Enhancing Land Use Classification through Deep Learning with UAV Imagery

MSc Research Project
Programme Name

Albin Saji
x21188777

School of Computing
National College of Ireland

Supervisor: Dr David Hamill

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Albin Saji
Student ID:	x21188777
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Dr David Hamill
Submission Due Date:	12/08/2024
Project Title:	Enhancing Land Use Classification through Deep Learning with UAV Imagery
Word Count:	8695
Page Count:	27

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:	Albin Saji
Date:	06/08/2024

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 placedinto the assignment box located outside the office.

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

AI Acknowledgement Supplement

1 [Insert Module Name]

2 [Insert Title of your assignment]

Your Name/Student Number	Course	Date

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

3 AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool

4 Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

[Insert Tool Name]	
[Insert Description of use]	
[Insert Sample prompt]	[Insert Sample response]

5 Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

6 Additional Evidence:

[Place evidence here]

7 Additional Evidence:

[Place evidence here]

Enhancing Land Use Classification through Deep Learning with UAV Imagery

Abstract

Land all around the world has various uses. In particular, tracking land use is one of the crucial tasks that a government can perform to track urban planning, environmental monitoring and even disaster management. The following task is both time-consuming and prone to errors when done traditionally. This research study proposes the adoption of deep learning models for land classification using the UAV imagery data. The study proposes the use of CNN networks and various architectures to detect the land using a pre-labelled dataset. The study has incorporated the CRISP-DM approach for land classification, and four different architectures are built, including, CNN, Alex-Net, VGg-16, and ResNet-50, where the ResNet-50 attained the maximum accuracy showcasing the advantage of deep learning and deep architectures for a given problem.

1 Introduction

7.1 Background

Land use worldwide is increasing rapidly with the increasing demand for food, livestock, and fuel. The "take, make, and dispose of" model fuels the global economy, where humans extract large quantities from the Earth, destroying the Land for animal farming, agriculture, and other activities. According to the McKinsey report, an additional cropland of 70 to 80 million hectares would be required by 2030 (Striking the Balance: Catalyzing a Sustainable (Land-Use Transition, 2023). In such scenarios, making sustainable efforts is the primary goal, and illegal use must be avoided. The IISD Project showcases how cattle grazing and soybean faring for animal farming are harming the Amazon rainforest, with almost 80% of deforestation in the Amazon. As further highlighted by Hannah Ritchie and Max Poser, the following affects rising sea levels and worsening conditions worldwide by creating arid environments and deserts. In such scenarios, tracking land use is the main aim of most governments and agencies worldwide.

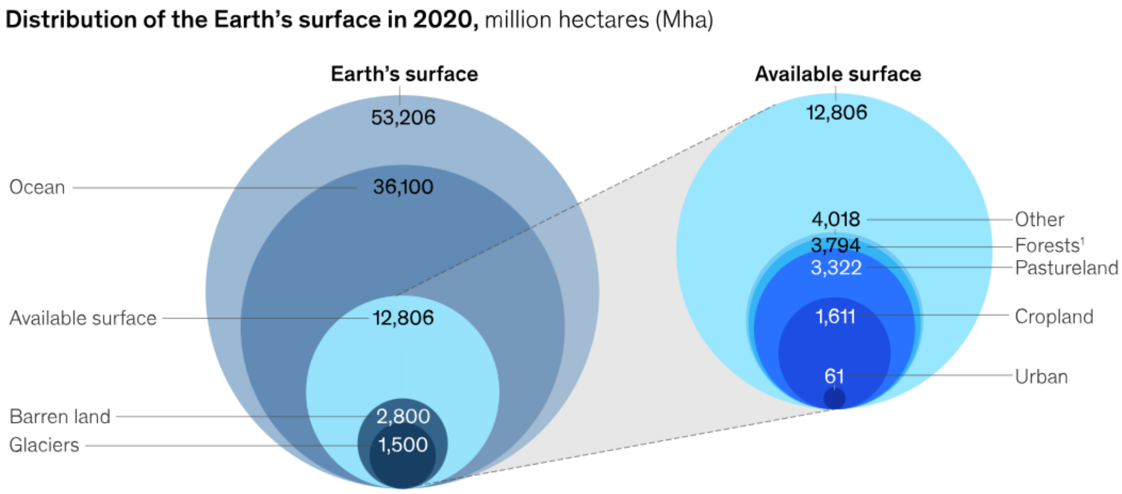


Figure 1 Land Distribution

(Hannah Ritchie and Max Roser, 2019)

7.2 Motivation

A satellite is an artificial object placed in orbit for various uses. Satellite imagery has increased interest from various users, businesses, and governments in earth observations, navigation, communication, weather, etc. Moreover, satellites can provide detailed photography and 'live' feeds of specific locations across Earth. With advancements in cloud computing and AI, the image data recorded by such satellites can be used to track various real-time activities occurring on our planet things on Earth. Governments are using it to track the forest and vegetation to protect them from wildfire. Scientists are using it to understand the Arctic glaciers and rivers by comparing it with past data. As with the advancement in AI, image classification has gained much popularity. One can easily classify the images with state-of-the-art models like CNN and LSTM. CNN and LSTM are deep learning models, where the CNN model deals with sequential data, and the LSTM model also deals with sequential data with one advantage of learning the context of the previous word, as the given model has memory inside it. Hence, with these models, one can implement them on the land dataset to categorize the Land and understand its usage over time. Gaining suitable remote sensing and deep learning knowledge would help build better prediction models in a given area. The study will focus on addressing the issue of land use using remote sensing data and implementing an AI model to categorize the Land using satellite data.

7.3 Research Objectives

1. How effectively are deep learning models classifying the Land using UAV imagery?
2. What is the performance difference between various model architectures when applied to a given dataset?
3. How can transfer learning and fine-tuning of the models improve their efficiency and reliability?

7.4 Research Questions

1. Can we evaluate the data quality and its effects on the deep learning architectures?
2. Can we evaluate the performance of various architectures based on the batch size of data?
3. Can we improve land use classification accuracy by more effectively capturing fine-grained spatial patterns alongside temporal changes with CNNs?
4. Can optimisations in neural networks improve Land use classification accuracy over base convolutional neural networks for high-resolution unmanned aerial vehicle imagery?
5. Can we understand the use of various data processing capabilities in building and refining image data?

7.5 Limitations

1. Cloud cover and shadows generated by tall ground objects can significantly affect the quality of UAV images, posing challenges to classification algorithms.
2. Random noise generated during image acquisition can also impact the quality of UAV images, affecting classification accuracy.
3. Deep learning models can suffer from overfitting or underfitting due to the complexity of the data and the limited training datasets.

4. Processing large amounts of UAV remote sensing data can be computationally intensive, requiring significant resources and processing time.
5. Accurate data labelling and annotation are crucial for deep learning model training. However, labelling large datasets can be time-consuming and labour-intensive, requiring careful planning and resource allocation.

7.6 Report Structure

The given Research is discussed in the report as follows.

a. **Literature Review:** The next chapter extensively reviews the past work done to build a roadmap for the research.
b. **Research Methodology:** Chapter three introduces the steps taken to solve the given problem based on the roadmap provided in Chapter Two by implementing the CRISP-DM methodology to classify the information as informative or not.
c. **Design Specification:** The design chapter introduces the new approach to solving the given problem, stating the novel idea and proposed solution.
d. **Implementation:** The fifth chapter discusses the methodology for implementing the data and subsequently provides an in-depth analysis of the Research.
e. **Evaluation:** The sixth chapter discusses the results obtained from the given Research and compares different models based on various evaluation metrics, discussing the final results.
f. **Discussion and Conclusion:** The last chapter provides the results obtained, evaluates and discusses the results of the given problem, provides limitations, and further elaborates on all the plans exploring the domains of machine learning and AI in the respective field, mainly focusing more on end-to-end deployment of the model.

2 Literature Review

2.1 Literature Surveys

Initial Research by (Osco et al. 2021) reviewed various deep learning capabilities in remote sensing, primarily focusing on Deep Neural Networks (DNN). The following neural networks are composed of neurons that quickly learn high-level features. These models are capable of progressive learning that happens in hidden layers, and based on that, the weights are calculated post-activation functions implementation. These models are implemented in the UAV remote sensing applications because of feature learning patterns. The Research shares the Web of Science Journal paper, and 190 papers were taken in the given Research to address the literature.

Moreover, some processing and reviews were taken to make the count at 232+ papers. These papers are then further categorized based on the data organization, such as sensor data, LIDAR, algorithms and architectures, and the business problem being solved there. The study highlights the use of deep learning models in the UAV, showcasing the different use cases such as Real-time processing of the images, Dimensionality Reduction, transfer learning, etc. These models explain how deep learning is used in the given UAV task. However, these are black-box models, and one cannot explain their interpretability. These papers are generally based on

object detection with some application in forest and agriculture. Also, the study shows that the most highlighted models are CNN, LSTM, and GAN. However, in the end, one has a systematic study and a standard dataset to compare different models.
As Deep learning models are famous for their black-box nature, the study examines various techniques used in explainable AI and how the following can be implemented in Sensory Image datasets.(Höhl et al., 2024) highlights how machine learning models have performed on the Earth Observation dataset. However, these models are hard to explain, and one cannot capture the relationship of various features and understand anything to draw scientific conclusions. The Research completed in the given field is increasing quickly with the number of papers explaining Artificial Intelligence. The PRISMA technique was initially adopted to capture 60 papers on XAI and find the available methods in each domain. The most used methods in the given approach are based on finding the most essential features. Deconvolutional networks are being constructed to understand how a CNN model learns in backpropagation. The layer-wise relevance propagation method was also introduced to showcase the relevant features of the modeling capabilities. However, the standard explanation methods used in remote sensing are Class Activation Mapping (CAM) and Grad-CAM, which introduce and find the critical region in the image visually used for classification. The next model, Occlusion sensitivity, is based on putting the patch on the image to understand model sensitivity. Then, the most important ones used in the modeling are SHAP and LIME, which incorporate model explainability based on essential features in simplified space.

Once the literature on the deep learning neural net and model explainability is available, we can use these deep learning models to classify Land based on Hyperspectral and Multispectral earth observation data. Research by (Vali et al., 2020) demonstrates how these channels collect data in one wavelength and multiple wavelengths. Multiple datasets are described in the given analysis, and different machine-learning techniques are discussed. The paper shifts towards a machine learning approach to understand the modern techniques involved in satellite imagery and advanced computational methods. The paper details the 3D hyperspectral data and how the following can be managed via deep learning and mostly CNN. The paper showcases the techniques that handle the unique challenges of addressing remote sensing data, data fusion, and feature engineering.

2.2 Deep Learning Capabilities and Transfer-Learning

Worldwide, mining is increasing quickly and is the leading cause of deforestation. The government adopts national manual monitoring activities every five years to capture deforestation and understand the mining capabilities. The following needs to be tracked in real-time using unmanned aerial vehicles to track such activities. Hence, the study by (Giang et al., 2020) further highlights using the deep learning model. Convolutional neural networks can efficiently work on such data, and the one discussed and highlighted in the Research U-Net model. The data used in the given Research is taken from the Phantom 4 RTK drone. The given dataset was cleaned using various data preprocessing steps to correct the distortions, collect the tie points, refine the points, and perform other steps to generate the best quality image. The problem statement here describes the classification of the Land into 6 classes mainly: open-cast mining, old permanent croplands, young permanent croplands, grasslands, bare soils, and water bodies. The U-Net model was compared with the Random Forest and SVM benchmark models. The U-Net model in the given study highlights the highest accuracy of 92.76% when used with Adam Optimizer, which outperforms the machine learning models.
The study shows that the given model can classify all the different land covers and can be used to cover up new technologies like environmental monitoring and other technologies. These models can capture low-level and high-level features to generate such results.

The previous Research highlights the U-Net model for classifying images. The following Research by (Kentsch et al., 2020) analyzed the data collected from Drones to analyze the data from the forest ecosystem and focus on evaluating the status and changes in the forest ecosystem in Japan using the UAV-acquired images and understanding how the transfer learning has significant impact on improving the deep learning model performance. The research datasets are the Seven Winter Mosaics Mountain Forest dataset and the pine tree plantation mixed broad leaf trees dataset. The trees are in Japanese mountain areas, which makes surveying difficult for drones. Also, the image recognition and segmentation tasks are more complicated in such areas. To target and classify such data, one needs much labeling power, and hence, there is a requirement for deep learning transfer learning models. The dataset was processed using various software like Meta Shape, GIMP, etc., to define the classes. The study uses two different architectures, ResNet 50 and U-Net, for classification and segmentation.

Transfer learning models are initialized with random weights and show great accuracy on satellite image datasets. The transfer learning technique increased the accuracy by 9.78%, and by using the more relevant dataset, the accuracy increased by 2.7%. The ResNet model could classify the winter mosaics with better True Positives, and UNet architecture provided better segmentation of the deciduous trees. The study underscores the importance of Transfer Learning in enhancing the performance of DL models, mainly when dealing with limited data. Using UAVs combined with DL techniques offers a powerful tool for forestry applications, providing detailed and accurate analysis of forest ecosystems. This approach can replace traditional, labour-intensive land surveys, offering a more efficient and less hazardous alternative.

The following literature matrix depicts the core studies within the domain.

Author/ Date	**Theoretical/ Conceptual Framework**	**Research Question(s)/ Hypotheses**	**Methodology**	**Analysis & Results**	**Conclusions**
Osco/ 2021	Literature Survey	Study the use of DNN in UAV Image Sensory	Extracted 236 Papers, Proceedings, and Journal From Web of Science Journal	Deep Learning Models Like LSTM and CNN are the most used algorithms.	Deep Learning models can give excellent results with the image dataset.
Höhl, A, Obadic, I/2024	Literature Review of XAI	How XAI is used in Remote Sensing	PRISMA scheme for detailed analysis	Find the essential Explainable Methods	SHAP and LIME are the most essential methods used in recent studies and literature.
Vali et al. 2020	Literature Review on Land Use and Land Cover (LULC)	Understand deep learning applications in LULC and cover the advancement	Categorize papers based on various aspects to find the best one.	Understand 2D and #d data and how it can be handled via Deep Learning	The paper summarises the current state of the art and highlights the gaps and future directions for Research in using deep learning for

					LULC classification.
(Giang et al., 2020)	U-Net in Land Cover Classification	The study focuses on Tan An quarry in Daknong province, a mining area undergoing significant topographical changes due to mining activities.	Use U-Net to classify the Land in 6 classes by preprocessing the data using multiple-stages	Study highlights the superiority of U-Net over ML Algorithms	The given method can be used in real-time analysis, which is crucial for environmental management and conservation efforts in regions affected by mining activities.
Kentsch et al. 2020	Deep Learning with CNN	How can transfer learning improve the accuracy of the deep learning models	Two Deep learning algorithms are built for classification and segmentation	Results show high accuracy with transfer learning models	Integrate transfer learning with forest datasets to provide accurate real-time monitoring of forests.

3 Research Methodology

3.1 Introduction

Cross-Industry Standard Process For Data Mining(CRISP-DM) is a defacto approach that is used for data mining. There are various approaches available, such as SEMMA(Sample, Explore, Modify, Model, Assess) and KDD (Knowledge Discovery In Databases). (SEMMA vs CRISP-DM, n.d.). However, the given methodology is the most widely used in data science projects and offers a step-wise implementation of data-based projects(CRISP-DM | KDnuggets, 2014). It provides a structured, iterative process of transforming a problem into a data-based problem and generates insights. The model consists of 6 phases, the task in each phase and the relationship of the phase with another phase. The given relationship between various tasks exists based on the type of problem and interest of the user. All the different phases are illustrated as follows, and generally, these phases are not rigid and move back and forth based on the type of problem. The following has been used in crop production(Rahmadi et al., 2023) to improve crop yield and understand the various changing factors so the environment can build better models. In health care, the model is redesigned for CRISP-MED-DM to address the challenges in healthcare and combine the data from various sources like EHR (Electronic Health Records), labs data to build better clinical support systems(CRISP Data Mining, n.d.; CRISP Data Mining Methodology Extension for Medical Domain, 2015). Right now, with the advancement in big data analytics, the following is used in the implementation of big data analytics systems. The approach is lightweight and general.(Montoya-Murillo et al., 2024)

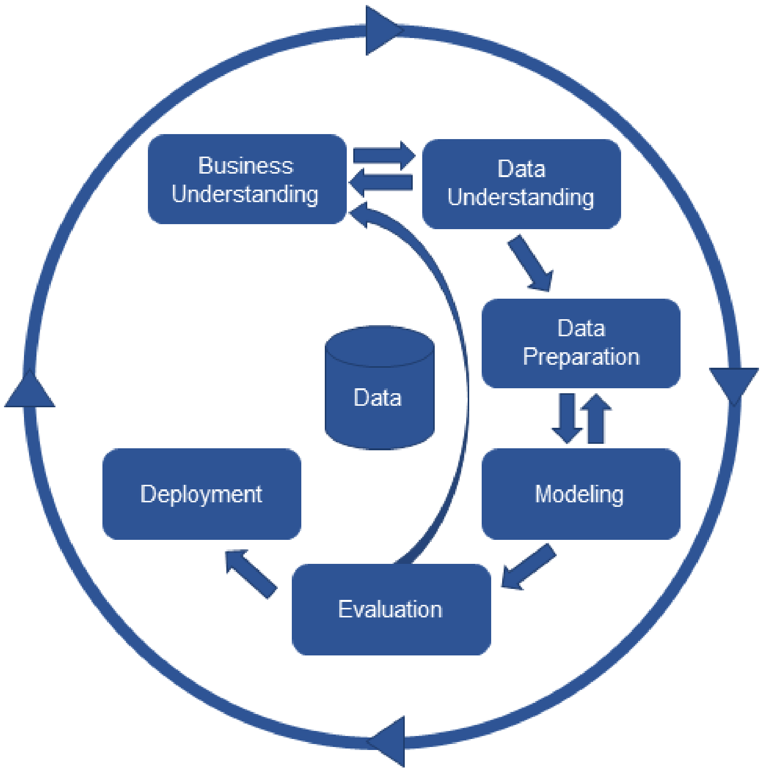


Figure 2 CRISP-DM

The given phases show the most frequent and frequently used dependencies and the outer circle highlights the cyclic nature of the approach. One has to encounter various new learnings that will be implemented in the next phase. A brief introduction of all the phases are outlined briefly:

Business Understanding: The first phase is to understand the objectives of the project and the requirements from a business perspective and convert the given knowledge into a data-based problem along with a design to achieve the objectives.

Data Understanding: Once the business is understood, Data understanding talks about all the initial steps taken to get familiar with the data, identify the problems and get a 360-degree view to form a hypothesis based on the information

Data Preparation: This step covers all the activities to construct a final dataset that would be fed to the model. Data preparation is the most time-consuming and iterative process, where the important features are selected and transformed for modelling.

Modelling: Once the data is ready, modelling is applied to the given dataset, and the given models are tuned as per the problem to attain the best results. Some techniques require a specific type of data, and that is why the step is necessary.

Evaluation: The built model is then tested with respect to the business as well as metrics. The given model is the best build model as per the data analysis phase and hence needs to be examined in terms of business objectives. Hence, at the end of this phase, a decision based on the data mining result needs to be made.

Deployment: Once the model is created, the following needs to be applied in the organization for better decision-making processes. The knowledge gained from the model needs to be presented so that customers can use it, and hence, the deployment phase offers the documentation and understanding of the model. Our study utilizes a CRISP-DM model to classify the different types of Land based on their satellite images and extract features from the land-related spatiotemporal dataset; the following is optimized for a comprehensive implementation. The given model is described as follows:

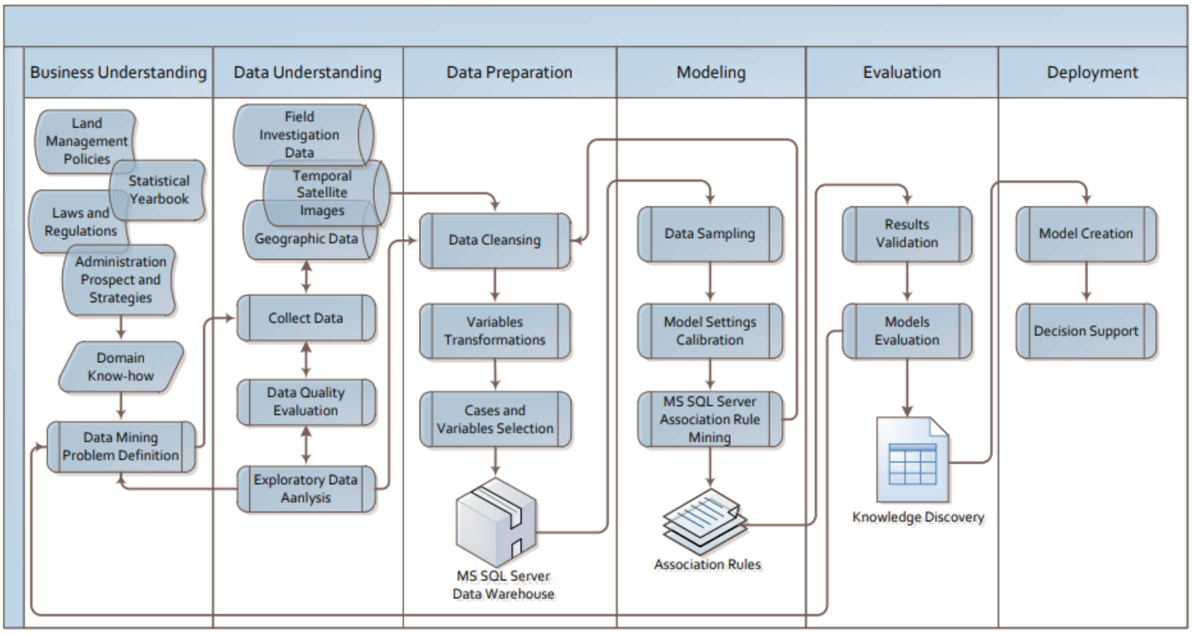


Figure 3 CRISP-DM for Land Classification

(Lin et al., 2011)

3.1.1 Business Understanding

The first step of the Research is business understanding, where all the objectives and requirements for the identified problem are discussed. The main aim of engaging with such Research is to overcome the traditional technology of geographic information systems. These technologies are not designed to find hidden patterns. With today's vast data knowledge, one can perform spatial analysis using state-of-the-art models to classify the Land based on their pattern. Hence, the first step, which shares the details of the problem, is identifying the type of Land with the help of deep learning models, which are known for their higher accuracy and efficiency as the Research can be used for urban planning and monitoring environmental activities. These models can reduce the cost for businesses relying on such models and sensing data (Campos-Taberner et al., 2020a; Jagannathan & Divya, 2021; Zhao et al., 2023)

3.1.2 Data Acquisition and Understanding

The steps of data understanding involve getting a 360-degree view of the data and checking its quality. The Research utilizes the data from [Kaggle under the name RSI-CB256](). The dataset has a total of 5641 images from 4 different classes.(Satellite Image Classification, n.d.). The classes given are Cloudy, Desert, Green Area, and Water. As can be seen, one cannot easily differentiate from these images, and high-level feature extraction is needed to identify the different land segments. One should have proper knowledge of satellite and aerial imagery datasets.

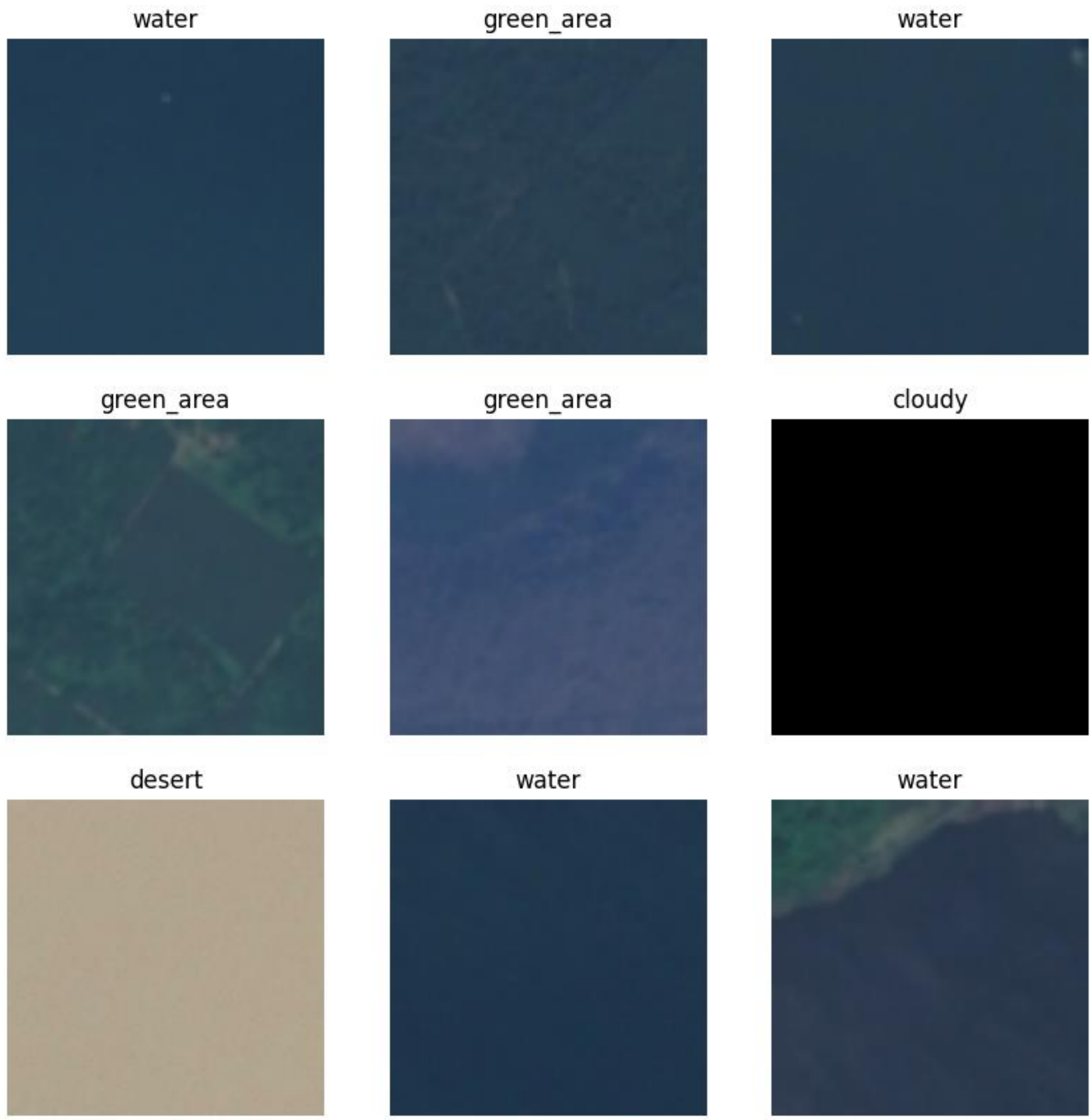


Figure 4 Sample Images

3.1.3 Data Preparation and Image Processing

The third step involves processing input data, where the given image data is resized into 180*180 pixels. The given step is the most crucial in any data mining project. Here, the data is split from the modelling perspective into training and testing in an 80:20 split. 4513 records were used for training and 1128 records for testing. The given dataset is less for a Deep learning model as they rely on large datasets. Hence, the study will incorporate data augmentation that will help make copies of the data from the data itself by some random transformations and help reduce the overfitting of models.

3.1.4 Modeling

The following process uses the preprocessed data for model building, where the optimal techniques are chosen and aligned with the data. Based on the given requirements, generally, one has to return to the data preparation steps. Four different architectures are trained on the given dataset. The most basic one used is CNN, a model known for best results on image datasets, followed by AlexNet, VGG-16, and ResNet-50. CNN are the models that show the best accuracy on image datasets and are designed for such data. AlexNet is the base CNN model with 5 layers that won the imageNet Competition in 2012 and shows the simple implementation of the CNN Architecture that is benchmarked. However, the following model has some drawbacks, and hence, the VGG16 model is used for processing large datasets with small convolutions and for better performance. The following model can capture land data with complex features. The RESNET-50 is used for residual learning and to showcase the concept of transfer learning, as all the other models stated, to showcase the problem of vanishing gradient. The capacity of convolutional neural networks (CNNs) such as ResNet-50, Alex Net, and VGG-16 to efficiently process and evaluate complicated visual data makes them ideal for land classification. They are well-suited for a variety of environmental monitoring and remote sensing applications due to their distinct characteristics, which include the speed and simplicity of Alex Net, the depth of VGG-16, and the advanced learning capabilities of ResNet-50. Urban planners, farmers, and environmentalists can all benefit from the automated land cover classifications made possible by these models.(CNN Architectures, n.d.)

3.1.5 Evaluation

The next step is the evaluation of these models, where the effectiveness of each model is checked to obtain a benchmark model where the training and validation accuracies are calculated for these models. The step also helps determine whether the models are able to achieve the objectives.

3.1.6 Deployment

Generally, the last step involves the deployment of the model over the cloud. Our study focuses on finding the best model that can be used for decision support and to classify the images with maximum accuracy.

3.2 Design Specification and Architectures

3.2.1 Two-Tier Architecture

A two-tier architecture was developed for the Research to classify the Land using the deep learning models. The top Layer is the user layer that receives the visualization and results of these classification models. The business logic layer is the second tier, where the classification models are trained using the collected data. Google Colab/ Jupyter Notebook is used to train the model. These tools are used to enhance and produce data information for training the models. Then, the models are tested and evaluated using deep learning libraries like TensorFlow and Keras to get the results.

3.2.2 Deep Learning Architectures

There are a lot of other deep learning architectures available that can be used for the same problem, are stated as follows:

RNN, a special version of LSSTM (Long Short-Term Memory), are the most effective model for analyzing the time series data and is suitable for analyzing the land cover over some time. The following is used to detect the land cover using a modified architecture BiLSTM.(Campos-Taberner et al., 2020b)

The model is known for its effective processing power for sequential data and can remember the context based on the previous inputs. The models are capable of handling sequences of variable lengths and are known for making dependencies from past data. However, the model has the problem of vanishing gradient descent and exploding gradient descent. The model is not built to carry information across multiple time stamps and may face issues with limited memory capacity. Hence, the model is complex and time-consuming in itself. Some variations of RNN can solve the given problem.

GANs are generative models known for generating synthetic data, and the following power can be used to improve classification. The models can enhance the training power by generating the Fake dataset and can help work on the problems with small datasets. The model works really well with producing photorealistic results and can be used for unsupervised machine learning. The model is versatile and can solve many problems, and it is mostly used for text generation.

However, training a GAN is highly unstable and may collapse at convergence. The model is computationally expensive and slow to train on large datasets. The model also faces the issue of overfitting, which can generate bias in training datasets.

Hybrid Models To maximize their effectiveness, hybrid models integrate multiple deep learning architectures. To improve classification accuracy, for instance, VGG19 networks can be mixed with other methods, such as transfer learning. These models can utilize pre-trained weights from established architectures, reducing the training time and improving performance. Hybrid models frequently outperform single models in terms of predicted accuracy because they combine many algorithms and take advantage of their strengths. Combining several methodologies helps strengthen the model, making it more able to withstand data noise and outliers, which in turn improves its generalization skills. By integrating deep learning and classical methods, they can automatically learn features from data, leading to a better representation of complicated patterns. Hybrid models are adaptable for a wide range of uses, including image recognition and natural language processing, because they can handle both structured and unstructured input efficiently. However, integration of different models can result in more complicated architectures, which in turn make them more difficult to design, develop, and maintain. This model training typically takes more time and computer resources, particularly when working with big datasets or complicated methods, leading to longer training times overall. The possibility of overfitting arises when combining numerous models; this is especially true when dealing with complicated individual models or when training data is inadequate. Complexity makes hybrid models less interpretable, which makes it hard to comprehend decision-making processes, which is important in many contexts.

3.3 Conclusion

The CRISP-DM model is tailored to fit all the requirements of a given business objective. The standard methodology was tweaked for deep learning models and image datasets. The study provides a comprehensive roadmap for developing a land classification model using deep learning, ensuring all critical aspects are covered, from data collection to model building. Next, the research will highlight the implementation of the given knowledge and showcase the novelty of the work.

4 Implementation of Neural Nets

4.1 Introduction

The chapter highlights how the CRISP-DM methodology would be followed in analyzing the Land Classification dataset and which model would be best for classifying the Land Type based on the image, showcasing the accuracy of the models and the best architecture out of all the models. The study adopted TensorFlow as the base library, which the Google Brain Team built. The following is the general-purpose framework to implement machine learning and deep learning architectures, which can be optimised for use on CPU and GPU. On the other hand, Keras is a high-level API used to build Neural Networks and is implemented with TensorFlow. The following is easy to use and is tightly integrated with TensorFlow, which can build and stack layers of Neural Networks and also help in the customization of Neural Nets.

4.2 Data Extraction and Augmentation

4.2.1 Data Extraction

The dataset taken from the Kaggle Repository has 4 different classes, and the following, with the help of the Keras library, is loaded in the directory. The main use of such API is to define the size of the image, split the data early, and define the number of images in each batch, as deep learning models take input in the form of batches.

4.2.2 Data Augmentation

The size of data in our Research is too small to train the deep learning models. These models are data-hungry models and require intensive training. Hence the models are fed new augmented data. Data augmentation is a technique where the size of the training dataset is increased by applying some modification to the training dataset. The following techniques make the model robust towards input data. The study has incorporated random zoom, rotation and flip of the images to make new samples that can be used to train the models.

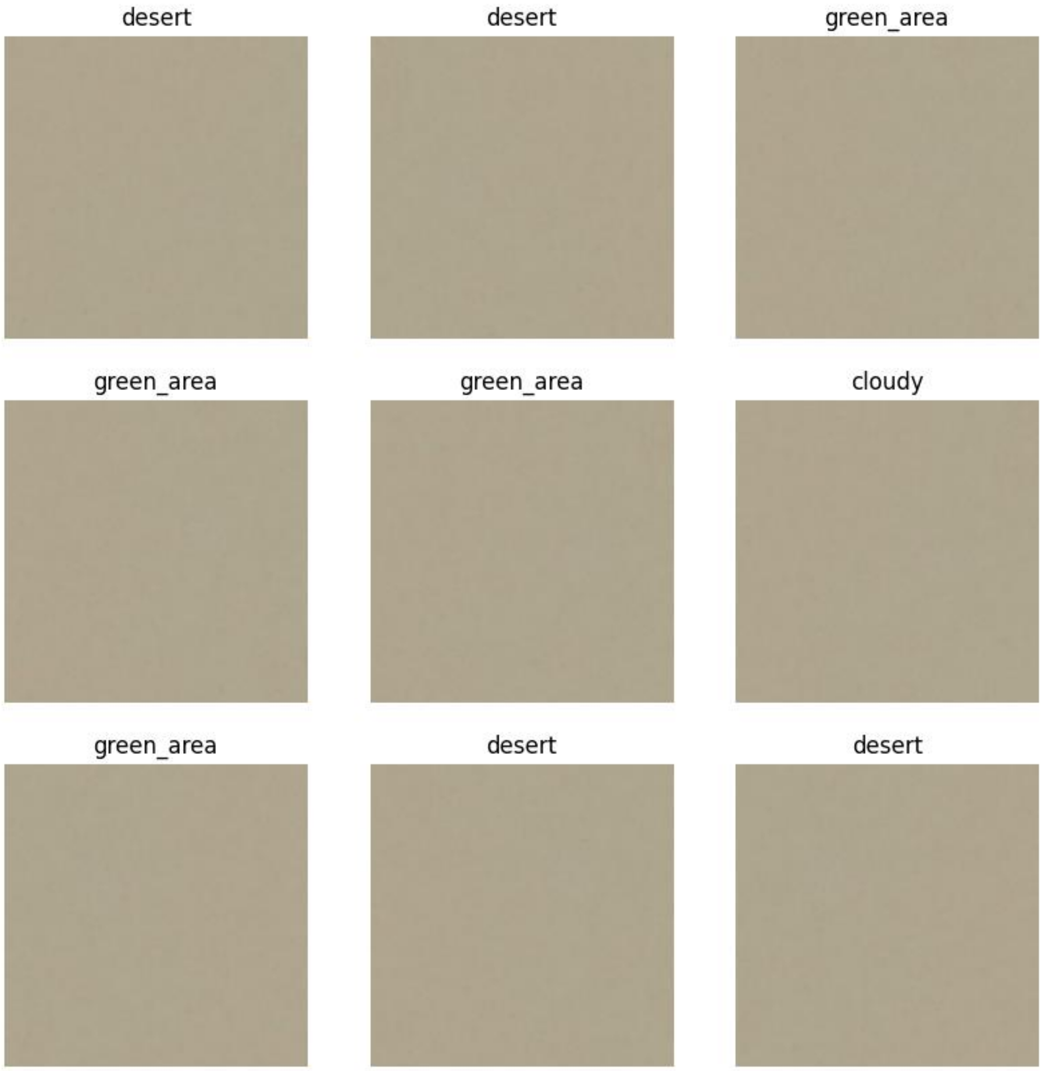


Figure 5 Data Augmentation

4.3 Model Training

Once the data is ready, the proposed methodology aims to enhance the accuracy of image classification models using various CNN architectures. Based on the performance, the models selected are:

4.3.1 CNN

CNN, known as Convolutional Neural Network, was the first model used in the Research. The model is famous for image and video recognition. The following network has an input layer, multiple hidden layers and an output layer. These hidden layers in the model are convolutional layers from where the name is derived and where all the features from an image are extracted. The term convolution refers to mathematical operations where the arrays are multiplied to generate new arrays. These arrays in convolutional neural networks are grayscale images (NumPy Array) and Kernels, also known as filters (NumPy Array), which are responsible for extracting features from an image. The most common layers used in our network are convolutional, activation ReLU, Pooling, Dropout and flattened layers.

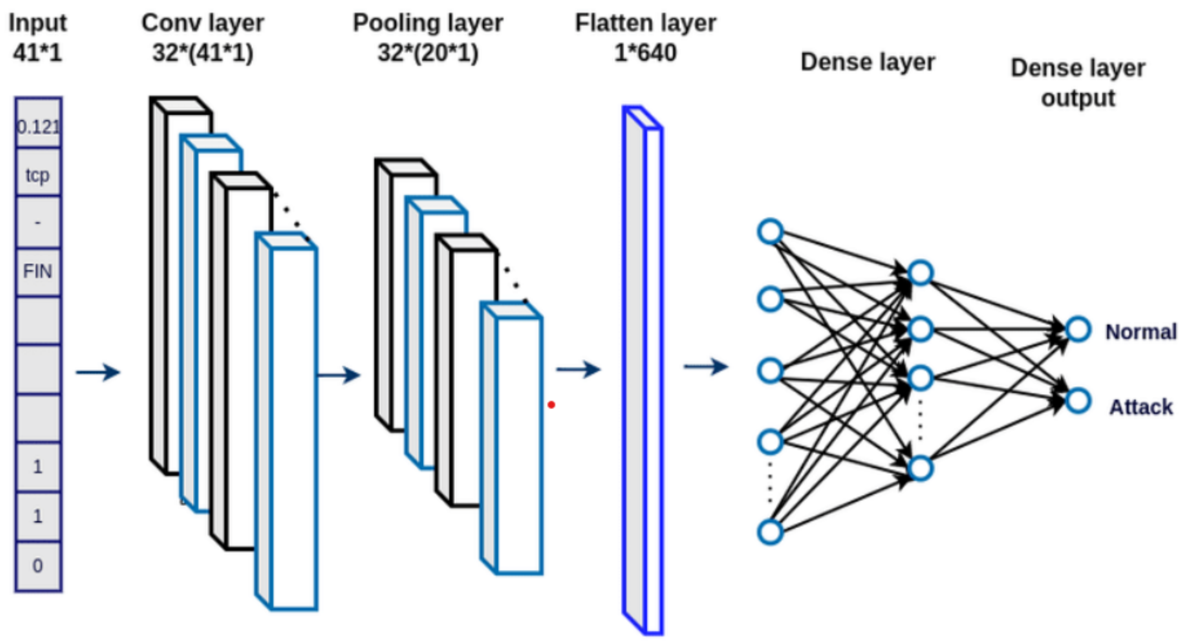


Figure 6 CNN

Convolutional Layer is the major layer that is used in CNN and is the fundamental component used for computer vision. It is one Layer responsible for extracting features from the data by applying a mathematical operation known as Convolution. The convolution layer uses a kernel, which is a matrix of weights that slides over the window and extracts the features in the form of edges, shapes and images. The model extracts the low-level features at the start, which are further combined to get the high-level features. The output of Convolution is also known as a feature map and showcases the presence of specific features.

ReLU Activation Function is the most used activation function in CNN responsible for adding non-linearity in the data. The function is simple and effective, and the non-linearity of the function helps find complex patterns in the data.

$$f(x)=\max\,(0, x)$$

The given equation denotes the function, and it showcases x greater than the output is x. Else, it is 0. The function is simple and is less costly computationally. The constant positive input reduces the problem of vanishing gradient in the next layers.

Pooling Layers are the essential Layers of CNN responsible for reducing the dimensions of the feature maps and controlling the number of parameters to be trained and computational complexity. There are different types of layers, like Max Pooling, Average Pooling, and Global Pooling, that give the maximum value average value from the feature map, reducing the size of the feature map.

Dropout Layer is a type of Layer responsible for dropping a fraction of neurons from the network. The following is a regularization technique that helps reduce the overfitting in the model and helps learn the robust features.

Flatten Layer is responsible for converting the output from the convolutional Layer and max-pooling layer into a one-dimensional vector that can be used as an input for a fully connected Layer, aka neural network.

Our study used a neural network where the data was first preprocessed by dividing the NumPy Array (Image) by 255 to rescale the image. Then, the data is fed into the Convolution Layer, which has 16 neurons, where padding is the same, which refers to adding a new pixel to maintain the image size. The Max-Pooling Layer follows this, and the same combination is applied multiple times, as stated in the given figure. Once the model is ready, the model is compiled using Adam Optimizer, which uses the concept of gradient descent to update the weights of the model based on the loss function, which is Sparse Categorical Cross Entropy, which is used to calculate the loss in multi-class classification model. the whole model is trained on 10 epochs so that model can learn the data extensively and see reduction in loss. The loss and accuracy while training the model are stated as follows:

Figure 7 Training Loss and Accuracy of CNN

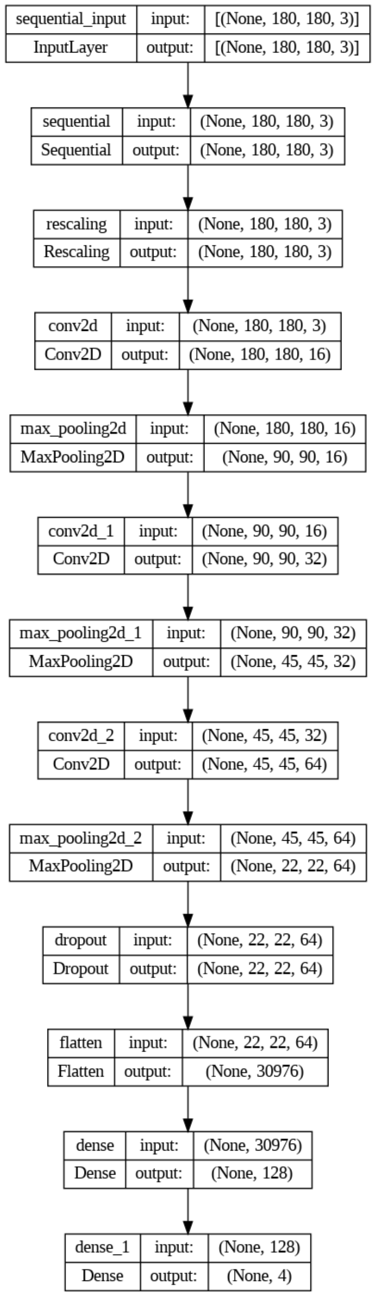


Figure 8 CNN Architecture

4.3.2 Alex-Net

Alex Net is again a CNN architecture that has some special architecture, and the following is famous for winning the ImageNet Large Scale Recognition Challenge. Developed by Alex Krizhevsky(Ciregan et al., 2012)and replacing the second-place entry model by 10%. The given model has an architecture that has eight layers where 3 fully connected layers follow 5 convolutional layers. The model is different from other architectures as it uses the Max-Pooling layer that improves the accuracy, and again, the model uses the same ReLU activation function. The model is trained on the same dataset, and the training accuracy and loss graphs are as

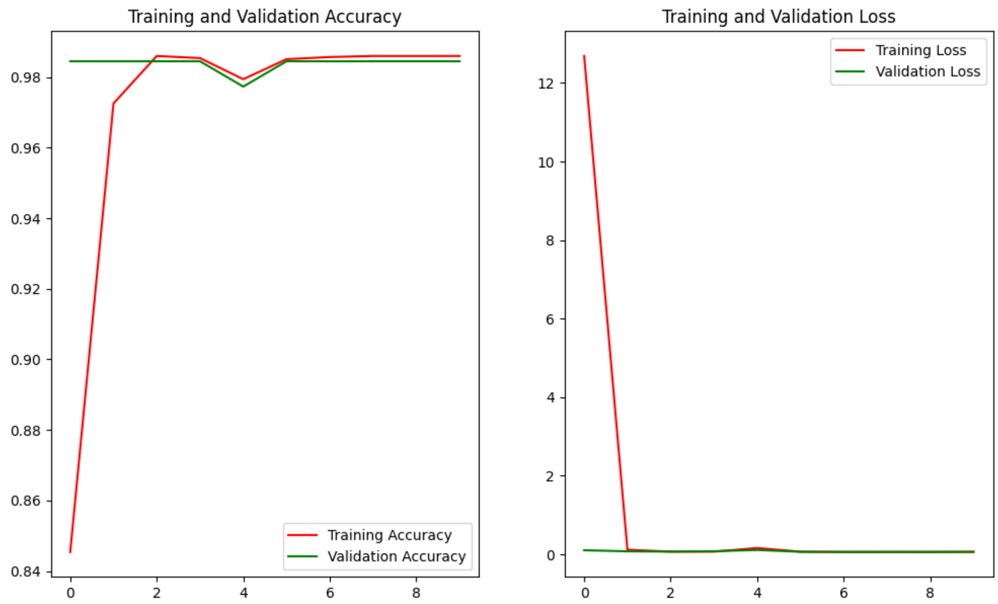


Figure 9 Training Accuracy and Validation Loss of Alex Net

From the very first epoch onward, the training accuracy steadily rises, eventually stabilizing at roughly 98%.In the same way, the validation accuracy settles at about 98%. By the end of the first period, the training loss has plummeted from its initial high point. The pattern is also present in the validation loss, which is somewhat greater than the training loss. Shows that the model is doing well on both the validation and training datasets. The Architecture of the given model is as follows:

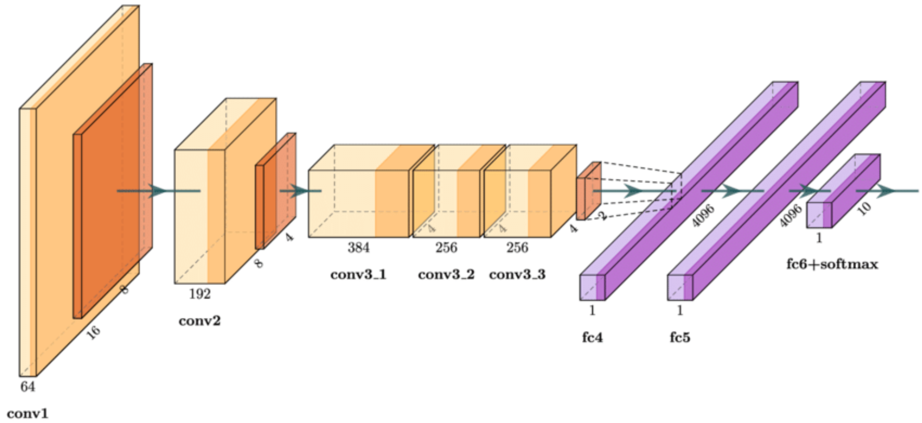


Figure 10 Alex Net

The modified version used in the Research is as follows: the input is of shape 180,180,3.

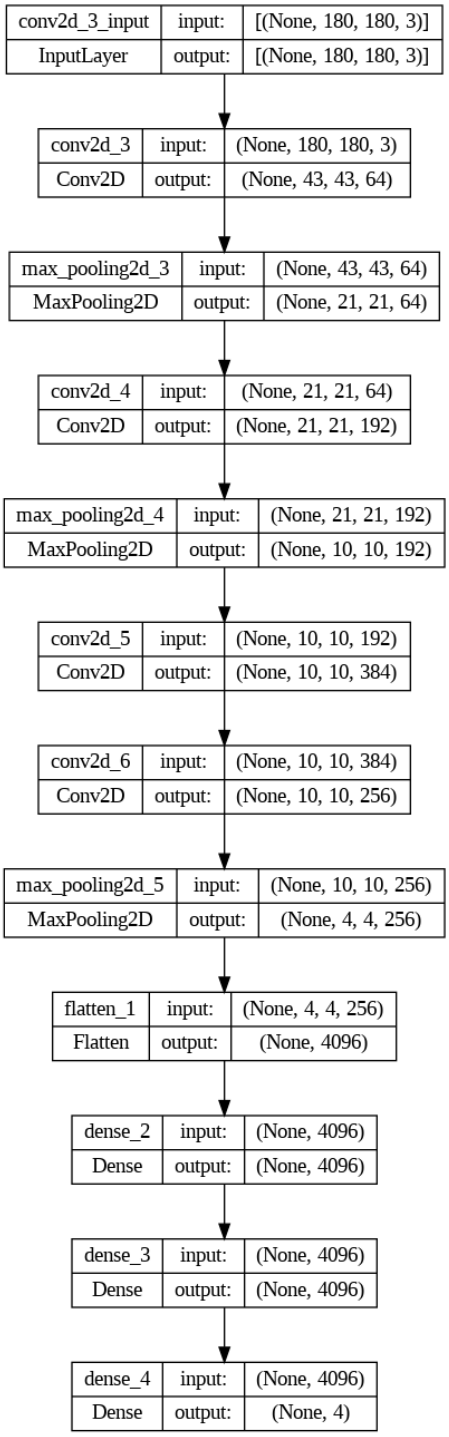


Figure 11 Alex Net Architecture

4.3.3 VGG-16

Oxford University's Visual Graphics Group (VGG) published a 2014 study called "Very Deep Convolutional Networks for Large-Scale Image Recognition" that introduced the widely used VGG16 CNN design. The model is known as VGG-16, as it has 16 layers. There are a total of 13 convolutional layers and 3 fully connected layers. The convolutional layers in the given model are of size 3*3 to capture the maximum information from the image. VGG16's most distinctive feature is its depth. The reason behind its exceptional performance is its deeper architecture compared to earlier networks. The Rectified Linear Unit (ReLU) activation function follows each fully linked and convolutional Layer. ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 results showed that VGG16 performed exceptionally well. A top-five test accuracy rate of 92.7% was attained. The study used the VGG-16 with a transfer earning concept, where the weights of the model were not updated as they were trained on the ImageNet dataset. This concept increases the training speed and reduces memory consumption.

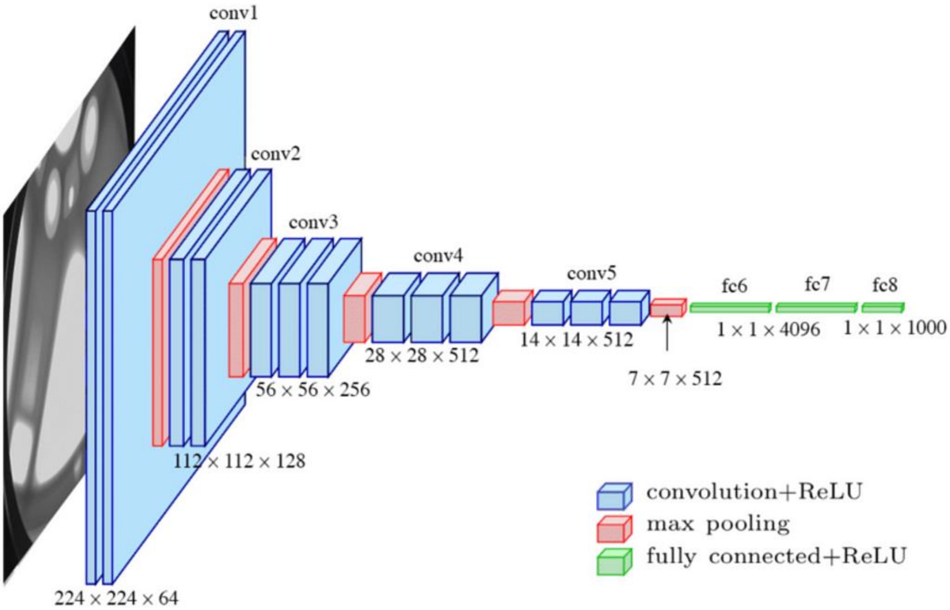


Figure 12 VGG- 16

The purpose of using VGG-16 as a transfer learning model is to get the best accuracy as the lowest layers of a VGG16 pre-trained model on a big dataset, such as ImageNet, capture general features, such as edges and textures, which allows for transfer learning. Countless activities can benefit from these features. The output of one Layer in VGG16 can be used as features in another model through feature extraction. Only the fully connected layers are trained to get information about the new dataset, making it the best choice for faster speed and high accuracy.

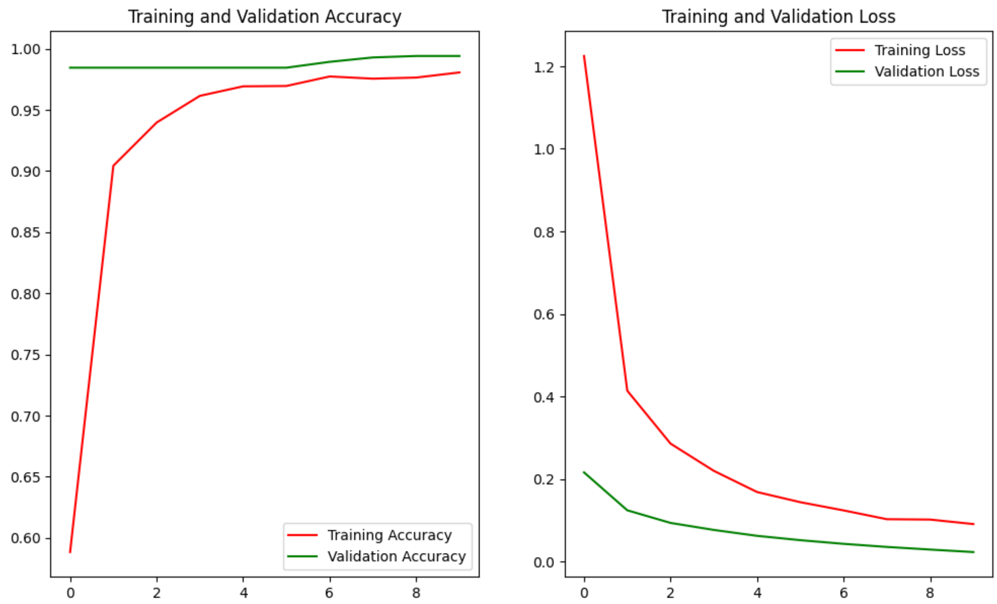


Figure 13 Training Accuracy and Validation Loss of VGG-16

The charts highlight high accuracy for training as well as validation datasets as an added benefit. The model architecture of the Research is as follows:

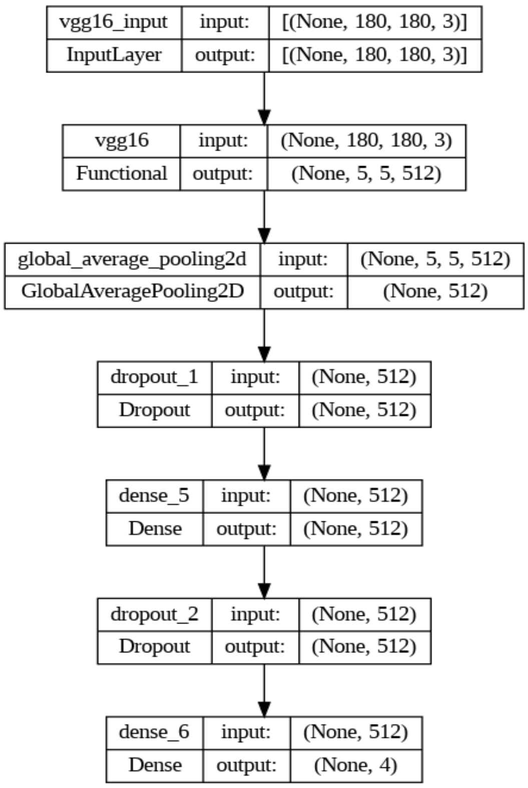


Figure 14 VGG-16 Architecture

4.4.4 ResNet-50

With the improvement in neural network architecture, ResNet-50, also known as Residual Network 50, having 50 layers, was introduced by (He et al., 2015). The given model, again, uses deep architecture and the residual block, which includes shortcut connections or skip connections that bypass one or more layers. These residual blocks help in reducing the vanishing gradient problem and train the model with deeper architecture. The model contains a total of 48 convolutional layers: 1 Max-Pool and 1 Average Pool Layer. The model is directly loaded from Keras with pre-trained weights from the ImageNet dataset, and the model is customized to the problem. With a very low learning rate, the model is trained on the dataset, and the accuracy and validation loss are recorded.

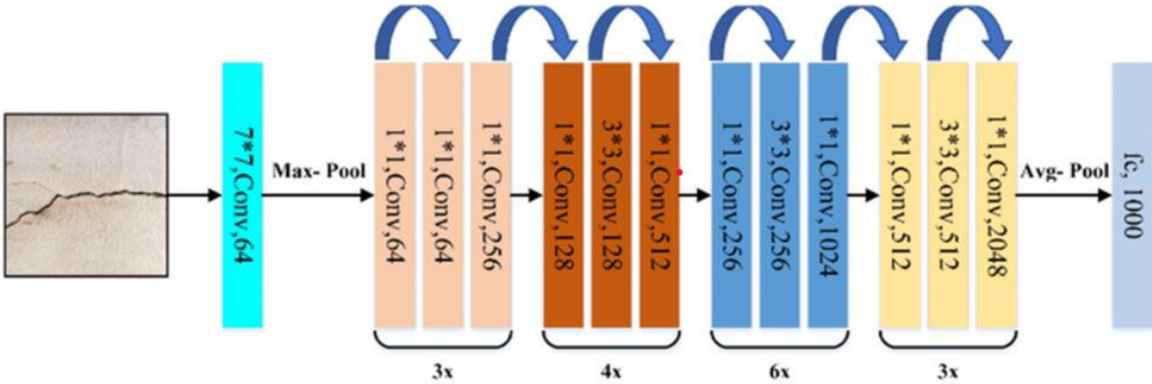


Figure 15 ResNet 50 Architecture

One can observe higher training accuracy and a sudden drop in the loss while training the model because of robust deep architecture.

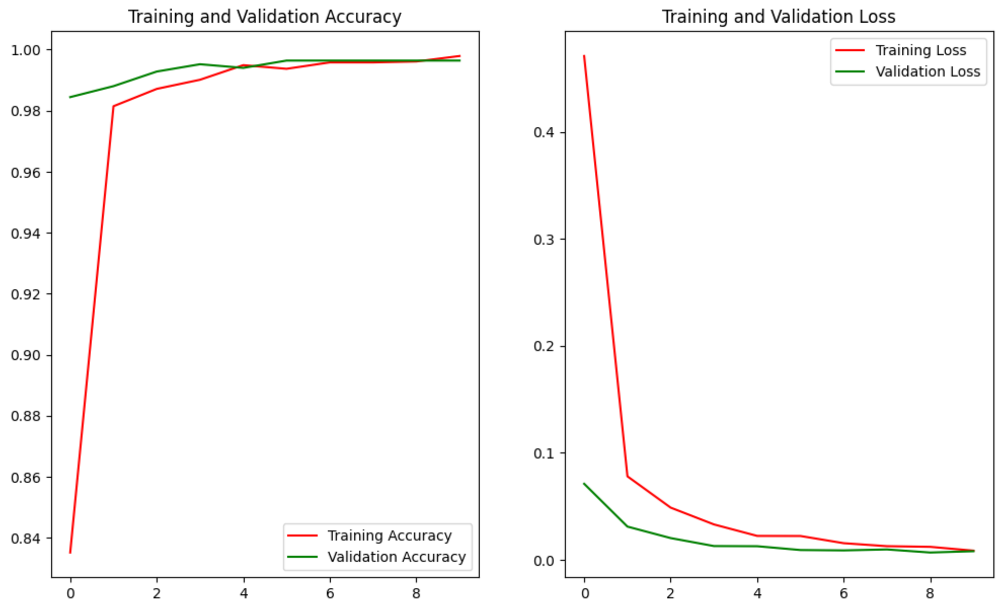


Figure 16 Training Accuracy and Validation Loss of ResNet-50

The original architecture available in the Research is stated as follows:

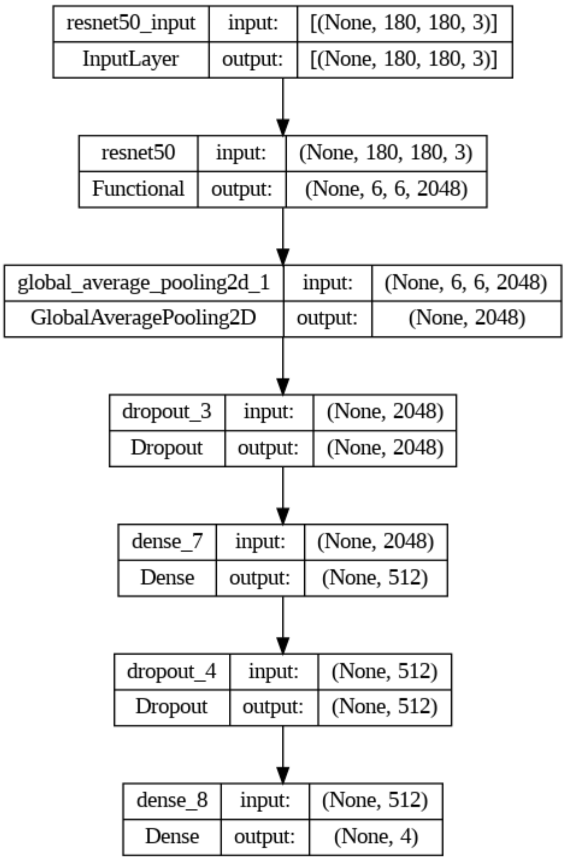


Figure 17 ResNet 50 Architecture

These four models used in the research are modified CNN architectures that researchers develop to solve problems on image datasets. The study has used these deep networks as an advantage of transfer learning, where the weights are already frozen. Now, the study will address which models are perfect for our research based on various evaluation metrics, as discussed in the next chapter.

5 Results

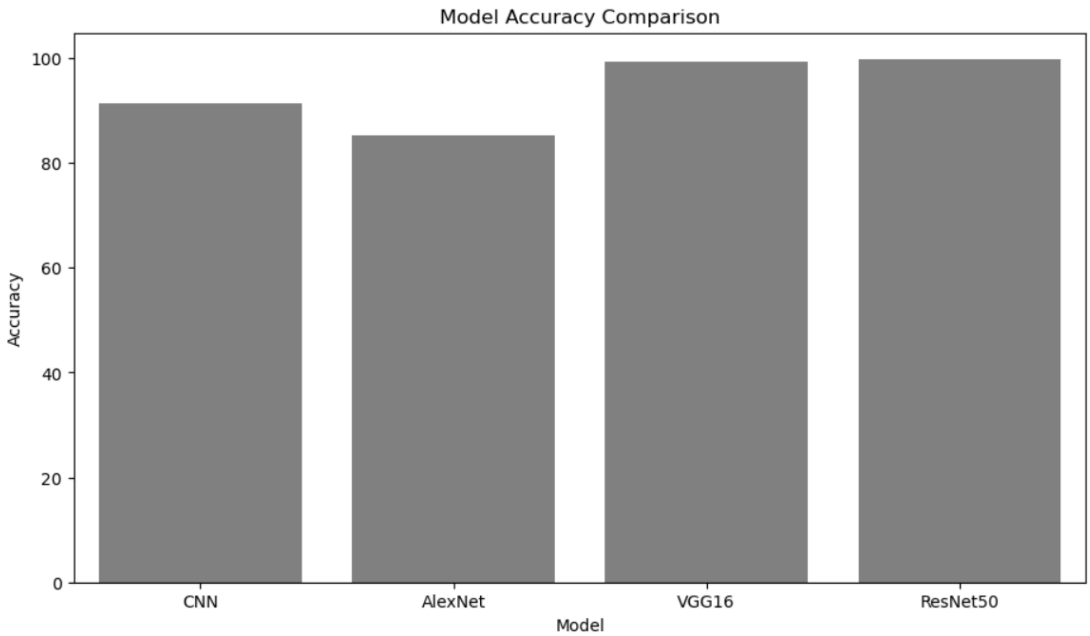
5.1 Results

All these models are evaluated based on accuracy scores. The accuracy score generally depicts how good the model is in predicting the actual label. The following is calculated by the number of correct predictions by the total number of predictions. The accuracy scores of all the models are as follows:

Table 1 Accuracy

Model	Accuracy
CNN	91.11
Alex-Net	89.87
VGG-16	99.48
ResNet-50	99.84

One can observe that the base CNN model and Alex-Net have the same accuracy. However, the best model is ResNet-50, which is based on the deeper architecture, making it a better choice for such activities with the usage of transfer learning. By increasing the depth of the model, the accuracy is increased. At the time of training, all the models had high training and validation accuracy, making it a generalized model on the given dataset. The loss also decreases gradually, showcasing the same characteristics.



To get the output understood in a better manner, our study has showcased the confusion matrix for all the models to get clarity on the class distributions. Here, the labels are cloudy, desert, green, and water, and one can see how the classification, as well as misclassifications, are happening on the dataset.

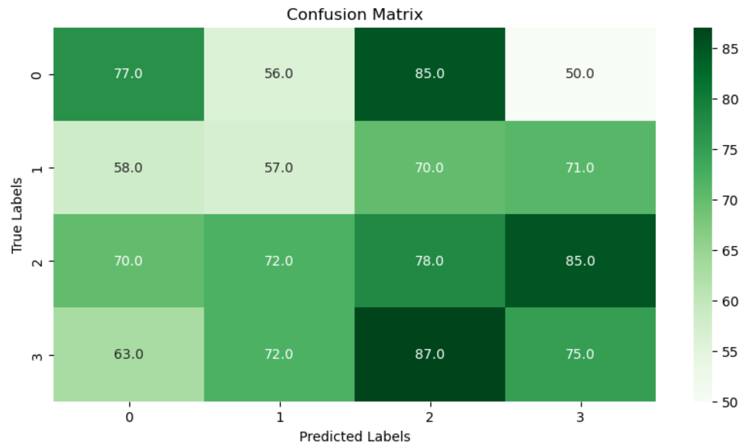


Figure 18 CNN Confusion Matrix

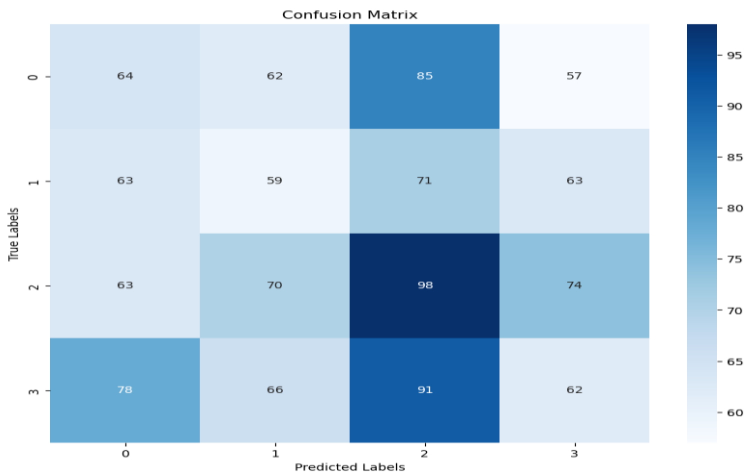


Figure 19 AlexNet Confusion Matrix

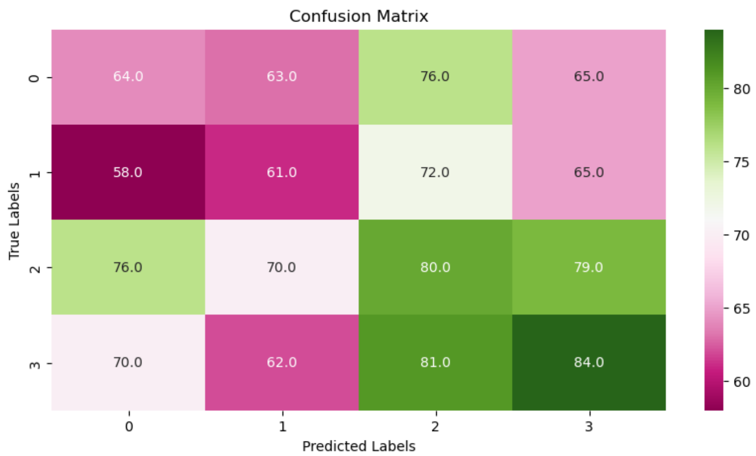


Figure 20 VGG 16 Confusion Matrix

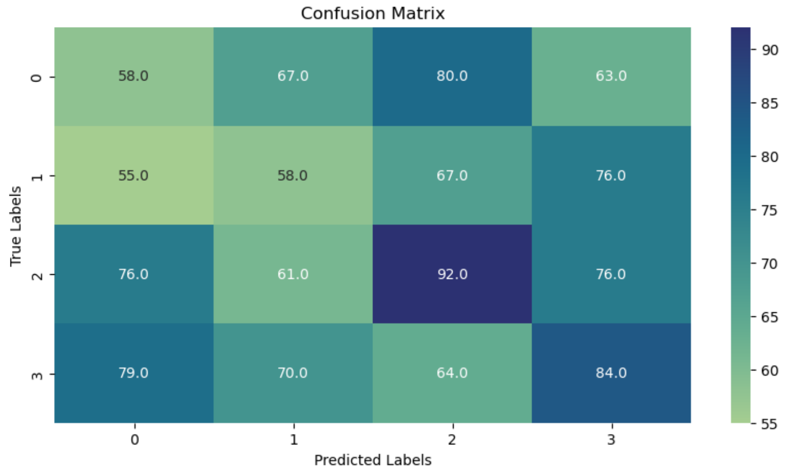


Figure 21 ResNet-50 Confusion Matrix

5.2 Critical Evaluations

Once the best results are obtained, the next step is to discover the best model. ResNet-50 is the best model used in the Research context, as it addresses all of the research objectives.

- ResNet50, with complex architecture, can extract the features from the satellite images with high accuracy as the dataset is a high-resolution dataset, and the model was able to find the features on its own.
- Simple CNN architecture is easy to implement and less expensive computationally. However, these deeper architectures, with the use of transfer learning, were able to capture most of the information and deliver maximum accuracy on complex tasks.
- Transfer learning reduced the time for model training in the Research, making it attain better accuracy on small datasets than larger ones like ImageNet.

As the models attain higher accuracy, our research provides answers to each of the research questions to better validate the problem.

RQ1: Can we evaluate the data quality and its effects on the deep learning architectures? Generally, data quality is the critical factor in any data-based problem and working with data that a UAV collects is highly critical and complex data. The data used in the existing research is of high quality and does not require much validation. The following is taken from Kaggle and is a benchmark database as confirmed by the study. (RSI-CB: A Large-Scale Remote Sensing Image Classification Benchmark, 2017). The data quality generally impacts the model accuracy and degrades the performance of the model, which is not seen in the research. Also, the models have attained higher training accuracy, which showcases the best quality data.

RQ2: Can we evaluate the performance of various architectures based on the batch size of data? In all the models, generally the batch size taken is 64; the following batch size helps in the proper utilization of GPU in the model building as the model is trained on Colab with constrained GPU power. Hence, for faster training, it is used, and it is seen in the literature that batch size 64 is the benchmark size and requires fewer updates per iteration in the model building. Furthermore, the convergence of models is faster with the given batch size; however, to attain the best results in one of the models, the learning rate is adjusted and reduced to learn better from the data. Hence, overall performance is the same for all the models except the Alex Net, and new batch sizes can be tried out there.

RQ3: Can we improve land use classification accuracy by more effectively capturing fine-grained spatial patterns alongside temporal changes with CNNs? For the land classification purpose, the models are improved following the pyramid structure, and only those models that have more pooling layers are chosen to capture most of the data. Also, the study has focused on using advanced architectures and techniques like data augmentation for diversity in training. The data is recorded over a period that satisfies the needs of temporal changes these models can capture.

RQ4: Can optimisations in neural networks improve Land use classification accuracy over base convolutional neural networks for high-resolution unmanned aerial vehicle imagery? Optimizations are part of improving the accuracy of deep learning models in distinguishing land types. All the models used in the research followed a pyramid structure. Also, the study has incorporated transfer learning as these models are trained on the ImageNet dataset and are fine-tuned for the land task. Our models have also incorporated the dropout layer, batch size, and learning rate to increase the accuracy of models and maintain stability in the learning process.

RQ5: Can we understand the use of various data processing capabilities in building and refining image data? Our study has incorporated various steps, like resizing the images to a standard size, as this would be helpful in batch processing of the images. Also, post-data augmentation is applied to increase the diversity in the data, considering the low data size for deep learning models. All these points validate our research questions and based on the accuracy of the models, make our research acceptable on the problem.

6 Discussion and Future Work

6.1 Discussion

Land classification is a crucial task in remote sensing that helps the government manage the Land. Deep learning, a growing field of AI, has changed land use and land classification tasks as these models can understand complex spatial and temporal data and find features automatically. Various deep learning architectures are used in the given field, and the following has changed the traditional approach of detecting patterns using pixels. The study has highlighted the use of deep learning architecture for classifying the Land based on image datasets, and the architectures have showcased high accuracies. With deep architecture, the accuracy of these models increases, and these large datasets with labels tend to increase the accuracy of the model. The study also highlights the use of transfer learning for better accuracy and gaining insights and combining this with GIS data will help in better classification in the future.

To get better clarity and to address the given problem, the study stated some research questions to see the validation of deep learning models on the given dataset. Our research question showcases the effectiveness of deep learning models as well as transfer learning models, and it is relevant to the research in identifying the problem. Also, the study focuses on addressing the core objectives of the research by focusing on data augmentation, data processing, and its effects on model building. It also highlights the use of various architectural styles in model buildings. As the image data is complex and cannot be understood by the basic machine learning models, the study approached deep learning architecture, especially the CNN models. It showed the ability of deep learning models to classify the land type that the government can use for better decision-making. The main contribution of our research in the given field is the incorporation of CRISP-DM methodology. The following process is an iterative process that puts the structure into a data mining process, and the following method is the most robust. Also, our study used the UAV dataset, which is a high-resolution image dataset that is better than satellite data. Also, the study incorporated a benchmark comparison of different models, which has valuable information in the given field. The comparison of base models with the transfer learning models also provides insights and advancement in deep architectures.

Also, the study aims to clarify all the research questions, and for deep learning models, hypotheses are not stated. Our study has compared benchmark models that state the proper structure of the research and its completeness. Looking at all the phases of the CRISP-DM model, our study has implemented a step-by-step approach and incorporated advanced techniques at the end to select the best model out of all, making it appropriate for research, and the best model can be deployed in the future so make it more of an implemented case study. The results obtained from the research are based on the data taken from UAV imagery, and all the data get preprocessed to remove any scope of bias available. Also, data preprocessing steps helped in that.

Designing a deep learning architecture is quite challenging. Moreover, to obtain the results the study also required suitable controls for experimentation purposes. First, the baseline model was defined to compare all the other models, and the data standardisation used in all the techniques was also validated to make the performance of the model easy. To understand the impact of fine-tuning and hyperparameters, all of the models are compared with transfer learning models to understand the impact and isolate its effects. In addition to mentioning how transfer learning and fine-tuning improved model performance, the final model architecture showed a test-set accuracy of 99.19%. The goals and questions posed by the research should be directly addressed in the study's conclusions. The research paper further addresses its claims that fine-tuning and transfer learning increased model accuracy and performance measures with ResNet-50 in outperforming other models. Also, the data was available openly, and there are no ethical concerns as it does not involve any human object or relevant information.

6.2 Potential Future Work

Potential future work will focus on building models that are more interpretable and explainable. All the deep learning models are black boxes, and understanding the updating of weights and biases is somewhat challenging. Furthermore, to achieve better efficiency, lightweight models need to be built that can potentially attain great accuracy and do the processing in real time. This would arguably be helpful at times of disaster, such techniques like semi-supervised learning and data augmentation would help improve the models. As such, such work can address a better understanding of the data, whilst using edge computing and processing of the data at the source location, would inherently improve the efficiency of such models.

References

Campos-Taberner, M., García-Haro, F. J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., & Gilabert, M. A. (2020a). Understanding deep learning in land use classification based on Sentinel-2 time series. *Scientific Reports 2020 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-74215-5

Campos-Taberner, M., García-Haro, F. J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., & Gilabert, M. A. (2020b). Understanding deep learning in land use classification based on Sentinel-2 time series. *Scientific Reports 2020 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-74215-5

Ciregan, D., Meier, U., & Schmidhuber, J. (2012). 2012 IEEE Conference on Computer Vision and Pattern Recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/CVPR.2012.6248110

CNN Architectures. (n.d.). Retrieved July 23, 2024, from https://www.researchgate.net/figure/CNN-architectures-of-a-AlexNet-b-VGG-16-c-VGG-19-d-ResNet-50-e-ResNet-101_fig2_345313795

CRISP Data Mining. (n.d.). Retrieved July 22, 2024, from https://www.proquest.com/docview/1722161251?sourcetype=Scholarly%20Journals

CRISP Data Mining Methodology Extension for Medical Domain. (2015). https://www.researchgate.net/publication/277775478_CRISP_Data_Mining_Methodology_Extension_for_Medical_Domain

CRISP-DM | KDnuggets. (2014). https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

Giang, T. L., Dang, K. B., Le, Q. T., Nguyen, V. G., Tong, S. S., & Pham, V. M. (2020). U-net convolutional networks for mining land cover classification based on high-resolution UAV imagery. *IEEE Access*, *8*, 186257–186273. https://doi.org/10.1109/ACCESS.2020.3030112

Hannah Ritchie and Max Roser. (2019). *"Half of the world's habitable land is used for agriculture."*

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Höhl, A., Obadic, I., Torres, M. Á. F., Najjar, H., Oliveira, D., Akata, Z., Dengel, A., & Zhu, X. X. (2024). *Opening the Black-Box: A Systematic Review on Explainable AI in Remote Sensing*. http://arxiv.org/abs/2402.13791

Jagannathan, J., & Divya, C. (2021). Deep learning for the prediction and classification of land use and land cover changes using deep convolutional neural network. *Ecological Informatics*, *65*, 101412. https://doi.org/10.1016/j.ecoinf.2021.101412

Kentsch, S., Caceres, M. L. L., Serrano, D., Roure, F., & Diez, Y. (2020). Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study. *Remote Sensing*, *12*(8). https://doi.org/10.3390/RS12081287

Lin, Y.-W., Wu, M.-C., Kuo, Y.-C., & Chen, C.-F. (2011). *INTEGRATION OF GIS AND DATA MINING TECHNOLOGY FOR LAND MANAGEMENT*.

Montoya-Murillo, D., Mora, M., Galvan-Cruz, S., & Muñoz-Zavala, A. (2024). *A Selective Conceptual Review of CRISP-DM and DDSL Development Methodologies for Big Data Analytics Systems*. 123–160. https://doi.org/10.1007/978-3-031-40956-1_5

Osco, L. P., Marcato Junior, J., Marques Ramos, A. P., de Castro Jorge, L. A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., & Li, J. (2021). A review on deep learning in UAV remote sensing. In *International Journal of Applied*

Earth Observation and Geoinformation (Vol. 102). Elsevier B.V. https://doi.org/10.1016/j.jag.2021.102456

RSI-CB: A Large Scale Remote Sensing Image Classification Benchmark. (2017). https://www.researchgate.net/publication/317240969_RSI-CB_A_Large_Scale_Remote_Sensing_Image_Classification_Benchmark_via_Crowdsource_Data

Rahmadi, L., Hadiyanto, Sanjaya, R., & Prambayun, A. (2023). Crop Prediction Using Machine Learning with CRISP-DM Approach. *Lecture Notes in Networks and Systems*, *787 LNNS*, 399–421. https://doi.org/10.1007/978-981-99-6550-2_31/TABLES/6

Satellite Image Classification. (n.d.). Retrieved July 9, 2024, from https://www.kaggle.com/datasets/mahmoudreda55/satellite-image-classification

SEMMA vs CRISP-DM. (n.d.). Retrieved July 22, 2024, from https://www.starburst.io/blog/semma-vs-crisp-dm/

Striking the balance: Catalyzing a sustainable land-use transition. (2023).

Vali, A., Comai, S., & Matteucci, M. (2020). Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. In *Remote Sensing* (Vol. 12, Issue 15). MDPI AG. https://doi.org/10.3390/RS12152495

Zhao, S., Tu, K., Ye, S., Tang, H., Hu, Y., & Xie, C. (2023). Land Use and Land Cover Classification Meets Deep Learning: A Review. *Sensors*, *23*(21), 8966. https://doi.org/10.3390/s23218966