REPUBLIC OF TURKEY
OSTİM TECHNICAL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF COMPUTER ENGINEERING
COMPUTER ENGINEERING
MASTER’S PROGRAM

CLOUD CLASSIFICATION USING RESIDUAL NETWORK

MASTER’S THESIS

PREPARED BY
MERYEM SENA BARK

THESIS SUPERVISOR
DR. OSMAN AKIN

ANKARA-2024

THESIS ACCEPTANCE AND APPROVAL

This study, titled “Cloud Classification using Residual Network” and submitted by Meryem Sena Bark on 13/06/2024, was found successful as a result of the thesis defense on 13/06/2024 and accepted as a Master's/PhD Thesis by our jury.

Date of Approval: 24/07/2024

Jury Member: Dr. Meltem Imamoğlu Yıldırım_________
Turkish Aeronautical Association University

Jury Member: Dr. Murat Şimşek_______________
Ostim Technical University

Thesis Supervisor: Dr. Osman Akın_______________
Ostim Technical University

APPROVAL

I hereby approve that this study, accepted by the jury, fulfills the requirements for being a Master's/PhD Thesis.

…../…../20….

____

Institute Director

DECLERATION

I hereby declare that I have given Ostim Technical University the permission to archive all or any part of my Master's/PhD thesis approved by the Institute in printed or digital format and make it accessible under the conditions specified below. With this permission, all intellectual property rights other than the usage rights granted to the University will remain with me and the usage rights of all or a part of my thesis for future studies (article, book, license, patent, etc.) will belong to me alone.

I declare and undertake that my thesis is entirely my own work, that I do not violate the rights of others and that I am the sole authorized owner of my thesis. I undertake that I use the copyrighted resources with written permission, which must be used with written permission from their owners, and to submit copies of the permissions to the University upon request.

Within the scope of the "Directive on Collecting, Organizing and Opening to Access The Theses and Dissertations in Electronic Environment"" published by the Council of Higher Education, my thesis is made available on the YÖK National Thesis Center and Ostim Technical University Open Access System, except for the following conditions.

☐ With the decision of the Institute / Faculty Administrative Board, the open access of my thesis has been postponed for 2 years from my graduation date.[1]

☐ With the reasoned decision of the Graduate School / Faculty Administrative Board, the opening of my thesis has been postponed for ………………from the date of my graduation.[2]

☐ A confidentiality decision has been made regarding my thesis.[34]

Date:

Signature:

---

[1] ARTICLE 6(1) In the event that a patent application is made for a graduate thesis or the patenting process is ongoing, upon the recommendation of the thesis advisor and the approval of the graduate school department, the graduate school or faculty executive board may decide to postpone the access to the thesis for two years.

[2] AR 6(2) For theses that use new techniques, materials and methods, that have not yet turned into articles or are not protected by methods such as patents, and that contain information and findings that may create unfair gain for third parties or institutions if shared on the internet, the thesis may be prevented from being accessed for a period not exceeding six months upon the recommendation of the thesis advisor and the approval of the institute department, with the reasoned decision of the institute or faculty board of directors.

[3] ARTICLE 7(1) The decision of confidentiality regarding graduate theses related to national interests or security, security, intelligence, defense and security, health, etc. is made by the institution where the thesis is conducted. The confidentiality decision regarding graduate theses prepared within the framework of a cooperation protocol with institutions and organizations is made by the university board of directors upon the recommendation of the relevant institution and organization and the approval of the institute or faculty. Theses for which a confidentiality decision is made are notified to the Council of Higher Education.

[4] ARTICLE 7(2) Theses for which a confidentiality decision has been made are kept by the institute or faculty within the framework of confidentiality rules during the confidentiality period, and are uploaded to the Thesis Automation System if the confidentiality decision is lifted.

OSTİM TECHNICAL UNIVERSITY

INSTITUTE OF NATURAL AND APPLIED SCIENCES MASTER'S / DOCTORAL THESIS ORIGINALITY REPORT

Thesis Title: Cloud Classification Using Residual Network

Student Name Surname: Meryem Sena Bark

Thesis Supervisor Title Name Surname: Dr. Osman Akın

Department: Graduate School of Engineering and Science

Program: Computer Engineering

Date: 17 / 07 / 2024.

The part of my Master's thesis, which consists of Introduction, Main Chapters and Conclusion and consists of 33 pages in total, was examined by my thesis advisor and myself on 16/02/2024 by the Turnitin plagiarism detection program.

According to the originality report, the similarity score of my thesis is 11%.

Applied filters:

1. Bibliography (excluding)

2. Citations (excluding)

3. Parts of text with less than five (5) words overlapping (excluding)

I declare that my thesis does not contain any plagiarism; that I accept all kinds of legal responsibility that may arise in the event that the contrary is detected and that the information I have given above is correct.

Student Signature:………………….

APPROVAL

Date: 17/07/2024

Dr. Osman Akın

ETHICAL DECLERATION

I declare that this study is an original study, that I act in accordance with scientific ethics and rules in the preparation, data collection, analysis, presentation of information and all other stages of the study, that I have obtained all the document information in the study within the framework of academic ethics and rules, that I have presented all visual, audio, and written information and results in accordance with the scientific rules of ethics, that I have not made any falsifications in the data I have used, that I have referred to the sources I have used in accordance with scientific norms, that my thesis has been written by me and is original, except for the cases where I cited sources, that it has been produced by me under the supervision of my thesis advisor Dr. Osman Akın and written in accordance with Ostim Technical University thesis writing guide.

09/07/2024

Meryem Sena Bark

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Osman Akın, for his invaluable guidance and knowledge throughout my master's thesis. His patience and support at every stage of this work have been crucial in completing this study.

I also extend my thanks to all my professors and the staff at Ostim Technical University who have supported me throughout my education. I am grateful for the knowledge and experiences they have imparted to me during my academic journey.

Furthermore, I would like to thank my family, friends, and loved ones for being my source of morale and motivation. Their support made this achievement possible.

09/07/2024

Meryem Sena Bark

ÖZ

Yazar Adı ve Soyadı : Meryem Sena Bark

Üniversite : OSTİM Teknik Üniversitesi

Enstitü : Fen Bilimleri Enstitüsü

Program Adı : Bilgisayar Mühendisliği

Tezin Türü : Yüksek Lisans Tezi

Sayfa Sayısı : 33

Tarihi : 2024

RESİDUAL NETWORK İLE BULUT GÖRÜNTÜLERİNİN SINIFLANDIRILMASI

Bulut oluşumlarını tanımlamak ve bu oluşumlara dayanarak hava durumu tahmini yapmak, birçok birey ve sektör için hayati öneme sahip görevlerdir. Uydu görüntülerinden elde edilen bulutların tanımlanması ve şekillerinin belirlenmesi, hava durumu tahmin süreçlerini önemli ölçüde kolaylaştırabilir. Bulut şekli tahmini için yapay zekanın kullanılması konusundaki son gelişmelere rağmen, bulutların karmaşık yapıları nedeniyle bu zorluk devam etmektedir. Bu tez, gelişmiş derin öğrenme algoritmalarını kullanarak bulut sınıflandırma performansını iyileştirmeyi hedeflemektedir. 2 farklı veri seti üzerinde çalıştığımız tezde öncelikle NASA Worldview tarafından oluşturulan 5546 fotoğraf üzerinde 4 farklı bulut türünün (balık, şeker, çakıl, çiçek) sınıflandırılması ve sonrasında Harvard Üniversitesi tarafından sağlanan 11 sınıflı 2543 tane fotoğrafı bulunan Cirrus Cumulus Stratus Nimbus (CCSN) üzerine ResNet50 derin öğrenme modeli, bulut görüntülerinin özelliklerini çıkarmak ve sınıflandırmak için kullanılmıştır. Kayıp fonksiyonu olarak binary cross entropy ve weighted binary cross entropy tercih edilmiştir, bu da sınıflandırma işlemlerinin doğruluğunu ve dengesini artırmak amacı taşımaktadır. Veriler çeşitli veri işleme teknikleriyle analiz edilerek sınıflandırma işlemi için uygun hale getirilmiş ve literatürdeki diğer çalışmaların sonuçları ile kıyaslanarak performans analizi yapılmıştır. Çalışma doğrultusunda 5546 fotoğraf üzerinden %69.1 başarı oranı ve CCSN veri seti üzerinden de eğitim sonucunda %91.6 başarı sonucu elde edilmiştir.

Anahtar Sözcükler: Bulut Görüntülerinin Sınıflandırılması, ResNet50, Kayıp Fonksiyonlar

ABSTRACT

Thesis	: Meryem Sena Bark
University	: OSTİM Technical University
Institute	: Graduate School of Natural and Applied Sciences
Program's Name	: Computer Engineering
Thesis Type:	: Master
Pages	: 33
Year	: 2024

CLOUD CLASSIFICATION USING RESIDUAL NETWORK

Identifying cloud formations and making weather predictions based on these formations are crucial tasks for many individuals and industries. The classification and delineation of clouds from satellite images significantly facilitate weather forecasting processes. Despite recent advancements in using artificial intelligence for cloud shape prediction, the complexity of cloud structures continues to pose challenges. This thesis aims to enhance cloud classification performance using advanced deep learning algorithms.

In this study, we worked with two different datasets. Firstly, we classified four types of clouds (fish, sugar, pebble, flower) on 5546 photographs sourced from NASA Worldview. Subsequently, we used the ResNet50 deep learning model on the CCSN dataset provided by Harvard University, which comprises 2543 photographs categorized into 11 classes (Cirrus, Cumulus, Stratus, Nimbus). The ResNet50 model was employed to extract features and classify cloud images.

Binary cross-entropy and weighted binary cross-entropy loss functions were chosen to enhance classification accuracy and balance. Various data preprocessing techniques were applied to prepare the data for classification, and the performance was analyzed by comparing it with results from other studies in the literature.

Through this study, we achieved a success rate of 69.1% on the 5546 photographs and 91.6% after training on the CCSN dataset. This underscores the effectiveness of our approach in cloud classification using deep learning techniques.

Keywords: Cloud Classification, RestNet50, Loss Functions

CONTENTS

THESIS ACCEPTANCE AND APPROVAL i

DECLERATION ii

ETHICAL DECLERATION iv

ACKNOWLEDGEMENTS v

ÖZ vi

ABSTRACT vii

LIST OF TABLES ix

LIST OF FIGURES x

SYMBOLS AND ABBREVIATIONS xi

1. INTRODUCTION 1

2.1. Classification 6

2.2. Deep Learning Based Approaches 8

2.2.1. Convolution Neural Networks (CCNs) 9

2.2.2. Residual Network (ResNet) 10

2.3. Related Works 12

3. PROPOSED METHOD 15

3.1. Loss Functions 16

3.2. Binary Cross-Entropy Loss Functions 17

3.3. Weighted Binary Cross-Entropy Loss Functions 18

3.4. ResNet50 19

4. EXPERIMENTS 22

4.1. Dataset 22

4.2. Experiments 25

5. CONCLUSION 30

REFERENCES 31

LIST OF TABLES

Table 4.1. Distribution of the cloud labels in CCSN dataset 25

Table 4.2. Comparisons of F1 Scores Between ResNet50 and EfficientUnet 28

Table 4.3. Comparisons Between ResNet50 and CloudRCNN 28

Table 4.4. Comparisons with CCSN datasets 29

LIST OF FIGURES

Figure 1.1. 4 types of clouds, (a) Sugar, (b) Gravel, (c) Flower, (d) Fish 2

Figure 2.1. Humans, object detection and segmentation classifications [9]......................... 5

Figure 2.2. Multi-labeled images, (a) Fish-Flower-Gravel, (b) Fish-Flower-Gravel Sugar. 8

Figure 2.3. Multi-label cloud shapes shown in the square plot [10].................................... 8

Figure 2.4. Deep Neural Network Architecture [11].. 9

Figure 2.5. Convolution Neural Network [12] 10

Figure 2.6. Residual Network.. 11

Figure 3.1. Example of Predicted Results 16

Figure 3.2. Keras application ResNet50 21

Figure 4.1. Here are examples of the clouds categories: Flower, Gravel, Fish and Sugar with labeled masks 23

Figure 4.2. Histogram table of data distribution (5546 images). Blue: Fish, Green: Flower, Red: Sugar, Light Blue: Gravel 24

Figure 4.3. Images with labels of the CCSN dataset 24

Figure 4.4. ResNet50 with BCE 26

Figure 4.5. ResNet50 with WBCE 27

SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
BCE	Binary Cross Entropy
WBCE	Weighted Binary Cross Entropy
ResNet	Residual Network
CNN	Convolution Neural Network
RNN	Recurrent Neural Network
GNN	Graph Neural Networ
SENet	Squeeze-and-Excitation Network
CCSN	Cirrus Cumulus Stratus Nimbus

1. INTRODUCTION

Clouds, being one of the most captivating yet elusive natural phenomena, have fascinated scientists and laypeople alike for centuries. Acting as both reflectors and absorbers of solar radiation, they exert a profound influence on the Earth's energy balance [1], [2]. In addition, the structure of clouds affects how precipitation is distributed as well as how storms form and fade. Moreover, clouds are essential to the global water cycle because they make it easier for water vapor to travel throughout the atmosphere [3]. Several variables can be used to classify clouds, such as their physical phase (liquid, ice, or mixed), height (low, mid-level, or high) within the atmosphere, and appearance (color, shape, and combination of these elements). Stratus, Cumulus, Altostratus, Cirrus, and Cirrocumulus are common forms of clouds. By comprehensively understanding the diverse characteristics of clouds, we can improve our capacity to foresee changes in the environments and forecast weather patterns [1]. Therefore, understanding the Earth’s climate system and creating precise predictions about future climatic changes depend heavily on the study of cloud structure.

Various methods and approaches exist for identifying clouds from satellite images. Segmentation and object detection are among the most employed techniques for accurately discerning different cloud types. These methodologies have been widely adopted across various studies, each tailored to address specific research objectives and inquiries. Through the utilization of cutting-edge technologies and novel approaches, researchers have been able to delve into cloud patterns and behaviors with unprecedented detail, thereby advancing our understanding of atmospheric dynamics. By employing innovative methodologies, such as remote sensing techniques and numerical modeling, researchers can extract valuable insights into the complexities of cloud-climate interactions. This progressive approach not only facilitates more accurate weather predictions but also holds promise for improving environmental forecasting, thereby contributing to our ability to adapt to and mitigate the impacts of climate change [4], [5].

Several scholarly inquiries within this discipline have made use of the dataset provided by Kaggle, specifically titled “Understanding Clouds from Satellite Images” [6]. While the EfficientUnet model addresses segmentation [7], another study employing the Mask R-CNN framework proposes enhancements for semantic segmentation of satellite cloud images [8]. However, despite these efforts, a universally sufficient degree of accomplishment in the literature that can be applied to all forms of clouds. Moreover, there is a lack of demonstrated proficiency in accurately locating and identifying cloud formations. Nevertheless, improving these identifications could lead to a significant enhancement in results.

In this thesis, our objective is to efficiently classify four diverse multi-labeled cloud formations -fish, flower, sugar, gravel- by employing one of the Residual Network models, specifically ResNet50. Our analysis incorporates two loss functions: Binary Cross-Entropy (BCE) and Weighted Binary Cross-Entropy (WBCE). The performance of the ResNet50 model in accurately distinguishing and categorizing diverse cloud formations was assessed, demonstrating the model's ability to provide probabilities ranging from 0 to 1 for correctly identifying each cloud type. To gauge this probability, we employed a training dataset that included images depicting the four specific cloud formations sourced from the 'Understanding Clouds from Satellite Images' dataset, which was obtained through the Kaggle competition.

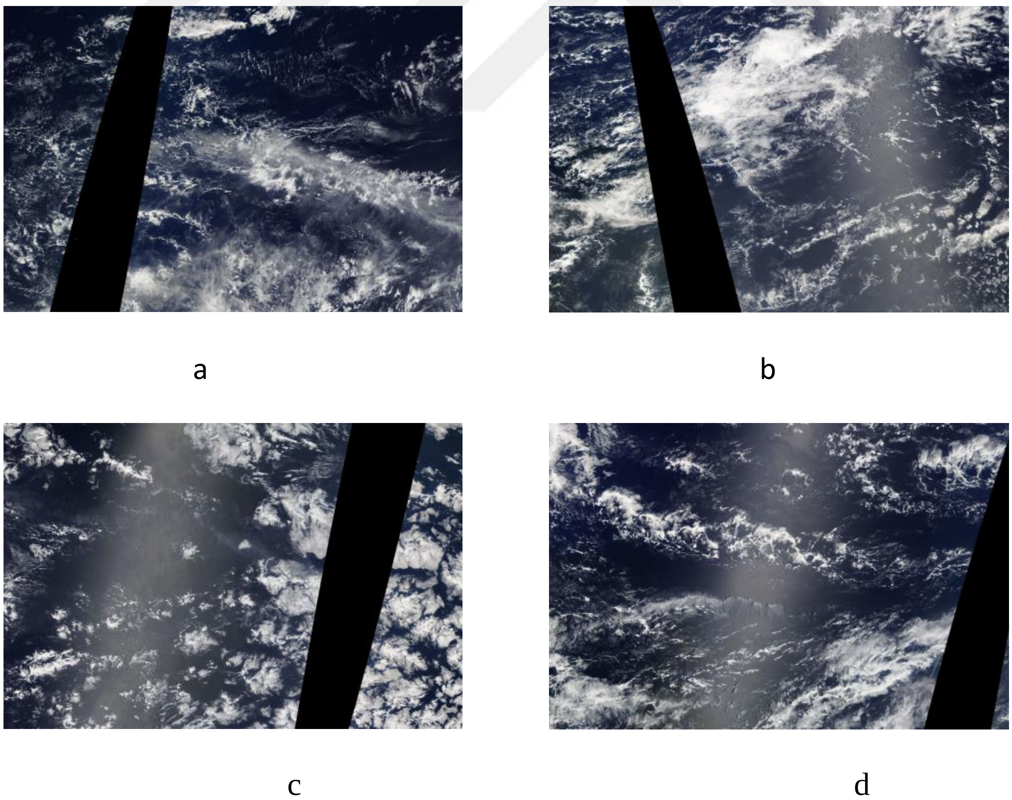


Figure 1.1. 4 types of clouds, (a) Sugar, (b) Gravel, (c) Flower, (d) Fish

Our first dataset consists of four types of clouds, as illustrated by the samples shown in Figure 1.1. In the context of this thesis, while an example of single-labeled classes per images has been provided, our dataset comprises a mixture of both multi-labeled and single-labeled instances. We have elaborated on this aspect under the dataset section, providing comprehensive details.

In our second dataset, there are 11 classes, with each image having a single label. The dataset comprises 2543 images, exhibiting a high class diversity despite the limited data. The challenge lies in assessing the performance of ResNet50 compared to other models under these conditions. Evaluating its success rate in such a scenario will provide insights into its effectiveness in handling datasets with significant class variability.

Through our research, we have made the following contributions:

- Development of a novel approach utilizing satellite images for the classification of four distinct cloud shapes.
- Exploration of the effects of two well-known loss functions (WBCE and BCE) on the classification results when used in conjunction with the ResNet50 model.

The remainder of this thesis is structured as follows: Chapter 2 provides a thorough exposition of the existing algorithms along with an in-depth discussion of related works pertinent to our own study. This segment delves into a meticulous examination of the algorithms currently in use, elucidating their underlying principles, strengths, and limitations. Moreover, it scrutinizes a spectrum of research endeavors that bear relevance to our own investigation, thereby furnishing a contextual backdrop for the subsequent chapters. By elucidating the theoretical underpinnings and practical applications of these algorithms, this chapter sets the stage for the subsequent sections wherein we present our novel approach and experimental findings. Chapter 3 clarifies our proposed model along with the evaluation metrics utilized. Section 4 examines the datasets we used and demonstrates our achievement results compared with another existing research. Finally, Chapter 5 presents the conclusion and summarization of the paper, along with avenues for future research.

2. BACKGROUND

Image processing is a science and technology used for the analysis, enhancement, and manipulation of digital images. images are acquired in digital form and processed using various algorithms. Image processing covers a wide range of both theoretical and practical aspects, and its primary goals are to improve image quality, extract specific features, or derive meaningful information from visual data.

The image processing workflow typically consists of several fundamental stages:

- Image Acquisition: Image acquisition starts with capturing the image using a sensor or camera and converting it into digital format. This step is the first link in the image processing chain, and the quality of the acquired image directly affects the success of subsequent processes.
- Preprocessing: The preprocessing stage includes techniques aimed at enhancing the quality of the image. During this stage, operations such as noise reduction, contrast enhancement, histogram equalization, and geometric corrections are performed. The goal is to make the image analyzable and meaningful.
- Image Transformation: Image transformations allow the representation of the image in different spaces. Methods such as Fourier transform, and wavelet transform are used to reveal frequency components or specific features within the image.
- Segmentation: The process of breaking an image up into distinct areas or objects is called segmentation. This step is typically achieved through techniques like edge detection, thresholding, and clustering. The aim is to identify objects or regions of interest.
- Feature Extraction: Feature extraction involves identifying and analyzing important features withing the image. Features such as edge, corners textures, and shapes are used for the further analysis of the image.
- Classification and Recognition: Identification and categorization of items or patterns within the image are part of the classification and recognition stage. Machine learning and artificial intelligence algorithms are frequently used in this stage. For example, facial recognition, handwriting recognition, and object classification are performed during these steps. This thesis also focused on the classification steps.

Figure 2.1. illustrates the distinction between object detection and segmentation-based classification on our own dataset.

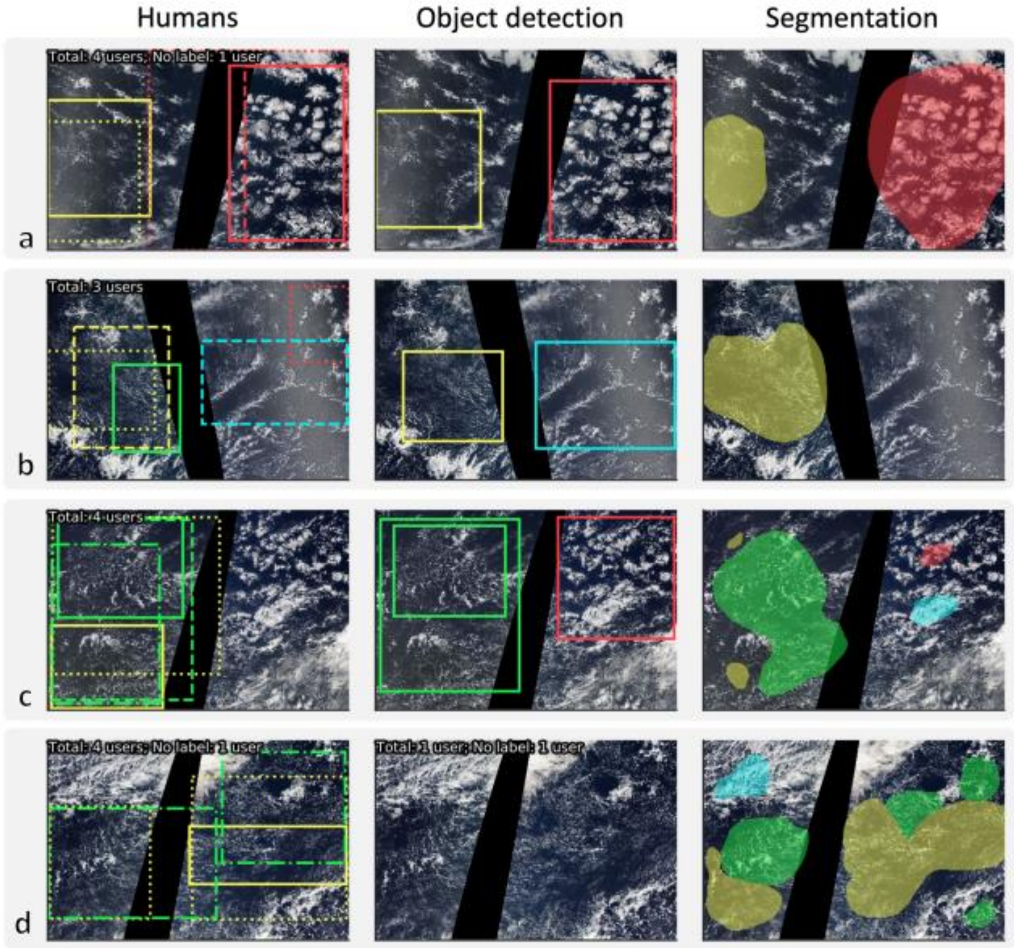


Figure 2.1. Humans, object detection and segmentation classifications [9]

Image processing is a crucial technology for understanding and processing visual data in the digital world. Advances in this field enable more precise and faster analyses, leading to significant innovations and improvements in many sectors. It is anticipated that in the future, machine learning and artificial intelligence-integrated image processing techniques will aid in the creation of more intelligent and independent systems.

The broad scientific topic of artificial intelligence (AI) studies computer systems' ability to carry out activities that normally require human intelligence. Problem-solving, language comprehension, visual perception, learning, and decision-making are some of these tasks. The main objective of AI is to create machines that can replicate human intellect in order to handle challenging tasks and change with the times.

Machine Learning (ML) is a portion of artificial intelligence where computer systems learn from data without needing to be explicitly programmed. According to Arthur

Samuel's 1959 definition, machine learning is "the field of study that gives computers the ability to learn without being explicitly programmed." ML works by recognizing patterns in large datasets and making predictions and decisions based on these patterns. It encompasses various methods, including supervised learning, unsupervised learning, and reinforcement learning.

Deep Learning (DL) is a portion of machine learning that relies on multi-layered artificial neural networks. DL, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has demonstrated outstanding achievements in areas such as image and speech recognition, natural language processing, and game playing. These networks automatically learn hierarchical representations of data and extract meaningful features from it.

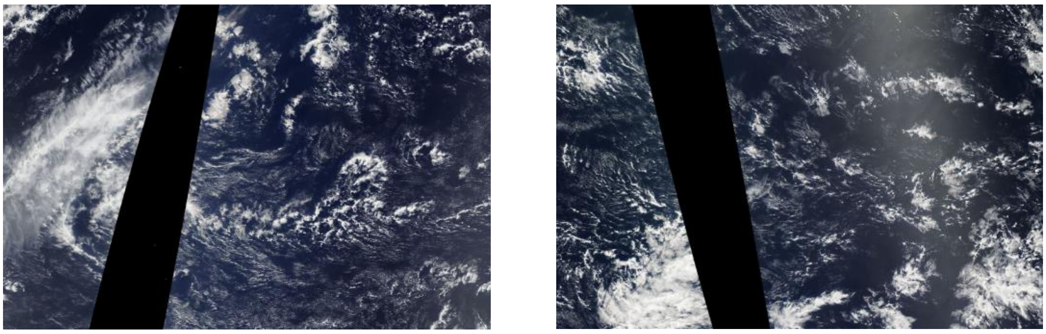
Artificial Intelligence, Machine Learning, and Deep Learning are interconnected disciplines that complement and reinforce each other. While AI provides the framework for creating intelligent systems, Machine Learning equips these systems with the ability to learn and adapt. Deep Learning, as an advanced subset of ML, enhances AI capabilities, particularly in handling complex data and tasks. The integration of these three fields leads to significant advancements in modern technology and innovation processes.

Elucidating methodologies that are closely aligned with our research objectives is paramount in our thesis. We focus on classification techniques that not only resonate with the scope of our study but also contribute substantially to advancing our specific field of inquiry. By prioritizing the discussion of these selected methodologies, our aim is to offer a nuanced understanding of their applicability and efficacy in addressing the inherent challenges within our research domain.

2.1. Classification

Classification, within the realm of deep learning, pertains to the task of categorizing data points into predefined classes or categories. This fundamental technique finds extensive applications in a variety of fields, including as computer vision and natural language processing. Deep learning models are particularly adept at classification tasks because they can automatically extract hierarchical features from unprocessed data. Several types of classification methodologies exist within the scope of deep learning:

- Supervised Classification: Supervised classification involves training a model on labeled data, where each data point is associated with a corresponding class label. Popular techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are commonly employed for tasks such as image classification and sequence labeling.
- Unsupervised Classification: Unsupervised classification, on the other hand, deals with clustering similar data points together without explicit class labels. Methods like k-means clustering and hierarchical clustering are frequently utilized to uncover latent structures within data.
- Semi-supervised Classification: Semi-supervised classification combines elements of both supervised and unsupervised learning paradigms. In scenarios where labeled data is scarce but unlabeled data is abundant, semi-supervised techniques aim to leverage the available labeled data alongside the unlabeled data to improve model performance.
- Multi-class Classification: Multi-class classification refers to scenarios where data points may belong to one of several possible classes. Deep learning models can be tailored to predict the probabilities of each class, enabling nuanced decision-making.
- Binary Classification: Binary classification involves categorizing data points into one of two exclusive classes. This type of classification is commonly encountered in applications such as spam detection and medical diagnosis.
- Multi-label Classification: Multi-label classification involves categorizing data points into multiple classes simultaneously. Unlike traditional classification tasks where each data point is associated with a single class, multi-label classification allows for multiple class labels per data point. This scenario commonly arises in applications such as image tagging, where an image may contain multiple objects or concepts that need to be identified. Deep learning models tailored for multi-label classification tasks typically output a set of probabilities or binary predictions for each class label, indicating the presence or absence of the respective label in the input data. Here in Figure 2.2 and Figure 2.3., there is an example of multi-labeled images from our dataset.



a b

Figure 2.2. Multi-labeled images, (a) Fish-Flower-Gravel, (b) Fish-Flower-Gravel Sugar

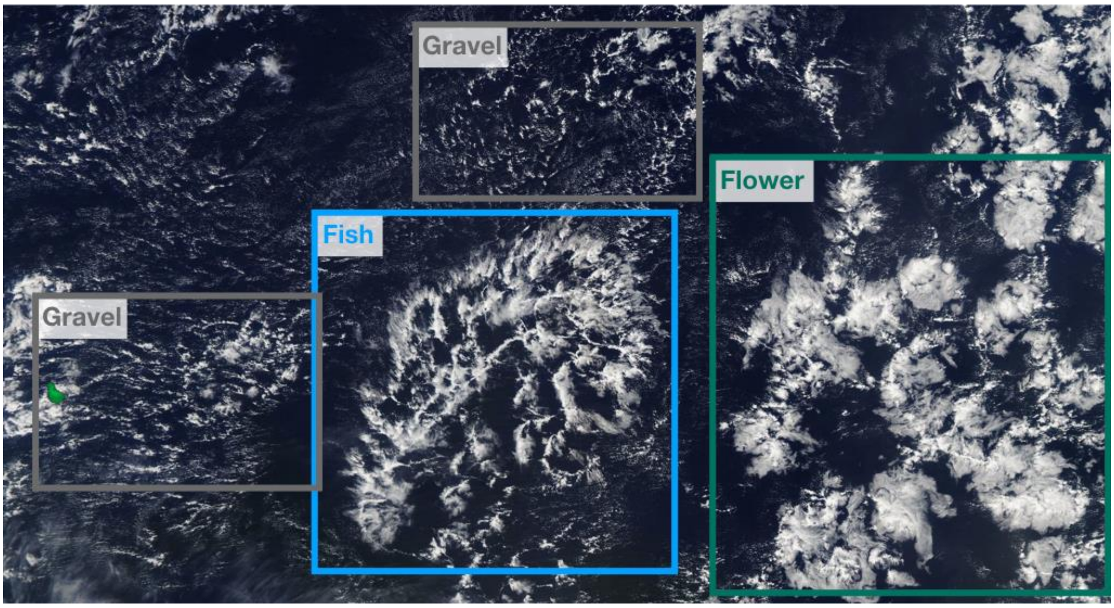


Figure 2.3. Multi-label cloud shapes shown in the square plot [10]

2.2. Deep Learning Based Approaches

Deep learning is a subfield of machine learning that makes use of intricate, multilayered neural networks to tackle challenging data problems. Deep learning aims to automatically extract patterns and relationships in data using these multi-layered neural networks trained on large datasets. This technique has achieved great success, especially in areas such as image and speech recognition, natural language processing, and time series analysis.

Deep learning performs the learning process using artificial neural networks consisting of many layers as shown on Figure 2.4. These layers process input data (such as an image or an audio recording) and gradually learn more abstract and complex features. In the final layer, typically referred to as an output layer, these features are used to produce results.

Training deep learning algorithms requires significant computational power and data. However, this approach can achieve human-level or superhuman performance in complex data tasks. Therefore, it has garnered significant interest in fields such as image recognition, speech recognition, automatic translation, and more.

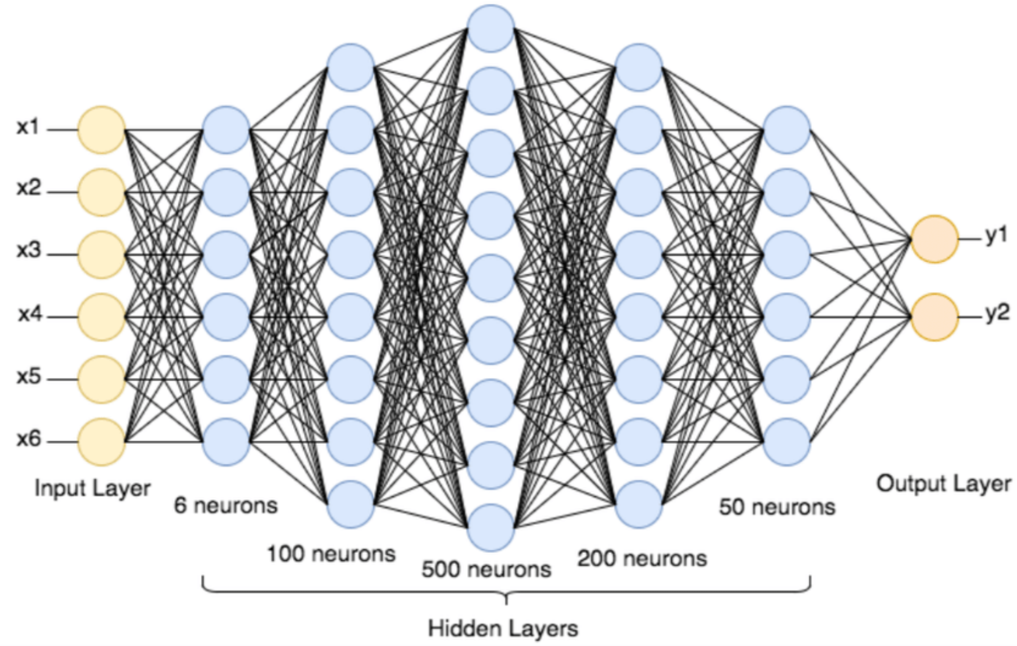


Figure 2.4. Deep Neural Network Architecture [11]

2.2.1. Convolution Neural Networks (CCNs)

Convolutional Neural Networks (CNNs) are a subset of deep learning models particularly effective in image processing and computer vision tasks. They are made up of several layers, such as fully connected, pooling, and convolutional layers. Convolutional layers employ filters to extract spatial hierarchies and patterns from the input data, whereas pooling layers lower the dimensionality of the data while maintaining crucial characteristics and enhancing computing effectiveness. Fully connected layers employ these features to perform classification or regression operations. CNNs are popular for their capacity to automatically extract pertinent features from unprocessed input, which makes them very useful for applications like object identification, facial recognition, and image classification. By leveraging shared weights and local connectivity, CNNs maintain spatial relationships in data, allowing them to achieve high accuracy with fewer parameters compared to traditional neural networks. Here in Figure 2.5. shows convolution neural network architecture. First step is the padding, this means losing data over borders so append a border of zeros and recalculate the convolution covering all the input values. In the max pooling steps take a small portion of the input and try to take average value referred to as an average pooling or take a maximum value termed as max pooling, so the max pooling on image is not taking out all values, just being summarized value over all values present. [12] fully connected layer consists of several hidden layers, each containing numerous neurons, with every neuron being completely connected to the neurons in the next layer. The input for a fully connected layer is a one-dimensional (1D) feature vector, created by flattening the feature maps generated by the preceding convolutional and pooling layers. [13]

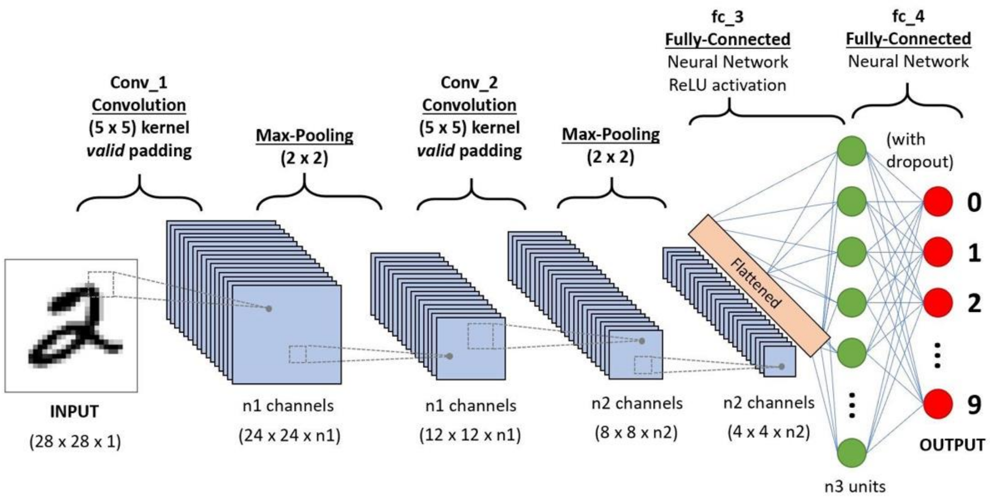


Figure 2.5. Convolution Neural Network [12]

2.2.2. Residual Network (ResNet)

A deep learning architecture called Residual Network (ResNet) was created to solve the "vanishing gradient" issue that arises during the training of extremely deep neural networks. Introduced by Kaiming He and colleagues in 2015, ResNet revolutionized the field of deep learning. Its key feature is the use of skip connections (or shortcuts) that pass the output of one layer directly to subsequent layers. These connections allow for effective weight updates and make it possible to train a very deep network. Here is the formula of the Residual Network which is S(x) is output of the algorithm, F is a residual function, W are parameters of the block and x is input also shown on Figure 2.6. [14]

$$S(x) = F(x, W) + x \tag{2.1}$$

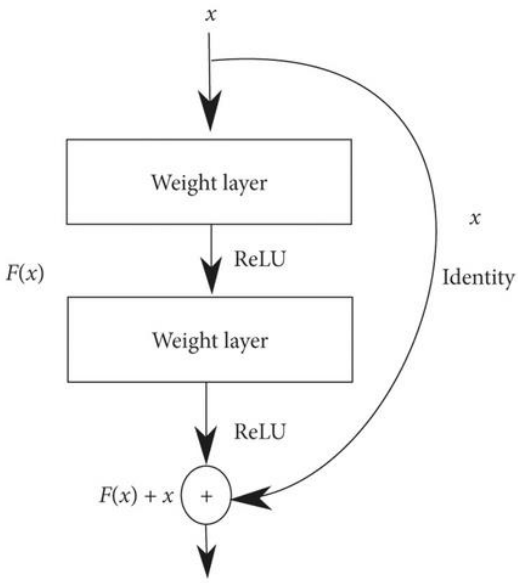


Figure 2.6. Residual Network

ReLU as shown on Figure 2.6. is the Rectified Linear Unit(ReLU) that serves as the activation function in each layer of the network. ReLU keeps positive input values unchanged while setting negative input values to zero. This functionality enhances the network's learning capacity and speeds up computations. ReLU is used in each layer within a residual block to introduce non-linearity. This nonlinearity allows the network to model complex functions. Additionally, it eliminates the issue of vanishing gradients, making deep network training more efficient. This feature is particularly important for very deep networks like ResNet.

ResNet architecture is a significant milestone in deep learning, addressing the "degradation problem" which occurs when training very deep networks. ResNet solves this issue using "residual blocks." Here are some of the popular ResNet variants:

ResNet-18: Consists of 18 layers. (17 convolution layers and 1 fully connected layer). It's lighter and faster due to its lower depth, suitable for small to medium-sized datasets.

ResNet-34: Consists of 34 layers. An intermediate depth network, providing better performance for more complex datasets.

ResNet-50: Consists of 50 layers, using "bottleneck" structures with smaller convolutional filters. Suitable for large datasets and applications requiring high performance

ResNet-101: Consist of 101 layers. A deeper capability of learning more complex data. Ideal for very datasets and applications demanding high accuracy.

ResNet-152: Consist of 152 layers, one of the deepest ResNet models. Suitable for very large-scale datasets and highly complex tasks, though it requires significant computational resources.

During the various stages of this research, different ResNet architectures were experimented with, and it was observed that the highest performance was achieved using ResNet-50. Extensive experimentation and analysis demonstrated that ResNet-50 outperformed other variants such as ResNet-18, ResNet-34, ResNet-101, and ResNet-152, in terms of accuracy and computational efficiency.

2.3. Related Works

Deep learning algorithms assist in classifying cloud images by recognizing patterns and features that signify different types of clouds or meteorological phenomena. Researchers have employed convolutional neural networks to classify cloud shapes with remarkable accuracy, enhancing understanding of atmospheric dynamics and improving weather forecasting capabilities.

While convolutional neural networks (CNNs) have demonstrated remarkable success in single-label image classification tasks, researchers have aimed to adapt CNNs for multi-label image classification scenarios. Jiang et al. proposed a hybrid approach combining recurrent neural networks (RNNs) with CNNs, forming the CNN-RNN framework for multi-label classification across three distinct datasets [15]. In this framework, a CNN is utilized to extract features from the image, while a recurrent layer integrates information from previously predicted labels. The final output label probability is then computed based on both the image representation and the output from the recurrent layer, showcasing promising performance compared to alternative models.

In another study, Ivo M. et al. investigated the effectiveness of ResNet50 when applied to large-sized X-ray images [16]. Their analysis encompassed considerations such as, position, age and gender in addition to the X-ray images, while also exploring the performance of alternative Residual Network models such as ResNet-38 and ResNet-101. The dataset utilized in this study comprised X-ray images categorized into 14 distinct classes, with results compared against existing literature. Their approach demonstrated notable success, particularly in five out of the 14 classes.

Mario and Simone et al. proposed a transfer learning approach utilizing pre-trained deep neural network architectures, including AlexNet, ResNet18, ResNet50, and GoogleNet, applied to various cloud datasets [17]. Their study involved comparisons with prior research efforts, showcasing the efficacy of transfer learning methodologies.

In a different vein, Xinrui and Juan et al. presented a Random Forest Model for cloud classification, incorporating spectral, texture, and color features within the dataset [18]. They observed improved results through post processing with a threshold method applied to the model.

Thibaut et al. emphasized the challenge of multi-label image tagging compared to single-label images, proposing a learning approach leveraging partial ethics for multi-label images [19]. Their study employed Binary Cross Entropy (BCE) and partial BCE loss methods using the Graph Neural Network (GNN), with notable enhancements observed in the study outcomes.

Furthermore, Ahmed and Noor et al. introduced the EfficientUnet model for the same dataset as ours, achieving satisfactory results [7]. EfficientNet, a novel CNN version, served as an encoder, coupled with the use of Unet as a decoder, forming an architecture of segmentation Their performance on the Kaggle Competition leaderboard [6] was highlighted, accompanied by precision-recall tables for each label.

Similarly, Gonghe and Baohe et al. introduced the CloudRCNN model, based on deep neural networks and utilizing the same dataset as ours [8]. Their study aimed to enhance the MaskRCNN framework, achieving improved success scores through the implementation of different loss functions, including Auxiliary loss and feature fusion techniques.

As a study utilizing different datasets, Emmanuel and Purushotham et al. proposed the Cloud-MobiNet model. They expanded the CCSN dataset to approximately 41,891 images and achieved a success rate of 97.45%.[20] Thanks to their shared code, we also tested the MobiNet model with the original CCSN dataset containing 2,543 images and included our comparison.

Li et al. proposed a new method of classifying cloud types using a combination of Convolution Neural Networks (CNNs) and Squeeze-and-Excitation (SENet) networks [21]. The study addresses the limitations of current satellite cloud classification methods, which typically analyze single pixel thresholds without considering the relationships between surrounding pixels. They worked on a CCSN dataset (2543 images) and got 94,50% accuracy.

Most recently, Mehmet, Muruvvet et al. conducted research on the classification of cloud images, employing various deep learning algorithms including MobileNet V2, Inception V3, EfficientNetV2L, VGG-16, Xception, ConvNeXtSmall, and ResNet-152 V2. Their study focused on the classification of twelve different types of clouds." These studies collectively contribute to the ongoing efforts to refine and innovate within the realm of multi-label image classification, each offering unique insights and methodologies to address the inherent challenges within the domain.[22]

Despite the advancements in cloud shape recognition algorithms in the literature, fully locating cloud formations within images remains a challenge. While segmentation-based algorithms have been proposed, the choice and implementation of the loss function also play a significant role.

3. PROPOSED METHOD

We aimed to improve cloud classification performance using advanced deep learning algorithms. We utilized the ResNet50 model to classify four types of clouds (Fish, Sugar, Gravel, Flower) from satellite images. Our proposed method aims to accurately classify four distinct cloud shapes, building upon the ResNet50 algorithm as a foundation and introducing a novel framework within its scope. This framework demonstrates that employing object detection for cloud shape classification leads to more precise results. Additionally, the careful selection and thorough evaluation of the loss function play pivotal roles in determining the success of the algorithmic framework. Through rigorous experimentation, we have discerned that the adoption of two distinct loss functions, namely Binary CrossEntropy and Weighted Binary Cross-Entropy, leads to notably enhanced outcomes, particularly in the realm of multi-label classification tasks. Our approach highlights the efficacy of leveraging the ResNet50 model for the precise identification and classification of various cloud formations, wherein the probability of accurately categorizing a specific type of cloud spans the spectrum from 0 to 1 shown on Figure 3.1. To ascertain and validate this probability, we meticulously curated a diverse training dataset encompassing an array of cloud formations, thereby ensuring the robustness and generalizability of our findings

This section commences with an elucidation of the concept of the loss function, which serves as a critical component in the evaluation and optimization of machine learning models. Subsequently, it provides a detailed exposition of the two loss functions utilized in this study, namely Binary CrossEntropy and Weighted Binary Cross-Entropy, shedding light on their respective formulations and utility in the context of multi-label classification tasks. Following this, the ResNet50 model, one of the pivotal residual network algorithms scrutinized in this thesis, is delineated, along with an explanation of its architectural framework and pertinent parameters.

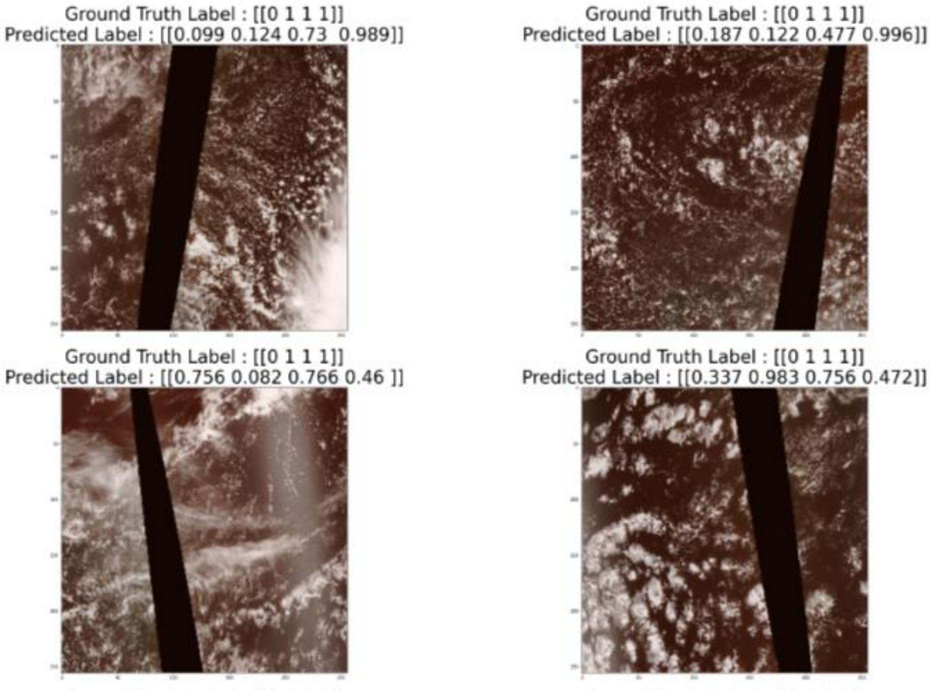


Figure 3.1. Example of Predicted Results

3.1. Loss Functions

The loss function plays a pivotal role in evaluating the performance of a neural network model during its training phase. It quantifies the disparity between the predicted outputs of the model and the actual ground truth labels. A low value of the loss function indicates successful predictions, while a high value signifies inaccuracies in the model's predictions. Throughout the training process, monitoring the loss function values is essential. A consistent decrease in these values suggests that the model is learning effectively, while a consistent increase indicates potential issues necessitating adjustments to the model.

By encapsulating various aspects of the model’s performance into a single metric, the loss function facilitates the optimization process, guiding the model towards improved performance. Therefore, selecting the most appropriate loss function is essential to maximizing the efficiency of the model. The selection of the loss function depends on the specific characteristics and requirements of the problem at hand whether it involves regression, classification, or multi-classification tasks. Here is a formula of the loss function in 3.1

$$Loss = (y_i - \widehat{y_i})^2$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2$$

(3.1)

y_i is the actual value, $\widehat{y_i}$ is the predicted value and N is the total number of the points. In this formula, the difference between the actual value and the predicted value for each data point is squared. These squared differences are then summed and divided by the number of data points to calculate the mean squared error. Squaring the differences ensures that both positive and negative errors contribute positively to the total error, and it also gives more weight to larger errors.

In summary, the term $(y_i - \widehat{y_i})^2$ represents the squared error for each data point, and MSE calculates the average of these squared errors.

Loss functions play a significant role in shaping the overall performance of the model [23]. In our thesis, we explore two types of loss functions to observe their impact on model performance: binary cross-entropy loss function and weighted binary cross-entropy loss function [24].

3.1. Binary Cross-Entropy Loss Functions

The Binary Cross-entropy (BCE) function stands out as one of the most popular choices for loss functions in multi-label classification tasks. It calculates the cross-entropy loss between the true labels and the predicted labels for each individual label. The total loss is computed as the summation of the binary cross-entropy losses across all labels. This approach enables the model to effectively evaluate and adjust its predictions for each label independently, optimizing its performance across the entire multi-label classification problem.

$$L = -\frac{1}{N}\sum_{i}^{N} y_i * \log\big(p(y_i)\big) + (1 - y_i) * \log(1 - p((y_i))$$

(3.2)

Here is N number of samples y_i is real value p (y_i) is the probability of class 1 (predicted value), and 1-p(y_i) is the probability of class 0.

3.2. Weighted Binary Cross-Entropy Loss Functions

The Weighted Binary Cross Entropy Loss Function is a specialized variant of the binary cross-entropy loss function, where positive and negative weights are assigned to each class. In this formulation, the presence of a label (1) is treated as the positive and minority class, while the non-presence (0) is considered the negative and majority class. The weights are assigned accordingly, with positive weights assigned to the occurrence class and negative weights assigned to the non-occurrence class. This weighting scheme aims to address class imbalances and prioritize the correct prediction of minority classes.

{ 'Fish' : 1.0009025270758, 'Flower': 1.186096256684,
'Sugar': 0.7349237905898, 'Gravel': 0.9418259023355 }

{ 'Fish' : 0.9990990990991, 'Flower': 0.8643803585347,
'Sugar': 1.56417489421721, 'Gravel' : 1.0658337337819}

The WBCE loss function is typically implemented in code according to the following formulation

$$L_{WBCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[w_{pos} * y_i * \log(\widehat{y_i}) + w_{neg} * (1 - y_i) * \log(1 - \widehat{y_i})\right]$$

(3.3)

Where:

- N is the total number of samples
- y_i is the true label (0 or 1) for the ith sample
- $\widehat{y_i}$ is the predicted probability of the positive class for the ith sample.
- w_{pos} is the weight assigned to the positive class (occurrence).
- w_{neg} is the weight assigned to the negative class (non-occurance).

This loss function is implemented directly in code rather than being provided by a standard library, allowing for flexibility in customizing the weighting scheme to suit specific classification tasks and address class imbalances effectively.

3.3. ResNet50

The ResNet50 model, alternative to the Residual Network (ResNet), is distinguished by its 48 convolutional layers, supplemented by 1 MaxPool layer and 1 Average Pool layer. It boasts a staggering 3.8 x 10^9 floating-point operations, underscoring its computational complexity and capabilities.

Deep convolutional neural networks, exemplified by ResNet50, have ushered in significant breakthroughs in image classification. Beyond image classification, these networks have also revolutionized various other visual recognition tasks, leveraging their depth to deal with growing complicated challenges and achieve higher accuracy rates. Nevertheless, as networks grow deeper, training them becomes more arduous, and accuracy gains may plateau or even diminish. Residual Learning offers a solution to these challenges by addressing them concurrently [25].

In configuring the ResNet50 model [26], various parameters can be specified, including pooling type, weights, inclusion of top layers, input shape, number of classes, and classifier activation functions [27]. These configurable options provide flexibility in tailoring the model to specific tasks and optimizing its performance for different applications.

The ResNet50 model is configured with the following arguments:

- Include top: This parameter is set to False, which means that the completely linked layer at the top of the network is excluded.
- Weights: The weights used in the model are from the 'ImageNet' dataset, which implies pre-training on ImageNet. These weights provide a starting point for the model's parameters.
- Pooling: The pooling mode selected for feature extraction is Average Pooling (Avg). This means that global average pooling will be applied to the output of the last convolutional block, resulting in a 2D tensor output.
- Input shape: The input shape specified for the model is (256, 256, 3), representing images with 256x256 pixels and 3 color channels (RGB).

For this multi-label classification task, the ResNet50 model is utilized with two different loss functions, Binary Cross-Entropy (BCE) and Weighted Binary Cross-Entropy (WBCE). The primary objective of this study is to discern and evaluate the effects of employing various loss functions on the classification accuracy of cloud labels. This investigation entails a comprehensive analysis aimed at elucidating how selection of loss function influences the performance and efficacy of the classification model, particularly in the domain of multi-label cloud classification. Furthermore, the study endeavors to juxtapose the outcomes derived from our proposed approach with those attained by other algorithms documented in the existing literature. By conducting such comparative analyses, we seek to discern any discernible advantages or limitations of our methodology vis-à-vis alternative approaches, thereby contributing to a deeper understanding of the efficacy and applicability of different algorithmic frameworks in the context of cloud classification tasks. As shown on Figure 9, The ResNet-50 architecture employs several convolutional layers, each followed by batch normalization and ReLU activation functions. These layers are crucial for extracting features from the input images (clouds) Subsequently, max pooling layers are used to down sample the feature maps, reducing their spatial dimensions while retaining the most significant features. ResNet-50's architecture relies on two primary types of blocks: the identity block and the convolutional block. The identity block straightforwardly routes the input through multiple convolutional layers and then merges it with the original input, allowing the network to learn residual functions that effectively map inputs to outputs. In contrast, the convolutional block begins with a 1x1 convolutional layer to reduce the number of filters before applying a 3x3 convolutional layer, distinguishing it from the identity block. The concluding segment of ResNet-50 consists of fully connected layers, crucial for the classification task. The output from the final fully connected layer undergoes softmax activation to produce the ultimate class probabilities in conjunction with the loss functions. The evaluation of the obtained results is discussed in the Experiments section.

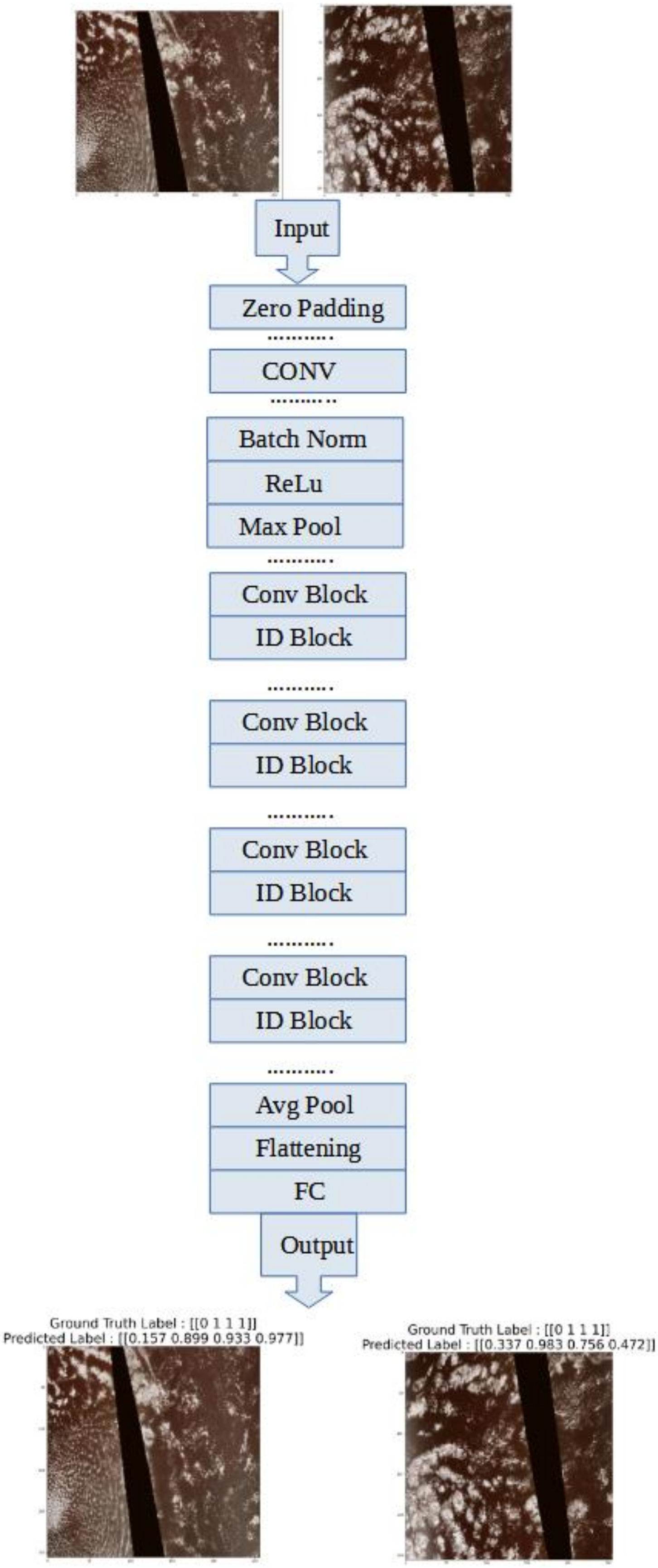


Figure 3.2. Keras application ResNet50

4. EXPERIMENTS

In this part, we delve into the characteristics of the dataset utilized for our study, emphasizing the distribution of labels within it. Additionally, we aim to delineate the disparities between the outcomes of our study and those of previous research endeavors.

The dataset employed in our study comprises a collection of satellite images annotated with cloud labels. These labels correspond to various cloud formations, encompassing distinct shapes and patterns. An analysis of the label distribution reveals insights into the prevalence of different cloud types within the dataset. By examining the frequency of occurrence for each label, we gain a comprehensive understanding of the dataset's composition and potential class imbalances.

Furthermore, we compare the results obtained from our study with those reported in prior research efforts. We scrutinize the methodologies, models, and evaluation metrics employed in these studies to discern any discrepancies in performance. Factors such as the choice of loss functions, model architectures, and dataset preprocessing techniques may contribute to variations in outcomes across different studies.

By elucidating the characteristics of the dataset and elucidating the discrepancies between our study's findings and those of prior research, we aim to provide valuable insights into the efficacy of our approach and identify areas for further improvement in future investigations.

4.1. Dataset

In the thesis, we utilized the image datasets sourced from the "Understanding Clouds from Satellite Images" challenge on Kaggle [6], which are generously provided by NASA Worldview. The dataset comprises a total of 5546 images, which were divided into training and validation sets consisting of 4436 and 1110 images, respectively. These images depict various cloud formations and are labeled with four types of cloud shapes: Fish, Flower, Gravel, and Sugar. Each image has been meticulously labeled by at least three scientists from a team of 68 data scientists involved in the project.

All images in the dataset have consistent dimensions, with a resolution of 1400 x 2100 pixels. Figure 4.1. illustrates an example image labeled with corresponding masks, showcasing the annotated cloud shapes.

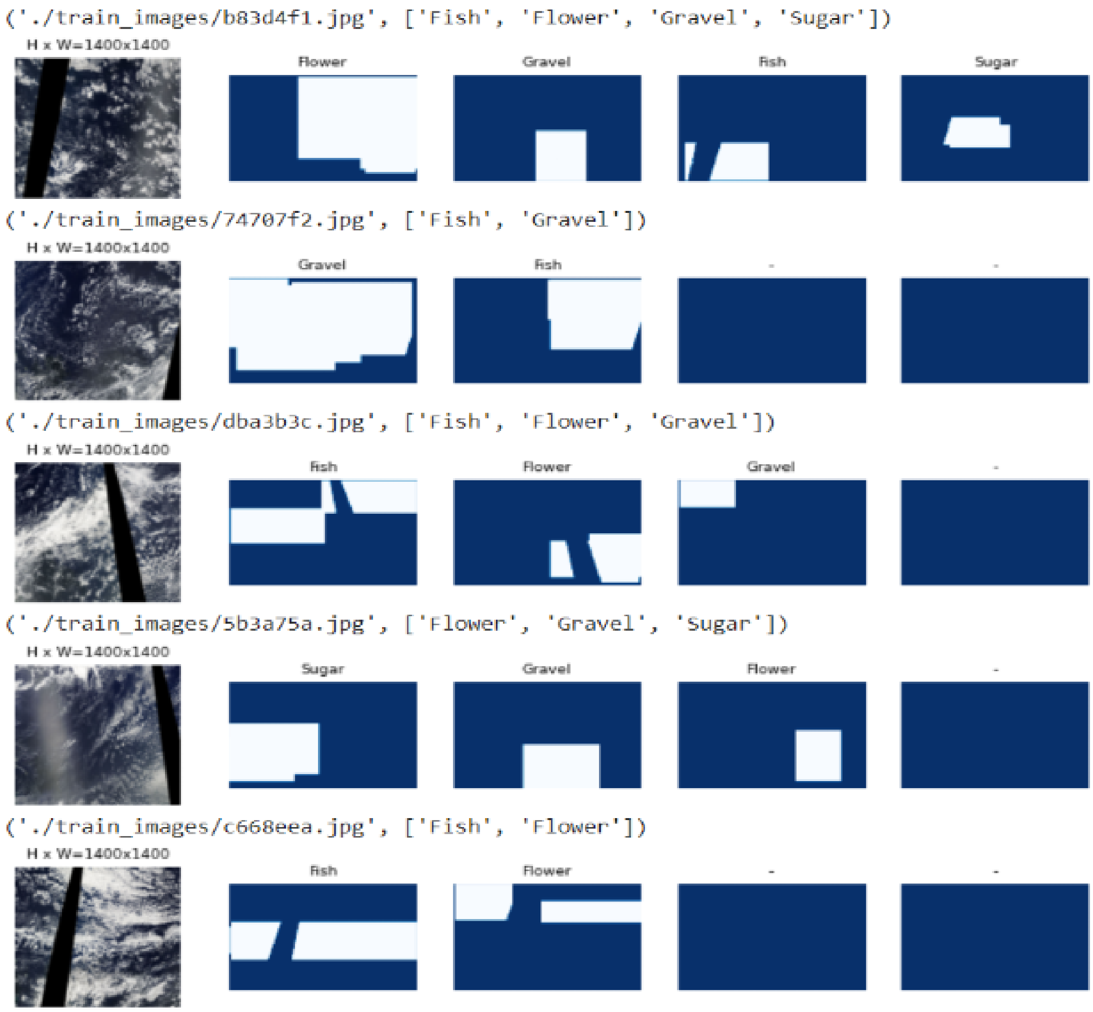


Figure 4.1. Here are examples of the clouds categories: Flower, Gravel, Fish and Sugar with labeled masks

Additionally, we conducted an analysis of the data distribution, as depicted in Figure 4.2. through a histogram table.

This analysis revealed that the distribution of data across different cloud shapes is not balanced. This imbalance in the dataset could potentially pose challenges during model training and may influence the performance of the classification algorithm.

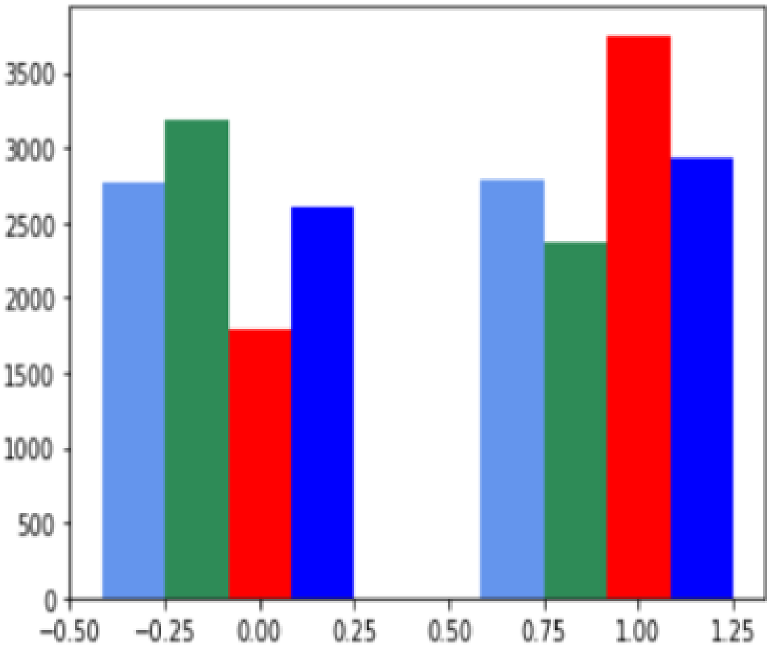


Figure 4.2. Histogram table of data distribution (5546 images). Blue: Fish, Green: Flower, Red: Sugar, Light Blue: Gravel

By providing insights into the dataset characteristics, including its origin, size, labeling process, and data distribution, we aim to provide a comprehensive understanding of the data used in our study and its potential implications on model performance.

As a second dataset, the Cirrus Cumulus Stratus Nimbus (CCSN) dataset consists of a total of 2543 cloud images and includes 11 classes owned by Harvard University.[28] The sample images of the classes are presented in Figure 4.3. while their numerical distribution within the dataset is detailed in Table 1. Each class represents a distinct type of cloud, providing a comprehensive overview of the dataset's diversity. This information is crucial for understanding the dataset's composition and the subsequent performance of the model.

Figure 4.3. Images with labels of the CCSN dataset

Table 4.1. Distribution of the cloud labels in CCSN dataset

Cloud Name	Short Name	Total Number
Stratocumulus	Sc	340
Cirrostratus	Cs	287
Nimbostratus	Ns	274
Cirrocumulus	Cc	268
Cumulonimbus	Cb	242
Altocumulus	Ac	241
Stratus	St	221
Contrails	Ct	200
Altostratus	As	188
Cumulus	Cu	182
Cirrus	Ci	139
Total		2543

4.2. Experiments

In our experimentation, we applied two different metrics as previously mentioned. The first metric involved training the ResNet50 model with a binary cross-entropy loss function, using images resized to 256x256 pixels over 10 epochs. This configuration is referred to as ResNet50BCE. For this model, we utilized the Keras application library to instantiate the ResNet50 architecture. The binary cross-entropy loss function was paired with the Adam optimizer in a learning rate of 0.00001, and accuracy was chosen as the evaluation metric.The second metric, ResNet50WBCE, also utilized images resized to 256x256 pixels, but employed a weighted binary cross-entropy loss function over 15 epochs. The weighted binary cross-entropy loss function was implemented by us, rather than being sourced from any existing Keras library. Similar to ResNet50BCE, the ResNet50WBCE model used the same Adam optimizer learning rate with accuracy as the evaluation metric.

Through these experiments, we aimed to evaluate the performance of the ResNet50 model under different loss functions and training epochs. By comparing the results of ResNet50BCE and ResNet50WBCE, we aimed to gain insights into the efficacy of the weighted binary cross-entropy loss function and its impact on model performance. Additionally, our custom implementation of the loss function allowed for flexibility in tailoring the model to our specific classification task.

Here is epoch accuracy compared between ResNet50 with binary cross entropy and weighted binary cross entropy in Figure 4.4. and Figure 4.5. We compared the performance of binary cross entropy (BCE) and weighted binary cross entropy (WBCE) loss functions using the ResNet50 algorithm. With BCE, we achieved satisfactory results during training over 10 epochs, reaching a training accuracy of up to 76%, while the validation accuracy remained around 69%. When employing WBCE, training accuracy also reached around 74%, with validation performance hovering around 68%.

These findings indicate that BCE and WBCE exhibit similar performance during training, with WBCE slightly outperforming BCE on the validation set. Despite achieving high training accuracy, its lower-than-expected performance on the validation set may suggest overfitting or an inability to address data imbalance. The superior performance of WBCE on the validation set may indicate the effectiveness of the weights used to address class imbalance.

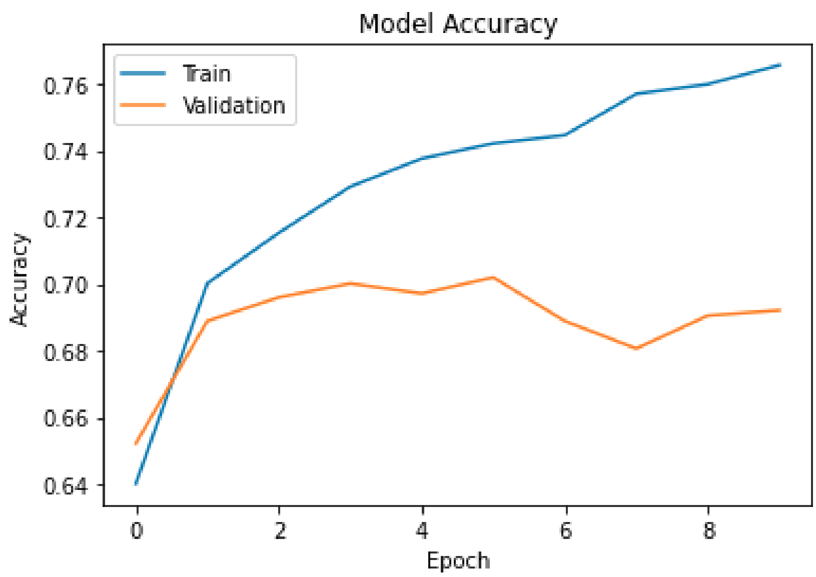


Figure 4.4. ResNet50 with BCE

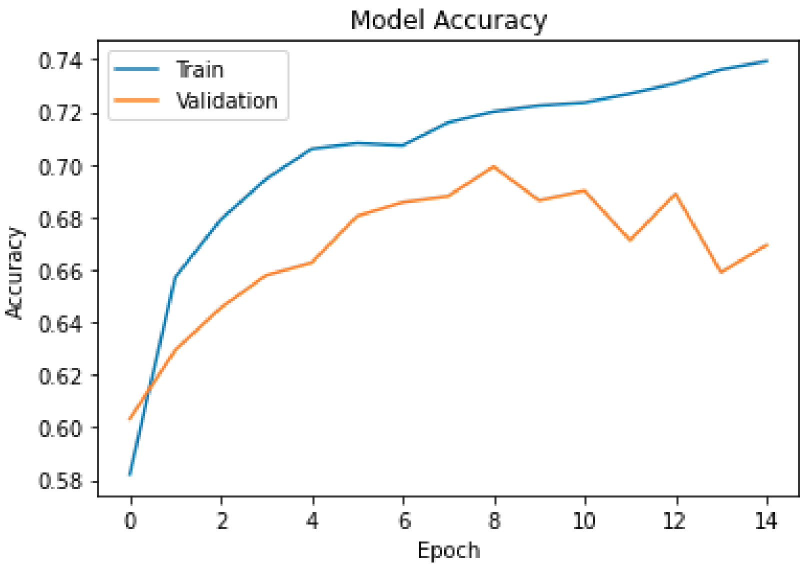


Figure 4.5. ResNet50 with WBCE

Utilizing the same dataset, Ahmed and Noor et al. [7] shared precision-recall tables for every class in their results. Drawing from the shared results, precision-recall tables were meticulously constructed for each individual class within the classification framework. These tables provide a comprehensive breakdown of the precision and recall metrics corresponding to each class, thereby offering insights into the classification performance at a granular level. Leveraging these shared findings, we then calculated the F1 scores for each label, which combines the precision and recall metrics using their harmonic mean for each class. Subsequently, a comparative analysis was conducted, juxtaposing these calculated F1 scores with our own experimental results. This comparative assessment allowed us to thoroughly evaluate the performance of our proposed methodology in relation to the benchmark results reported in the literature, thus providing valuable insights into the effectiveness and robustness of our approach in the domain of cloud label classification. However, concurrently with our research, Sunveg and Kunal collaborated with us on the dataset, sharing their F1 scores [29] Their study focused on segmentation, employing a CNN-based model featuring a U-shaped encoder-decoder network and residual blocks. The comparison is summarized in Table 4.2. These F1 scores provide insights into the model's performance in terms of precision and recall for each cloud type. By comparing our results with those reported by Ahmed and Noor et al., we can gauge the effectiveness of our approach and identify areas for potential improvement.

Table 4.1. Comparisons of F1 Scores Between ResNet50 and EfficientUnet

Algorithm	Fish	Flower	Gravel	Sugar	Avg
ResNet50 BCE	**0,71**	0,69	**0,73**	**0,80**	**0,73**
ResNet50 WBCE	**0,71**	**0,70**	0,71	0,70	0,70
Efficient-Unetb0-b5[7]	0,31	0,56	0,45	0,59	0,49
EffResUNet[22]	-	-	-	-	0,57

*Average of the all clouds f1 scores

In a study by Gonghe and Zuo et al. [8], they introduced CloudRCNN as an enhancement of the MaskRCNN model. They provided label-based scores and compared their model with others. Our comparison with their results is presented in Table 4.2. It is important to note that while Gonghe and Zuo et al.'s work is centered around segmentation for classification, this thesis approach focuses on object detection, and our calculations are based on predictions between 0 and 1 results. Despite this difference in methodology, our comparison allows us to assess the relative performance of our model compared to theirs.

Table 4.2. Comparisons Between ResNet50 and CloudRCNN

Algorithm	Fish	Flower	Gravel	Sugar	Avg
ResNet50 BCE	0,66	**0,77**	**0,65**	**0,69**	**0,70**
ResNet50 WBCE	**0,68**	**0,77**	0,63	0,61	0,67
CloudRCNN(auxiliary) [8]	0,381	0,329	0,525	0,420	0,41
CloudRCNN(feature) [8]	0,352	0,277	0,467	0,460	0,49

Based on the F1 score averages for the four clouds in ResNet50-BCE, ResNet50-WBCE, and EfficientUnet, the values are 0.732, 0.704, and 0.478, respectively. This indicates an increase in success rate of approximately 26% when comparing ResNet50-BCE with EfficientUnet.

Furthermore, the average success scores for the four clouds, as shown in Table 4.3. are 0.691, 0.671, 0.414, and 0.389, respectively. This demonstrates a 31% increase in success rate across the board when comparing ResNet50-BCE with EfficientUnet. These findings highlight the effectiveness of ResNet50-BCE and ResNet50-WBCE models in accurately classifying cloud formations compared to EfficientUnet. The improvement in success rates underscores the utility of employing ResNet50 models for cloud classification tasks.

In the classification of cloud images, we also conducted our own experiments using our second dataset, CCSN, with the ResNet50, MobiNet, and SENet algorithms. The performance results of these algorithms, applied to the CCSN dataset, are summarized in Table 4.4. These experiments provide valuable insights into the effectiveness of different models for cloud image classification, highlighting their strengths and weaknesses when applied to a diverse set of cloud images.

We obtained our results by running the source code shared in the study by Emmanuel and Purushotham.[20] We believe that the reason for the significantly lower performance of MobiNet in our case is since we ran our experiments on a dataset consisting of only 2,543 images without augmentation. In contrast, they extended their dataset to approximately 41,000 images through augmentation and shared their code accordingly.

We drew insights about SENet from the Li et al.[21] study and, unable to access their exact source code, we developed our own SENet algorithm and integrated it into our work. Due to differences in model architecture, our results likely differ from those reported by Li et al.

Table 4.3. Comparisons with CCSN datasets

Algorithm	Training Score	Test Score
ResNet50	**0.9165**	**0,51**
SENet	0.9089	0.33
MobiNet	0.2620	0.16

5. CONCLUSION

This thesis focuses on classifying cloud structures using two datasets. The first dataset consists of satellite images categorized into four types: Sugar, Fish, Gravel and Flower. The second dataset, CCSN, comprises 11 labels. The classification task employed the ResNet50 model, with images resized to 256x256 pixels. To evaluate their impact on performance, both binary cross-entropy (BCE) and weighted binary cross-entropy (WBCE) loss functions were utilized. The results indicate that the ResNet50 model with BCE loss function achieved superior performance for the Sugar and Gravel cloud shapes, whereas the WBCE loss function yielded better results for the Flower and Fish shapes. This underscores the significance of employing tailored loss functions for specific classes to enhance overall classification accuracy. For the second dataset, RestNet50 with BCE achieved the best results.

Furthermore, our approach was benchmarked against state-of-the-art methods to discern disparities. The findings underscore substantial advancements achieved through the approach we suggest, notably highlighting the ResNet50 model employing BCE loss is the most effective strategy for the cloud structure classification problem at hand.

For prospective investigations, it's crucial to acknowledge the imbalance in label distribution within the dataset utilized for this thesis. In future studies, exploring adjustments aimed at achieving a more equitable dataset could be beneficial, along with using larger datasets similar to other studies to achieve better results. By mitigating these inherent imbalances, enhanced performance in the classification endeavor is expected, thus opening up intriguing avenues for further exploration in this field

REFERENCES

[1] M. Quante, "The role of clouds in the climate system," *Journal de Physique IV (Proceedings)*, vol. 121, EDP Sciences, 2004.

[2] D. D. Turner, et al., "Thin liquid water clouds: Their importance and our challenge," *Bulletin of the American Meteorological Society*, vol. 88, no. 2, pp. 177-190, 2007.

[3] S. Graham, C. Parkinson, and M. Chahine, "The water cycle," *NASA Earth Observatory*, 2010.

[4] P. Schultz, "An explicit cloud physics parameterization for operational numerical weather prediction," *Monthly Weather Review*, vol. 123, no. 11, pp. 3331-3343, 1995.

[5] C. Bai, et al., "Multimodal Information Fusion for Weather."

[6] "Understanding Clouds from Satellite Images." Available online: https://www.kaggle.com/c/understanding-cloud-organization.

[7] T. Ahmed and N. H. N. Sabab, "Classification and understanding of cloud structures via satellite images with EfficientUNet," *SN Computer Science*, vol. 3, pp. 1-11, 2022.

[8] G. Shi and B. Zuo, "CloudRCNN: A Framework Based on Deep Neural Networks for Semantic Segmentation of Satellite Cloud Images," *Applied Sciences*, vol. 12, no. 11, p. 5370, 2022.

[9] H. Schulz, S. Rasp, B. B. Stevens, and S. Bony, "Combining Crowd-sourcing and Deep Learning to Understand Meso-scale Organization of Shallow Convection," in *AGU Fall Meeting Abstracts*, vol. 2019, pp. GC33A-03, 2019.

[10] "Understanding Cloud Organization from Satellite Images," Available online: https://mpimet.mpg.de/en/communication/news/understanding-cloud-organization-from-satellite-images.

[11] M. Bahi and M. Batouche, "Deep Learning for Ligand-Based Virtual Screening in Drug Discovery," in *PAIS 2018*, pp. 1-5, 2018.

[12] "What is the Convolutional Neural Network Architecture," Available online: https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/.

[13] J. Song, S. Gao, Y. Zhu, and C. Ma, "A survey of remote sensing image classification based on CNNs," *Big Earth Data*, vol. 3, no. 3, pp. 232-254, 2019.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, Available online: http://arxiv.org/abs/1512.03385.

[15] J. Wang, et al., "Cnn-rnn: A unified framework for multi-label image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[16] I. M. Baltruschat, et al., "Comparison of deep learning approaches for multi-label chest X-ray classification," *Scientific Reports*, vol. 9, no. 1, pp. 1-10, 2019.

[17] M. Manzo and S. Pellino, "Voting in transfer learning system for ground-based cloud classification," *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 542-553, 2021.

[18] X. Wan and J. Du, "Cloud Classification for Ground-Based Sky Image Using Random Forest," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 43, 2020.

[19] T. Durand, et al., "Learning a Deep ConvNet for Multi-label Classification with Partial Labels-Supplementary."

[20] E. K. Gyasi and P. Swarnalatha, "Cloud-MobiNet: an abridged Mobile-Net convolutional neural network model for ground-based cloud classification," *Atmosphere*, vol. 14, no. 2, p. 280, 2023.

[21] G. Li, "A novel computer-aided cloud type classification method based on convolutional neural network with squeeze-and-excitation," in *Journal of Physics: Conference Series*, vol. 1802, no. 3, p. 032051, IOP Publishing, 2021.

[22] M. Guzel, M. Kalkan, E. Bostanci, K. Acici, and T. Asuroglu, "Cloud type classification using deep learning with cloud images," *PeerJ Computer Science*, vol. 10, p. e1779, 2024. Available online: https://doi.org/10.7717/peerj-cs.1779.

[23] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2020.

[24] V. Patela, S. Shuklaa, and M. Gyanchandania, "Analysis of Different Loss Function for Designing Custom CNN for Traffic Sign Recognition," 2022.

[25] K. He, et al., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[26] X. Wang, et al., "Multi-label classification of pelvic organ prolapse using stress m magnetic resonance imaging with deep learning," *International Urogynecology Journal*, vol. 33, no. 10, pp. 2869-2877, 2022.

[27] "ResNet50," Available online: https://keras.io/api/applications/resnet/#resnet50-function.

[28] J. L. Zhang, P. Liu, F. Zhang, and Q. Q. Song, "CloudNet: Ground-based cloud classification with deep convolutional neural network," *Geophysical Research Letters*, vol. 45, pp. 8665-8672, 2018. Available online: https://doi.org/10.1029/2018GL077787.

[29] S. Nalwar, K. Shah, R. V. Bidwe, B. Zope, D. Mane, V. Jadhav, and K. Shaw, "EffResUNet: Encoder Decoder Architecture for Cloud-Type Segmentation," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 150, 2022.