree can perform with reduced dimensions.

The CART decision tree used in this research utilizes the Gini method (criterion: "gini") to create the split points in the data. Gini impurity indicates the frequency of misclassified labels occurred in the random chosen element in the subset.

The pruning of decision tree may occur by excessive growth of branches of the decision tree, which may increase the computational power and reduces the performance. The pruning has been prevented in the model by setting the parameter value of 'max_leaf_nodes=360' in this experiment. Usually, the data can be visualized as a flowchart showing the decision split that occurred at each stage, but in this research since the data has been already converted into numerical format by extracting the features. The visualization of the data will not help in interpreting the process of classification.

3.3.2 Random Forest

Random forest is said to be ensembled model, which works by integrating cumulative decision provided numerous decision trees utilized in the network. Integration of multiple decision classifiers can reduce the variance and increase the performance of the forest classifier. The final decision of labeling the data with appropriate classes will be done depending on the voting from the random sampling of the unit tree classifiers embedded

[1]https://www.h5py.org/

[2]https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

together. The Random Forest Classifier utilized for this research has been imported from the 'sklearn.ensemble'package [3] from the python environment. The decision split of the decision tree depends on the Gini Impurity value, which has been set in default for the model. This method of random sampling and integration of output of results is called Bootstrap Aggregation (Bagging).

The algorithm for the function of Random Forest is as follows, (i) random selection of 'n' training samples from ground dataset, using Bootstrap, (ii) 'k' round of extraction is performed, and training sets were obtained, (iii) Training the individual decision trees with 'k' training sets, (iv) The average of results from each tree model provides the result of the ensembled model.

The parameter of the model has been defined with n_estimator = 10, which the minimum value, which defines the number of trees in the forest. Parameters such as criterion for splitting the branch (criterion: "gini") is set to default, which is gini. 'minimum_sample_split' is kept as 2, which defines the minimum samples required to make a split. The base_estimator is defined as the 'DecisionTreeClassifier'.

3.4 User Interface

This User Interface has been developed to utilize the trained classifier models and feature extractor layers, to classify the land-use class of the test image. The test images were cropped from the Google Earth website [4] and provided as input to GUI. The feature extraction layers and their operations were defined in different functions in the program, which can be utilized to extract the features from the image provided by the user. The user interface developed is a desktop graphical user interface. This application has been created by utilizing the different packages from the PyQT5 graphical user interface framework as shown in the figure below. The trained classifier models were stored as pickle files, individually. Pickling is used to store the python objects by serializing it. The trained classifiers have been dumped into individual files, which are be utilized in the user interface to classify the test aerial image provided by the user. The functions in GUI takes place in three steps, the user imports the test image, and extracts the feature and label from the image, and classify the scene class category of the image.

4 Design Specification

The overall architecture of Aerial Scene classification approach that has been followed in this research is shown in the Figure 8. The research has been conducted in three phases, Feature Extraction Layer, Feature Classification layer and User Interface Layer.

- **Feature Extraction Layer:** In Feature Extraction layer, the process of extracting deep features from aerial images has been performed and the features and corresponding labels were stored in separate weights.
- **Feature Classification Layer:** The extracted features and labels were imported into the Feature Classification layer, and utilized to train the classifier models and the trained models were stored as separate executable files.

[3] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[4] https://www.google.com/earth/

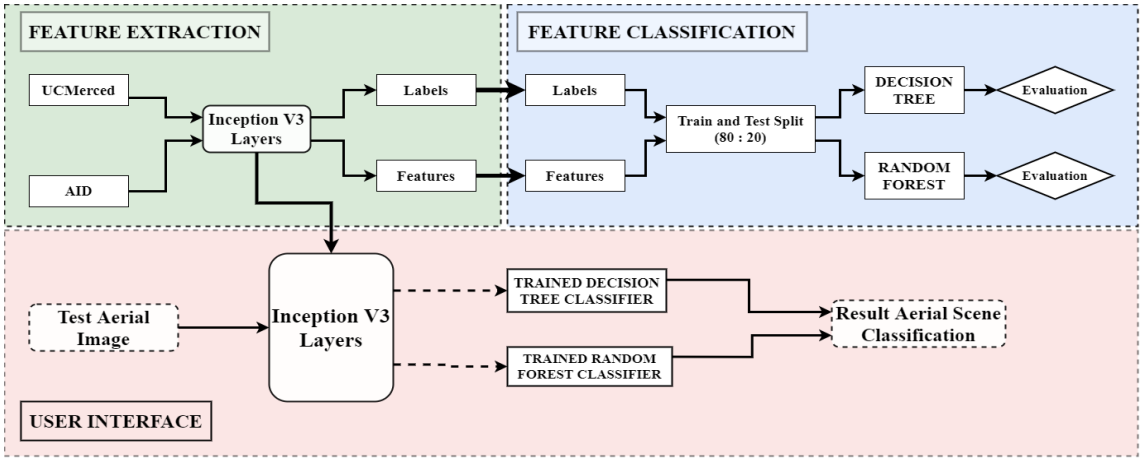

Figure 8: Design Specification of Scene Classification framework

- **User Interface Layer:** In User Interface layer, a desktop user interface has been developed which could help the user to utilize the trained classifiers to classify the test image. This layers enables the utilization of inception layers to extract the features of the image and to classify them using the trained classifier models.

5 Implementation

The implementation of the research, has been conducted by following the procedure mentioned in the Design Specification Section 4.

5.1 Feature Extraction

The extraction of features from the aerial images has been done by utilizing the layers of the Inception V3 model. The phase of implementation is done by importing the dataset and extracting the features and its corresponding labels and storing the weights in different files, which can be utilized to train the classifier. The Inception V3 network utilized for this project has been imported from the 'keras.applications'[5] library, and pre-trained with primary weight file. The pixel dimensions of the images in datasets were different, UCM dataset has images of 256 x 256 pixel dimension and AID dataset has aerial images of 600 x 600 pixel dimension, these images were reshaped to 299 x 299 dimension and imported in input layer of the network in RGB channel format (299x299x3). The images were converted in the array and manipulated by expanding the dimension of the array to one sample. The arrays were pre-processed into machine readable format to import the data into the layers of the model. The image data in the input layer has passed through the series of deep convolution network, to extract the reduced feature parameters. The extracted features were flattened in single array without losing its parameters in order to fit the input layer of machine learning classifiers. This process has been designed to operate as loop to extract features from every image from every classes and its labels. The process of importing and extracting the features from image has been visually presented in the console, with the help of progress bar using 'tqdm' package[6].The

[5]https://keras.io/api/applications/
[6]https://tqdm.github.io/

extracted features and its corresponding labels were appended on the individual arrays. Later a dataset is created using the 'h5py'[7] package and the appended feature values and label values were dumped as individual files.

5.2 Feature Classification

The feature classification phase of the research has been performed by tree-based machine learning classifiers, Decision Tree and Random Forest. The weights of features and labels extracted from the previous phase has been loaded to this phase. Two empty arrays have been created and the features and labels values were appended on the arrays. The data is now splited in to training and testing data in the proportion of 80:20, to evaluate the performance of the classifiers, by utilizing the 'train_test_split' library of 'sklearn.model_selection'[8] package. Decision Tree classifier has been defined with the parameter of max_leaf_nodes = 360, which could limit the growth of the branches of the tree, in order to prevent pruning and increase the accuracy without degrading. Gini Index is considered here to measure the impurity in the random samples considered and to make the branch split.

Random Forest classifier has been imported and the number of decision trees to be included in the forest is set to 10 (n_estimators = 10), which is the minimum value, in order to prevent the over-fitting of the model. The classifiers with the parameters were imported into an array and the classifiers were trained with the training set of the data and evaluated with the test set. The results of classification such as Accuracy, F1-score, Precision and Recall were appended in a dataframe and printed on file and exported in '.csv' format. And the results of the confusion matrix were exported as image in '.png' format. The trained classifier models were serialized and saved as pickle file, which can be imported in the user interface for testing the performance of the classifiers.

5.3 User Interface

The user interface has been developed which would make the process of classifying the land-use class of aerial images, by utilizing the deep feature extractor and trained tree-based classifiers, more feasible. The interface has been developed with the help of PyQT5 graphical interface framework[9]. The front-end libraries of the application have been imported from 'PyQT5.QTGui' and 'PyQT5.QTWidgets' packages. The stored classifiers were imported into this module. Once the user is imported the test image, features extraction will be done by utilizing the inception layers. And the extracted features were stored in an array. To perform classification of the test image features the user has to choose anyone of the classifiers mentioned in the drop-down button. Further the classification will be done by utilizing the weights of the trained classifier and the results will be displayed. The whole layout is divided into two parts, one deals with getting the arguments from the user and other deals with visualizing the input image and output of the classifiers chosen to perform scene classification.

[7] https://www.h5py.org/

[8] https://scikit-learn.org/stable/modules/generated/sklearn.model$_s$*election.train*$_t$*est*$_s$*plit.html*

[9] https://pypi.org/project/PyQt5/

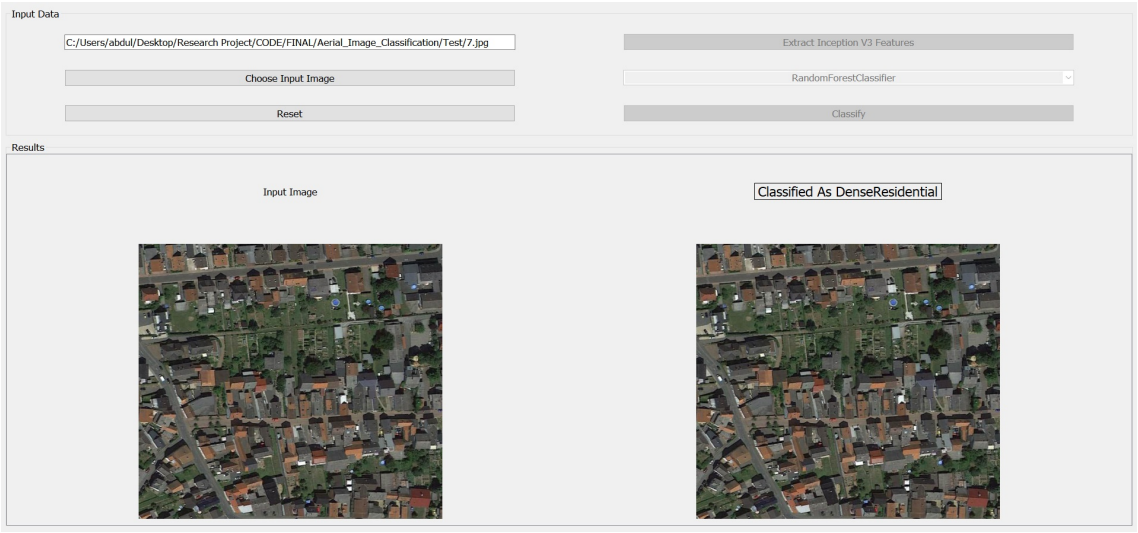

Figure 9: User Interface

6 Evaluation

The Evaluation phase of a project is as important as the execution of the project, where the experiment conducted has been critically evaluated to measure the performance using the metrics. Since the task perform in this research is classification, the metrics followed in this research are calculation of Accuracy, F1-Score, Recall and Precision, and in addition to this the confusion matrix has been plotted for the multiple classes in the model. The results of the evaluation were appended on a dataframe and exported as '.csv' file. The evaluation of performance of the classifiers has been conducted with two different datasets of different pixel resolution. The first phase of evaluation is to compare the performance of the Random Forest in UCM dataset and the AID dataset in the Section 6.1 and second phase of evaluation is to compare the performance of Decision tree with UCM and AID datasets in the Section 6.2. The models were trained with the UCM data and the performance is evaluated, in order to test the robustness of the model the AID dataset (benchmark dataset) has been used. The values of the evaluation metrics were multiplied with 100 and the decimal points were reduced to round-off to two value, in order to show the values in percentage.

6.1 Evaluation of performance of Random Forest

The Random forest classifier has been trained with the train set of data and evaluted with the test set of data. It has been observed that the Random Forest has perform good with the both datasets UCM dataset and AID dataset. The Random Forest model has proved to be robust to the high resolution of input data, with the integration of Inception V3 model for feature extraction. The Figure 10 provides the comparison of the evaluation metrics of different datasets with Random Forest classifier.

Accuracy provides the proportion of right prediction of land-uses class among the other prediction of classes, it has been observed that Random forest model has performed better in both the dataset in correctly classifying the classes, with 98.52% accuracy and 98% accuracy for UCM dataset and AID dataset, respectively. Recall, which is also called as Sensitivity of the model, provides the correctly predicted positive land-use classes among

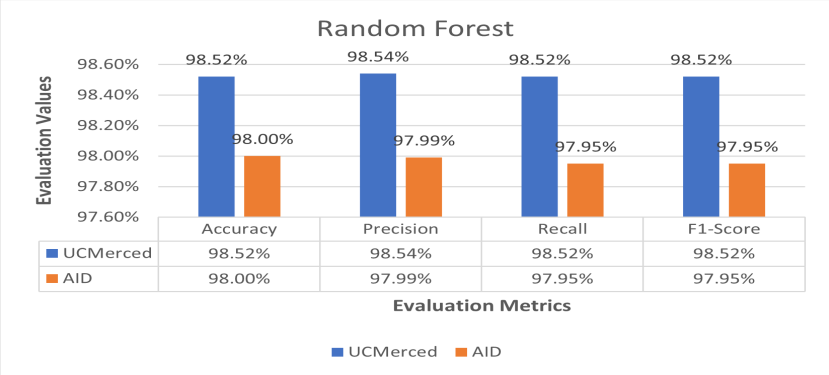

Figure 10: Evaluation of Random Forest Classifier.

the other positive land-use classes which is observed to be 98.52% for UCM dataset and 97.95% for AID dataset. Precision helps to evaluate the ability of the model to predict only the correct instances of each land-use class, and Random forest classifier has proved to have 98.54% in Precision to predict the correct instances in UCM dataset and 97.99% Precision in AID dataset. F1-score can help to evaluate the robustness of the model irrespective to the multiple classes provided for classification, Random Forest has proved to be robust when trained with both datasets, with F1-Score of 98.52% for UCM dataset and 97.95% for AID dataset.

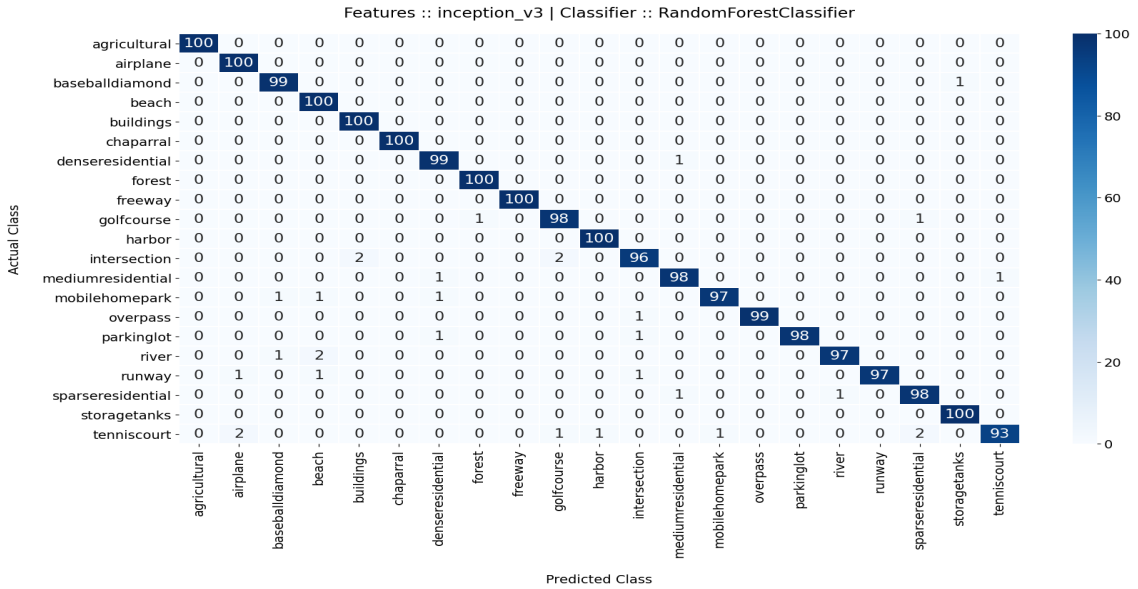

Figure 11: Confusion Matrix of Random Forest with UCMerced dataset

The confusion matrix has been plotted and visualized using the heatmap, the prediction of different land-use classes of UCM dataset and AID dataset has been shown in the Figure 11 and Figure 12, respectively. The actual land-use classes or the ground truth data has been provided along the y-axis and the output of predictions of the classifiers are plotted along the x-axis as Predicted classes. The dark pixels along the diagonal of the matrix shows the correct prediction of actual classes and the predicted classes. The graph plotted shows the dark boxes along the diagonal shows that almost all the classes in the test image set as been correctly predicted by the model.

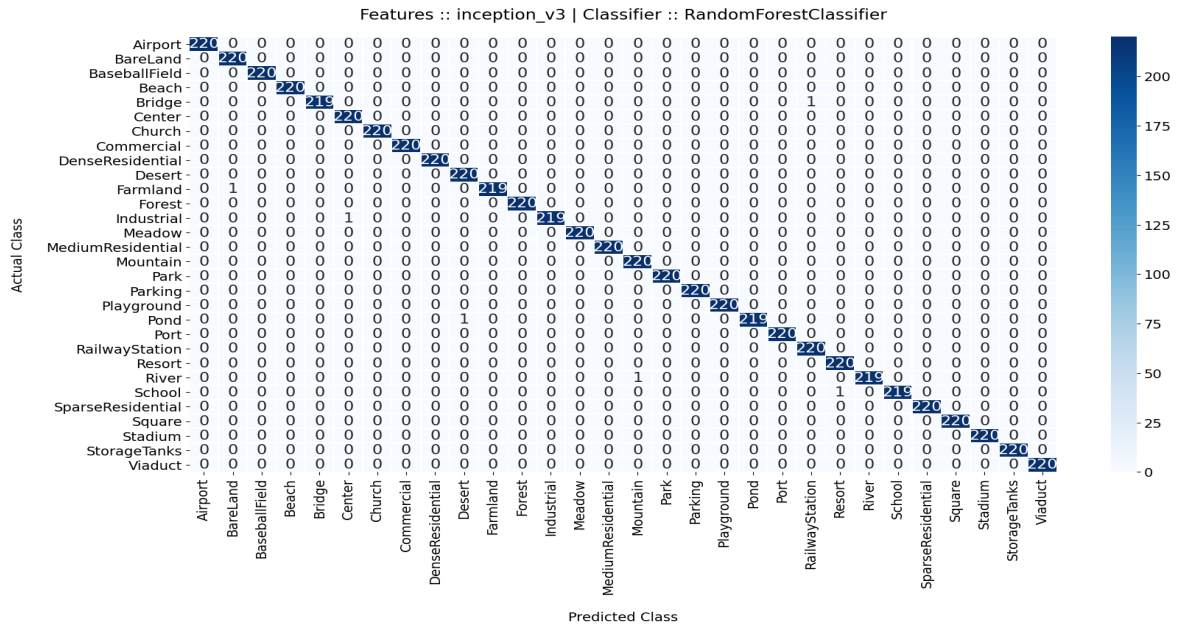

Figure 12: Confusion Matrix of Random Forest with AID dataset

6.2 Evaluation of performance of Decision Tree

The datasets have been passed through Inception V3 model for feature extraction, the Decision Tree classifier model has been utilized to classify the multiple land-use classes from the test data. The Decision Tree classifier has performed with accuracy of 94.76% when trained with UCM dataset, but failed to perform well in AID dataset, where it performed with accuracy of 62.17%. This proved that the Decision Tree classifier cannot perform better when it comes to higher resolution aerial image dataset and datasets with high correlation of features between classes. The evaluation metrics of Decision Tree classifier with UCM and AID datasets has been shown in the Figure 13.

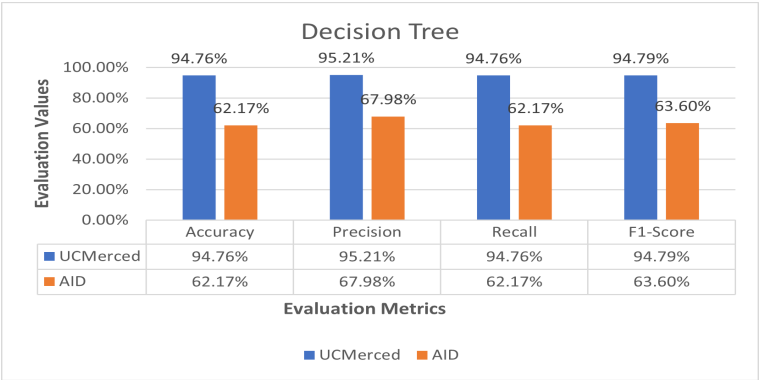

Figure 13: Evaluation of Decision Tree Classifier.

The decision tree classifier has performed with 95.21% Precision in UCM dataset and it reduced to 67.98% in AID dataset, identifying the correct instances among each land-use classes. Decision Tree classifier has performed with 94.76% and 67.98% Recall, in the UCM and AID datasets. As mentioned, the Decision Tree has failed to be robust with respect to higher-dimension of the data, which has been shown with the F1-Score value of 94.79% for UCM dataset, low-resolution aerial image dataset, and F1-Score value for

63.6% in AID dataset, which has aerial images of high resolution.

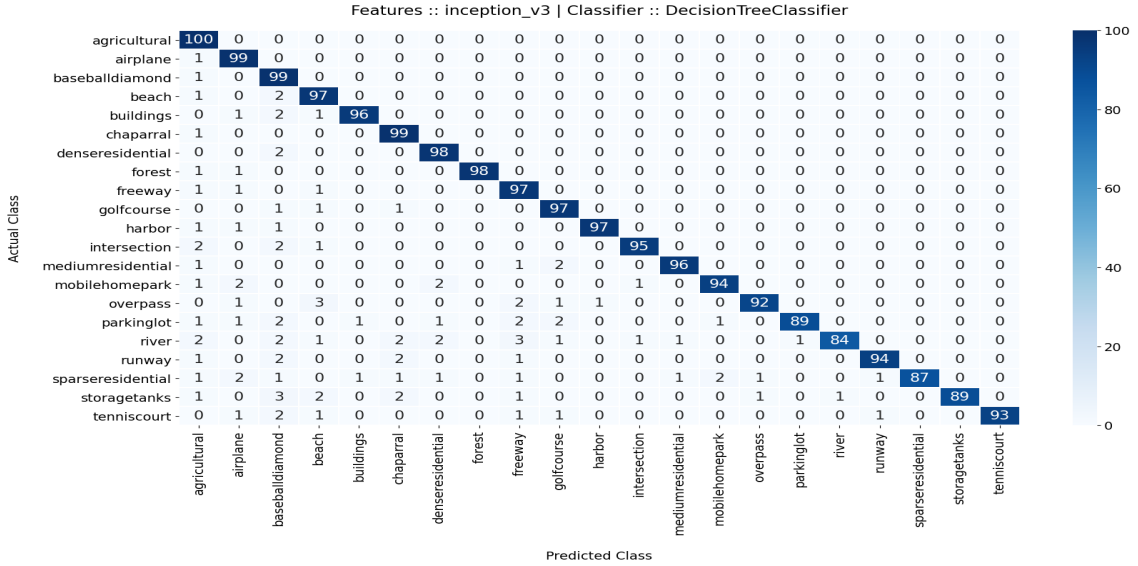

Figure 14: Confusion Matrix of Decision Tree with UCMerced dataset

The confusion matrix has been plotted with heatmap as shown in the figures below for both the datasets. The confusion matrix of the UCM dataset has been shown in Figure 14, has dark pixels along the diagonal of the plot which shows the boxes intersecting the same land-use class in actual class and predicted class has the maximum value of correctly predicted land-use classes. But in the confusion matrix plotted for AID dataset classified using the decision tree is shown in Figure 15, the diagonal boxes show that the values on the diagonal boxes were between 180 to 100 which shows that model has averagely performed in predicting the different land-use classes in AID dataset.

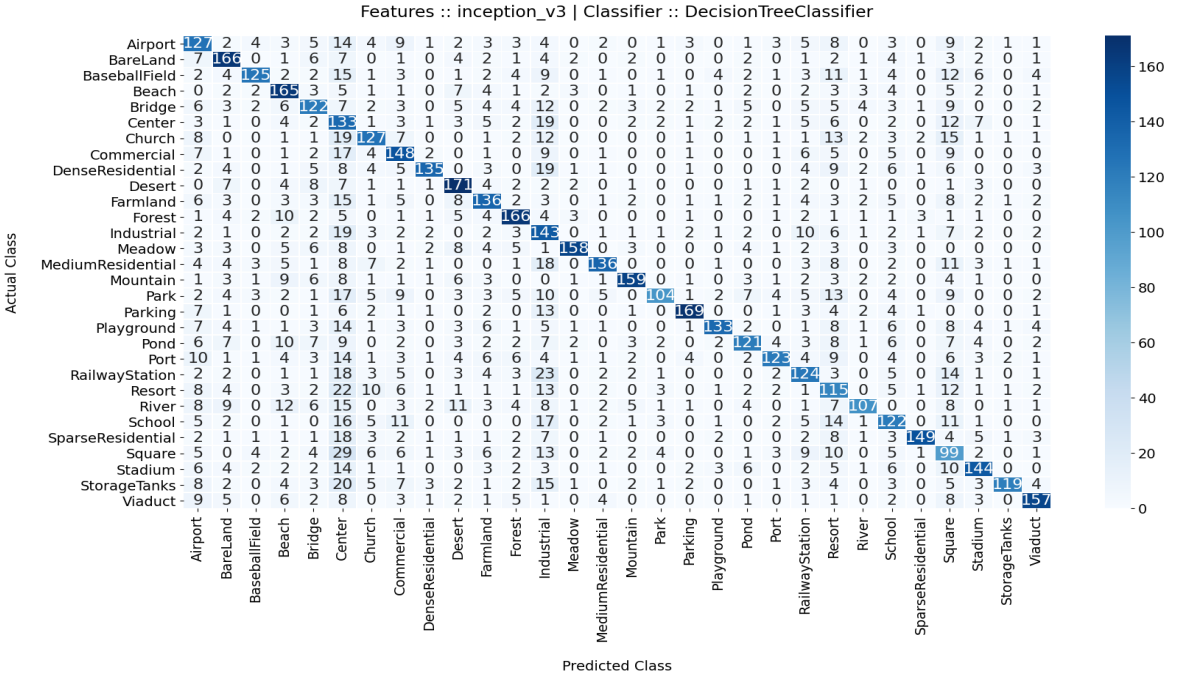

Figure 15: Confusion Matrix of Decision Tree with AID dataset

6.3 Discussion

From the above evaluation of the performance of tree-based classifier, integrating with Inception V3 feature extractor, it has been evident that Ensembled model, Random Forest has performed better than the Decision Tree model, in classifying the scenes from the high resolution aerial images. The Random Forest classifier has proved robustness by performing with 98% accuracy in AID dataset, whereas Decision Tree was not robust to the high-resolution dataset, as observed with the accuracy of 62.17% for AID dataset. The parameter for Decision Tree has been set with 'max_leaf_node = 360' for UCM dataset, increasing further reduce the accuracy. In case of AID dataset, increasing the 'max_leaf_node' parameter has improved the performance of the Decision Tree, but this increases the risk of pruning of the Decision Tree, which will drastically affect the performance of the Decision Tree classifier.

Further the performance of the trained tree-based classifiers with respect to different aerial images can be done by deploying them to the user interface. The test image has been extracted from Google Earth website[10]. The test image has been imported into the User Interface, and the features were extracted by using the Inception V3 layers and one of the trained classifiers as shown in the Figure 9. And it has been observed that the Random Forest classifiers has classified all the different resolutions aerial images provided for testing the trained classifier. Decision Tree has classified the all the low resolution aerial images, whereas it failed to classify the high resolution images.

This experiment proved that the utilization of pre-trained Deep Convolutional Feature Extractor and Random Forest classifier can yield better performance, irrespective to the higher dimensionality of the data.

7 Conclusion and Future Work

With the development in the field of utilization of deep learning approach on aerial scene classification field, this research could provide an insight of utilization of deep convolution feature extractor and Tree-based classifiers on aerial landscape datasets with different resolutions. And the evaluation conducted in Section 6 it has been observed that Random Forest classifier has performed with 98.52% and 98% accuracy in UCM dataset and AID dataset, respectively, which helps prove the robustness Random Forest classifier irrespective to the dimension of the aerial image data. Decision Tree has performed with accuracy of 94.76% in UCM dataset, which proved that decision tree classifier can perform good with the low-dimensional aerial image dataset, whereas this classifier failed to perform equally good in case of AID dataset, the model performed with 62.17% accuracy which proves that the decision tree classifiers could not perform better with high-dimensional aerial image dataset. Hence, it has been concluded that Random Forest classifier is robust to different resolutions of aerial images, provided the deep features has been extracted with deep convolution network, which makes this classifier suitable to be used along with applications running in different platform.

To mention as future development of this research, the process of training the classifiers, can be performed with hyper-parameter optimization using algorithm like GridSearchCV,etc., instead to hard-coding the parameters to the classifiers. The implemented approach can be further developed to extract the features from hyper-spectral images

[10]https://www.google.com/earth/

which has been captured with different bandwidth. With respect to User Interface, the application can be developed into executable file, which may help the researchers to explore by implementing different trained classifier models and different features extractors and observe the performance.

References

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing* **114**: 24–31.

Chen, S. and Tian, Y. (2015). Pyramid of spatial relatons for scene-level land use classification, *IEEE Transactions on Geoscience and Remote Sensing* **53**(4): 1947–1957.

Cheriyadat, A. M. (2014). Unsupervised feature learning for aerial scene classification, *IEEE Transactions on Geoscience and Remote Sensing* **52**(1): 439–451.

Hu, F., Xia, G.-S., Hu, J. and Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery, *Remote Sensing* **7**(11): 14680–14707.

Hu, F., Xia, G.-S., Yang, W. and Zhang, L. (2018a). Recent advances and opportunities in scene classification of aerial images with deep models, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp. 4371–4374.

Hu, F., Xia, G., Yang, W. and Zhang, L. (2018b). Recent advances and opportunities in scene classification of aerial images with deep models, *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4371–4374.

Li, E., Xia, J., Du, P., Lin, C. and Samat, A. (2017). Integrating multilayer features of convolutional neural networks for remote sensing scene classification, *IEEE Transactions on Geoscience and Remote Sensing* **55**(10): 5653–5665.

Nogueira, K., Miranda, W. O. and Dos Santos, J. A. (2015). Improving spatial feature representation from aerial scenes by using convolutional networks, *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 289–296.

Nogueira, K., Penatti, O. A. and Dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification, *Pattern Recognition* **61**: 539–556.

Penatti, O. A. B., Nogueira, K. and dos Santos, J. A. (2015). Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?, *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 44–51.

Pires de Lima, R. and Marfurt, K. (2020). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis, *Remote Sensing* **12**(1): 86.

Polat, K. and Güneş, S. (2009). A novel hybrid intelligent method based on c4. 5 decision tree classifier and one-against-all approach for multi-class classification problems, *Expert Systems with Applications* **36**(2): 1587–1592.

Qi, K., Guan, Q., Yang, C., Peng, F., Shen, S. and Wu, H. (2018). Concentric circle pooling in deep convolutional networks for remote sensing scene classification, *Remote Sensing* **10**(6): 934.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P. and Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **13**: 6308–6325.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* .

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016). Rethinking the inception architecture for computer vision, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Wang, C., Chen, D., Hao, L., Liu, X., Zeng, Y., Chen, J. and Zhang, G. (2019). Pulmonary image classification based on inception-v3 transfer learning model, *IEEE Access* **7**: 146533–146541.

Wang, G., Fan, B., Xiang, S. and Pan, C. (2017). Aggregating rich hierarchical features for scene classification in remote sensing imagery, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**(9): 4104–4115.

Xia, G., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L. and Lu, X. (2017). Aid: A benchmark data set for performance evaluation of aerial scene classification, *IEEE Transactions on Geoscience and Remote Sensing* **55**(7): 3965–3981.

Yang, Y. and Newsam, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification, *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pp. 270–279.

Zhang, B., Chen, Z., Peng, D., Benediktsson, J. A., Liu, B., Zou, L., Li, J. and Plaza, A. (2019). Remotely sensed big data: evolution in model development for information extraction [point of view], *Proceedings of the IEEE* **107**(12): 2294–2301.